

---

# Proceedings of the 9th Graz Brain-Computer Interface Conference 2024

Join Forces - Increase Performance

September 9-12, 2024  
Graz University of Technology, Austria

---

Edited by  
Gernot R. Müller-Putz, Kyriaki Kostoglou, Markus E. Oberndorfer, Selina C. Wriessneger

---



© Harry Schiffer

---

Partnered Event



Verlag der Technischen Universität Graz 2024



---

2024

Verlag der Technischen Universität Graz

[www.tugraz-verlag.at](http://www.tugraz-verlag.at)

ISSN 2311-0422

ISBN 978-3-99161-014-4

DOI 10.3217/978-3-99161-014-4



This work is licensed under the Creative Commons  
Attribution 4.0 International (CC BY 4.0) license.

<https://creativecommons.org/licenses/by/4.0/deed.en>

This CC license does not apply to the cover, third party material  
(attributed to other sources) and content noted otherwise.

## Welcome Note

---

### Join Forces – Increase Performance

We chose this year's conference title to concisely reflect the current state of the BCI research field. Researchers from both, the invasive and non-invasive communities, have increasingly worked together, forming a unified community. Techniques from the non-invasive field are now being applied in invasive research and vice versa. Additionally, we are at a point where the definition of a BCI is being questioned and needs to be reformulated. These questions and many more are crucial and need to be addressed achieving progress in BCI research.

The 9th Graz Brain-Computer Interface Conference (GBCIC2024) provides a platform for extensive discussions and exchanges among BCI experts from over 22 countries. We have received nearly 100 scientific contributions from approximately 476 authors, all peer-reviewed by at least two different reviewers. Accepted papers will be openly accessible and published by Verlag der TU Graz. The present conference proceedings are the result of this rigorous review process.

As a partnered event of the BCI Society, we have assembled a diverse and multifaceted program. We have organized several workshops as Satellite Events before the conference. During the conference, researchers will present their work either as talks or posters. We are fortunate that renowned experts in the field such as Dr. Andrea Kübler, Dr. Jennifer Collinger, Dr. Camille Jeunet-Kelway, Dr. Nick Ramsey, and Dr. Henri Lorach accepted our invitation to present keynote addresses at the conference. After a break of several years, GBCIC2024 will conclude with a tour to the South Styrian Vine Yards.

The BCI conferences held in Graz, Austria, are considered an international initiative that fosters stronger scientific cooperation in the BCI field.

We wish all participants an exciting and stimulating Graz BCI Conference 2024.



Gernot R. Müller-Putz  
Conference Chair

## Editorial Board

---

Prof. Dr. **Gernot Rudolf Müller-Putz** is head of the Institute of Neural Engineering and its associated Laboratory of Brain-Computer Interfaces. He received his MSc in electrical and biomedical engineering in 2000, his PhD in electrical engineering in 2004 and his habilitation and “*venia docendi*” in medical informatics from Graz University of Technology in 2008. Since 2014 he is full professor for semantic data analysis. He has gained extensive experience in the field of biosignal analysis, brain-computer interface research, EEG-based neuroprosthesis control, communication with BCI in patients with disorders of consciousness, hybrid BCI systems, the human somatosensory system, and BCIs in assistive technology over the past 24 years. He has also managed several national projects (State of Styria) and international projects (Wings for Life, EU Projects) and he recently coordinated the EU Horizon 2020 project MoreGrasp. Furthermore, he organized and hosted seven international Brain-Computer Interface Conferences over the last 17 years in Graz and chairing the 9th Conference in Sept. 2024. Since August 2019 he is Speciality Chief Editor of *Frontiers in Human Neuroscience: Brain-Computer Interfaces*. He has authored more than 200 peer reviewed publications which were cited more than 16000 times (h-index 77). Recently he was awarded with an ERC Consolidator Grant “Feel your Reach” from the European Research Council. In May 2017 he received the Ludwig-Guttmann Award from the German Medical Spinal Cord Injury Association (DMGP). In May 2018 he was elected into the Board of Directors of the International Brain-Computer Interface Society. In May 2019 he received the Science Award from the State of Styria.

**Selina Christin Wriessnegger** is Associate professor at the Institute of Neural Engineering (BCI-Lab), Graz University of Technology, Austria. From 2001 to 2005 she was PhD student at the Max-Planck-Institute for Human Cognitive and Brain Sciences and received her PhD from the Ludwig-Maximilians University in Munich, Germany. During that time, she spent one year in Rome as research assistant at IRCCS (Fondazione Santa Lucia), Laboratory for Human Psychophysiology. From 2005 to 2008 she was university assistant at the Karl-Franzens-University Graz, section neuropsychology. From 2009 until May 2016 she was senior researcher at the Institute of Neural Engineering (BCI-Lab). In 2017 she was visiting professor at SISSA (Scuola Internazionale Superiore di Studi Avanzati), Trieste. Her research interests are, neural correlates of covert actions, novel applications of BCIs for healthy users, passive BCIs, VR-based neuroadaptive systems and mental state detection.

**Kyriaki Kostoglou** received her diploma degree in Electrical and Computer Engineering from Aristotle University of Thessaloniki (AUTH), Greece and her M.Sc. degree in Computer Engineering from University of Cyprus (UCY), Cyprus. In 2017, she completed her Ph.D. studies and received her Ph.D. degree from the Department of Electrical and Computer Engineering, McGill University, Canada. The topic of her Ph.D. thesis was the identification of multiple-input time-varying systems and binary response systems for biomedical applications. She worked as a postdoc researcher in medical ultrasound imaging at the Institute of Signal Processing, Johannes Kepler University, Linz, Austria. Currently she is a postdoc at the Institute of Neural Engineering, Graz University of Technology, Graz, Austria. Her current research interests include system identification and signal processing for biomedical applications and brain computer interfaces.

**Markus Erwin Oberndorfer** is university assistant at the Institute of Neural Engineering (BCI-Lab), Graz University of Technology, Austria. He received his M.Sc. in Biomedical Engineering, specializing in Computational Neuroscience, from the Graz University of Technology in 2024. His research primarily addresses the forward and inverse problems in EEG, as well as the study of electric potentials originating from the spinal cord. Currently he is working towards his PhD degree in Biomedical Engineering.

## Organizing Committee I

---

### Conference Chair

Gernot R. Müller-Putz  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*  
(BCI Society Member)

### Conference Secretary

Irmgard B. Humenberger  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

### Industrial Sponsoring

Gernot R. Müller-Putz  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*  
(BCI Society Member)

### Papers and Proceedings

Kyriaki Kostoglou  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Irmgard B. Humenberger  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Markus E. Oberndorfer  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

### Poster Sessions

Markus E. Oberndorfer  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

### Program Book

Kyriaki Kostoglou  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Patrick Suwandjieff  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Johanna Egger  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Shayan Jalilpour  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

## Organizing Committee II

---

### Social Media & Graphics

Markus Crell  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

### Volunteers

Selina C. Wriessnegger  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

### Exhibition & Infrastructure

Shayan Jalilpour  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Markus Crell  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Patrick Suwandjieff  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Johanna Egger  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

### Workshops

Adyasha Dash  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

### Talk & Poster Award

Selina C. Wriessnegger  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

### Additional Local Staff

Hannah Pulferer  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Nitikorn Srisrisawang  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

Michael Wimmer  
*Know-Center GmbH*  
*Austria*

Fu Xi  
Centre for Brain-Computing Research  
*Nanyang Technological University*  
Singapore

Christoph Sabitzer  
*Institute of Neural Engineering*  
*Graz University of Technology*  
*Austria*

## International Program Committee and Review Board I

---

We are very grateful to all reviewers for their help, to make this conference a success!

### A

Aarnoutse Erik  
Arvaneh Mahnaz

University Medical Center Utrecht  
University of Sheffield

### B

Berezutskaya Julia  
Berger Lisa  
Brunner Clemens

University Medical Center Utrecht  
University of Graz  
University of Graz

### C

Cisotto Giulia  
Collinger Jennifer  
Coyle Damien  
Crell Markus

University of Padova  
University of Pittsburgh  
University of Ulster  
Graz University of Technology

### D

Dash Adyasha

Graz University of Technology

### E

Egger Johanna

Graz University of Technology

### G

Grosse-Wentrup Moritz

University of Vienna

### H

Halder Sebastian  
Herff Christian

University of Essex  
Maastricht University

### J

Jalilpour Shayan  
Jeunet-Kelway Camille

Graz University of Technology  
University Bordeaux

### K

Kanoh Shin'Ichiro  
Kober Silvia  
Kobler Reinmar

Shibaura Institute of Technology  
University of Graz  
Advanced Telecommunications Research Institute Intern.

## International Program Committee and Review Board II

---

König Peter  
Kostoglou Kyriaki  
Krusienski Dean

University Osnabrück  
Graz University of Technology  
Virginia Commonwealth University

### L

Lopes Dias Catarina  
Lotte Fabien

Medical University of Graz  
INRIA Bordeaux Sud-Ouest

### M

Martinez-Cagigal Victor  
Matran-Fernandez Ana  
Mattia Donatella  
Müller-Putz Gernot R.  
Millán José del R.

University of Valladolid  
University of Essex  
Fondazione Santa Lucia, IRCCS  
Graz University of Technology  
University of Texas at Austin

### O

Oberndorfer Markus E.  
Ofner Patrick  
Ortner Rupert

Graz University of Technology  
University of Freiburg  
g.tec medical engineering GmbH

### P

Papadopoulos Sotiris  
Perdikis Serafeim  
Pereira Joana  
Pillette Léa  
Pulferer Hannah

University Lyon  
University of Essex  
University of Freiburg  
Université de Rennes  
Graz University of Technology

### R

Raggam Philipp  
Ron-Angevin Ricardo  
Roy Raphaëlle  
Rutkowski Tomasz M.

University of Vienna  
University of Málaga  
Université de Toulouse  
RIKEN, AIP, Japan

### S

Savic Andrej  
Sburlea Andreea  
Srisrisawang Nitikorn  
Suwandjeff Patrick

University of Belgrade  
University of Groningen  
Graz University of Technology  
Graz University of Technology

## International Program Committee and Review Board III

---

### **T**

Tangermann Michael  
Thielen Jordy  
Tonin Luca

Radboud University  
Radboud University  
University of Padova

### **V**

Valeriani Davide  
Vansteensel Mariska  
Vourvopoulos Athanasios

Technogym  
University Medical Center Utrecht  
ISR-Lisboa

### **W**

Wimmer Michael  
Wriessnegger Selina Christin

Know-Center GmbH  
Graz University of Technology

## List of Authors I

---

List of authors in alphabetical order with start pages of their respective contributions.

### A

Abdelhafez, Norhan .....	484
Ahmadi, Sara .....	337, 343
Aksenova, Tetiana.....	68, 80
Ali, Rushna .....	507
Amadiou, Franck.....	179
Amaunam, Idorenyin.....	531
Amigó-Vega, Joaquín.....	478
Ammendola, Lidia .....	331
Annicchiarico, Côme .....	472
Arpaia, Pasquale.....	47, 331
Ayyoubi, Amir Hossein .....	425

### B

Baker, Matthew R.....	127, 513
Banks, Samantha A. ....	127
Bannier, Elise .....	254
Barłoga, Aneta .....	272
Baum, Jonathan.....	443
Baumgarten, Daniel .....	289
Bellicha, Angelina.....	80
Berezutskaya, Julia.....	58
Berger, Lisa.....	403
Bertoni, Tommaso .....	319
Besheli, Behrang Fazli .....	425
Bhattacharyya, Saugat.....	295
Bidgoli, Seyed Javad .....	92
Biktimirov, Artur .....	381
Bonaiuto, James J. ....	236
Bouet, Romain .....	420
Bougrain, Laurent .....	92
Branco, Mariana P.....	150, 391, 495
Brock, Anke M.....	212
Brunner, Peter.....	132, 190, 507

## List of Authors II

---

### C

Cabestaing, François .....	249
Cancino-Fuentes, Nathalia.....	489
Cantürk, Atilla.....	41
Carrara, Igor .....	431
Cebolla Alvarez, Ana Maria .....	92
Chabardès, Stéphan.....	80
Chang, Su-Youne.....	190
Charvet, Guillaume .....	68, 80
Chavez, Mario.....	98
Cheron, Guy .....	92
Christopoulos, Alexandros .....	224
Ciuffini, Roberta .....	366
Coenen, Volker .....	507
Congedo, Marco .....	236
Corsi, Marie-Constance .....	98
Coudroy, Elina.....	466
Covelo, Joana.....	489
Crell, Markus.....	139
Cropano, Maria .....	331
Cueva, Valérie Marissens .....	92
Cunningham, Andrew.....	63
Cury, Claire .....	254

### D

Daly, Ian.....	525
Darnet, Ludovic.....	236
Dash, Adyasha.....	543
De Blasiis, Paolo.....	331
de Jong, Ivo .....	86, 156
De Luca, Matteo .....	331
De Vico Fallani, Fabrizio .....	174
Debroize, Rene-Paul.....	254
Della Calce, Anna .....	331
Desain, Peter .....	337, 343
Desbois, Arthur .....	174

---

## List of Authors III

---

Di Matteo, Alessandro .....	355, 366
Diserens, Karin .....	319
Dreyer, Pauline .....	448
Driessens, Léa .....	278
Dumas, Cassandra .....	375
Duque-Lopez, Andrea .....	190
Dussard, Claire .....	168, 375
Dürschmid, Stefan .....	207
<b>E</b>	
Eder, Manuel .....	145
Edlinger, Guenter .....	489
Egger, Johanna .....	5
Eidel, Matthias .....	115
Ekramy, Nora .....	484
Elsayed, Neven .....	63
Engelhardt, Will .....	190
Esposito, Antonio .....	47
<b>F</b>	
Farne, Alessandro .....	420
Fernández-Rodríguez, Álvaro .....	230
Forin, Paolo .....	360
Fragueiro, Agustina .....	254
Freudenburg, Zachary .....	58, 391, 460
Fu, Zhichun .....	53
Fugger, Peter .....	145
<b>G</b>	
Galdieri, Fortuna .....	47
Gao, Xin .....	53
Gargiulo, Ludovica .....	331
Gasq, David .....	179
Gattaz, Lucie .....	420
George, Nathalie .....	168, 375
Gherman, Diana E. ....	307
Grechukhin, Natalia .....	466
Grevet, Elise .....	179

---

## List of Authors IV

---

Grosse-Wentrup, Moritz .....	145, 272
Guerci, Philippe .....	92
Guetschel, Pierre .....	11, 349, 437
Guger, Christoph.....	284, 301, 489
<b>H</b>	
Halder, Sebastian.....	319, 525
Hashemi, Iraj.....	92
Herff, Christian .....	478
Hermes, Dora .....	132, 190, 195, 501, 518
Hinrichs, Hermann .....	207
Hinss, Marcel F. ....	212
Hons, Manuel.....	266
Hornero, Roberto .....	230, 409
Huang, Harvey .....	195
Hugueville, Laurent.....	375
<b>I</b>	
Ince, Nuri F.....	132, 190, 195, 425, 507
Izac, Margaux .....	179, 466
<b>J</b>	
Jalilpour, Shayan.....	23
Jensen, Michael A.....	127, 132, 195, 501, 518
Jeunet-Kelway, Camille.....	168, 179, 375, 466
Juillard, Violaine .....	80
Jöhr, Jane .....	319
<b>K</b>	
Kamada, Kyosuke.....	284
Kanoh, Shin'Ichiro.....	29, 35
Kapeller, Christoph.....	284
Karakas, Serpil .....	80
Kasprzak, Hubert .....	201
Keller, Dirk .....	495
Kerezoudis, Panagiotis .....	195
Kim, Inyong.....	190, 507
Kim, Jiwon .....	190, 507
Klassen, Bryan T.....	127, 195, 513

---

## List of Authors V

---

Klein, Guido .....	437
Klug, Marius.....	218, 243
Kober, Silvia Erika.....	185, 266, 403
Kojima, Simon.....	29, 35
Komendziński, Tomasz .....	201
Korostenskaja, Milena.....	284
Kostoglou, Kyriaki .....	5, 17, 74
Kremen, Vaclav.....	190, 507
Krol, Laurens R. ....	372
Kromm, Maria .....	150
Kubben, Pieter L. ....	478
Kübler, Andrea .....	115, 162, 260
<b>L</b>	
Lafaye de Micheaux, Hugo .....	68
Lampert, Frederik.....	190, 507
Lau, Brian .....	375
Le Jeune, François .....	278, 397
Lebedev, Mikhail .....	381
Lecuyer, Anatole .....	249
Leeb, Robert.....	531
Lekhnitskaya, Polina .....	1
Lopes da Silva, Marina .....	319
Lorach, Henri .....	68, 80
Lotte, Fabien.....	92, 385, 443, 448, 472
Lozzi, Daniele .....	355, 366
Lus, Giacomo.....	331
Lécuyer, Anatole .....	313, 397
<b>M</b>	
Maby, Emmanuel .....	420
Macé, Marc J-M. ....	278, 313, 397
Maffei, Luigi .....	331
Malangone, Daniela .....	331
Manes, Costanzo .....	355
Manivannan, Prithviraj .....	86
Mannino, Camilla .....	98

---

## List of Authors VI

---

Marcos-Martínez, Diego.....	409
Marrelli, Alfonso .....	366
Martel, Félix .....	68, 80
Martinez, Jesus Casal.....	443
Martín-Fernández, Ana .....	230
Martínez-Cagigal, Víctor .....	230, 409
Masson, Eva .....	260
Matran-Fernandez, Ana .....	319, 525
Matsoukis, Stratis.....	489
Mattei, Enrico .....	355, 366
Mattout, Jérémie .....	236, 420, 472
Maurel, Pierre .....	397
Maurer, Magdalena .....	289
Mehrkanoon, Siamak .....	495
Meistelman, Claude .....	92
Menegatti, Emanuele .....	360
Mignosi, Filippo.....	355
Mihić Zidar, Lucija .....	272
Miller, Kai J. ....	127, 132, 190, 195, 425, 501, 507, 513, 518
Mirehkoohi, Mehdi Javani .....	460
Mivalt, Filip.....	190, 507
Mizukami, Naoki.....	35
Moccaldi, Nicola.....	331
Mohammadian, Farhad .....	272
Mohammadpour, Mostafa .....	284
Moreau, Thomas.....	11
Muñoz-Montes de Oca, Jenny Noemí .....	414
Müller-Putz, Gernot R. ....	5, 17, 23, 63, 74, 104, 121, 139, 537
<b>N</b>	
N'Kaoua, Bernard .....	385, 466
Narayanan, Shekhar .....	343
Natalizio, Angela .....	47
Nawaz, Rab .....	319
Neumann, Amira .....	460
Niewińska, Nina .....	201

---

## List of Authors VII

---

Noel, Jean-Paul .....	319
Nour-Eldin, Mohammed .....	484
<b>O</b>	
Oberndorfer, Markus .....	104
Offenberg, Elena Charlotte.....	58
Ojeda Valencia, Gabriela .....	127, 513
Ojemann, Jeffrey.....	195
Okkabaz, Jhan L.....	425, 507
Ornello, Raffaele .....	366
Otake-Matsuura, Mihoko .....	201
Ottenhoff, Maarten C.....	478
<b>P</b>	
Palatella, Alessio.....	360
Pan, Yanzhao.....	218, 243
Papadopoulo, Theodore.....	431
Papadopoulos, Sotirios .....	236
Parvis, Marco .....	47
Pascual-Roa, Beatriz .....	409
Pepicelli, Alex.....	63
Perdikis, Serafeim.....	319, 531
Permezel, Fiona.....	518
Petieau, Mathieu.....	92
Petit, Jimmy .....	115
Pfeiffer, Maria.....	260
Pierrieau, Emeline.....	168, 466
Piliugin, Nikita .....	381
Pillette, Léa .....	168, 278, 313, 375, 397, 466
Placidi, Giuseppe .....	355, 366
Pollastro, Andrea.....	47
Polsinelli, Matteo.....	355, 366
Polyanskaya, Arina .....	109
Popa, Alexia-Theodora .....	145
Porubcová, Natália.....	109
Prasad, Girijesh .....	295
Pretl, Harald.....	301

---

## List of Authors VIII

---

Pulferer, Hannah .....	74
Py, Jacques .....	179
Pérez-Velasco, Sergio .....	409
<b>Q</b>	
Quach, Michael .....	425
Querry, Ambre .....	420
<b>R</b>	
Rabe, Lea .....	218, 243
Raemaekers, Mathijs .....	150
Raggam, Philipp .....	145, 272
Raimo, Simona .....	331
Ramsey, Nick F. ....	58, 150, 391, 460
Raslan, Ahmed .....	501
Redmond, Erin .....	443
Reichert, Christoph .....	207
Reintsema, Lars H. ....	207
Rimbert, Sébastien .....	92
Rizzo, Lorianna .....	420
Roc, Aline .....	448
Rockhill, Alex .....	501
Rodrigues, Johannes .....	260
Rodriguez-Herreros, Borja .....	531
Romero-Morales, Héctor .....	414
Rosignoli, Chiara .....	366
Rosipal, Roman .....	109
Rossignol, Eléa .....	466
Rouillard, José .....	115
Roy, Raphaëlle N. ....	212, 448
Rošt'Áková, Zuzana .....	109
Rutkowski, Tomasz M. ....	201
<b>S</b>	
Sacco, Simona .....	366
Samanta, Kaniska .....	295
Sanchez-Vives, Maria V. ....	489
Santamaría-Vázquez, Eduardo .....	230, 409

---

## List of Authors IX

---

Sauter-Starace, Fabien .....	68, 80
Savalle, Emile .....	278, 397
Sayed, Abdelrahman.....	484
Sburlea, Andreea I. ....	86, 156, 224, 301
Schalk, Gerwin.....	132, 190, 507
Scharinger, Josef .....	284, 489
Schellander, Sophia .....	150
Scheppink, Hanneke.....	337
Scherer, Reinhold .....	319
Schmartz, Denis.....	92
Schomaker, Pauline .....	301
Schreiner, Leonhard.....	301
Schwarzgruber, Michael.....	284, 301
Seguin, Perrine Rose.....	420
Serino, Andrea .....	319
Settgast, Tomko .....	162
Shevtsova, Yulia G.....	454
Shishkin, Sergei L. ....	454
Si-Mohammed, Hakim .....	249
Sieghartsleitner, Sebastian .....	301
Signoriello, Elisabetta .....	331
Silvestri, Gianluigi .....	437
Sintsov, Mikhail .....	381
Sobolová, Gabriela .....	109
Soghoyan, Gurgen.....	381
Sorrentino, Pierpaolo .....	98
Sosulski, Jan.....	325
Srisrisawang, Nitikorn .....	74, 121
Struber, Lucas.....	80
Sultana, Mushfika .....	531
Suwandjieff, Patrick .....	537
Swamy, Chandra Prakash.....	425
Swann, Nicole .....	501
Sweeney-Reed, Catherine M. ....	207
Szul, Maciej J.....	236

---

## List of Authors X

---

### T

Tadi, Tej .....	531
Tangermann, Michael.....	11, 325, 337, 349, 437
Tantawy, Manal .....	484
Tates, Alberto .....	525
Thielen, Jordy .....	325, 337, 343
Thomas, Bruce H. ....	63
Tonin, Luca .....	360
Torres-García, Alejandro Antonio .....	414
Tortora, Stefano .....	360
Trocellier, David .....	385, 448

### V

Valdenegro-Toro, Matias .....	86, 156, 224
van den Boom, Max A. ....	190, 507
van den Wittenboer, Lüke .....	156
Van Der Lee, Gael .....	249
Vansteensel, Mariska .....	495
Vasilyev, Anatoly N.....	454
Veas, Eduardo E. ....	63
Venot, Tristan .....	174
Verwoert, Maxime .....	478
Villaseñor-Pineda, Luis .....	414
Vitale, Vincenzo Maria .....	212
Vitkova, Viktoriya .....	92
Volmer, Ben .....	63
Volosyak, Ivan.....	41
Vorwerk, Johannes .....	289

### W

Ward, Tomas.....	443
Wassenaar, Peter .....	349
Welter, Marc.....	443, 448
Wimmer, Michael .....	63
Won, Kyungho .....	313
Wong-Lin, Kongfatt .....	295
Wood, Guilherme .....	185, 266, 403

---

## List of Authors XI

---

Worrell, Gregory A. ....	190, 425, 507
Wriessnegger, Selina Christin .....	266, 543
Wu, Xiaolong .....	53
<b>Y</b>	
Yashin, Artem S. ....	454
<b>Z</b>	
Zander, Thorsten O. ....	307, 372
Zhang, Dingguo .....	53
Zhong, Walker.....	391

## Table of Contents I

---

1. WORD PREDICTION DURING NATURALISTIC SPEECH PERCEPTION .....	1
Polina Lekhnitskaya	
DOI: 10.3217/978-3-99161-014-4-001	
2. INVESTIGATING TEMPORAL VARIATIONS IN MRCPS AND THEIR INFLUENCE ON CLASSIFICATION: A 10-HOUR EEG STUDY .....	5
Johanna Egger, Kyriaki Kostoglou, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-002	
3. S-JEPA: TOWARDS SEAMLESS CROSS-DATASET TRANSFER THROUGH DYNAMIC SPATIAL ATTENTION .....	11
Pierre Guetschel, Thomas Moreau, Michael Tangermann	
DOI: 10.3217/978-3-99161-014-4-003	
4. OPTIMIZING TIME-VARYING AUTOREGRESSIVE MODELS FOR BCI APPLICATIONS ....	17
Kyriaki Kostoglou, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-004	
5. RECOGNITION OF PERTURBATION EVOKED POTENTIAL BY USING MIXED-DEPTHWISE CONVOLUTIONS .....	23
Shayan Jalilpour, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-005	
6. INTRODUCING THE ASME-SPELLER, AUDITORY BCI SPELLER UTILIZING STREAM SEGREGATION: A PILOT STUDY .....	29
Simon Kojima, Shin'Ichiro Kanoh	
DOI: 10.3217/978-3-99161-014-4-006	
7. A NEW AUDITORY BRAIN-COMPUTER INTERFACE BASED ON STREAM SEGREGATION UTILIZING ASSR .....	35
Shin'Ichiro Kanoh, Naoki Mizukami, Simon Kojima	
DOI: 10.3217/978-3-99161-014-4-007	
8. A NOVEL CHATGPT-DRIVEN COMMUNICATION AID BASED ON CODE-MODULATED VISUAL EVOKED POTENTIALS (CVEP) .....	41
Atilla Cantürk, Ivan Volosyak	
DOI: 10.3217/978-3-99161-014-4-008	
9. A STUDY OF PERFORMANCE VARIABILITY IN DEEP NEURAL NETWORKS FOR MOTOR IMAGERY CLASSIFICATION: TOWARDS A ZERO-CALIBRATION APPROACH.....	47
Pasquale Arpaia, Antonio Esposito, Fortuna Galdieri, Angela Natalizio, Marco Parvis, Andrea Pollastro	
DOI: 10.3217/978-3-99161-014-4-009	

## Table of Contents II

---

10. DEEP LEARNING FOR MOTOR IMAGERY-BASED BCIS USING SEEG SIGNALS.....	53
Zhichun Fu, Xiaolong Wu, Xin Gao, Dingguo Zhang	
DOI: 10.3217/978-3-99161-014-4-010	
11. HIGH-PERFORMANCE NEURAL DECODING OF 14 DUTCH KEYWORDS .....	58
Elena Charlotte Offenberg, Julia Berezutskaya, Zachary Freudenburg, Nick F. Ramsey	
DOI: 10.3217/978-3-99161-014-4-011	
12. PROCESSING OF INCONGRUENT INFORMATION CAN BE DECODED FROM SINGLE- TRIAL EEG: AN AR-STUDY .....	63
Michael Wimmer, Alex Pepicelli, Ben Volmer, Neven Elsayed, Andrew Cunningham, Bruce H. Thomas, Eduardo E. Veas, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-012	
13. AUTO-ADAPTATION OF ECOG-BASED MOTOR BCI USING NEURAL RESPONSE DE- CODER: A CROSS-PATIENT STUDY .....	68
Hugo Lafaye de Micheaux, Félix Martel, Fabien Sauter-Starace, Guillaume Charvet, Henri Lorach, Tetiana Aksenova	
DOI: 10.3217/978-3-99161-014-4-013	
14. CORRECTING TRAJECTORY-DECODING ERRORS VIA CORTICAL SUBSTRATES OF CONTINUOUS ERRONEOUS FEEDBACK PROCESSING.....	74
Hannah Pulferer, Kyriaki Kostoglou, Nitikorn Srisrisawang, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-014	
15. PREDICTORS OF ECOG-BCI PERFORMANCES ACROSS SUBJECTS AND SESSIONS DE- RIVED FROM IDLE STATE CHARACTERISTICS .....	80
Lucas Struber, Félix Martel, Serpil Karakas, Violaine Juillard, Angelina Bellicha, Fabien Sauter-Starace, Stéphan Chabardès, Henri Lorach, Guillaume Charvet, Tetiana Aksenova	
DOI: 10.3217/978-3-99161-014-4-015	
16. UNCERTAINTY QUANTIFICATION FOR CROSS-SUBJECT MOTOR IMAGERY CLASSIFI- CATION .....	86
Prithviraj Manivannan, Ivo de Jong, Matias Valdenegro-Toro, Andreea I. Sburlea	
DOI: 10.3217/978-3-99161-014-4-016	
17. TOWARDS RIEMANNIAN EEG CLASSIFIERS TO DETECT AWAKE AND ANESTHETIZED STATES USING MEDIAN NERVE STIMULATION .....	92
Valérie Marissens Cueva, Sébastien Rimbert, Ana Maria Cebolla Alvarez, Mathieu Petieau, Viktoriya Vitkova, Iraj Hashemi, Guy Cheron, Claude Meistelman, Philippe Guerci, Denis Schmartz, Seyed Javad Bidgoli, Laurent Bougrain, Fabien Lotte	
DOI: 10.3217/978-3-99161-014-4-017	

## Table of Contents III

---

18. NEURONAL AVALANCHES FOR EEG-BASED MOTOR IMAGERY BCI .....	98
Camilla Mannino, Marie-Constance Corsi, Pierpaolo Sorrentino, Mario Chavez	
DOI: 10.3217/978-3-99161-014-4-018	
19. LOCALIZING NEURAL SOURCES IN THE CERVICAL SPINAL CORD .....	104
Markus Oberndorfer, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-019	
20. A SMALL STEP TOWARDS THE DETECTION OF MENTAL FATIGUE INDUCED BY BCI- HMD TRAINING .....	109
Arina Polyanskaya, Roman Rosipal, Gabriela Sobolová, Zuzana Rošt'Áková, Natália Po- rubcová	
DOI: 10.3217/978-3-99161-014-4-020	
21. RECORDING THE SSSEP WITH THE CEEGRID .....	115
Jimmy Petit, Matthias Eidel, José Rouillard, Andrea Kübler	
DOI: 10.3217/978-3-99161-014-4-021	
22. INVESTIGATING COORDINATES REPRESENTATION DURING REACHING VIA LOW-FRE- QUENCY EEG: A PRELIMINARY STUDY .....	121
Nitikorn Srisrisawang, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-022	
23. IDENTIFYING NEW FEATURES FOR BCI CONTROL: SPECTRAL CHANGES IN THE MO- TOR THALAMUS REVEAL HAND REPRESENTATION DURING OVERT AND IMAGINED MOVEMENT .....	127
Matthew R. Baker, Bryan T. Klassen, Michael A. Jensen, Gabriela Ojeda Valencia, Samantha A. Banks, Kai J. Miller	
DOI: 10.3217/978-3-99161-014-4-023	
24. FEASIBILITY OF STEREO EEG BASED BRAIN COMPUTER INTERFACING IN AN ADULT AND PEDIATRIC COHORT .....	132
Michael A. Jensen, Gerwin Schalk, Nuri F. Ince, Dora Hermes, Peter Brunner, Kai J. Miller	
DOI: 10.3217/978-3-99161-014-4-024	
25. DETECTION OF MOTION TERMINATION FROM EEG DURING THE EXECUTION OF CON- TINUOUS HAND MOVEMENT .....	139
Markus Crell, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-025	
26. AN EMG-BASED BRAIN-COMPUTER INTERFACE FOR COMMUNICATION-IMPAIRED PA- TIENTS: A CASE STUDY .....	145
Philipp Raggam, Manuel Eder, Alexia-Theodora Popa, Peter Fugger, Moritz Grosse-Wentrup	

---

## Table of Contents IV

---

DOI: 10.3217/978-3-99161-014-4-026

27. FINDING THE OPTIMAL SIX: DECODING FROM A LARGE SET OF HAND GESTURES WITH 7T FMRI FOR IMPROVED BCI CONTROL..... 150  
Maria Kromm, Sophia Schellander, Mariana P. Branco, Mathijs Raemaekers, Nick F. Ramsey  
DOI: 10.3217/978-3-99161-014-4-027
28. TRANSFERRING BCI MODELS FROM CALIBRATION TO CONTROL: OBSERVING SHIFTS IN EEG FEATURES..... 156  
Ivo de Jong, Lüke van den Wittenboer, Matias Valdenegro-Toro, Andreea I. Sburlea  
DOI: 10.3217/978-3-99161-014-4-028
29. RESTING-STATE BRAIN CRITICALITY AND PERFORMANCE WITH P300-BASED BCIS.. ..... 162  
Tomko Settgast, Andrea Kübler  
DOI: 10.3217/978-3-99161-014-4-029
30. BIDIRECTIONAL NEUROFEEDBACK: A CONTROL CONDITION COMPLEMENTARY TO SHAM? ..... 168  
Emeline Pierrieau, Léa Pillette, Claire Dussard, Nathalie George, Camille Jeunet-Kelway  
DOI: 10.3217/978-3-99161-014-4-030
31. DYNAMIC BRAIN NETWORKS IN MOTOR IMAGERY-BASED BCI ..... 174  
Tristan Venot, Arthur Desbois, Fabrizio De Vico Fallani  
DOI: 10.3217/978-3-99161-014-4-031
32. WHICH FACTORS AFFECT THE ACCEPTABILITY OF BCIS FOR FUNCTIONAL REHABILITATION AFTER STROKE AMONG PATIENTS? A QUESTIONNAIRE STUDY AMONG 140 PATIENTS AND A COMPARISON WITH THE GENERAL PUBLIC. .... 179  
Elise Grevet, Margaux Izac, Franck Amadieu, Jacques Py, David Gasq, Camille Jeunet-Kelway  
DOI: 10.3217/978-3-99161-014-4-032
33. BREAKING OUT OF THE FEEDBACK LOOP: TRANSFERRING MASTERY OF SELF-REGULATION DURING NEUROFEEDBACK TO OTHER CONTEXTS..... 185  
Silvia Erika Kober, Guilherme Wood  
DOI: 10.3217/978-3-99161-014-4-033
34. INTEGRATING CORTEC BRAININTERCHANGE DEVICE AND BCI2000 WITH A CLOUD INTERFACE ..... 190  
Filip Mivalt, Frederik Lampert, Max A. van den Boom, Jiwon Kim, Andrea Duque-Lopez, Will Engelhardt, Inyong Kim, Su-Youne Chang, Dora Hermes, Peter Brunner, Vaclav Kremen, Nuri F. Ince, Gerwin Schalk, Gregory A. Worrell, Kai J. Miller  
DOI: 10.3217/978-3-99161-014-4-034

## Table of Contents V

---

35. SPATIAL AND SPECTRAL CHANGES IN CORTICAL POTENTIALS DURING PINCHING VERSUS THUMB AND INDEX FINGER FLEXION .....	195
Panagiotis Kerezoudis, Michael A. Jensen, Harvey Huang, Jeffrey Ojemann, Bryan T. Klassen, Nuri F. Ince, Dora Hermes, Kai J. Miller	
DOI: 10.3217/978-3-99161-014-4-035	
36. PASSIVE OLFACTORY BRAIN-COMPUTER INTERFACE PARADIGM FOR AWARENESS LEVEL PREDICTION .....	201
Tomasz M. Rutkowski, Hubert Kasprzak, Nina Niewińska, Mihoko Otake-Matsuura, Tomasz Komendziński	
DOI: 10.3217/978-3-99161-014-4-036	
37. SSVEP-BASED COVERT COMMUNICATION USING HYPERSCANNING.....	207
Lars H. Reintsema, Catherine M. Sweeney-Reed, Stefan Dürschmid, Hermann Hinrichs, Christoph Reichert	
DOI: 10.3217/978-3-99161-014-4-037	
38. EEG-BASED PERFORMANCE ESTIMATION DURING A REALISTIC DRONE PILOTING TASK .....	212
Marcel F. Hinss, Vincenzo Maria Vitale, Anke M. Brock, Raphaëlle N. Roy	
DOI: 10.3217/978-3-99161-014-4-038	
39. LESS IS MORE: ADVANCING EEG-BASED ONLINE CONTINUOUS MACHINE ERROR DETECTION WITH THE LIGHTWEIGHT MAX-MIN AMPLITUDE NOISE FILTERING TECHNIQUE .....	218
Yanzhao Pan, Lea Rabe, Marius Klug	
DOI: 10.3217/978-3-99161-014-4-039	
40. ANA-E: A NOVEL APPROACH FOR PRE-TRAINED ERROR DETECTION MODELS IN BRAIN-COMPUTER INTERFACES .....	224
Alexandros Christopoulos, Matias Valdenegro-Toro, Andreea I. Sburlea	
DOI: 10.3217/978-3-99161-014-4-040	
41. ASSESSING CALIBRATION DURATIONS FOR C-VEP-BASED BCIS: INSIGHTS FROM NON-BINARY PATTERNS AND SPATIAL FREQUENCY VARIATIONS .....	230
Víctor Martínez-Cagigal, Álvaro Fernández-Rodríguez, Eduardo Santamaría-Vázquez, Ana Martín-Fernández, Roberto Hornero	
DOI: 10.3217/978-3-99161-014-4-041	
42. IMPROVED MOTOR IMAGERY DECODING WITH SPATIOTEMPORAL FILTERING BASED ON BETA BURST KERNELS.....	236
Sotirios Papadopoulos, Ludovic Darnet, Maciej J. Szul, Marco Congedo, James J. Bonaiuto, Jérémie Mattout	

## Table of Contents VI

---

DOI: 10.3217/978-3-99161-014-4-042

43. EEG-BASED STIMULUS CLASSIFICATION IN A FULL-BODY MOVEMENT, VIRTUAL REALITY PARADIGM ..... 243  
Lea Rabe, Yanzhao Pan, Marius Klug  
DOI: 10.3217/978-3-99161-014-4-043
44. EEG MARKERS OF ACCELERATION PERCEPTION IN VIRTUAL REALITY ..... 249  
Gael Van Der Lee, Anatole Lecuyer, François Cabestaing, Hakim Si-Mohammed  
DOI: 10.3217/978-3-99161-014-4-044
45. EYE-TRACKING AND SKIN CONDUCTANCE TO MONITOR TASK ENGAGEMENT DURING NEUROFEEDBACK SESSIONS ..... 254  
Agustina Fragueiro, Rene-Paul Debroize, Elise Bannier, Claire Cury  
DOI: 10.3217/978-3-99161-014-4-045
46. RELIABILITY OF INDIVIDUAL TASK-RELATED FRONTAL-MIDLINE-THETA FREQUENCY FOR NEUROFEEDBACK TRAINING ..... 260  
Maria Pfeiffer, Eva Masson, Andrea Kübler, Johannes Rodrigues  
DOI: 10.3217/978-3-99161-014-4-046
47. MOTOR IMAGERY VIVIDNESS AND NATURALISTIC INNER SPEECH HABITS IN SPEECH IMAGERY CLASSIFICATION ..... 266  
Manuel Hons, Silvia Erika Kober, Selina Christin Wriessnegger, Guilherme Wood  
DOI: 10.3217/978-3-99161-014-4-047
48. REVIRE: A VIRTUAL REALITY PLATFORM FOR BCI-BASED MOTOR REHABILITATION.. ..... 272  
Lucija Mihić Zidar, Philipp Raggam, Farhad Mohammadian, Aneta Barłoga, Moritz Grosse-Wentrup  
DOI: 10.3217/978-3-99161-014-4-048
49. WHICH IMAGINED SENSATIONS MOSTLY IMPACT ELECTROPHYSIOLOGICAL ACTIVITY ? ..... 278  
Emile Savalle, François Le Jeune, Léa Driessens, Marc J-M. Macé, Léa Pillette  
DOI: 10.3217/978-3-99161-014-4-049
50. ONLINE DETECTION OF EPILEPTIC SPIKES FOR USE IN EPILEPSY MONITORING .... 284  
Mostafa Mohammadpour, Christoph Kapeller, Kyosuke Kamada, Josef Scharinger, Michael Schwarzgruber, Milena Korostenskaja, Christoph Guger  
DOI: 10.3217/978-3-99161-014-4-050
51. COMPARISON OF CNN-BASED EEG CLASSIFICATION IN SENSOR AND SOURCE SPACE ..... 289
-

## Table of Contents VII

---

Magdalena Maurer, Daniel Baumgarten, Johannes Vorwerk  
DOI: 10.3217/978-3-99161-014-4-051

52. IMPACT OF MENTAL FATIGUE ON REGAINING MOTOR FUNCTIONALITY: A PRELIMINARY EEG STUDY ON STROKE SURVIVORS ..... 295  
Kaniska Samanta, Kongfatt Wong-Lin, Girijesh Prasad, Saugat Bhattacharyya  
DOI: 10.3217/978-3-99161-014-4-052
53. MAPPING NEUROMUSCULAR REPRESENTATION OF GRASPING MOVEMENTS USING ULTRA-HIGH-DENSITY EEG AND EMG ..... 301  
Leonhard Schreiner, Pauline Schomaker, Sebastian Sieghartsleitner, Michael Schwarzgruber, Harald Pretl, Andreea I. Sburlea, Christoph Guger  
DOI: 10.3217/978-3-99161-014-4-053
54. DECODING MORAL JUDGEMENT FROM TEXT: A PILOT STUDY ..... 307  
Diana E. Gherman, Thorsten O. Zander  
DOI: 10.3217/978-3-99161-014-4-054
55. REAL-TIME NEUROFEEDBACK ON INTER-BRAIN SYNCHRONY: CURRENT STATES AND PERSPECTIVES..... 313  
Kyungho Won, Léa Pillette, Marc J-M. Macé, Anatole Lécuyer  
DOI: 10.3217/978-3-99161-014-4-055
56. TO REPEAT OR NOT TO REPEAT? ERP-BASED ASSESSMENT OF THE LEVEL OF CONSCIOUSNESS - A CASE STUDY ..... 319  
Sebastian Halder, Ana Matran-Fernandez, Rab Nawaz, Marina Lopes da Silva, Tommaso Bertoni, Jean-Paul Noel, Jane Jöhr, Andrea Serino, Karin Diserens, Reinhold Scherer, Serafeim Perdakis  
DOI: 10.3217/978-3-99161-014-4-056
57. EXPLORING NEW TERRITORY: CALIBRATION-FREE DECODING FOR C-VEP BCI ..... 325  
Jordy Thielen, Jan Sosulski, Michael Tangermann  
DOI: 10.3217/978-3-99161-014-4-057
58. MACHINE LEARNING-BASED IDENTIFICATION OF TES-TREATMENT NEUROCORRELATES ..... 331  
Pasquale Arpaia, Lidia Ammendola, Maria Cropano, Matteo De Luca, Anna Della Calce, Ludovica Gargiulo, Giacomo Lus, Luigi Maffei, Daniela Malangone, Nicola Moccaldi, Simona Raimo, Elisabetta Signoriello, Paolo De Blasiis  
DOI: 10.3217/978-3-99161-014-4-058
59. TOWARDS AUDITORY ATTENTION DECODING WITH NOISE-TAGGING: A PILOT STUDY ..... 337  
Hanneke Scheppink, Sara Ahmadi, Peter Desain, Michael Tangermann, Jordy Thielen
-

## Table of Contents VIII

---

DOI: 10.3217/978-3-99161-014-4-059

60. TOWARDS GAZE-INDEPENDENT C-VEP BCI: A PILOT STUDY.....	343
Shekhar Narayanan, Sara Ahmadi, Peter Desain, Jordy Thielen	
DOI: 10.3217/978-3-99161-014-4-060	
61. APPROXIMATE UMAP ALLOWS FOR HIGH-RATE ONLINE VISUALIZATION OF HIGH-DI- MENSIONAL DATA STREAMS .....	349
Peter Wassenaar, Pierre Guetschel, Michael Tangermann	
DOI: 10.3217/978-3-99161-014-4-061	
62. ANALYSIS OF THE EEG RESTING-STATE SIGNALS FOR BCI .....	355
Enrico Mattei, Daniele Lozzi, Alessandro Di Matteo, Costanzo Manes, Filippo Mignosi, Matteo Polsinelli, Giuseppe Placidi	
DOI: 10.3217/978-3-99161-014-4-062	
63. AN ALTERNATIVE TRAINING PROTOCOL FOR A MOTOR IMAGERY BMI BASED ON A COLLABORATIVE APPROACH .....	360
Alessio Palatella, Paolo Forin, Stefano Tortora, Emanuele Menegatti, Luca Tonin	
DOI: 10.3217/978-3-99161-014-4-063	
64. THE CHALLENGE OF DRIVING BCI WITH EMOTIONAL SIGNALS COLLECTED BY EEG .....	366
Daniele Lozzi, Enrico Mattei, Roberta Ciuffini, Alessandro Di Matteo, Alfonso Marrelli, Raffaele Ornello, Matteo Polsinelli, Chiara Rosignoli, Simona Sacco, Giuseppe Placidi	
DOI: 10.3217/978-3-99161-014-4-064	
65. PROJECT NAFAS: ANNOUNCEMENT AND BRIEF OVERVIEW.....	372
Laurens R. Krol, Thorsten O. Zander	
DOI: 10.3217/978-3-99161-014-4-065	
66. NEUROFEEDBACK PERFORMANCE UNDER CHALLENGING CONDITIONS: THE THETA- AGENCY INTERPLAY .....	375
Claire Dussard, Léa Pillette, Cassandra Dumas, Laurent Hugueville, Brian Lau, Camille Jeunet-Kelway, Nathalie George	
DOI: 10.3217/978-3-99161-014-4-066	
67. PERIPHERAL NERVE STIMULATION AND AUDITORY SIMULATION CLOSED LOOP SYS- TEM FOR SENSORY DECISION MAKING IN TRANSHUMERAL AMPUTEES.....	381
Gurgen Soghoyan, Artur Biktimirov, Nikita Piliugin, Mikhail Sintsov, Mikhail Lebedev	
DOI: 10.3217/978-3-99161-014-4-067	
68. VALIDATING NEUROPHYSIOLOGICAL PREDICTORS OF BCI PERFORMANCE ON A LARGE OPEN SOURCE DATASET .....	385

## Table of Contents IX

---

David Trocellier, Bernard N'Kaoua, Fabien Lotte  
DOI: 10.3217/978-3-99161-014-4-068

69. THE GOOD, THE BAD, AND THE UGLY OF IEEG SIGNALS: IDENTIFYING ARTIFACTUAL CHANNELS USING CONVOLUTIONAL NEURAL NETWORKS ..... 391  
Zachary Freudenburg, Walker Zhong, Mariana P. Branco, Nick F. Ramsey  
DOI: 10.3217/978-3-99161-014-4-069
70. INTRODUCING THE USE OF THERMAL NEUROFEEDBACK ..... 397  
François Le Jeune, Emile Savalle, Anatole Lécuyer, Marc J-M Macé, Pierre Maurel, Léa Pillette  
DOI: 10.3217/978-3-99161-014-4-070
71. DOUBLE-BLIND AND SHAM-CONTROLLED AUGMENTED REALITY EEG-NEUROFEEDBACK STUDY ..... 403  
Lisa Berger, Guilherme Wood, Silvia Erika Kober  
DOI: 10.3217/978-3-99161-014-4-071
72. INTER-TASK TRANSFER LEARNING BETWEEN UPPER-LIMB MOTOR EXECUTION AND MOTOR IMAGERY ..... 409  
Sergio Pérez-Velasco, Diego Marcos-Martínez, Eduardo Santamaría-Vázquez, Víctor Martínez-Cagigal, Beatriz Pascual-Roa, Roberto Hornero  
DOI: 10.3217/978-3-99161-014-4-072
73. PURSUING THE IMPLEMENTATION OF A NEUROTUTOR: AN EEG-BASED CLASSIFICATION OF READING TYPES..... 414  
Héctor Romero-Morales, Jenny Noemí Muñoz-Montes de Oca, Alejandro Antonio Torres-García, Luis Villaseñor-Pineda  
DOI: 10.3217/978-3-99161-014-4-073
74. ASSESSMENT OF SEVERAL EEG ACTIVE PARADIGMS IN LOCKED-IN SYNDROME ... 420  
Perrine Rose Seguin, Emmanuel Maby, Romain Bouet, Lucie Gattaz, Ambre Querry, Lorianna Rizzo, Alessandro Farne, Jérémie Mattout  
DOI: 10.3217/978-3-99161-014-4-074
75. AN ONLINE SPIKE DETECTION AND MONITORING FRAMEWORK IN IEEG RECORDED USING BRAIN INTERCHANGE DEVICE ..... 425  
Behrang Fazli Besheli, Amir Hossein Ayyoubi, Jhan L. Okkabaz, Chandra Prakash Swamy, Michael Quach, Kai J. Miller, Gregory Worrell, Nuri F. Ince  
DOI: 10.3217/978-3-99161-014-4-075
76. ENHANCING MOTOR IMAGERY BCI CLASSIFICATION WITH BLOCK-TOEPLITZ AUGMENTED COVARIANCE MATRICES AND SIEGEL METRIC..... 431  
Igor Carrara, Theodore Papadopoulos

## Table of Contents X

---

DOI: 10.3217/978-3-99161-014-4-076

77. SYNTHESIZING EEG SIGNALS FROM EVENT-RELATED POTENTIAL PARADIGMS WITH  
CONDITIONAL DIFFUSION MODELS..... 437  
Guido Klein, Pierre Guetschel, Gianluigi Silvestri, Michael Tangermann  
DOI: 10.3217/978-3-99161-014-4-077
78. EEG SINGLE-TRIAL DECODING OF VISUAL ART PREFERENCE..... 443  
Marc Welter, Jesus Casal Martinez, Erin Redmond, Jonathan Baum, Tomas Ward, Fabien  
Lotte  
DOI: 10.3217/978-3-99161-014-4-078
79. EXPLORING EOG MARKERS OF FATIGUE DURING MOTOR IMAGERY BCI USE ..... 448  
Pauline Dreyer, Aline Roc, David Trocellier, Marc Welter, Raphaëlle N. Roy, Fabien Lotte  
DOI: 10.3217/978-3-99161-014-4-079
80. SHOULD ATTEMPTED MOVEMENTS REPLACE MOTOR IMAGERY IN BCI? THE ISSUE OF  
COMPATIBILITY WITH GAZE USE ..... 454  
Sergei L. Shishkin, Artem S. Yashin, Yulia G. Shevtsova, Anatoly N. Vasilyev  
DOI: 10.3217/978-3-99161-014-4-080
81. USING A CNN-LSTM ARCHITECTURE WITH DATA AUGMENTATION TO IMPROVE HD-  
ECOG SPOKEN SYLLABLE CLASSIFICATION ..... 460  
Mehdi Javani Mirehkoohi, Zachary Freudenburg, Amira Neumann, Nick F. Ramsey  
DOI: 10.3217/978-3-99161-014-4-081
82. NEURAL CORRELATES OF EXPERTISE DURING KINESTHETIC MOTOR IMAGERY:  
SHOULD WE REWARD MAXIMUM SMR-ERD? ..... 466  
Margaux Izac, Eléa Rossignol, Emeline Pierrieau, Natalia Grechukhin, Elina Coudroy, Bernard  
N'Kaoua, Léa Pillette, Camille Jeunet-Kelway  
DOI: 10.3217/978-3-99161-014-4-082
83. BAYESIAN MODEL OF INDIVIDUAL LEARNING TO CONTROL A MOTOR IMAGERY BCI  
.....472  
Côme Annicchiarico, Jérémie Mattout, Fabien Lotte  
DOI: 10.3217/978-3-99161-014-4-083
84. USING TRANSFORMER NETWORKS FOR STREAMING SPEECH SYNTHESIS FROM IN-  
TRACRANIAL EEG ..... 478  
Joaquín Amigó-Vega, Maxime Verwoert, Maarten C. Ottenhoff, Pieter L. Kubben, Christian  
Herff  
DOI: 10.3217/978-3-99161-014-4-084

## Table of Contents XI

---

85. NEUROPHONE: REAL-TIME BRAIN-MOBILE PHONE INTERFACE .....	484
Norhan Abdelhafez, Manal Tantawy, Abdelrahman Sayed, Nora Ekramy, Mohammed Nour-Eldin	
DOI: 10.3217/978-3-99161-014-4-085	
86. NOVEL MATERIALS FOR BRAIN COMPUTER INTERFACES: PERSPECTIVES AND ASPECTS OF COMBINATION OF A MAGNETOELECTRIC STIMULATOR AND A GRAPHENE MICROTRANSISTOR ARRAY RECORDING SYSTEM .....	489
Stratis Matsoukis, Josef Scharinger, Joana Covelo, Nathalia Cancino-Fuentes, Maria V. Sanchez-Vives, Guenter Edlinger, Christoph Guger	
DOI: 10.3217/978-3-99161-014-4-086	
87. COMPARING FINGERS AND GESTURES FOR BCI CONTROL USING AN OPTIMIZED CLASSICAL MACHINE LEARNING DECODER .....	495
Dirk Keller, Mariska Vansteensel, Siamak Mehrkanon, Mariana P. Branco	
DOI: 10.3217/978-3-99161-014-4-087	
88. REFERENCING SCHEMES AND THEIR EFFECT ON OSCILLATIONS AND BROADBAND POWER SPECTRAL SHIFTS IN STEREOELECTROENCEPHALOGRAPHY .....	501
Alex Rockhill, Michael A. Jensen, Nicole Swann, Ahmed Raslan, Dora Hermes, Kai J. Miller	
DOI: 10.3217/978-3-99161-014-4-088	
89. FUNCTIONAL REPRESENTATION OF SOMATOSENSORY, VISUAL, AND REINFORCEMENT PROCESSING ON THE CANINE BRAIN SURFACE .....	507
Frederik Lampert, Filip Mivalt, Inyong Kim, Nuri F. Ince, Jiwon Kim, Jhan L. Okkabaz, Max A. van den Boom, Vaclav Kremen, Rushna Ali, Volker Coenen, Gerwin Schalk, Peter Brunner, Gregory A. Worrell, Kai J. Miller	
DOI: 10.3217/978-3-99161-014-4-089	
90. MOVEMENT ASSOCIATED INCREASE IN THALAMIC BROADBAND SPECTRAL POWER IS A POTENTIAL FEATURE FOR BCI CONTROL.....	513
Bryan T. Klassen, Matthew R. Baker, Gabriela Ojeda Valencia, Kai J. Miller	
DOI: 10.3217/978-3-99161-014-4-090	
91. DYNAMIC SUPPRESSION OF THE CORTEX THROUGH SYNCHRONISATION DURING BRAIN COMPUTER INTERFACING .....	518
Fiona Permezel, Michael A. Jensen, Dora Hermes, Kai J. Miller	
DOI: 10.3217/978-3-99161-014-4-091	
92. WAVELET PACKET DECOMPOSITION TO EXTRACT FREQUENCY FEATURES FROM SPEECH IMAGERY .....	525
Alberto Tates, Ana Matran-Fernandez, Sebastian Halder, Ian Daly	
DOI: 10.3217/978-3-99161-014-4-092	

## Table of Contents XII

---

93. EEG CORRELATES OF ERROR-RELATED ACTIVITY DURING BALLISTIC COMPUTER MOUSE MOVEMENTS .....	531
Idorenyin Amaunam, Mushfika Sultana, Borja Rodriguez-Herreros, Tej Tadi, Robert Leeb, Ser- afeim Perdakis	
DOI: 10.3217/978-3-99161-014-4-093	
94. FROM CUE-BASED TO SELF-PACED MOVEMENT DETECTION: INFLUENCE OF THE CUE ON TRAINING DATA .....	537
Patrick Suwandjieff, Gernot R. Müller-Putz	
DOI: 10.3217/978-3-99161-014-4-094	
95. TOWARDS A MODEL-BASED PERSONALIZATION APPROACH FOR DRIVING A BCI ...	543
Adyasha Dash, Selina Christin Wriessnegger	
DOI: 10.3217/978-3-99161-014-4-095	

# WORD PREDICTION DURING NATURALISTIC SPEECH PERCEPTION

P.A. Lekhnitskaya<sup>1</sup>

<sup>1</sup> Kazan (Volga region) Federal University, Kazan, Russia

E-mail: lekhnitskaya.polina@gmail.com

**ABSTRACT:** The mechanisms of word prediction have not been studied in the natural speech perception paradigm, which formed the aim of the study: to explore the connection between the function of the EEG responses and the omitted words during naturalistic speech perception, confidence score of trained language model. 14 neurotypical subjects (mean age - 23,5 years; 5 males) participated in the research. EEG included 24 channels. It was proposed to listen to the story and comprehend it. The obtained results show differences in listening to omitted and non-omitted words in T3, T5, P3 electrodes. For modelling the connection between neural signals and naturalistic speech stimuli, mTRF was applied. One of the possible future directions of the research is to explore the communication processes in this paradigm.

## INTRODUCTION

The human brain is a complex dynamical system that continuously processes the input information. For acoustic stimuli, as with other types of sensory information, it is important to distinguish signal from noise; and by understanding the features of signal, a person can easily perceive the speech. In recent years, researchers have started to shift their attention to the use of continuous, natural speech to explore the ways the brain assesses auditory stimuli [3]. One of the possible approaches, known as system identification, is to model the obtained data based on the speech stimuli [3]. In this vein, the brain is treated as a "black box", in which there are some mappings between the features of the input speech and neurophysiological responses. Such a black box may be represented as a linear time-invariant system with obtaining a so-called temporal response function (TRF) by the connections between EEG and both acoustic and linguistic features [3].

To the best of our knowledge, the mechanisms of word prediction have not been studied in this paradigm. During speech perception, words are embedded in a broader context which facilitates meaning interpretation. Recipients can also make predictions about specific lexemes that can appear in the upcoming discourse. This task is similar to masked language modeling, where a pre-trained model predicts a masked token in a sentence (usually it is marked as [MASK]), by attending to tokens bidirectionally. In this case, a model also makes predictions about the word by its context [7]. Now the neuroscience of perception and

language widely uses an integrative modeling approach in which computation and brain function are reflected in computational models [6]. Moreover, direct evidence for alpha, beta and gamma bandwidth in predictive coding is accumulated from observations of increased gamma power to stimuli with prediction errors, but differences in these rhythms with stimulus predictability are not well known [1]. However, the brain responses and the possible link between their reactions and the reactions of trained language models in word prediction have not been simulated.

Thus, the aim of the following study is to explore the connection between the function of the EEG responses during naturalistic speech perception, confidence score of trained language model. It is hypothesized that the link between EEG signals and a trained language model naturalistic speech perception exists. The expected outcome is the approximation of the mentioned link.

## MATERIALS AND METHODS

14 neurotypical subjects (mean age - 23,5 years; 5 males) participated in the research. EEG recording was performed by a portable neuro-headset Mitsar-EEG-SmartBCI (Mitsar LLC, St. Petersburg) in a soundproofed room shielded from electromagnetic fields. EEG included 24 channels, in the international 10–20 system; impedance devices are maintained at a level below 10 kOhm. The experiment was implemented in the NeuroBureau program. During experiment, it was proposed to listen to the story about cosmonauts (duration = 5 min 2 sec, language - Russian), in which 48 words in word combinations were omitted (content words without functional ones). The words were chosen randomly, the main criterion was compliance with the context. The task was to understand the whole story. After listening to the recording, subjects were asked to complete a test with questions about the content of the story. Since all participants completed this task without mistakes, we can say with some probability that the missing words were recovered correctly during listening.

The time periods with omitted words were taken by the duration of the omitted word, and the time periods with non-omitted words were taken by the making shift in one second. Tools from "MNE" Python library with integrated methods were used for following EEG analysis [10]. Data preprocessing included the

filtering, interpolation, artefact removal, re-referencing, and time frequency analysis (tfr\_morlet). The high-pass filter is at 1 Hz, low-pass filter is at 40 Hz. During the recording of the study, participants sat motionless with their eyes closed. Artifacts associated with minimal movement were removed using independent component analysis. For each participant was performed this sequence of actions, after that the result data was aggregated in one dataset.

Mann-Whitney U Test (w/ continuity correction) with Bonferroni correction for multiple comparisons and Machine learning classifiers ("Scikit-learn" Python library) were applied to explore differences in EEG responses to the text (TP) and omitted words (OW) comprehension. For training and testing the data was chosen randomly; the size of testing set = 0.3, the size of training one = 0.7 respectively. The preprocessing stage included only applying Standard Scaler for train and test data. The reported results were not cross-validated. The aim of applying binary classification and using so many different classifiers is to additionally prove found by Mann-Whitney U Test differences.

Spearman rank order correlations analysis was used to explore the link between separate electrodes. Next, the transformer python library was utilized (pipeline is fill-mask) [8] with a ruBert-base pre-trained model [9]. The omitted words of the text were marked by [MASK]. Separately, the score reflecting the model's confidence about the selected word was added to the new dataset. Next, the results obtained from the model with the predicted EEG responses were compared. Underlying the computational modeling framework, implemented in the language domain, is the idea that the pre-trained language model can serve as hypotheses of the computations conducted in the brain. Time domain data and other used frequency-band signals analysis is used to investigate the data in terms of complex reactions.

## RESULTS

The statistically significant differences were obtained in T3 ( $p = 0,00$ ,  $z = -13,97$ ), T5 ( $p = 0,00$ ,  $z = 17,47$ ), P3 ( $p = 0,02$ ,  $z = 10,91$ ) electrodes in EEG responses to TP and omitted words OW comprehension. Additionally, spearman rank order correlations in OW show connections between T3 and T5 electrodes ( $r = 0,60$ ,  $p = 0,00$ ). This first finding shaded light on what electrodes are informative in terms of omitted word prediction.

Next, to explore possibility of the clear distinction between EEG activity while TP and OW phases, machine learning algorithms ("Scikit-learn" Python library) were applied. Data was previously preprocessed by Standard Scaler. Random Forest Classifier, K-Neighbors Classifier, Gradient Boosting Classifier, Logistic Regression, Decision Tree Classifier, MLPClassifier, and Gaussian NB showed high accuracy results among phases (table 1).

Table 1: Machine learning classification results in distinguishing omitted (OW) and non-omitted (NW) words listening

Algorithm	Words	F-score	Accuracy
Random Forest Classifier	OW	.99	.99
	NW	.99	
K-Neighbors Classifier	OW	.99	.99
	NW	.99	
Gradient Boosting Classifier	OW	.98	.98
	NW	.98	
Logistic Regression	OW	.95	.95
	NW	.95	
Decision Tree Classifier	OW	.90	.90
	NW	.89	
MLPClassifier	OW	.90	.90
	NW	.89	
Gaussian NB	OW	.87	.85
	NW	.81	

As distinct differences were found, the next aim was to model EEG responses and by this model try to predict the omitted word. For this purpose, mTRF [2] was used as a forward or encoding model to predict brain responses as the weighted sum of various acoustic and linguistic speech features. Continuous data was analyzed by dividing it into delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (> 30 Hz) rhythms. Initially, the linguistic speech feature was the frequency of the audio. Correlation between actual and predicted response for delta rhythm  $r_{fwd}=0.51$ , alpha rhythm  $r_{fwd}=0.508$ , beta rhythm  $r_{fwd}=0.734$ , gamma rhythm  $r_{fwd}=0.786$ . The best results are beta and gamma rhythms (fig.2), for them predicted EEG responses were obtained. Although correlation results in this mTRF analysis realization does not have p-values to reveal the significance of findings, it gives possible the connection between EEG activity and the core input (audio story), approximation of the response.

Transformers Python library predicted omitted words and in the model's confidence score and EEG activity correlations were found (table 2).

Table 2: Spearman rank order correlation results between predicted activity in P3, T5 and T3 electrodes and token score given by masked language modeling model.

Rhythm	Electrode	r
Gamma	P3	-0.0005
	T5	-0.2530
	T3	0.1546
Beta	P3	0.1083
	T5	0.0639

Rhythm	Electrode	r
	T3	-0.3162*

\* - statistically significant effect with  $p < 0.05$  marked

Statistically significant correlation was observed in beta-rhythm T3 electrode ( $r = -0.3162$ ,  $p = 0.00$ ), but not in P3 ( $r = -0.0005$ ,  $p > 0.05$ ), T5 ( $r = -0.2530$ ,  $p > 0.05$ ), T3 ( $r = -0.1546$ ,  $p > 0.05$ ) gamma and P3 ( $r = -0.1083$ ,  $p > 0.05$ ), T5 ( $r = -0.0639$ ,  $p > 0.05$ ) beta electrodes. From the table 1 we can observe the negative connection between the language model confidence and modeled EEG human language processing.

## DISCUSSION

The purpose of this study was to investigate the possibility of connecting modeled EEG response and pre-trained language model to the word prediction task during naturalistic speech perception. For this purpose, the T3 electrode was the most informative, T5, P3 electrodes showed statistically significant differences.

The T3 electrode is proximate to BA44, which might be linked with the prediction of the functional elements (determiners, prepositions, morphological particles) retained within the stimuli [4]. Increased activity in the left inferior frontal gyrus (T3) has been also reported as a function of word integration in the syntactic context [4]. Furthermore, increased directed connectivity from BA44 (T3) to the posterior left middle temporal gyrus (T5 is near this zone) is observed when two-word phrases start with a function word compared to a non-predictive element, possibly reflecting the top-down transmission of a categorical expectation [4]. Machine learning results also reflect the clear difference between OW and TP trials.

For modelling the connection between neural signals and naturalistic speech stimuli, mTRF was applied. The obtained correlation between predicted and real responses denotes neural function, a generalization of the event potential obtained from averaging responses to repetitions of stimuli for continuous data. The proposed EEG model is able to accurately predict activity across neuronal populations in the human cortex during the processing of sentences with omitted words. The proposed idea is similar with the concept of predictive coding, which suggests that the brain has an internal world model. This model encodes causes of sensory inputs as parameters of a generative model. Determining which combination of the many possible causes best fits the current sensory data is achieved through a process of minimizing the error between the sensory data and the sensory inputs predicted by the expected causes [11]. Regarding the current study, the results obtained should be refined in the future to create a more accurate model.

The highest correlation between actual and predicted responses was obtained in beta and gamma rhythms. As the next step, we applied the transformers python library to a similar prediction task with marked

omitted words. The model confidence was compared with predicted EEG gamma and beta responses. The predicted EEG response was used to correlate with the language model instead of the actual EEG data to explore the quality of modelling and further possibility to apply predicted EEG response to language modelling domain. We assume that correlation in this case means that model results with prediction of EEG and transformers results have common base of the language perception, and such an approach could give fruitful direction for further investigation both cortical brain organization and the large language models domains.

Statistically significant, but not strong correlation was observed in the beta-rhythm T3 electrode. This is strong evidence for the involvement of beta oscillations across grammatical and semantic processing [5]. Power decreases in beta bandwidth occurring before speech onset within a picture naming task can be provoked by the semantic context provided by a preceding sentence [5]. In our study, we got a negative correlation between beta EEG response and confidence of the language model. The possible explanation is that more predictable words by the language model may be reflected in beta oscillations in modeled human EEG responses.

Such a match between modeled EEG human language processing and the language model may be the first step to creating a semantical network for speech rehabilitation among patients with some types of aphasia. In future it may be of interest to study the communication processes in the proposed paradigm. The main limitation of this research is the sample size.

## CONCLUSION

An attempt was made to explore the connection between the function of the EEG responses and the omitted words during naturalistic speech perception. The statistically significant differences were obtained in T3, T5, and P3 electrodes. Machine learning classification algorithms also show distinct differences in EEG signals during audio text comprehension. Anticipatory, likelihood-driven processes are to contribute to lexical, syntactic, and discourse processing, which were studied by mTRF method. We got the modeled brain responses for gamma and beta rhythms as the highest correlation was obtained. This model was compared with the language model. The obtained result may be regarded as the possible solution for developing a semantical network for speech rehabilitation among patients with some types of aphasia. One of the possible future directions of the research is to explore the communication processes in this paradigm and to increase the sample size.

## REFERENCES

- [1] Bastos, A. M., Lundqvist, M., Waite, A. S., Kopell, N., Miller, E. K. (2020). Layer and rhythm specificity for predictive routing. Proceedings of the

- National Academy of Sciences, 117(49), 31459-31469.
- [2] Bialas, Ole, Jin Dou, and Edmund C. Lalor. mTRFpy: A Python package for temporal response function analysis. *Journal of Open Source Software* 8.89 (2023): 5657.
- [3] Lindboom, Elsa, et al (2023). Incorporating models of subcortical processing improves the ability to predict EEG responses to natural speech. *Hearing Research* 433: 108767.
- [4] Maran, Matteo, et al (2022). Online neurostimulation of Broca's area does not interfere with syntactic predictions: A combined TMS-EEG approach to basic linguistic combination. *Frontiers in psychology* 13: 968836.
- [5] Scaltritti, M., Suitner, C., Peressotti, F. (2020). Language and motor processing in reading and typing: Insights from beta-frequency band power modulations. *Brain and Language*, 204, 104758. doi:10.1016/j.bandl.2020.104758
- [6] Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Fedorenko, E., et al (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45), e2105646118.
- [7] Sinha, K., Jia, R., Hupkes, D., Pineau, J., Williams, A., Kiela, D. (2021). Masked language modeling and the distributional hypothesis: Order word matters pre-training for little. *arXiv preprint arXiv:2104.06644*.
- [8] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, F., et al. (2020). Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- [9] Zmitrovich, D., Abramov, A., Kalmykov, A., Tikhonova, M., Taktasheva, E., Astafurov, D., et al. (2023). A family of pretrained transformer language models for Russian. *arXiv preprint arXiv:2309.10931*.
- [10] Gramfort A., Luessi M., Larson E., Engemann D.A., Strohmeier D., et al. (2013) MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, doi:10.3389/fnins.2013.00267
- [11] Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and cognition*, 112, 92-97.

# INVESTIGATING TEMPORAL VARIATIONS IN MRCPS AND THEIR INFLUENCE ON CLASSIFICATION: A 10-HOUR EEG STUDY

J. Egger<sup>1</sup>, K. Kostoglou<sup>1</sup>, G. R. Müller-Putz<sup>1,2</sup>

<sup>1</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>2</sup>BioTechMed Graz, Austria

E-mail: gernot.mueller@tugraz.at

**ABSTRACT:** Locked-in patients rely on stable performance of BCIs to provide them with a means of communication. To build a robust BCI, we demonstrate the need for adaptive decoding that accounts for temporal variations in electroencephalogram (EEG) dynamics. We analyzed six consecutive EEG sessions recorded between 2p.m. (afternoon) and 12a.m. (midnight) of 15 healthy participants engaged in a four-right-hand gesture task. We employed four-class classifiers trained on movement-related cortical potentials of different sessions and applied the decoders to the same session to evaluate the impact of temporal fluctuations in EEG on decoding capabilities. As a step towards adaptive decoding, we developed constantly updated classifiers by training on the most recently collected data and compared these to a stationary classifier trained once on the first session. Our findings revealed that temporal variations in EEG during movement tasks influence classification performance. In this context, we demonstrated that adaptive decoding provides a remedy to build a robust BCI usable for patients in the home-environment.

## INTRODUCTION

A brain-computer interface (BCI) is a system that establishes a means of communication between the human brain and external devices by capturing and interpreting bioelectrical signals such as non-invasive electroencephalography (EEG) or invasive electrocorticography (ECoG) that are modulated by the user's intention [1,2]. Such a BCI system provides an alternative way of communication for patients suffering from severe motor neuron disorders such as amyotrophic lateral sclerosis, trauma or stroke that risk losing complete muscle control and the ability to communicate while still being conscious leading to locked-in syndrome (LIS) [3,4]. The EU project INTRECOM aims for the development of a novel, fully implantable BCI technology to allow for real-time motor and speech decoding to provide LIS patients with a means of communication in the home environment. Communication enabled by motor decoding shall be realized by movement attempt and the usage of four to five different gestures for discrete cursor control to permit the selection of characters or words presented in matrix-format on a screen. In this study, the execution of four different right-hand gestures in healthy individuals

is investigated as a preliminary work towards decoding of movement-related cortical potentials (MRCPs) for a four-directions cursor control in a BCI system. A prerequisite for BCIs integrated into the home environment is the stable and robust performance that enables the user to interact with their surroundings whenever necessary, e.g., to call a caregiver. Variations in the EEG directly influence the performance of such BCI systems, thereby affecting the communication abilities of users dependent on these systems. Changes in concentration, attentiveness, motivation [5], and fatigue [6,7], or the influence of direct or indirect feedback [8,9], are possible factors contributing to alterations in EEG. Previous literature has reported temporal variations in the delta [10], theta, alpha and beta [11,12] frequency bands during resting states that follow a diurnal pattern. We hypothesize that such temporal alterations also manifest in EEG signals during movement tasks and furthermore influence decoding capabilities of BCI systems based on MRCPs. Adaptive decoding has proven to be a useful tool in the context of alterations in EEG due to various factors [8,9], therefore we introduce adaptive decoding to enhance the performance stability of the BCI system. In this paper, we aim to capture changes in movement-related EEG patterns throughout the day and night by recording six EEG sessions during gesture tasks at 2-hour intervals over a 10-hour period with fifteen healthy participants. Further, we demonstrate a preliminary approach towards adaptive EEG decoding by introducing a continuously adaptive classifier and hypothesize that decoders including most recent data for training purposes significantly outperform decoders that are not updated throughout the course of a day.

## MATERIALS AND METHODS

### A. EEG recordings throughout the day and night

We recruited twenty-two healthy, right-handed participants (13 female, 9 male) that agreed with the inclusion criteria targeting a narrow age group from 20 to 40 years and an early morning routine starting between 5a.m. and 7a.m. each day. Additionally, we focused on a stable sleeping pattern by excluding candidates regularly working night shifts or feeling a physical or psychological effect in the absence of caffeine for more than 24 hours. On the day of the measurement, participants arrived at the laboratory of the Institute of

Neural Engineering of Graz University of Technology at 12p.m. They were clarified about the study procedure, had the opportunity to ask questions, and then provided their written informed consent. The study was approved by the local ethics review board. Subsequently, we equipped every participant with an EEG cap holding 60 active, gel-based electrodes (actiCAP Brain Products GmbH, Germany) according to the 10-10 international electrode standard setup. For simultaneous recording of EEG and electrooculogram (EOG), four additional active electrodes were positioned at the outer canthi of the eyes as well as on the inferior and superior of the left eye. The ground and reference electrode were positioned on the forehead at the position of FPz and the right mastoid, respectively. The signals were sampled at 500Hz and amplified using biosignal amplifiers (BrainAmp, Brain Products GmbH, Germany). To monitor hand movements, we used a motion capture system developed at the institute. A green marker was glued to the participant's right index finger, and a video camera recorded the movement at a sampling rate of 30Hz. Each participant performed six recording sessions every two hours starting at 2p.m. until 12a.m. on the measurement day, each one lasting approximately one hour. Between the recordings, the participants followed a strict experimental schedule and performed prespecified tasks that imitated a usual workday. These tasks involved demanding geometric and linguistic games during the first two breaks, followed by a standardized dinner after the third recording at 7p.m. During the last two breaks, participants were tasked with activities such as watching a documentary and listening to music to induce fatigue. At the beginning of each recording session, the electrode impedance was checked, and gel was applied if necessary. Then, the participant was asked to perform a psychomotor-vigilance task and answer questionnaires regarding emotions, hunger level, and tiredness' symptoms. Further, 2min of resting EEG were recorded. To remove eye artifacts, a 6-min EEG measurement was performed to simultaneously record EEG and EOG while the participant was asked to blink or move the eyes vertically or horizontally. After the main paradigm, another 2min of resting EEG were recorded. The main paradigm involved four right hand gestures (fist, pistol, pincer grasp and "Y"-gesture of the American sign language). Participants were seated in front of a computer screen positioned 50 to 60cm away, with their right hand on a table inside a wooden box equipped with the video camera. They were asked to follow on-screen instructions and to refrain from blinking and swallowing during each trial. The paradigm followed the procedure outlined by Patrick Ofner et al. [13]. Each trial began with a 1-s presentation of a class cue, including a fixation cross displayed after the cue for 0.5 to 1s. Participants were asked to focus on the fixation cross to avoid eye movements. A 2- to 3-s preparation period followed, during which a filled green circle shrank to match the inner white circle. Participants performed the instructed gesture when the circles overlapped and kept the position for about 3s until the screen went black, signifying the

end of a trial. A 1.5-s break between trials allowed participants to rest. The total trial duration ranged from 8 to 9.5s. Each participant performed 8 movement runs of approximately 5min each with a 30-s break in between. In total, 64 trials per gesture and session were recorded for each participant.

### B. Processing of recordings

The recorded signals were processed using MATLAB R2022b (Mathworks, Massachusetts, USA) and EEGLAB [14]. Initial steps included visual inspection, interpolation of noise-contaminated channels, and removal of 50Hz line noise and its first harmonic using a Butterworth bandstop filter of 2<sup>nd</sup> order. A Butterworth highpass filter of 5<sup>th</sup> order at 0.3Hz addressed the issue of drifts and a Butterworth lowpass filter of order 70<sup>th</sup> at 70Hz attenuated high-frequency noise. An eye artifact attenuation model was applied as described by Kobler et al. [15], and the most frontal electrodes were excluded. Pops and drifts were attenuated using the HEAR algorithm [16] and noisy temporal electrodes were removed. MRCPs were extracted using a Butterworth lowpass filter of 4<sup>th</sup> order at 3Hz. Movement-triggered epoching using the motion capture system produced 5.5s trials (-2.5s to 3s around movement onset). Trials exceeding a threshold of  $\pm 100\mu V$  were rejected, and the remaining trials were downsampled to 9Hz and re-referenced to a common average reference. To address the issue of unbalanced classes within each session, between sessions and subjects, the number of trials per gesture and session to include participants for further evaluation was set to 46 trials. Fifteen out of twenty-two participants fulfilled the criteria and were therefore included in subsequent analysis.

### C. Analysis of MRCPs

To evaluate significant changes in the MRCP shape, we employed a Wilcoxon rank sum test to compare the MRCP patterns from each session with session 6, which served as the reference. We combined trials of all four gestures across all participants. Statistical analysis was performed for each channel and each timepoint within a movement trial, therefore to correct for multiple comparisons, we applied the Benjamini and Hochberg [17] procedure that controls the false discovery rate and yields greater power than the commonly used Bonferroni technique [18].

### D. Classification of gestures

For classification of the four gestures, we employed a multiclass shrinkage linear discriminant analysis (sLDA) [19,20]. The input consisted of causal 1-s windows of all remaining electrodes that were shifted along movement trials at a sampling rate of 9Hz. Classification was performed offline on participants and sessions individually.

### E. Analysis of temporal changes in classification

To show whether potential temporal changes in the EEG during movement tasks affect decoding capabilities, we investigated the performance of five classifiers trained on each of the first five recording sessions and evaluated on the last (Fig. 1). First, we implemented a trial-based 5-fold cross-validation within each training/session set (Fig. 1) to see the general performance of the corresponding set (herein referred to as single session results). Then, as a second step, a classifier was trained on the whole training session and directly applied to session 6 recorded at 12a.m. This procedure was repeated for each one of the first five recording sessions and is outlined in Fig. 1.

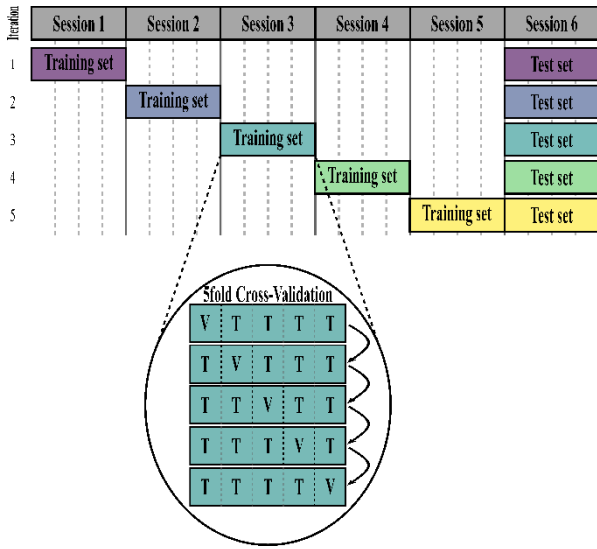


Figure 1: Classification procedure of single classifiers tested on session 6. Additionally, the trial based 5-fold cross-validation procedure for the single session results on session 3 is depicted.

### F. Comparison between adaptive and unrevised classification

As a preliminary step towards adaptive decoding, we investigated the difference in classification accuracy when employing an adaptive classifier in contrast to an unrevised decoder. Therefore, as indicated in Fig. 2, we shifted a window containing 46 trials per gesture across the six recording sessions that were used for training of the adaptive classifier. The subsequent 46 trials per gesture served as a test set. This procedure was performed in steps of one quarter of a session (12 trials), resulting in a total number of 17 trained classifiers along the duration of the study. For means of comparison, we implemented an unrevised classifier trained once on the very first window of 46 trials per gesture corresponding to the first recording session (see Fig. 2 as indicated in turquoise) that was further applied to every test set obtained in the previous approach. To assess whether the difference in decoding performance between the two classifiers was statistically

significant, we employed a Wilcoxon signed rank test on the classification accuracies obtained by every pair of classifiers. In order to correct for multiple comparisons (number of classifiers), we made use of the procedure developed by Benjamini and Hochberg [17].

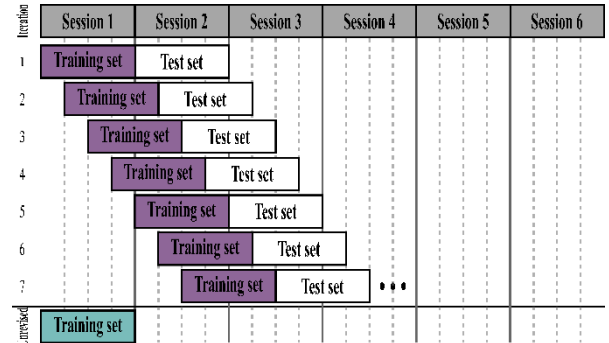


Figure 2: Adaptive (violet) and unrevised (turquoise) classification approach. As an example, only the first seven iterations of the adaptive classifiers are depicted. The test sets were the same for both classification approaches.

## RESULTS

### A. Analysis of MRCPs

In Fig. 3 we illustrate the temporal changes in MRCPs by depicting the averaged MRCPs across participants for measurement sessions 1 (at 2p.m.), 5 (at 10p.m.) and 6 (at 12a.m.), at electrode positions C1, Cz and C2 above the sensorimotor areas. For comparison purposes, session 6 served as a reference. Timepoints exhibiting significant ( $p < 0.05$ ) differences between the compared sessions are highlighted in color. As sessions 1 and 6 lie the furthest apart from each other, MRCPs of both sessions demonstrate greater difference in progression than MRCPs obtained during sessions 5 and 6.

### B. Analysis of temporal changes in classification

The classification results when investigating the impact of temporal EEG changes on movement classification performance can be seen in Fig. 4. Fig. 4a depicts the evolution of the cross-validated classification accuracies of the five decoders trained within different measurement sessions (Fig. 1). The temporal MRCP fluctuations were captured by the variation in maximum classification accuracy across time. The maximum accuracy at 2p.m. (session 1) increased from  $37.5\% \pm 5.6\%$  gradually to  $39.7\% \pm 3.2\%$  at 8p.m. and declined by 10p.m. (session 5) to  $37.4\% \pm 5.4\%$ . In comparison, Fig. 4b visualizes the performance of the five decoders when tested on the data of session 6. Apart from the decoder trained on session 4, recorded at 8p.m., which exhibited a decrease in accuracy ( $34.6\% \pm 5.1\%$ ) compared to the classifier trained on session 3 ( $36.3\% \pm 5.9\%$ ), we observed an increase in maximum classification accuracy as the time interval between training and test set recordings decreased.

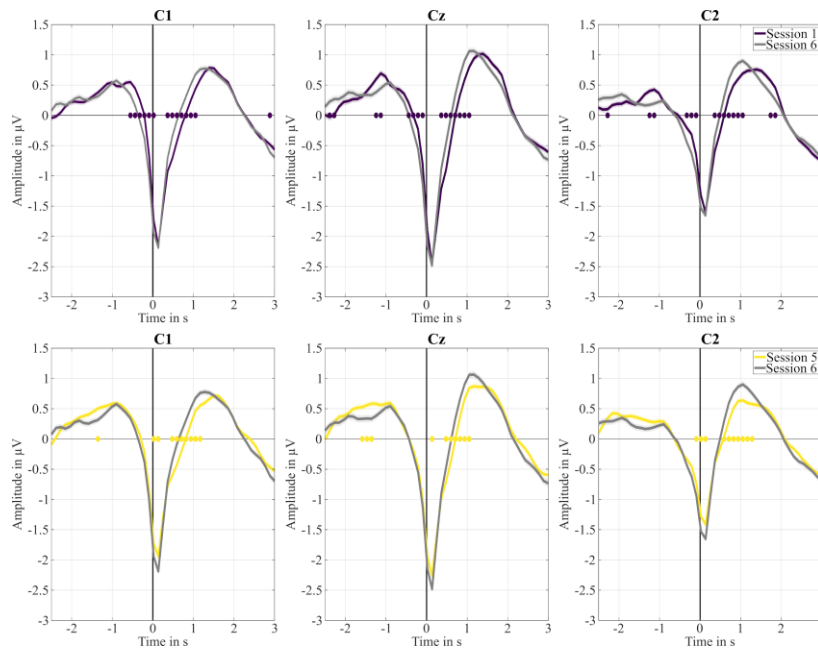


Figure 3: Average MRCPs across all participants for sessions 1 and 6 (top panel) and sessions 5 and 6 (bottom panel). The movement onset occurred at  $t=0$ s. Statistically significant differences ( $p < 0.05$ ) between sessions at each time point within a trial are indicated with color-coded dots on the zero-axis. In the top panel, we compared the MRCPs between session 1 and session 6. In the bottom panel, we compared the MRCPs between session 5 and session 6.

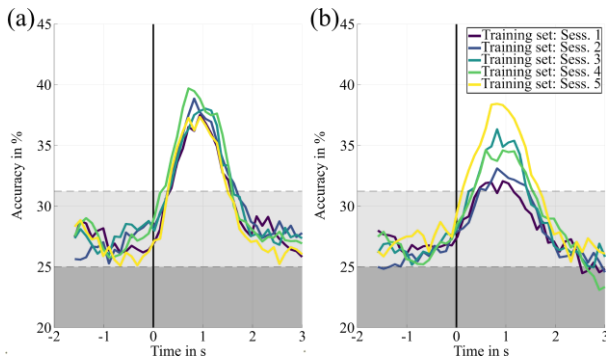


Figure 4: Classification results of different sessions. (a) Single session results. (b) Results of classification when tested on session 6. Indicated by the horizontal dashed lines are the theoretical chance level (25%) and the level of statistical significance (31.25%) as estimated using a permutation-based approach [21].

For example, the decoder trained on the first session achieved a maximum classification accuracy of  $32.1\% \pm 5.6\%$  whereas the classifier trained on the fifth session closest to session 6 used for testing yielded a maximum accuracy of  $38.4\% \pm 4.7\%$ .

### C. Comparison between adaptive and unrevised classification

Fig. 5 presents the variation in maximum classification accuracy across time for both the adaptive and unrevised classification model averaged across participants. In Fig. 5, one can observe that the adaptive decoder being trained on the most recent data outperforms the unrevised classifier which was kept constant throughout the process

at every shift along the time axis. This difference reaches statistical significance at some points, with a p-value less than 0.05.

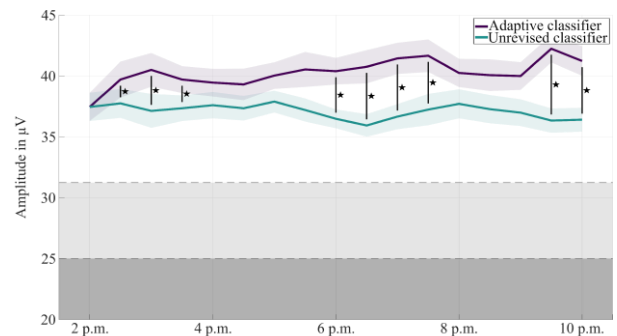


Figure 5: Comparison of the maximum classification accuracies obtained from both the adaptive (violet) and unrevised decoders (turquoise) shifted along the time axis. Depicted are the averages across participants ( $\pm$  standard error). The horizontal dashed line at 25 % indicates the theoretical chance level, the dashed line at 31.25% illustrates the level of statistical significance [21]. The seven vertical lines marked (\*) indicate statistical significance ( $p < 0.05$ ) differences between adaptive and unrevised decoder accuracies.

## DISCUSSION

We showed that throughout the day and night, MRCPs varied, hence movement classification performance was restricted, raising the necessity for adaptive classifiers that proved to outperform unrevised decoders. These findings are crucial for the development of BCI systems

used in the home-environment that need to be functioning at every day and nighttime to enhance the patient's independence.

#### A. Analysis of MRCPs

Analysis of MRCPs revealed that the frequency of timepoints exhibiting statistically significant deviations increased as more time elapsed between recording sessions. This was shown by comparing the MRCPs between sessions 1 (2p.m.) and 6 (12a.m.) and sessions 5 (10p.m.) and 6. Additionally, a variation in amplitude of MRCPs across time was observed. Session 5 showed a reduction in amplitude, especially highlighted by the statistically significant deviation at the timepoint of the motor potential when being compared to session 6. This change can be attributed to the increasing level of mental fatigue causing a decrease in amplitude of MRCPs [22]. Another factor influencing the amplitude of MRCPs is long-time training [23–25] meaning that experts require a reduced amount of effort, resulting in reduced activity at motor cortex sites involved in motor task preparation and execution. The long-time training effect observed in this study can be attributed to participants performing the same task repeatedly, hence leading to a decrease in MRCP amplitude over time. As this study was conducted in an open-loop manner, learning processes associated with controlling a BCI system could not be taken into account due to the absence of neurofeedback [26]. To account for the increase in MRCP amplitude observed during the transition from session 5 to session 6, previous studies have investigated the role of motivation [25,27]. It was shown that with rising levels of motivation accompanied by an increase in interest and excitement, P300 amplitudes increased. This phenomenon can also be observed in session 6, where the MRCP amplitude increases compared to session 5 possibly indicating the rise in motivation of participants to finish the last measurement. In general, we can eliminate the possibility of gel drying to be responsible for the observed variations in EEG dynamics as the gel was still wet after more than 12 hours when the cap was removed.

#### B. Analysis of temporal changes in classification

As described previously, the variations in classification accuracy across classifiers for the validation set (see Fig. 4a) arise due to temporal variation in the EEG dynamics during movement tasks. Recordings that are chronologically closer together exhibit less variability in terms of MRCP patterns than recordings that have a longer time interval between them. Therefore, as depicted in Fig. 4b, the classifier trained on session 5 at 10p.m. performs the best on the data recorded at 12a.m. in contrast to the other decoders trained on other sessions. These findings strongly emphasize the importance of adaptive decoding in the context of robust and stable performance of BCIs at all times.

#### C. Comparison between adaptive and unrevised classification

The maximum classification accuracies of the adaptive classifiers evaluated on temporally shifted test sets consistently outperformed the unrevised classifier at every time step. This superiority arises from the influence of MRCPs on decoding capabilities, and as these signals fluctuate over time, a classifier trained only once is incapable of capturing the evolving temporal dynamics inherent in EEG signals. Conversely, when constructing a classifier that incorporates the most recent data for training, a noticeable improvement in classification is observed. This underscores the positive impact of adaptive decoding on overall classification performance.

### CONCLUSION

In this preliminary work towards adaptive decoding for temporal dynamics in EEG signals, we showed that due to changes of MRCPs across time decoding needs to adapt to build a robust and stable BCI system that delivers reliable output for patients in their home-environment. We demonstrated that the usage of most recently collected data for means of training of a decoder significantly improved decoding performance. This paper using supervised adaptation which requires task labels as ground truth serves as preparatory work for future in-depth investigations regarding online adaptations of decoders. Since in real autonomous BCI use in the home-environment labels will not be available, unsupervised adaptation could be realized by a trail-wise update of the model's parameters, as proposed by Vidaurre et al. [28,29] or Hehenberger et al. [5].

### ACKNOWLEDGEMENTS

This project is funded by the European Union's HORIZON-EIC-2021-PATHFINDER CHALLENGES program under grant agreement No 101070939 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00198.

### REFERENCES

- [1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol.* 2002 Jun 1;113(6):767–91.
- [2] Wolpaw JR. Brain-computer interfaces. *Handb Clin Neurol.* 2013;110:67–74.
- [3] Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, et al. A spelling device for the paralysed. *Nature.* 1999 Mar 25;398(6725):297–8.
- [4] Chaudhary U, Vlachos I, Zimmermann JB, Espinosa A, Tonin A, Jaramillo-Gonzalez A, et al. Spelling interface using intracortical signals in a

- completely locked-in patient enabled via auditory neurofeedback training. *Nat Commun.* 2022 Mar 22;13(1):1236.
- [5] Hehenberger L, Kobler RJ, Lopes-Dias C, Srisrisawang N, Tumfart P, Uroko JB, et al. Long-Term Mutual Training for the CYBATHLON BCI Race With a Tetraplegic Pilot: A Case Study on Inter-Session Transfer and Intra-Session Adaptation. *Front Hum Neurosci.* 2021 Feb 26;15:635777.
  - [6] Millan JR. On the need for on-line learning in brain-computer interfaces. In: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat No04CH37541) [Internet]. IEEE; 2005. Available from: <http://ieeexplore.ieee.org/document/1381116/>
  - [7] Buttfeld A, Ferrez PW, Millán J del R. Towards a robust BCI: error potentials and online learning. *IEEE Trans Neural Syst Rehabil Eng.* 2006 Jun;14(2):164–8.
  - [8] Vidaurre C, Schlögl A, Cabeza R, Scherer R, Pfurtscheller G. A fully on-line adaptive BCI. *IEEE Trans Biomed Eng.* 2006 Jun;53(6):1214–9.
  - [9] Vidaurre C, Schlögl A, Cabeza R, Scherer R, Pfurtscheller G. Adaptive on-line classification for EEG-based brain computer interfaces with AAR parameters and band power estimates. *Biomed Tech.* 2005 Nov;50(11):350–4.
  - [10] Cacot P, Tesolin B, Sebban C. Diurnal variations of EEG power in healthy adults. *Electroencephalogr Clin Neurophysiol.* 1995 May;94(5):305–12.
  - [11] Cummings L, Dane A, Rhodes J, Lynch P, Hughes AM. Diurnal variation in the quantitative EEG in healthy adult volunteers. *Br J Clin Pharmacol.* 2000 Jul;50(1):21–6.
  - [12] Toth M, Kiss A, Kosztolanyi P, Kondakor I. Diurnal alterations of brain electrical activity in healthy adults: a LORETA study. *Brain Topogr.* 2007 Oct 11;20(2):63–76.
  - [13] Ofner P, Schwarz A, Pereira J, Wyss D, Wildburger R, Müller-Putz GR. Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury. *Sci Rep.* 2019 May 9;9(1):7134.
  - [14] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods.* 2004 Mar 15;134(1):9–21.
  - [15] Kobler RJ, Sburlea AI, Lopes-Dias C, Schwarz A, Hirata M, Müller-Putz GR. Corneo-retinal-dipole and eyelid-related eye artifacts can be corrected offline and online in electroencephalographic and magnetoencephalographic signals. *Neuroimage.* 2020 Sep;218:117000.
  - [16] Kobler RJ, Sburlea AI, Mondini V, Muller-Putz GR. HEAR to remove pops and drifts: the high-variance electrode artifact removal (HEAR) algorithm. *Conf Proc IEEE Eng Med Biol Soc.* 2019 Jul;2019:5150–5.
  - [17] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc.* 1995 Jan;57(1):289–300.
  - [18] Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J Educ Behav Stat.* 2002 Mar;27(1):77–83.
  - [19] Peck R, Van Ness J. The use of shrinkage estimators in linear discriminant analysis. *IEEE Trans Pattern Anal Mach Intell.* 1982 May;4(5):530–7.
  - [20] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components--a tutorial. *Neuroimage.* 2011 May 15;56(2):814–25.
  - [21] Mueller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random? A closer look on BCI results. *Aquat Microb Ecol.* 2008;10:52–5.
  - [22] Dirnberger G, Duregger C, Trettler E, Lindinger G, Lang W. Fatigue in a simple repetitive motor task: a combined electrophysiological and neuropsychological study. *Brain Res.* 2004 Nov 26;1028(1):26–30.
  - [23] Wright DJ, Holmes PS, Smith D. Using the movement-related cortical potential to study motor skill learning. *J Mot Behav.* 2011;43(3):193–201.
  - [24] Wright DJ, Holmes P, Di Russo F, Loporto M, Smith D. Reduced motor cortex activity during movement preparation following a period of motor skill practice. *PLoS One.* 2012 Dec 14;7(12):e51886.
  - [25] Baykara E, Ruf CA, Fioravanti C, Käthner I, Simon N, Kleih SC, et al. Effects of training and motivation on auditory P300 brain-computer interface performance. *Clin Neurophysiol.* 2016 Jan;127(1):379–87.
  - [26] Neuper C, Pfurtscheller G. Neurofeedback Training for BCI Control. In: *Brain-Computer Interfaces.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 65–78. (Frontiers Collection).
  - [27] Goldstein RZ, Cottone LA, Jia Z, Maloney T, Volkow ND, Squires NK. The effect of graded monetary reward on cognitive event-related potentials and behavior in young healthy adults. *Int J Psychophysiol.* 2006 Nov;62(2):272–9.
  - [28] Vidaurre C, Kawanabe M, von Büna P, Blankertz B, Müller KR. Toward unsupervised adaptation of LDA for brain-computer interfaces. *IEEE Trans Biomed Eng.* 2011 Mar;58(3):587–97.
  - [29] Vidaurre C, Sannelli C, Müller KR, Blankertz B. Co-adaptive calibration to improve BCI efficiency. *J Neural Eng.* 2011 Apr;8(2):025009.

# S-JEPA: TOWARDS SEAMLESS CROSS-DATASET TRANSFER THROUGH DYNAMIC SPATIAL ATTENTION

Pierre Guetschel<sup>1</sup>, Thomas Moreau<sup>2</sup>, Michael Tangermann<sup>1</sup>

<sup>1</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

<sup>2</sup>Université Paris-Saclay, Inria, CEA, Palaiseau 91120, France

E-mail: pierre.guetschel@donders.ru.nl

**ABSTRACT:** Motivated by the challenge of seamless cross-dataset transfer in EEG signal processing, this article presents an exploratory study on the use of Joint Embedding Predictive Architectures (JEPAs). In recent years, self-supervised learning has emerged as a promising approach for transfer learning in various domains. However, its application to EEG signals remains largely unexplored. In this article, we introduce Signal-JEPA for representing EEG recordings which includes a novel domain-specific spatial block masking strategy and three novel architectures for downstream classification. The study is conducted on a 54 subjects dataset and the downstream performance of the models is evaluated on three different BCI paradigms: motor imagery, ERP and SSVEP. Our study provides preliminary evidence for the potential of JEPAs in EEG signal encoding. Notably, our results highlight the importance of spatial filtering for accurate downstream classification and reveal an influence of the length of the pre-training examples but not of the mask size on the downstream performance.

## INTRODUCTION

Electroencephalography (EEG) allows capturing neural activity directly from the scalp, offering a high temporal resolution signal to investigate brain functions. The interpretation of EEG signals with machine learning methods opens the door to build brain-computer interfaces (BCIs). Despite the potential of BCIs, their practical application is hindered by the intensive requirement for calibration data, which is both time-consuming and demanding for participants. Recently, transfer learning has been explored to mitigate the constraints imposed by data scarcity and calibration demands. Self-supervised learning (SSL) presents itself as a pivotal strategy to tackle transfer learning, enabling models to learn rich representations from unlabeled data that can be used to efficiently solve downstream tasks. A particular approach to SSL consists in strategically masking parts of the input data and training a model to predict these masked elements. A key advantage of masking-based SSL methods is that they can be applied to virtually any type of data.

*Masking Strategies* proved a determining factor for the success of SSL methods in various domains, including image [2–4], speech [2, 5], text [2], and video [1] pro-

cessing. For instance, *random masking* strategies, where the regions to be masked are sparsely selected, typically deliver inferior results compared to *block masking* strategies, where larger continuous regions are masked, requiring the model to gain a deeper understanding of the data distribution [2].

*Masked Autoencoders (MAEs)* stand as the quintessential entry point to masking-based SSL [3]. These models aim to reconstruct the masked sections of the input directly, and the training objective is computed by comparing the original input sections to the reconstructed ones. Unfortunately, reconstructing the input and comparing elements in the original space is not without challenges. When the original input space has a high dimensionality, the reconstruction can be computationally expensive, and necessitate the use of domain-specific constraints to produce valid signals. The original signals can also be noisy, increasing the difficulty of encoding meaningful parts of the signal. Moreover, the reconstruction's difficulty can vary significantly across different areas of the input data: with images, reconstructing a monochromatic, non-structured sky is less difficult than reconstructing a structured object like a hand. Such structural disparities are one of the reasons, why conventional metrics like mean-square error often fail to assess the reconstruction quality. These challenges question the scalability and adaptability of MAEs as a universal SSL approach.

*Joint-Embedding Predictive Architectures (JEPAs)* offer a promising alternative to address the limitations associated with direct reconstruction. JEPA-like methods avoid reconstructing the input in its original space and focus on predicting latent representations, or *embeddings*, of the data [4]. This approach confers two major benefits: first, it is computationally efficient, especially with high-dimensional input spaces as the embeddings can reduce the dimensionality; second, the metric's selection in the embedding space is less critical, as the embeddings are learned adaptively to the chosen metric. However, as the "ideal" embedding vectors are unknown, the reconstruction objective is undefined a priori. This challenge is addressed by constructing target embedding vectors during the training by using a bootstrapping procedure which will be further explained in the S-JEPA Framework section. The potential of JEPA-like frameworks has been highlighted by their promising results with images [2, 4],

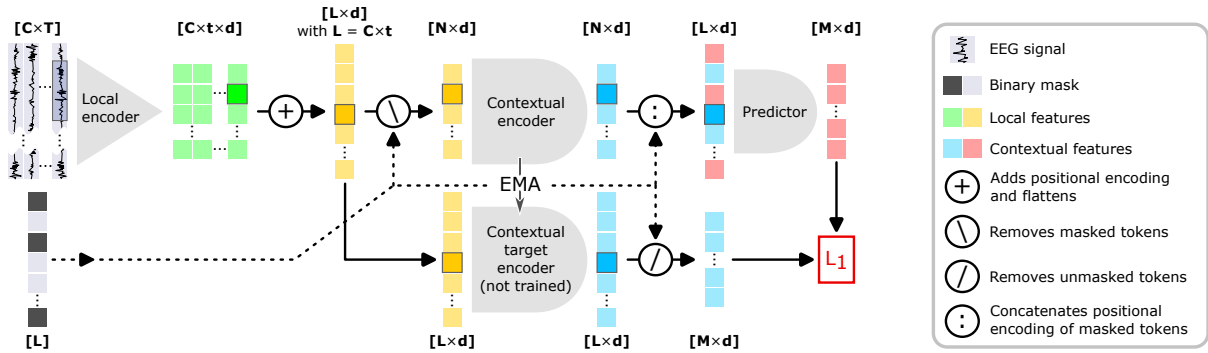


Figure 1: **S-JEPA** training procedure. The framework takes as input EEG recordings with  $C$  channels and  $T$  time samples, and binary masks of length  $L$ . First, the *Local encoder* independently transforms  $t$  windows from each channel into  $C \times t = L$  embedding vectors, called *tokens*, of dimensionality  $d$ . Then, the tokens are marked according to their originating channel and temporal position, and are flattened into a sequence of length  $L$ . Subsequently, only the unmasked tokens are passed to the *Contextual encoder*, while the full tokens sequence is given to the *Contextual target encoder* to generate training targets. Finally, the *Predictor* attempts to reconstruct the masked tokens and its predictions are compared with the targets using an  $L_1$  loss. During the optimisation, the parameters of the Contextual target encoder are not trained via gradient backpropagation but follow those of the Contextual encoder by Exponential Moving Average (EMA). Figure inspired from [1].

speech [2, 5], text [2], and videos [1].

*Applications to the EEG Domain* of masking-based SSL techniques have started to emerge. Pérez-Velasco and colleagues used a masked autoencoder approach, with random masking over the spatial and temporal dimensions [6]. Chien and colleagues experimented with MAEs with block masking over the temporal dimension [7]. Kostas et al. and Foumani et al. also used block masking over the temporal dimension but coupled it with JEPA-like training strategies [8, 9].

*Motivation.* Despite the advancements in applying SSL to EEG data, the exploration of block masking strategies over EEG channels remains uncharted territory. Such an approach holds the potential for developing robust channel attention mechanisms, and thus could facilitate dynamic spatial filtering. This capability could prove instrumental in adapting to recordings with varying channel sets, thereby facilitating cross-dataset transfer learning or tackling corrupted channels. This paper seeks to bridge this gap by investigating the implications of a channel-based block masking strategy on SSL efficacy in EEG signal processing.

The main application of models trained with SSL is the following fine-tuning on the actual task of interest, the so-called downstream task. While most research within the EEG domain has focused on fine-tuning for sleep stage classification tasks, their application to a BCI context remains largely untapped, with only two studies by Kostas et al. and Pérez-Velasco et al. exploring the impact of SSL on BCI tasks [6, 8]. As BCI systems suffer from data scarcity there is a consistent goal to minimize the amount of calibration data required before each online session, specifically as the calibration phase requires sustained attention from the participant and thus is tiring. This highlights a significant opportunity to explore the effectiveness of SSL models across various BCI paradigms, including but not limited to motor imagery protocols, thereby contributing to the broader understanding and ap-

plication of SSL in enhancing BCI performance.

Furthermore, the application of pre-trained SSL models for solving downstream tasks often involves the addition of a linear layer atop the embedding dimension. This practice, however, may not be optimal in high-dimensional embedding spaces. Through a comparative analysis of six different strategies for leveraging pre-trained architectures in downstream tasks, this work aims to support our understanding and application of SSL in EEG data processing.

*Research Questions and Plan.* This manuscript is an explorative study investigating what approaches should be adopted for training SSL algorithms on EEG signals, and what domain-specific considerations warrant attention. We approach this through three research questions: 1) What constitutes the most efficacious masking strategy for SSL when applied to EEG data? 2) How does the temporal length of examples used influence the SSL training process? 3) What fine-tuning strategies lead to the best downstream performance?

To answer these, we first propose a novel masking strategy as part of the Signal-based Joint-Embedding Predictive Architecture (S-JEPA) framework in the S-JEPA Framework section. Then, we introduce fine-tuning strategies tailored for S-JEPA in the Downstream Evaluation section. The datasets used for pre-training and for downstream evaluation are presented in the Datasets section. Finally, the Results and Discussion sections will report and critically discuss the outcomes and implications of the experiments we conducted.

## S-JEPA FRAMEWORK

The Signal-based Joint-Embedding Predictive Architecture (S-JEPA) framework is illustrated in Figure 1. It is used to pre-train models. Its architecture is inspired by BENDR and MAEEG, introduced in the pioneering studies by respectively Kostas et. al. [8] and Chien et. al. [7].

A key modification is introduced in the design of the local encoder to enable our novel masking strategy. This section details the architecture's components, the spatial masking strategy, and the training process.

**Local encoder.** The local encoder is implemented as a convolutional neural network (CNN) with five layers, each formed of a convolution with a Gaussian error linear unit (GELU) non-linearity. Contrary to the encoders in BENDR and MAEEG, which accept multi-channel input windows, our encoder processes windows from a singular channel. Each window is encoded into a 64-dimensional embedding vector, hereafter referred to as a *token*. The windows are 1.19 s long, with a stride of 1.0 s. The first convolutional kernel covers 0.25 s while the following ones simply combine feature vectors in pairs, i.e., both the kernel temporal lengths and strides are 2.

**Contextual encoder.** This encoder consists of a transformer architecture with eight layers, as introduced by Vaswani and colleagues [10]. It processes the unordered sequence of tokens generated by the local encoder, necessitating the addition of position-encoding information to localize them temporally and spatially. The temporal positioning of each token is defined using a cosine encoding [10] over the first 34 dimensions, whereas spatial positioning is achieved through trainable embeddings for each channel, initialized using cosine encoding based on their three-dimensional coordinates. The contextual encoder receives tokens that only contain local information; its role is to establish relationships between them.

**Spatial Block Masking strategy.** Unique to our methodology is the independent encoding of each channel by the local encoder. While it avoids the learning of spatial filters at early stages, it paves the way for innovative spatial dimension masking strategies. Literature in both, image processing [2] and EEG signal analysis [9] suggests that block masking yields superior results compared to random masking. Motivated by these findings, our novel approach extends block masking to the spatial domain of EEG channels.

Given the irregular distribution of EEG channels, the concept of a contiguous block of tokens does not translate di-

rectly from its application in images or temporal signals, where the pixels and time samples are regularly spaced. Thus our approach masks all channels within a predetermined radius of a randomly chosen central channel, as illustrated in Figure 2. In this work, we compare three mask sizes with diameters approx. 40 %, 60 % and 80 % of the head size. This strategy inherently introduces variability in the number of masked tokens.

**S-JEPA pre-training.** With the operational principles of both the local and contextual encoders established, we introduce two ancillary components exclusively utilized during the training phase. The first, termed the *Contextual target encoder*, is a non-trainable duplicate of the contextual encoder which serves to generate the training targets. Its parameters are updated via exponential moving average (EMA). The second component, the *Predictor*, is a transformer decoder architecture with four layers as delineated by Vaswani et al. [10]. The comprehensive training methodology is illustrated in Figure 1.

The models are trained until no improvement of the validation loss is observed for 10 epochs (i.e., complete pass through the entire dataset), at which point the best model is saved for subsequent fine-tuning.

## DOWNSTREAM EVALUATION

Upon completion of the network's pre-training through the SSL task, which holds no intrinsic value beyond training purposes, we proceed to evaluate its efficacy on practical *downstream* classification tasks which, in our case, are BCI tasks. This step is crucial for determining the real-world applicability of the pre-trained model.

**Downstream classification architectures.** Using pre-trained models for downstream classification tasks necessitates altering their architecture to allow predicting class probabilities. The most widely adopted modification, linear probing, consists of adding a linear classification layer directly above the embedding space [4]. However, the individual tokenization of each channel in our approach leads to a high-dimensional latent space, close to the dimensionality of the raw input examples, which would make linear probing inefficient. In response to this challenge, we enrich the architecture with two layers instead which are *Spatial aggregation* and *Fully-Connected*, as explained in Figure 3.

The integration of these layers is explored in three distinct configurations. (a) The **Contextual** downstream architecture places both layers after the contextual encoder as depicted in Figure 3a. (b) The **Post-local** downstream architecture discards the pre-trained contextual encoder and adds the novel layers atop the local encoder as shown in Figure 3b. (c) The **Pre-local** downstream architecture also discards the pre-trained contextual encoder but then places the spatial averaging layer *before* the local encoder as illustrated in Figure 3c. This third alternative allows the network to perform a spatial EEG filtering step, as commonly present in BCI architectures.

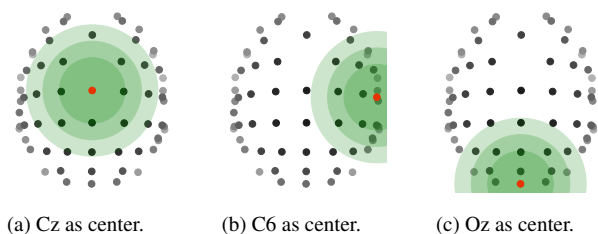


Figure 2: Visualisation of the **spatial block masking** strategy for three example mask centres (red electrodes). The dark to light green spheres represent masks of diameters 40 %, 60 % and 80 % of the head size, as used in our experiments. Assuming a top-down view upon the scalp, the depth of the electrodes is denoted by their intensity (black: close, grey: distant). For a given mask, all electrodes within the corresponding sphere are hidden from the contextual encoder and must be predicted by the predictor.

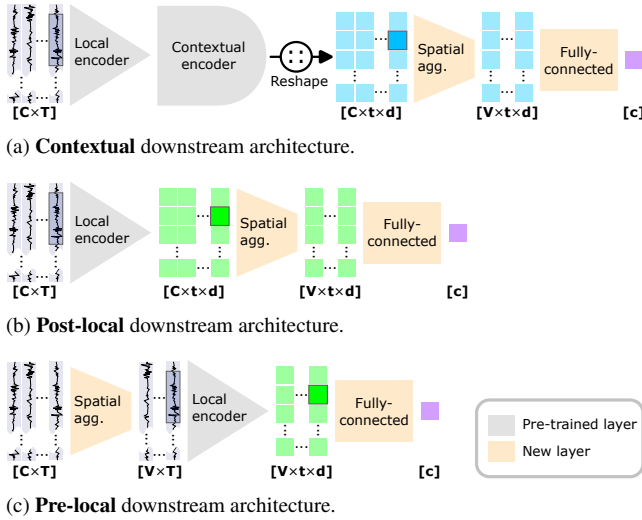


Figure 3: **Downstream classification architectures.** In each of the three alternative alterations of the pre-trained networks, two new layers are added. 1) *Spatial aggregation* is a convolutional layer that realizes weighted combinations of the elements in the channels dimension into  $V \ll C$  "virtual" channels. 2) *Fully-connected* is a linear layer that predicts  $c$  class probabilities.

**Fine-tuning.** We examine two distinct fine-tuning strategies for these downstream architectures. The first method, indicated by the prefix *new*-, involves exclusively training the newly introduced layers, keeping the pre-trained components frozen. The alternative strategy, denoted with the prefix *full*-, consists in fine-tuning the entire network. This second strategy starts with a warm-up phase of 10 epochs where only the newly added layers undergo training, preventing the deterioration of the pre-trained layers' performance due to irrelevant feedback [11], before the previously existing layers are included into the training.

For both strategies, the model is fine-tuned until no improvement of the validation loss is observed for 50 epochs, at which point the best model is restored for testing. Additionally, it should be noted that the temporal length of examples used during the fine-tuning phase is determined by the requirements of the downstream task, which is independent of the length of examples utilized during the SSL pre-training phase.

## DATASETS

For the exploratory investigation in this work, we used the dataset introduced by Lee and colleagues [12], subsequently referred to as the *lee2019* dataset. It contains EEG recordings from 54 subjects, each undergoing three distinct BCI paradigms: steady-state visual evoked potentials (SSVEP) with four classes, visual event-related potentials (ERP), and left vs. right hand motor imagery (MI). The recordings are from  $C = 62$  EEG channels, the spatial distribution of which is detailed in Figure 2. The dataset was loaded from the MOABB framework [13], bandpass filtered at 0.5 - 40 Hz and downsampled to 128 Hz.

We used the first 40 subjects to pre-train any model. The subsequent 7 subjects were used for the validation during this pre-training phase. The remaining 7 subjects were reserved for the downstream performance evaluation.

**Pre-training data.** SSL methods do not necessitate labels. As such, the training and validation examples are slices of the continuous recordings taken at a fixed interval of 16.9 seconds. This study compared three example lengths  $T$  approximately distributed on a logarithmic scale, namely 1, 4, and 16 seconds. All allocated subjects and all paradigms were used collectively during the pre-training phase which yielded a total of 36,576 training examples and 6,528 validation examples. No artefact rejection method was applied.

**Downstream evaluation data.** For each subject allocated for downstream evaluation and each paradigm, the fine-tuning performance of the different pre-trained models is assessed using a 5-folds within-subject stratified cross-validation procedure. The examples used are 4.19 seconds long for MI and SSVEP, and 1.19 seconds long for ERP. These lengths are chosen to respect the durations defined in the original dataset [12] while maintaining compatibility with our tokenization process.

## RESULTS

**Experimental details.** For pre-training, our setup compares all combinations of signal durations (1s, 4s, and 16s) and mask sizes (40 %, 60 %, and 80 % of head size), along with a *no-pre-training* baseline.

For downstream performance evaluation, we assess each pre-trained model using the three downstream architectures—*contextual*, *post-local*, and *pre-local*—and the two fine-tuning approaches—*full* and *new*. The *no-pre-training* is only assessed with the *full* fine-tuning approach. A given pairing of pre-training and fine-tuning configurations is referred to as a *pipeline*.

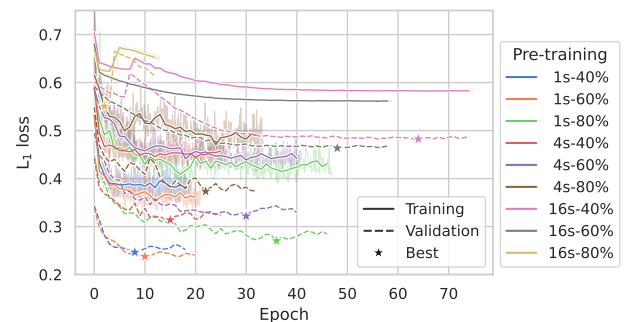


Figure 4: **Pre-training curves** of the different configurations tested. The solid and dashed lines indicate the loss on the training and validation sets. While the validation loss was tested once per epoch only, the training loss was logged after every optimisation step. The train loss on individual optimisation steps is visible in the background, corresponding epoch-wise averages are outlined in white. A star marks the lowest validation loss per curve, the early stopping time point and consequently the checkpoint from which any fine-tuning started.

**Pre-training dynamics.** The pre-training phase, see Figure 4, provides insights into the training process under the S-JEPA framework. It reveals that training curves under the 16s condition are significantly smoother than the other configurations. Additionally, the early stopping mechanism concludes the training of the 16s-80% configuration prematurely at 12 epochs due to an early trough in the loss curve. Similarly, the 16s-40% setup also encounters an early trough yet manages to recover and complete a longer training. The longest training durations are observed in the 16s-40% and 16s-60% configurations, enduring for 74 and 58 epochs, respectively, translating to approximately 12 and 10 hours of training.

**Pipelines ranking on downstream performances.** The comprehensive ranking of all tested pipelines is detailed in Figure 5, offering a comparative overview of the performance of the different combinations over the experimental protocols. Notably, the top-performing pipelines in the downstream tasks are the 16s-60% and 16s-40%

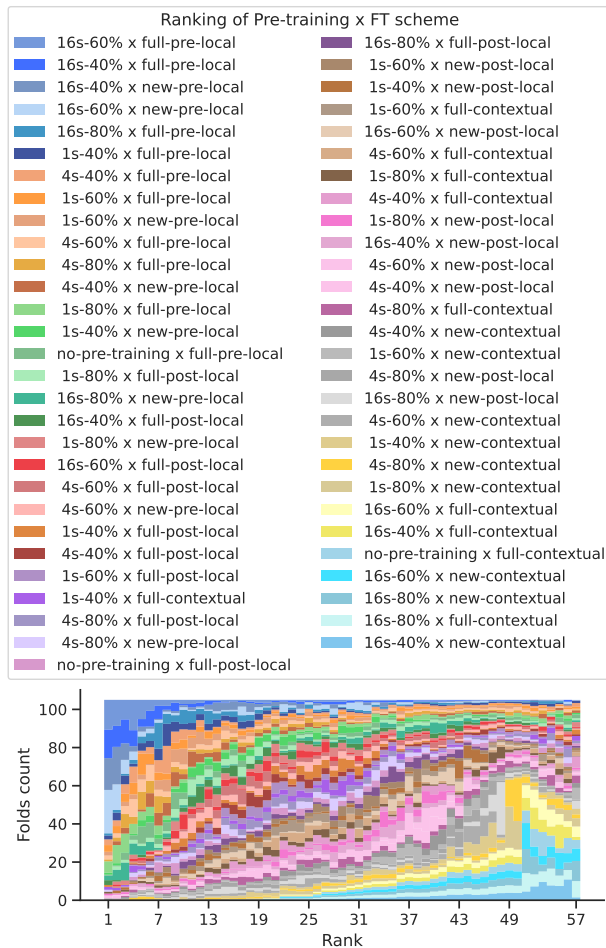


Figure 5: Global downstream classification ranking of all the combinations of pre-training configurations and fine-tuning schemes. Each of the three test datasets has 7 subjects and 5 folds per subject, which makes a total of 105 folds. In the legend, the combinations are ordered according to their average rank over all folds. The vertical span of a coloured "pixel" in the plot represents the number of folds in which this configuration has obtained the rank indicated by the x-axis.

models, especially when paired with any pre-local fine-tuning strategy. They notably occupy the first rank in two-thirds of the cases. In contrast, lower-performing pipelines feature new-contextual or full-contextual fine-tuning, particularly when combined with 16s and no-pre-training configurations.

**Paradigm-wise downstream performances.** on the individual paradigms is reported in Figure 6. For the subsequent analysis, we discern several key observations: 1) The pipelines obtaining the best score on the ERP, SSVEP and MI tasks are respectively 16s-40%-full-pre-local with a 97% AUC, 16s-60%-new-pre-local with a 94% accuracy, and 16s-40%-new-pre-local with a 65% accuracy. 2) Pipelines combining 16s or no-pre-training with contextual architectures frequently result at chance level on average. 3) Pipelines combining 1s or 4s with contextual architectures also perform at chance level on the SSVEP paradigm but above chance level on the MI and ERP ones. 4) Most pipelines manage to achieve respectable scores on the ERP dataset. 5) Only a select subset of pipelines excel on the SSVEP dataset. 6) The MI paradigm scores exhibit notable variability, as indicated by the considerable standard deviation.

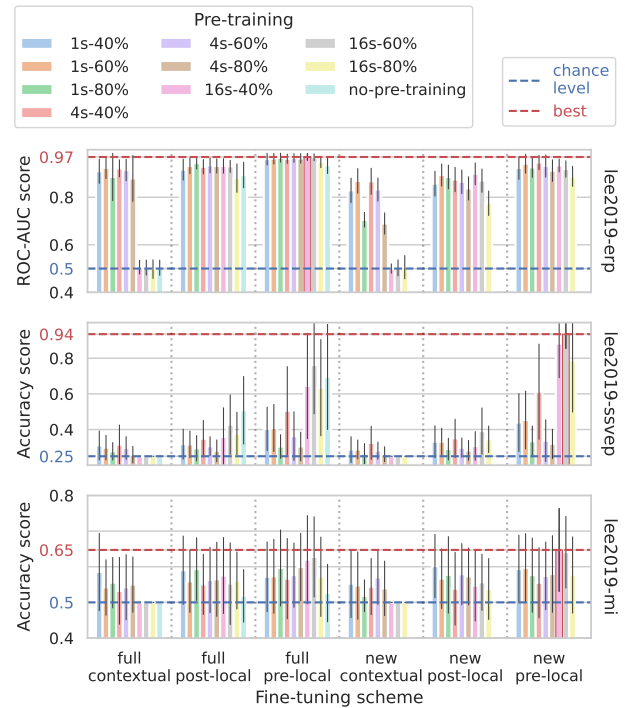


Figure 6: Downstream classification scores of all the pipelines on the three test datasets. The height of the coloured bars corresponds to the average classification score over all the test subjects and cross-validation folds, while the thin black bars correspond to their standard deviation.

## DISCUSSION

In the light of observations made in the Results section, we aim to answer the research questions posed, draw conclusions, and provide guidelines for future research.

*Examples' length largely influences downstream performance.* In particular, the 16s pipelines consistently show the best performance, highlighting the advantage of a longer context during pre-training. Conversely, when considering solely the results using the contextual downstream architecture, the 16s pipelines often perform at chance level. On the other hand, the 1s and 4s pipelines yield better results, which may be attributed to their better alignment of the attention mechanism's training with the short signal windows of the downstream tasks.

*Mask radius' impact on downstream performance uncertain,* as our results do not demonstrate a clear trend between the masks compared. It is possible that the range of masks we compared is not optimal or that other factors are influencing the results. Future work should consider comparing our spatial masking strategy with temporal masking to better understand the relative strengths and weaknesses of these approaches.

*The best fine-tuning strategy implements spatial filtering* The pre-local architecture emerges as the best for downstream classification. This architecture's approach to linearly combine channels before computing features enables effective spatial filtering, thereby enhancing the signal-to-noise ratio. This finding underscores the critical role of spatial filtering in boosting model performance by leveraging the inherent spatial properties of EEG data.

*State-of-the-art comparison.* According to Chevallier and colleagues [14] who benchmarked numerous decoding algorithms across all the datasets available in MOABB [13], the current state-of-the-art (SOTA) performances for within-session classification on the Lee2019 dataset are: ERP at  $98.41 \pm 2.03\%$ , SSVEP at  $89.44 \pm 13.84\%$ , and MI at  $84.74 \pm 13.19\%$ . Notably, all pipelines establishing the SOTA utilize Riemannian geometry. Our approach matches the SOTA for ERP, enhances it for SSVEP, but falls short on MI. A critical difference in our evaluation methodologies should be noted: our downstream evaluation only focuses on the last 7 subjects, unlike their analysis on all 54 subjects. Specifically, Lee and colleagues have identified 6 out of these 7 subjects as hard to classify on the MI task [12]. We believe this exceptionally high rate of challenging subjects might explain our low MI performance.

*Choice of Dataset.* The need for large datasets is paramount when training transformers, potentially explaining the underperformance of the contextual downstream strategy. Although this exploratory study on the Lee2019 dataset provides valuable insights, future research should pivot towards larger datasets to fully harness the capabilities of contextual architectures.

*Conclusion.* This exploratory work introduces a novel masking strategy and three fine-tuning approaches, positioning our method competitively within the realm of BCI tasks. We achieve SOTA performance on two out of three evaluated downstream tasks. Our findings suggest that long pre-training windows favor the local features encoder, while short windows benefit the contextual encoder. Therefore, future research should aim at suc-

cessfully training both the local and contextual encoders. However, no influence of the mask radius on the downstream performance was found. Finally, the best downstream architecture includes a spatial filtering step and discards the contextual encoder.

## ACKNOWLEDGEMENTS

This work is in part supported by the Donders Center for Cognition (DCC) and is part of the project Dutch Brain Interface Initiative (DBI2) with project number 024.005.022 of the research programme Gravitation which is (partly) financed by the Dutch Research Council (NWO).

## REFERENCES

- [1] Bardes A *et al.* Revisiting Feature Prediction for Learning Visual Representations from Video. Feb. 2024. arXiv: 2404.08471 [cs].
- [2] Baevski A, Babu A, Hsu WN, Auli M. Efficient Self-supervised Learning with Contextualized Target Representations for Vision, Speech and Language. Dec. 2022. arXiv: 2212.07525 [cs, eess].
- [3] He K, Chen X, Xie S, Li Y, Dollar P, Girshick R. Masked Autoencoders Are Scalable Vision Learners. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: New Orleans, LA, USA, Jun. 2022, 15979–15988.
- [4] Assran M *et al.* Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. Apr. 2023. arXiv: 2301.08243 [cs, eess].
- [5] Baevski A, Zhou Y, Mohamed A, Auli M. Wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., 2020, 12449–12460.
- [6] Pérez Velasco S, Marcos Martínez D, Santamaría Vázquez E, Ruiz Gámez R, Hornero R. Evaluación del Impacto del Aprendizaje Auto-Supervisado en la Precisión de Interfaces Cerebro-Ordenador basadas en Imagenación Motora. Universidad Politécnica de Cartagena, Nov. 2023.
- [7] Chien HYS, Goh H, Sandino CM, Cheng JY. MAEEG: Masked Auto-encoder for EEG Representation Learning. Oct. 2022. arXiv: 2211.02625 [cs, eess].
- [8] Kostas D, Aroca-Ouellette S, Rudzicz F. BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data. Frontiers in Human Neuroscience. 2021;15:653659.
- [9] Foumani NM, Mackellar G, Ghane S, Irtza S, Nguyen N, Salehi M. EEG2Rep: Enhancing Self-supervised EEG Representation Through Informative Masked Inputs. Feb. 2024. arXiv: 2402.17772 [cs, eess].
- [10] Vaswani A *et al.* Attention is All you Need. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., 2017.
- [11] Kumar A, Raghunathan A, Jones R, Ma T, Liang P. Fine-Tuning can Distort Pretrained Features and Underperform Out-of-Distribution. In: ICLR. arXiv, Feb. 2022.
- [12] Lee MH *et al.* EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. GigaScience. 2019;8(5):giz002.
- [13] Aristimunha B *et al.* Mother of all BCI Benchmarks. Zenodo. Oct. 2023.
- [14] Chevallier S *et al.* The largest EEG-based BCI reproducibility study for open science: The MOABB benchmark. Apr. 2024. arXiv: 2404.15319 [cs, eess, q-bio].

# OPTIMIZING TIME-VARYING AUTOREGRESSIVE MODELS FOR BCI APPLICATIONS

Kyriaki Kostoglou<sup>1</sup>, Gernot Müller Putz<sup>1,2</sup>

<sup>1</sup> Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>2</sup> BioTechMed Graz, Graz, Austria

E-mail: kkostoglou@tugraz.at

**ABSTRACT:** Monitoring the spectral characteristics of brain signals can provide insights into the underlying processes responsible for their generation. In brain-computer interface (BCI) applications, this is relatively important in decoding neural activity as it can provide a means to differentiate between various tasks or mental states. To capture spectral variations, herein, we focus on time-varying autoregressive models (TVAR). We introduce a framework designed to efficiently optimize and apply these models to multi-trial and multi-channel data, including electroencephalography (EEG) signals. Our approach was validated using EEG data from motor imagery tasks.

## INTRODUCTION

Time-varying autoregressive (TVAR) models are widely utilized in brain-computer interface (BCI) research, serving various purposes including time-varying power spectral density estimation to analyze shifts in brain dynamics [1], [2], [3], as well as feature extraction crucial for online BCI applications [4], [5], [6], [7], [8], [9], [10]. AR models, known for their ability to capture prominent frequency components in the signals, provide a powerful tool for brain activity analysis. The integration of TV estimation techniques further enhances their effectiveness, enabling real-time monitoring of temporal variations in these components.

The central focus of these models revolves around the autoregressive (AR) coefficients, which play a fundamental role in shaping the power spectrum characteristics of the analyzed signal. By segmenting the data into overlapping quasi-stationary windows, one can monitor the temporal evolution of these coefficients. This approach facilitates the detection of changes in the signal dynamics associated with alterations in the user's cognitive processes or task-related activities. The estimated coefficients can subsequently be used for classification purposes. Alternatively, to streamline the process and eliminate the need for data segmentation, recursive techniques such as Recursive Least Squares (RLS) and Kalman Filtering (KF) can be employed. These methods provide at each time sample an estimate of the AR coefficients without the necessity of dividing the data into separate windows, which typically results in increased processing time.

Despite the widespread application of TVAR models,

limited studies have focused on optimizing their use, particularly in the context of feature extraction and classification within BCIs. Many studies resort to windowing approaches due to the absence of clear guidelines on employing recursive techniques. They also rely on predefined model structures derived from previous literature, potentially compromising model performance. One of the primary questions that we will try to answer here is how to effectively tune and integrate these models into multi-channel signals and apply them to unseen datasets and online BCI scenarios. Schlögl et al. [4], Pfurtscheller et al. [5] and Brunner et al. [7] were among the first to provide a comprehensive framework on TVAR models for single-trial electroencephalography (EEG) classification. They proposed methods to optimally tune the model hyperparameters, as well as strategies for integrating recursive techniques, as these techniques also impact the estimation results.

Inspired from [4], [5] and [7], we herein, explore further the application of these models and we present a concise methodological approach that can be readily implemented and extended to other TVAR model variants such as TV multivariate AR models [11], [12] and root tracking techniques [13], [14] for offline and online BCI applications. We furthermore propose the incorporation of an additional feature derived from the tracking process, in addition to the commonly employed TVAR coefficients, for classification purposes. This recommendation arises from our observation of increased accuracy when incorporating this additional feature. To validate our approach, we used a publicly available EEG dataset that consists of four different motor imagery tasks [7], [15].

## MATERIALS AND METHODS

*Time-varying Autoregressive Model (TVAR):* In a TVAR model, the current value of a time-series  $\mathbf{y}$  is expressed as a linear combination of its past values [16],

$$y(n) = \sum_{k=1}^p a_k(n)y(n-k) + e(n), \quad e \sim N(0, R) \quad (1)$$

where  $\mathbf{a}(n) = [a_1(n) \dots a_p(n)]^T$  are the AR coefficients at time point  $n$ ,  $p$  denotes the model order which specifies the number of past lags considered and  $e(n)$  is zero mean, white gaussian noise with variance  $R$ .

In practical terms, Eq. 1 assumes that the analyzed signal is the output of a TV filter driven by white gaussian noise. The characteristics of the filter as well as its temporal variations are captured by the TVAR coefficients.

**Kalman Filter (KF):** One common technique used to estimate and track the coefficients of Eq. 1 is the KF. The KF models the temporal evolution of the AR coefficients as a random walk driven by white gaussian noise (also known as process noise) with variance  $Q$ ,

$$\mathbf{a}(n) = \mathbf{a}(n-1) + \mathbf{w}(n), \quad \mathbf{w} \sim \mathcal{N}(0, Q) \quad (2)$$

$Q$  essentially dictates the magnitude of the expected coefficient variations. Eq. 1 can thus be expressed as,

$$y(n) = \mathbf{a}^T(n)\boldsymbol{\varphi}(n) + e(n), \quad e \sim \mathcal{N}(0, R) \quad (3)$$

where  $\boldsymbol{\varphi}(n) = [y(n-1) \dots y(n-k)]^T$  is the regressor vector, at time point  $n$ , containing past lags of the time-series. Using the AR state-space representation of Eqs. 2-3, the KF algorithm can be employed to estimate the AR coefficients (i.e., state variables) at each time point,

$$\hat{e}(n) = y(n) - \boldsymbol{\varphi}^T(n)\hat{\mathbf{a}}(n-1) \quad (4)$$

$$\mathbf{K}(n) = \frac{\mathbf{P}(n-1)\boldsymbol{\varphi}(n)}{R + \boldsymbol{\varphi}^T(n)\mathbf{P}(n-1)\boldsymbol{\varphi}(n)} \quad (5)$$

$$\mathbf{P}(n) = \mathbf{P}(n-1) + \mathbf{Q}\mathbf{I} - \mathbf{K}(n)\boldsymbol{\varphi}^T(n)\mathbf{P}(n-1) \quad (6)$$

$$\hat{\mathbf{a}}(n) = \hat{\mathbf{a}}(n-1) + \mathbf{K}(n)\hat{e}(n) \quad (7)$$

where  $\hat{e}(n)$  is the one-step ahead prediction error,  $\hat{\mathbf{a}}(n)$  are the tracked TVAR coefficients and  $\mathbf{K}(n)$  is the Kalman gain matrix which minimizes the a posteriori error covariance  $\mathbf{P}(n)$ . The combination of KF and AR models will be referred to, herein, as KF-TVAR.

**KF-TVAR hyperparameters:** As indicated in [4], the performance of the KF-TVAR approach depends on several factors. Here, we focus on the following model hyperparameters:

- The AR model order  $p$  as it impacts the representation of the captured underlying dynamics.
- The values  $R$  and  $Q$  of the measurement and process noise, respectively.  $Q$  defines the magnitude of the AR coefficient variations, whereas  $R$  represents the variance of the underlying noise.
- The initial value of the covariance matrix  $\mathbf{P}$ . A common practice is to set the initial covariance matrix  $\mathbf{P}(0)$  to a diagonal matrix  $\mathbf{P}(0) = P_0\mathbf{I}$  where  $P_0$  is typically assigned a large value. This choice determines the initial uncertainty associated with the estimated coefficients and affects the early KF tracking behavior.
- The initial coefficient estimates  $\hat{\mathbf{a}}(0)$ . If the initial coefficients approximate the true values at the analyzed time point, the KF-TVAR model is more likely to quickly converge or adapt to changes in the AR coefficients over time.

**Other factors influencing the KF-TVAR performance:**

In addition to the aforementioned hyperparameters, the performance of the KF-TVAR model can be influenced by various signal preprocessing steps. For instance, the choice of sampling rate has been demonstrated to impact

the KF-TVAR tracking accuracy [17]. Here, we focus on spatial filtering methods and particularly on the common average reference (CAR) filtering technique which is widely applied in BCI research.

**Adapting and tuning the KF-TVAR method on multi-channel and multi-trial signals:** BCI systems typically rely on multi-channel signals. These systems are built upon a training dataset to establish associations between the features and the desired target tasks. Once adequately trained, they can make predictions or classifications on unseen data. This study focuses on extracting TVAR coefficients as key features for classification purposes. Since the data includes multi-channel and, typically, multi-trial signals we propose a two-step approach. The first step involves optimizing the KF hyperparameters, namely  $R, Q, P_0, \hat{\mathbf{a}}^T(0)$  in a data-driven manner for varying AR model orders. The second step includes extraction of the TVAR coefficients and classification. During this step, the optimal AR model order is selected based on cross-validation (CV). Our proposed approach can be summarized as follows,

**Step 1)** Select the first, in chronological order, trial from each class of the training set.

**Step 2)** For an ascending model order  $p$  (e.g.,  $p = 1 \dots 12$ ), apply the KF-TVAR approach to each channel and tune the model hyperparameters  $X_p = [R, Q, P_0, \hat{\mathbf{a}}^T(0)]$  using a genetic algorithm (GA) [18] or any other global optimization technique. As objective function, based also on the work of Schlögl et al. [4], we propose the average normalized mean squared error (NMSE) within the selected trials defined as,

$$J(X_p) = \frac{1}{C} \sum_{k=1}^C \frac{\|\hat{\mathbf{e}}_k\|_2^2}{\|\mathbf{y}_k\|_2^2} \quad (8)$$

where  $C$  is the number of classes (and therefore trials used for model optimization),  $\hat{\mathbf{e}}_k$  is the a priori error of Eq. 4 and  $\mathbf{y}_k$  the corresponding channel signal belonging to the  $k^{\text{th}}$  class/trial. The optimization process can be performed separately for each channel, yielding different sets of hyperparameters. However, an alternative strategy involves averaging Eq. 8 across all channels to obtain a unified set of hyperparameters.

**Step 3)** For each model order  $p$ , use the obtained hyperparameter set/sets  $X_p$  and apply the KF-TVAR technique to all subsequent training trials to extract the TVAR coefficients from each channel. Furthermore, as an additional feature we propose the TV trace of the covariance matrix  $\mathbf{P}(n)$  (Eq. 6). Within each trial, the feature vector at each time point consists of the concatenated AR coefficients and KF covariance traces from all channels resulting into a vector of dimension  $M \cdot p + M$ , where  $M$  is the number of channels. If the covariance trace is excluded, the vector's dimensionality becomes  $M \cdot p$ .

**Step 4)** For each model order  $p$ , employ a machine learning algorithm to map the relationship between TVAR coefficients and the various target classes. The optimal AR order  $p_{opt}$  can be selected through cross-validation within the training set. Note that the same AR

model order is applied across all channels.

**Step 5)** Use  $p_{opt}$  and the set of hyperparameters  $X_{p_{opt}}$  obtained in step 2 on a new dataset or for online tracking.

*Data:* The proposed KF-TVAR methodology was applied to dataset 2a of the BCI competition IV (<https://www.bbc.de/competition/iv/>) [15]. This dataset consists of EEG recordings from nine subjects during four cue-based motor imagery tasks, namely movement imagination of the tongue, the left and the right hand and both feet. The recordings were obtained on two different days. Each session consisted of 72 trials from each class. At the start of each session, a recording lasting approximately 5 minutes was conducted to assess the influence of the electrooculogram (EOG). The EEG data comprised 22 channels, sampled at a rate of 200 Hz, and bandpass-filtered within the range of 0.5 to 100 Hz.

*Signal preprocessing:* All the analysis was conducted in Matlab (The Mathworks Inc.). The EOG from the initial 5-minute recordings was utilized to perform linear regression on the EEG. The coefficients obtained from this regression were then applied to remove the influence of EOG artifacts from all subsequent EEG recordings during the session. The EEG signals were then resampled to 64Hz and were temporally aligned around the cue onset, with a window spanning from -2 to 7 seconds. We analyzed the data with and without CAR filtering.

*Applying the KF-TVAR methodology:* Initially, the KF-TVAR optimization was employed to each channel separately (single-channel optimization) and we extracted as features for classification only the TV-AR coefficients. To examine the effect of the initial coefficient estimates  $\hat{\mathbf{a}}(0)$ , we first set them to 0 and then we allowed the GA to optimize them. We then included the TV trace of the  $\mathbf{P}(n)$  matrix as an extra feature for classification. Finally, we examined the approach of obtaining one set of KF-TVAR hyperparameters for all channels (multi-channel optimization). The different approaches were categorized as follows:

#### Single-channel optimization and feature extraction:

For each channel and each investigated model order, the GA provided an optimal hyperparameter set  $X_p$  by minimizing Eq. 8. These values were then used to extract KF-TVAR features. We examined the following scenarios,

- **sC0W0:**  $\hat{\mathbf{a}}(0)$  set to 0, TV-AR coefficients extracted for classification.
- **sC0W1:**  $\hat{\mathbf{a}}(0)$  optimized by the GA, TV-AR coefficients extracted for classification.
- **sC1W0:** CAR rereferencing,  $\hat{\mathbf{a}}(0)$  set to 0, TV-AR coefficients extracted for classification.
- **sC1W1:** CAR rereferencing,  $\hat{\mathbf{a}}(0)$  optimized by the GA, TV-AR coefficients extracted for classification.
- **sC1W1+:** CAR rereferencing,  $\hat{\mathbf{a}}(0)$  optimized by the GA, both TV-AR coefficients and TV  $\mathbf{P}(n)$  trace extracted for classification.

#### Multi-channel optimization and feature extraction:

For each model order, the GA provided a unified set of hyperparameters  $X_p$  by minimizing the average of Eq. 8 across all channels. The scenarios we examined are summarized below.

- **mC1W1:** CAR rereferencing,  $\hat{\mathbf{a}}(0)$  optimized by the GA, TVAR coefficients extracted for classification.
- **mC1W1+:** CAR rereferencing,  $\hat{\mathbf{a}}(0)$  optimized by the GA, both TVAR coefficients and TV  $\mathbf{P}(n)$  trace extracted for classification.

Finally, for both single-channel and multi-channel approaches, the optimal model order  $p_{opt}$  was selected based on the CV performance. Subsequently, the KF-TVAR  $X_p$  set provided by the GA, corresponding to  $p_{opt}$ , (i.e.,  $X_{p_{opt}}$ ) was used to estimate KF-TVAR features.

*Classification:* By applying the KF-TVAR methodology within each trial of the first session (Session 1) we obtained features at each time sample (see Step 3). Since all trials were aligned to the cue onset, we utilized a sample-by-sample classification approach wherein a shrinkage linear discriminant analysis (sLDA) model was trained on each individual time sample relative to the cue onset (covariance shrinkage was applied using the Matlab toolbox covShrinkage [19]). Additionally, we employed a trial-based 10x1 fold CV scheme. This enabled us to estimate the CV accuracy within a trial over time (referred to as **Session 1 CV**).

As described earlier, the dataset included also a second session of EEG measurements obtained on a different day (Session 2). Using the optimized KF-TVAR hyperparameters from the first session we extracted TV features from trials of the second session (without reapplying the KF-TVAR optimization procedure). We investigated two scenarios. In the first scenario, for each participant we predicted the motor imagery task within each trial at the time point of maximum accuracy identified in the first session. The sLDA classifier was derived at that specific time point using all the data from the first session (referred to as **Session 2 Prediction**). In the second scenario, we retrained the sLDAs through 10x1 fold CV (similarly as Session 1 CV) on the second session (referred to as **Session 2 CV**). These scenarios were both analysed to understand whether decreases in accuracy resulting from session transfer stem from the features extracted or indicate the need for recalibration of the sLDA classifier. A similar approach was also followed in Brunner et al. [7].

*Statistics:* For statistical testing we employed Wilcoxon's signed-rank test, along with Benjamini-Hochberg [20] correction for multiple comparisons.

## RESULTS

Fig. 1 depicts the average runtime (over all participants) in seconds for the single-channel and the multi-channel optimization approaches as a function of the model order  $p$ . As the model order increases, runtime increases in both methods, with a slightly lesser impact observed for the multi-channel approach. Nevertheless,

the overall runtime remains under 1 second. It's important to note, however, that runtime, is also influenced by factors such as the number of channels (here  $M = 22$ ), sampling rate and trial length.

In Fig. 2, we present the classification accuracies (%) obtained for different scenarios using different optimization approaches (single-channel vs multi-channel optimization). Fig. 2a depicts the maximum value of the TV CV accuracy in Session 1 for all participants (Session 1 CV). In Fig. 2b, we identified, for each participant, the time point of maximum accuracy on Session 1 and utilized this time point to predict imagined movement in trials of Session 2. The KF-TVAR features were extracted from the EEG signals of Session 2 using the optimal hyperparameter sets obtained in Session 1 (Session 2 Prediction). Fig. 2b illustrates the resulting prediction accuracies. In Fig. 2c, KF-TVAR features were extracted from the EEG signals of Session 2 using the optimal hyperparameter set obtained in Session 1; however, the sLDA models were retrained using a 10x1 fold CV approach on Session 2 (Session 2 CV). The maximum accuracies acquired for each participant are depicted in Fig. 2c.

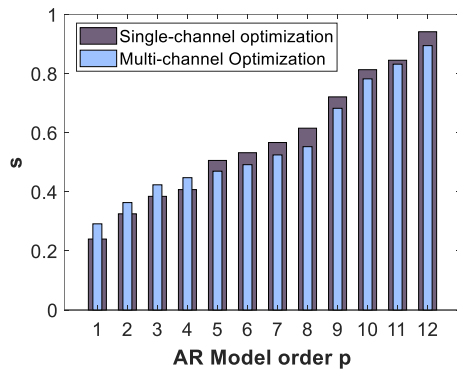


Figure 1: Average runtime (over all participants) for the single-channel and multi-channel KF-TVAR optimization approaches as a function of the AR model order  $p$  (on a 13<sup>th</sup> Gen Intel(R) Core TM i7-1355U using MEX files and a parallel pool of 10 workers). The GA was executed for 50 generations using the default *ga* Matlab settings. The upper and lower bounds for the hyperparameters were set as  $Q: [0, \inf]$ ,  $R: [0, \inf]$ ,  $P_0: [0, \inf]$ ,  $\hat{\mathbf{a}}(0): [-2, 2]$ .

First, we observed that CAR rereferencing led to increased accuracies (see results from sC0W0/ sC0W1/ sC1W0/ sC1W1 – significant increases ( $p < 0.027$ ) were found in Session 1 CV and Session 2 Prediction). Optimizing the initial AR coefficients  $\hat{\mathbf{a}}(0)$ , resulted into significant increases only in Session 1 CV, suggesting a possible dependence on session specific characteristics. Second, multi-channel optimization frequently resulted in higher predictive performance compared to single-channel optimization (see sC0W0/ sC0W1/ sC1W0/ sC1W1 vs vs mC1W1 and sC1W1<sup>+</sup> vs mC1W1<sup>+</sup>). Third, the augmented feature set containing both the TVAR coefficients as well as the TV KF covariance trace led to significantly higher accuracies compared to considering only the TVAR coefficients (see sC0W0/ sC0W1/ sC1W0/ sC1W1 vs sC1W1<sup>+</sup> and mC1W1 vs mC1W1<sup>+</sup>). Overall, mC1W1<sup>+</sup> exhibited superior performance

compared to all other methods ( $p < 0.05$  except for the scenario Session 1 CV, where sC1W1<sup>+</sup> and mC1W1<sup>+</sup> had similar performance).

By applying session transfer from Session 1 to Session 2 (Session 2 Prediction) we observed an anticipated statistically significant decrease ( $p = 0.007$ ) in accuracy in all methods. However, after retraining the sLDA models on Session 2 while maintaining the same extracted KF-TVAR features as before, the differences were no longer statistically significant.

To provide a more holistic view of the temporal evolution of the classification results within each trial, in Fig. 3, we present the TV accuracies obtained for the various scenarios across each participant. Here, we used the proposed multi-signal mC1W1<sup>+</sup> approach. The optimal AR model orders were found to be 5, 5, 10, 7, 12, 3, 10, 4 and 2 for Participants P1, P2, P3, P4, P5, P6, P7, P8 and P9, respectively.

## DISCUSSION

We presented a framework for optimal application of the KF-TVAR models on cue-based motor imagery tasks for the purposes of synchronous classification and prediction. Schlögl et al. [4] and Brunner et al. [7] have extensively examined the performance of these models and outlined a detailed process for optimizing them. Additionally, Brunner et al. [7], applied their methodology to the same motor imagery dataset analysed here. The difference of our study lies in the number of channels included in the analysis, the speed of the optimization process, as well as the methods applied for it. Brunner et al. [7] focused solely on channels C3, Cz, and C4, optimizing all relevant hyperparameters based on CV classification outcomes. In our approach, we allowed the inclusion of multiple channels, and we decoupled the KF optimization, enabling it to operate independently of the classification process (Fig. 1). Moreover, the utilization of GAs facilitates faster processing by eliminating the need to iterate through various hyperparameter values. The only hyperparameter that was selected based on CV was the AR model order  $p$ .

We decided to use single trials to optimize the KF hyperparameters in order to reduce computation runtime. We specifically selected the first trials from each class of the training set, assuming they were relatively free from significant artifacts. Our main goal during optimization was to identify the most suitable initialization hyperparameters for the KF. While the initialization phase influences tracking performance, the KF's adaptability and recursive nature, enables it to adjust to variations in the data. Thus, incorporating additional trials into the optimization process is unlikely to lead to significant changes in the final results. Regarding including single trials from all classes, the decision aimed to identify appropriate initialization hyperparameters that equally accommodate the signal characteristics of all classes, without favoring any specific class over others.

We further observed that using a single set of hyperparameters for all channels increased classification accuracies. We hypothesize that imposing uniform rate

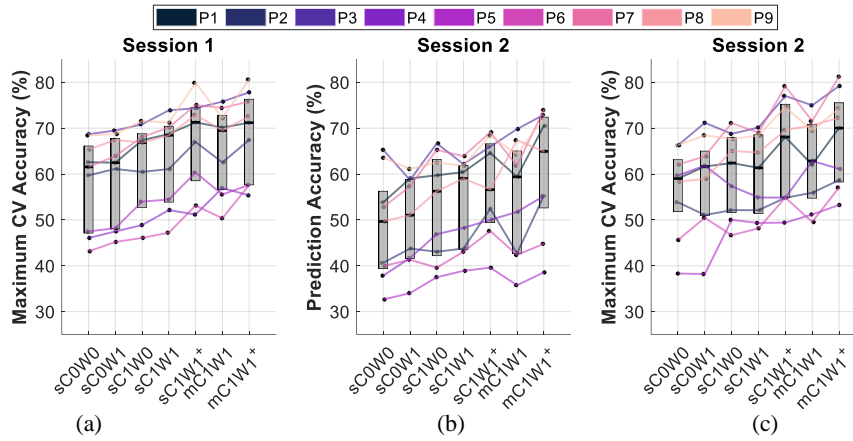


Figure 2: Boxplots depicting the (a) maximum CV accuracy (%) across participants on Session 1 using different KF-TVAR optimization approaches (*Session 1 CV*), (b) prediction accuracy (%) on Session 2, defined as the accuracy obtained on the time point of maximum accuracy identified in Session 1, along with the classifier derived at that specific time point (*Session 2 Prediction*) and (c) maximum CV accuracy (%) across participants on Session 2 (*Session 2 CV*). Features were extracted using the KF-TVAR hyperparameter sets obtained from Session 1. The sLDA was trained on Session 2 using a 10x1 fold CV procedure. The various colored lines represent different participants and depict the accuracy changes resulting from the application of different optimization approaches within the specific participant.

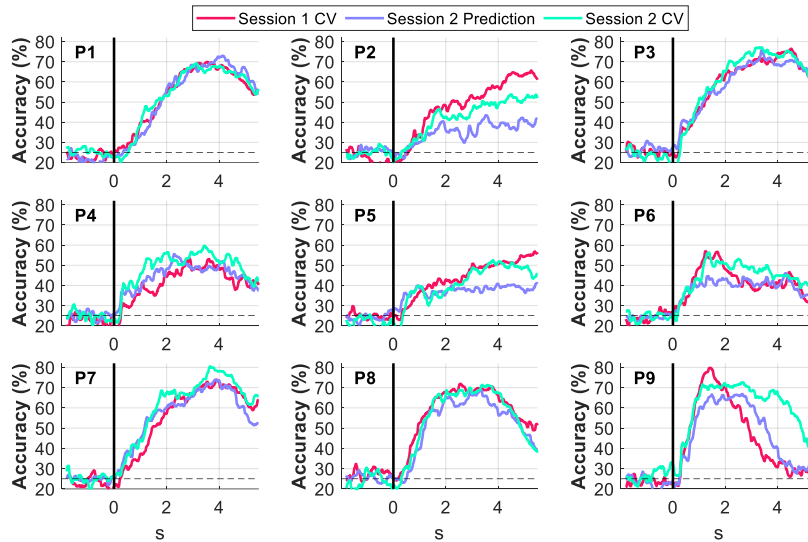


Figure 3: Temporal evolution of the accuracy (%) for each participant based on the proposed mC1W1<sup>+</sup> optimization method. The plotted curves depict the instantaneous accuracy achieved using a sample-by-sample classification approach. The black vertical line at  $n = 0s$  denotes the cue onset. *Session 1 CV* (red line) refers to the TV accuracy obtained through CV in Session 1. *Session 2 Prediction* (purple line) represents the TV accuracy estimated by extracting KF-TVAR features using the optimal hyperparameter sets obtained from Session 1, as well as the sLDA models trained on Session 1. *Session 2 CV* (cyan line) refers to the TV accuracy estimated by extracting KF-TVAR features using the hyperparameter sets obtained from Session 1, and training the sLDA models through CV on Session 2. The theoretical chance level was 25% (dashed horizontal black line at 25%).

of AR changes across all channels enhances discriminability in classification tasks and improves robustness against noisy channels. We also opted to apply the same AR model order across all channels, as assigning a unique order to each channel resulted in inferior classification performance (not shown here). Instead of relying on traditional model selection criteria such as the Akaike or Bayesian information criterion, we chose the optimal model order based on CV classification results. In [9], conventional AR model selection methods, typically applied in time-series analysis, were found inadequate for capturing discriminative EEG features related to motor imagery tasks. This implies that traditional system identification and signal analysis

approaches may not always translate effectively for classification purposes. In some cases, complex tasks may necessitate a higher-order AR model to capture informative temporal patterns, whereas simpler tasks may be adequately represented by a lower-order model.

Lastly, we propose including the TV trace of the KF covariance matrix  $P(n)$  (Eq. 6) as an additional feature alongside the TVAR coefficients. This recursively estimated feature contributes positively to the classification performance. Changes in the trace of this matrix over time indicate fluctuations in the variability of the estimated coefficients. Large trace variations may correspond to periods of significant changes in the underlying EEG signals, such as transitions between

different states or tasks. Based on our results (Fig. 2), this feature augmentation led to significantly higher classification accuracies. In terms of signal preprocessing, we found that CAR filtering generally led to improved classification performances compared to no spatial filtering.

We observed a significant decrease in accuracy when predicting Session 2 motor imagery tasks using optimal KF-TVAR hyperparameter sets and sLDA models from Session 1 (Session 2 Prediction). This outcome was anticipated, considering Session 2 was conducted on a separate day. To determine whether this decrease stemmed from the TVAR estimation procedure or the classification algorithm, we extracted KF-TVAR features on Session 2 using the hyperparameters of Session 1 and retrained the sLDA models. The CV accuracy was found to be similar to that of Session 1, suggesting no necessity to readjust the KF-TVAR tracking, but rather the sLDA algorithm (e.g., using an adaptive sLDA).

While direct comparisons may not be feasible due to variations in CV strategies and the number of channels employed, we reference the results obtained in [7]. For scenarios Session 1 CV, Session 2 Prediction and Session 2 CV the average, across participants, 0.9 quantile of the classification accuracy was 54.28%, 39.3% and 51.12%, respectively. In contrast, the proposed mC1W1+ algorithm achieved 63.4%, 58.3% and 63.4%, respectively.

Our approach, initially designed for synchronous classification, can be readily adapted to asynchronous BCI applications. This can be achieved by either optimizing KF-TVAR on continuous data or by segmenting the data and extracting optimal KF-TVAR hyperparameters, as described here. Once the optimal KF-TVAR hyperparameters are determined, the TVAR coefficients can be continuously tracked. Finally, the method can be extended to other TVAR variants such as the multivariate TVAR models [7], [11]. Rather than using TVAR coefficients for classification, future work will explore AR-based root tracking techniques [13], [14], [21]. These techniques directly track the poles and zeros of the signal-generating system, capturing its dominant spectral components. This transition from AR coefficients to poles and zeros, could offer additional predictive value in discriminating various EEG tasks.

## CONCLUSION

In conclusion, the methods discussed offer a robust framework for effectively applying TVAR models on BCI tasks. Future work will focus on further optimizing, speeding up and improving their application.

## REFERENCES

- [1] D. J. McFarland and J. R. Wolpaw, "Sensorimotor rhythm-based brain-computer interface (BCI): model order selection for autoregressive spectral analysis," *J Neural Eng*, vol. 5, no. 2, p. 155, 2008.
- [2] Y. Li, Q. Liu, S.-R. Tan, and R. H. M. Chan, "High-resolution time-frequency analysis of EEG signals using multiscale radial basis functions," *Neurocomputing*, vol. 195, pp. 96–103, 2016.

- [3] Z. G. Zhang, Y. S. Hung, and S.-C. Chan, "Local polynomial modeling of time-varying autoregressive models with application to time-frequency analysis of event-related EEG," *IEEE Trans Biomed Eng*, vol. 58, no. 3, pp. 557–566, 2010.
- [4] A. Schlögl, D. Flotzinger, and G. Pfurtscheller, "Adaptive autoregressive modeling used for single-trial EEG classification-verwendung eines Adaptiven Autoregressiven Modells für die Klassifikation von Einzeltrial-EEG-Daten," 1997.
- [5] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lügner, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 316–325, 1998.
- [6] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *J Neural Eng*, vol. 2, no. 4, p. L14, 2005.
- [7] C. Brunner, M. Billinger, C. Vidaurre, and C. Neuper, "A comparison of univariate, vector, bilinear autoregressive, and band power features for brain-computer interfaces," *Med Biol Eng Comput*, vol. 49, pp. 1337–1346, 2011.
- [8] H. Nai-Jen and R. Palaniappan, "Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004, pp. 507–510.
- [9] A. Atyabi, F. Shic, and A. Naples, "Mixture of autoregressive modeling orders and its implication on single trial EEG classification," *Expert Syst Appl*, vol. 65, pp. 164–180, 2016.
- [10] B. Graimann, J. E. Huggins, A. Schlogl, S. P. Levine, and G. Pfurtscheller, "Detection of movement-related patterns in ongoing single-channel electrocorticogram," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 11, no. 3, pp. 276–281, 2003.
- [11] L. Astolfi *et al.*, "Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators," *IEEE Trans Biomed Eng*, vol. 55, no. 3, pp. 902–913, 2008.
- [12] K. Kostoglou, A. D. Robertson, B. MacIntosh, and G. D. Mitsis, "A novel framework for estimating time-varying multivariate autoregressive models and application to cardiovascular responses to acute exercise," *IEEE Trans Biomed Eng*, 2019.
- [13] K. Kostoglou and M. Lunglmayr, "Root tracking using time-varying autoregressive moving average models and sigma-point Kalman filters," *EURASIP J Adv Signal Process*, vol. 2020, no. 1, p. 6, Dec. 2020, doi: 10.1186/s13634-020-00666-7.
- [14] L. T. Mainardi, A. M. Bianchi, G. Baselli, and S. Cerutti, "Pole-Tracking Algorithms for the Extraction of Time-Variant Heart Rate Variability Spectral Parameters," *IEEE Trans Biomed Eng*, vol. 42, no. 3, pp. 250–259, 1995, doi: 10.1109/10.364511.
- [15] M. Tangermann *et al.*, "Review of the BCI competition IV," *Front Neurosci*, p. 55, 2012.
- [16] L. Ljung and T. Söderström, *Theory and practice of recursive identification*. MIT press, 1983.
- [17] M. F. Pagnotta and G. Plomp, "Time-varying MVAR algorithms for directed connectivity analysis: Critical comparison in simulations and benchmark EEG data," *PLoS One*, vol. 13, no. 6, p. e0198846, 2018.
- [18] K. Kostoglou and G. D. Mitsis, "Modelling of multiple-input, time-varying systems with recursively estimated basis expansions," *Signal Processing*, 2018, doi: S0165168418303207.
- [19] Olivier Ledoit, "covShrinkage (<https://github.com/oledoit/covShrinkage/releases/tag/1.1.0>), GitHub."
- [20] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
- [21] G. Müller-Putz, M. Crell, J. Egger, P. Suwandjjeff, and K. Kostoglou, "Towards Implantable Brain-Computer Interface for Communication in Locked-In Syndrome patients: An introduction to INTRECOM," in *Current Directions in Biomedical Engineering*, 2023, pp. 1–4.

# RECOGNITION OF PERTURBATION EVOKED POTENTIAL BY USING MIXED-DEPTHWISE CONVOLUTIONS

Shayan Jalilpour<sup>1</sup>, Gernot R. Müller-Putz<sup>1,2</sup>

<sup>1</sup> Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>2</sup> BioTechMed, Graz, Austria.

E-mail: gernot.mueller@tugraz.at

**ABSTRACT:** Prior studies have explored the capability of decoding balance perturbations using electroencephalography (EEG) in single-trial classifications. The potential for real-time detection of perturbation-evoked potentials (PEPs) could facilitate the implementation of brain-computer interfaces (BCIs) in everyday assistive systems. Achieving the detection of these potentials in a subject-independent manner is crucial for this advancement. A key step towards this objective is the development of a model capable of identifying balance loss without requiring individual calibration for each subject and enabling online analysis. Deep neural networks have recently achieved significant milestones and have been successfully applied in neural engineering. In this study, we propose a lightweight neural network to assess the viability of single-trial classification of PEPs in a subject-independent manner. Our model was tested on three balance perturbation datasets, demonstrating superior performance in subject-independent classification compared to EEGNet, rLDA, and RBF-SVM classifiers.

## INTRODUCTION

Early detection of balance loss offers a promising avenue for preventing falls by enabling brain-computer interfaces (BCIs) as an assistive technology. In recent years, progress has been made in investigating EEG studies related to balance loss [1], [2], [3], [4], [5]. These studies have shown that perturbation-evoked potentials (PEPs) appear in brain signals during balance perturbations [2]. Such event-related potentials (ERPs) comprise different EEG components, including the N1 amplitude -a large negative potential in the fronto-central electrodes, with PEPs primarily characterized by this component. N1 is followed by a positive component, P2, and finally a negative wave called N2. Numerous studies have examined how the brain responds to different balance perturbations and have explored the effects of various stimuli on PEPs. Traditionally, research aimed to uncover the neuroscientific characteristics of PEPs by studying the grand average signals of PEPs [1], [4], [6], [7]. However, few studies have explored the feasibility of incorporating these brain potentials into BCI systems [8], [9]. One of the main steps to achieve this goal is being able to predict balance perturbation in single trials prior to muscle activation to maintain balance. Previously, single trial classification of

PEPs was investigated from spontaneous EEG data [9], [10], [11]. In another study, we attempted to classify PEPs in single trials in a simulated asynchronous task and further evaluated the detection of different types of perturbation such as angle and direction by using the brain signals [8], [12].

Given the low signal-to-noise ratio (SNR) in EEG recordings and differences between persons, a classification model trained with EEG data from one person is not transferable to another. This requirement for individual calibration of the BCI system for each user can be both time- and energy- consuming, as it involves collecting sufficient data for every participant. This challenge motivated researchers to develop methods that can mitigate this issue.

Domain adaptation and transfer learning are two techniques developed to address the differences in distribution between target and source domains [13]. However, these methods face challenges that limit their practical application in real-world scenarios [14]. For instance, these techniques rely on the offline adaptation of feature distribution and typically apply adjustments to the pre-built model. To address the challenge of subject-independent classification, we propose a novel neural network designed to detect balance perturbation through subject-independent classification using single trials.

Our model is designed based on neurophysiological principles in a manner that keeps the number of parameters low while simultaneously extracting subject-independent PEP features. We evaluated our model on three EEG balance perturbation datasets: two open-access datasets, and data collected during our two previous studies. We then compared the performance of our model against traditional classification methods, including rLDA and SVM, in addition to well-established neural networks like EEGNet.

## MATERIALS AND METHODS

*Datasets:* We evaluated the models' performance using three datasets. In the first and second datasets, thirty healthy participants were instructed to stand and walk on a treadmill-mounted balance beam [15], [16]. Two electromechanical motors, positioned on the left and right sides of the treadmill, and they were connected to the participants' waists via steel cables. These motors were programmed to rotate a bar attached to the cables

by 90 degrees, inducing mediolateral pull perturbations. During both the standing and walking scenarios, each participant experienced 150 perturbations over a session lasting 10 minutes. Brain activity throughout these sessions was recorded using a 128-electrode EEG system (BioSemi ActiveTwo, BioSemi) with a sampling rate of 512 Hz.

The third dataset includes neural recordings from 30 healthy participants, collected through two separate experiments, with each study involving 15 participants [8], [17]. The experimental setup was similar across both studies, where participants were seated in a glider that was tilted in both left and right directions in a simulated aviation scenario. An industrial robot was used to impose these perturbations by tilting the glider at angles of 5 and 10 degrees to simulate balance disruptions. In both experiments, participants completed six blocks, with each block consisting of 40 perturbations in the first study and 50 in the second study, respectively. Brain activity was measured at a sampling rate of 512 Hz with 63 electrodes using an EEGO amplifier (ANT-neuro, Enschede, Netherlands).

To simplify the discussion of dataset-specific findings and analyses, we will adopt specific notations. "Dataset 1" and "Dataset 2" will be used to denote the stand and walking waist perturbation conditions, respectively. "Dataset 3" will represent data from whole-body balance perturbations in simulated aviation scenarios. This notation will ease the discussion of dataset-specific findings and analyses.

*Pre-processing:* For each dataset, the preprocessing steps were consistent: data was first bandpass-filtered from 0.5 to 30 Hz and subsequently downsampled to 64 Hz. PEP epochs were extracted from the time range of 0 to 1.5 seconds following the perturbation's start. Moreover, rest epochs were segmented before each perturbation onset. Bad EEG channels were detected and eliminated from the rest of analysis. To filter out noise and artifacts, we initially applied Artifact Subspace Reconstruction (ASR) with a threshold value of 30 [18]. This step was followed by the application of independent component analysis (ICA) [19] combined with ICLabel [20] for the removal of eye and muscle artifacts from the EEG data. The processed data were then re-referenced using the common average reference (CAR) technique, and the removed channels were reconstructed through interpolation.

*Proposed model:* The proposed model begins with a two-dimensional convolutional neural network (CNN) to capture the initial temporal and spatial dependencies of EEG data. This CNN layer uses 15 filters with size of (2,3), followed by batch normalization and the activation

function of exponential linear units (ELUs) and a dropout layer with rate of 0.3. Drawing inspiration from the mixed depthwise convolutions introduced by Google researchers [21], the model divides the 15 filters into three tensors, with each tensor comprising 5 filters. Subsequently, depthwise convolutions are applied to the spatial dimension of each tensor individually, with a kernel size denoted by  $C$ , which corresponds to the number of channels. Following the extraction of spatial features, the model applies depthwise convolutions to the temporal dimension of the data, employing varying temporal kernel sizes. The sizes of these temporal kernels are 4, 8, and 16, representing temporal durations of 62, 125, and 250 milliseconds, respectively. These temporal kernels were chosen to enable the model to extract the long and short temporal representation of the data. The output features are then subjected to adaptive averaging with a size of 40 and a dropout layer with a rate of 0.65. Afterward, pointwise convolutions with filter size of 5 were applied to the concatenated output from the three tensors, and followed by a dropout layer of 0.5.

All depthwise and pointwise convolutions incorporate batch normalization, the ELU activation function, and a novel attention mechanism known as simAM [22]. Inspired by findings in visual neuroscience, simAM is based on the observation that informative neurons exhibit distinct firing patterns compared to their neighbors, leading to the spatial suppression of surrounding neurons.

In our research, we aim for the model to focus on key electrodes and time periods that are critical for PEP detection. In EEG balance studies, the N1 component appears with high negative amplitude in the frontocentral part of the brain, creating a contrast with other time points and brain areas. This feature allows the simAM mechanism to highlight these crucial periods and electrodes, thereby enhancing PEP identification.

The simAM method compute the importance of each neuron by using the energy function to obtain the linear separability among neurons

$$e_t^* = \frac{4(\beta^2 + \lambda)}{(t - \alpha)^2 + 2\beta^2 + 2\lambda} \quad (1)$$

In the above formula,  $t$  represents the target neuron, and  $\lambda$  denotes a coefficient with a value of  $1e-4$ .

$\alpha$  and  $\beta$  indicate the mean and variance, respectively, obtained by the formula provided below:

$$\alpha = \frac{1}{M} \sum_{i=1}^M x_i \quad (2)$$

$$\beta^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \alpha)^2 \quad (3)$$

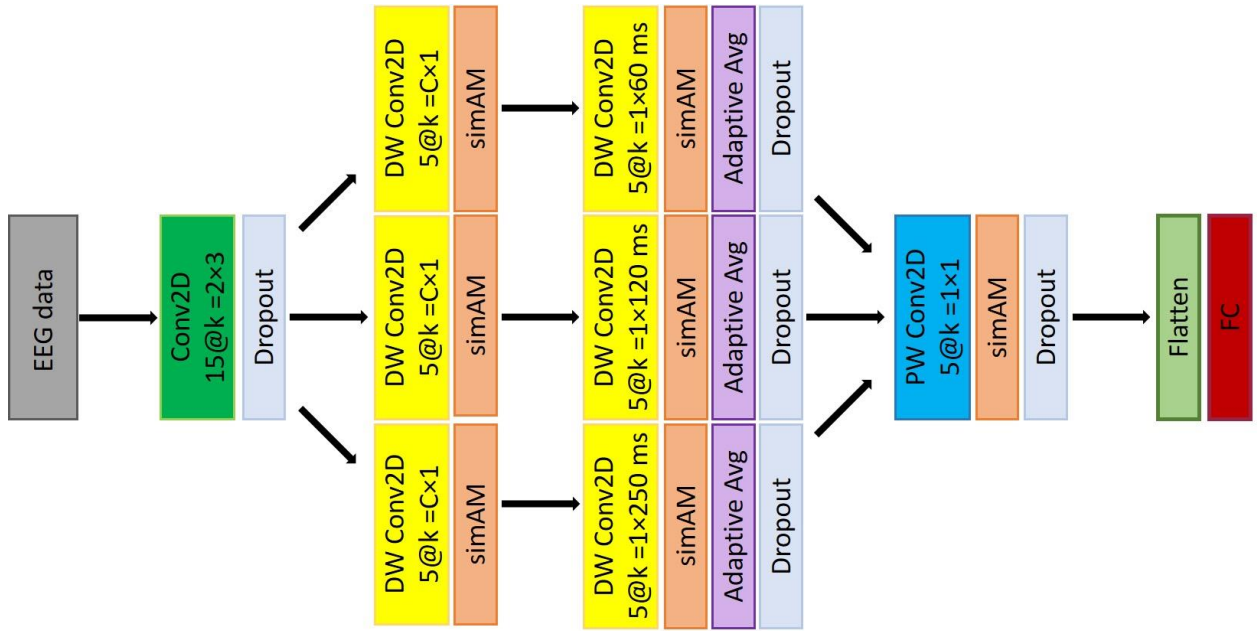


Figure 1. The architecture of proposed model

Where  $i$  is the index of the neuron, and  $M$  indicates all the neurons within each filter.  
Finally, simAM transforms the input feature map into a new feature of the same size by applying the sigmoid function to the energy function.

$$X = \text{sigmoid} \left( \frac{1}{E} \right) \odot X \quad (4)$$

The last layer of the model consists of a flatten layer and a fully connected layer to generate prediction scores for the PEP and non-PEP classes. The architecture of the model is illustrated in Fig. 1.

**Training and Evaluation Approaches:** The model was implemented in PyTorch on a GeForce RTX 3080 GPU. AdamW optimizer with a learning rate of 0.001 and weight decay of 1e-2 were used to optimize the parameters. Additionally, we utilized a batch size of 50, and the model was trained for 250 epochs. The evaluation of each model's (classifier's) performance was conducted through the accuracy metric for binary classification tasks.

## EXPERIMENTAL RESULTS

**ERP Analysis:** In Figure 2, we displayed the grand averaged EEG potentials in relation to the onset of perturbations across the three datasets. The analysis revealed that in the first and second datasets, the Perturbation Evoked Potentials (PEPs) consisted of three distinct components: P1, N1, and P2. Notably, the standing condition exhibited higher PEP amplitudes compared to the walking condition. For the third dataset, the PEPs were predominantly characterized by the N1 component located in the brain's central region. These observed differences can be linked to the task designs'

variations; specifically, perturbations in the first and second datasets targeted the participants' waist, while the third dataset involved perturbations affecting the entire body. Additionally, the N1 component in the third dataset demonstrated the highest negative amplitude compared to those in the first and second datasets.

**Classification:** For the first and second dataset, we utilized a leave one subject out approach with 30 participants to assess our model's PEP detection. In this approach, data from one subject served as the test set while data from the remaining 29 subjects constituted the training set. For the third dataset, we employed the recorded data from the first study as the training set and the data collected from the second study as the test set. As a foundation for comparison, we selected shrinkage LDA, RBF-SVM, and EEGNet [23] as baseline models to evaluate the performance of our proposed model. In addition to accuracy, we compared each model's number of trainable parameters. For the sLDA and SVM classifiers, we limited the number of features to 600 using the Fisher algorithm to select the best 600 features. This was done to prevent overfitting, maintaining a feature-to-sample ratio of 1/10. Tab. 1 and 2 shows the obtained accuracy for the standing and walking conditions for 4 models respectively. In the standing condition, it can be seen that neural network models can improve the accuracy substantially in comparison with traditional machine learning algorithms such as sLDA and SVM. The model exhibited superior performance compared to sLDA and SVM, with improvements of 9.2% and 8%, respectively. The proposed model achieved the highest performance, with an average accuracy of 86.9%, and it outperformed the EEGNet model by achieving a 1.6% increase in accuracy while having 16% fewer parameters.

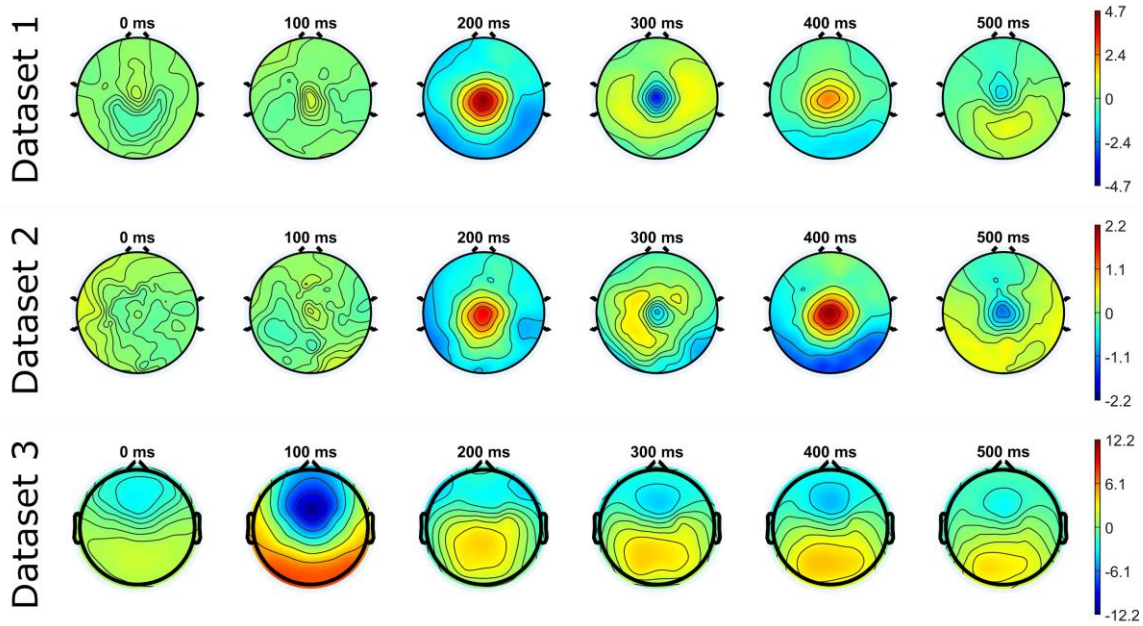


Figure 2. The scalp topography of three datasets in the time range of 0, 100, 200, 300, 400 and 500 ms. For improved visualization, we adjusted the amplitude scales differently for each dataset. The scales range from  $[-4.7, 4.7]$ ,  $[-2.2, 2.2]$ , to  $[-12.2, 12.2]$  microvolts.

Table 1: Results of Dataset 1 (standing condition)

Model	Accuracy (%)	#of parameters
sLDA	77.7	600
SVM	78.9	600
EEGNet	85.3	3248
Our model	86.9	2727

Tab. 2 presents the average subject-independent classification accuracy for the walking condition, noting that due to the smaller PEP amplitude compared to the standing condition, the accuracy in detection of perturbation was lower than the standing condition. For this dataset as well, the proposed model led in performance, attaining an accuracy of 71.8%. EEGNet came in second with a 70% accuracy rate. Following these, SVM and LDA demonstrated accuracies of 66.7% and 65.3%, respectively.

Table 2: Results of Dataset 2 (walking condition)

Model	Accuracy (%)	#of parameters
sLDA	65.3	600
SVM	66.7	600
EEGNet	70	3248
Our model	71.8	2727

Lastly, the model's performance was assessed on the third dataset, which involved participants experiencing balance perturbations in a simulated cockpit scenario. Our model demonstrated superior performance

compared to other models. Within neural networks, our model achieved 3.2% higher accuracy than EEGNet while maintaining 20% fewer parameters. Additionally, SVM showed comparable performance to the neural network models, with only 4% difference in accuracy.

Table 3: Results of Dataset 3 (whole body perturbation)

Model	Accuracy (%)	#of parameters
sLDA	87.8	600
SVM	92.4	600
EEGNet	92.8	2208
Our model	96	1752

## DISCUSSION

In this research, we investigated the potential of detecting PEPs in a subject-independent manner through EEG single trials, by employing a lightweight CNN-based model. Our approach initially involves extracting spatio-temporal patterns from the data, followed by the implementation of temporal mixed depthwise convolutions using three distinct temporal kernel sizes to accommodate the variability of ERP durations. Additionally, we incorporated a novel attention mechanism, simAM, designed to focus on discriminative features. This attention mechanism is unique in its ability to introduce 3D weights into the model without increasing its parameter count, thereby maintaining a low parameter structure for the model. To assess our model's efficacy, we applied it to three balance perturbation datasets and compared its performance against traditional models like sLDA and SVM, as well as advanced neural networks such as EEGNet. Our findings indicate that

neural networks surpass traditional machine learning methods in subject-independent classification across all datasets. Moreover, our model achieved accuracy improvements over EEGNet by 1.3%, 1.8%, and 3.2% for the first, second, and third datasets, respectively. Future endeavors will focus on evaluating our model across additional ERP datasets, also subsequent research will aim at enhancing the model's capability to extract complex spatial patterns of the data.

## CONCLUSION

We introduced a novel CNN model that utilizes mixed depthwise convolutions and the simAM attention module to enhance the detection of PEPs from spontaneous EEG data. The efficacy of this model was evaluated across three distinct datasets and its performance was benchmarked against both traditional machine learning techniques and advanced neural networks. Our findings demonstrate that our model achieved the highest accuracy rates when compared with the other models examined.

## REFERENCES

- [1] V. Dietz, J. Quintern, W. Berger, and E. Schenck, "Cerebral potentials and leg muscle e.m.g. responses associated with stance perturbation," *Exp. Brain Res.*, vol. 57, no. 2, Jan. 1985, doi: 10.1007/BF00236540.
- [2] J. P. Varghese, R. E. McIlroy, and M. Barnett-Cowan, "Perturbation-evoked potentials: Significance and application in balance control research," *Neurosci. Biobehav. Rev.*, vol. 83, pp. 267–280, Dec. 2017, doi: 10.1016/j.neubiorev.2017.10.022.
- [3] A. L. Adkin, A. D. Campbell, R. Chua, and M. G. Carpenter, "The influence of postural threat on the cortical response to unpredictable and predictable postural perturbations," *Neurosci. Lett.*, vol. 435, no. 2, pp. 120–125, Apr. 2008, doi: 10.1016/j.neulet.2008.02.018.
- [4] A. R. Sipp, J. T. Gwin, S. Makeig, and D. P. Ferris, "Loss of balance during balance beam walking elicits a multifocal theta band electrocortical response," *J. Neurophysiol.*, vol. 110, no. 9, pp. 2050–2060, Nov. 2013, doi: 10.1152/jn.00744.2012.
- [5] S. M. Peterson and D. P. Ferris, "Group-level cortical and muscular connectivity during perturbations to walking and standing balance," *NeuroImage*, vol. 198, pp. 93–103, Sep. 2019, doi: 10.1016/j.neuroimage.2019.05.038.
- [6] R. Goel, S. Nakagome, N. Rao, W. H. Paloski, J. L. Contreras-Vidal, and P. J. Parikh, "Fronto-Parietal Brain Areas Contribute to the Online Control of Posture during a Continuous Balance Task," *Neuroscience*, vol. 413, pp. 135–153, Aug. 2019, doi: 10.1016/j.neuroscience.2019.05.063.
- [7] R. Goel, R. A. Ozdemir, S. Nakagome, J. L. Contreras-Vidal, W. H. Paloski, and P. J. Parikh, "Effects of speed and direction of perturbation on electroencephalographic and balance responses," *Exp. Brain Res.*, vol. 236, no. 7, pp. 2073–2083, Jul. 2018, doi: 10.1007/s00221-018-5284-5.
- [8] S. Jalilpour and G. Müller-Putz, "Toward passive BCI: asynchronous decoding of neural responses to direction- and angle-specific perturbations during a simulated cockpit scenario," *Sci. Rep.*, vol. 12, no. 1, p. 6802, Apr. 2022, doi: 10.1038/s41598-022-10906-5.
- [9] J. C. Ditz, A. Schwarz, and G. R. Müller-Putz, "Perturbation-evoked potentials can be classified from single-trial EEG," *J. Neural Eng.*, vol. 17, no. 3, p. 036008, Jun. 2020, doi: 10.1088/1741-2552/ab89fb.
- [10] A. S. Ravindran *et al.*, "Interpretable Deep Learning Models for Single Trial Prediction of Balance Loss," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Toronto, ON, Canada: IEEE, Oct. 2020, pp. 268–273. doi: 10.1109/SMC42975.2020.9283206.
- [11] A. Sujatha Ravindran, C. A. Malaya, I. John, G. E. Francisco, C. Layne, and J. L. Contreras-Vidal, "Decoding neural activity preceding balance loss during standing with a lower-limb exoskeleton using an interpretable deep learning model," *J. Neural Eng.*, vol. 19, no. 3, p. 036015, Jun. 2022, doi: 10.1088/1741-2552/ac6ca9.
- [12] S. Jalilpour and G. R. Muller-Putz, "Direction decoding of physical and visual perturbations from EEG," in *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)*, Rome, Italy: IEEE, Oct. 2022, pp. 427–431. doi: 10.1109/MetroXRINE54828.2022.9967637.
- [13] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, Jan. 2021, doi: 10.1016/j.neucom.2020.09.017.
- [14] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *J. Neural Eng.*, vol. 15, no. 3, p. 031005, Jun. 2018, doi: 10.1088/1741-2552/aab2f2.
- [15] S. M. Peterson and D. P. Ferris, "Human electrocortical, electromyographical, ocular, and kinematic data during perturbed walking and standing," *Data Brief*, vol. 39, p. 107635, Dec. 2021, doi: 10.1016/j.dib.2021.107635.
- [16] S. M. Peterson and D. P. Ferris, "Differentiation in Theta and Beta Electrocortical Activity between Visual and Physical Perturbations to Walking and Standing Balance," *eneuro*, vol. 5, no. 4, p. ENEURO.0207-18.2018, Jul. 2018, doi: 10.1523/ENEURO.0207-18.2018.
- [17] S. Jalilpour and G. Müller-Putz, "Balance

- perturbation and error processing elicit distinct brain dynamics,” *J. Neural Eng.*, vol. 20, no. 2, p. 026026, Apr. 2023, doi: 10.1088/1741-2552/acc486.
- [18] T. Mullen *et al.*, “Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG,” in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka: IEEE, Jul. 2013, pp. 2184–2187. doi: 10.1109/EMBC.2013.6609968.
- [19] S. Makeig, A. Bell, T.-P. Jung, and T. J. Sejnowski, “Independent Component Analysis of Electroencephalographic Data,” in *Advances in Neural Information Processing Systems*, D. Touretzky, M. C. Mozer, and M. Hasselmo, Eds., MIT Press, 1995. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/1995/file/754dda4b1ba34c6fa89716b85d68532b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/1995/file/754dda4b1ba34c6fa89716b85d68532b-Paper.pdf)
- [20] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, “ICLabel: An automated electroencephalographic independent component classifier, dataset, and website,” *NeuroImage*, vol. 198, pp. 181–197, Sep. 2019, doi: 10.1016/j.neuroimage.2019.05.026.
- [21] M. Tan and Q. V. Le, “MixConv: Mixed Depthwise Convolutional Kernels,” 2019, doi: 10.48550/ARXIV.1907.09595.
- [22] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, “SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks”.
- [23] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces,” *J. Neural Eng.*, vol. 15, no. 5, p. 056013, Oct. 2018, doi: 10.1088/1741-2552/aace8c.

# INTRODUCING THE ASME-SPELLER, AUDITORY BCI SPELLER UTILIZING STREAM SEGREGATION: A PILOT STUDY

Simon Kojima<sup>1</sup>, Shin'ichiro Kanoh<sup>1,2</sup>

<sup>1</sup>Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan

<sup>2</sup>College of Engineering, Shibaura Institute of Technology, Tokyo, Japan

E-mail: nb21106@shibaura-it.ac.jp

**ABSTRACT:** The auditory BCI spellers are considered the only means of communication for late-stage patients with severe neurological disorders such as amyotrophic lateral sclerosis (ALS). To date, several auditory BCI spellers have been proposed. However, they require multiple steps, visual support, or multi-channel audio systems. In this study, we proposed an ASME-speller, which stands for Auditory Stream segregation, Multiclass, ERP speller, that uses an auditory BCI paradigm based on auditory stream segregation to detect the target of the user's selective attention by presenting a QWERTY keyboard-like audio stimuli. The 64-channel electroencephalogram was measured while the six subjects carried out 15-character ASME-speller paradigms. Offline simulation using dynamic stopping showed that the ASME speller achieved an average accuracy of 0.73 and an average ITR of 3.78 bits/min. The best results were achieved with an accuracy of 0.97 and an ITR of 7.61 bits/min. These results indicate that the ASME speller can be used as a new auditory BCI speller. This study provides more users with a high-accuracy and intuitive new speller option.

## INTRODUCTION

Brain-computer interfaces (BCIs) give their users communication and control channels that do not depend on the brain's normal output channels of peripheral nerves and muscles [1]. Many BCIs aimed to restore communication for locked-in patients suffering from progressive motor diseases such as amyotrophic lateral sclerosis (ALS) [2]. Many spelling protocols using visual stimuli have been proposed [3] to realize the application of BCI in communication. However, it is known that patients with late-stage ALS have unreliable gaze control [4], and the BCIs using visual stimuli are not adequate for those patients. On the other hand, auditory BCIs do not occupy their sight and can be used by visually impaired patients. Thus, it is meaningful to realize the auditory BCIs for spelling application.

Furdea et al. [5] proposed an auditory speller BCI similar to a visual P300 speller [6]. In this system, a display of a  $5 \times 5$  matrix containing 25 alphabet characters and voices two number words coded with each character's position in the matrix was presented. One corresponded to the row, and one corresponded to the column. The system

detected which character the users paid attention to with two steps. The target row was detected in the first step, and the target column was detected in the second step. Klobassa et al. did a similar study but with a  $6 \times 6$  matrix containing all 26 alphabet characters and miscellaneous [7]. Also, they changed the human voice to environmental sounds. Schreuder et al. [8] utilized the AMUSE paradigm [9] for a spelling application. This system also detected the target character using a two-step procedure. They divided alphabet characters into six groups. The target group was detected in the first step, and the target character was detected in the second step. Each character group and character was presented from one of the six loudspeakers surrounding the subjects' heads.

Some auditory speller BCIs have been proposed; however, these studies had one of the following issues. —(1) One trial cannot determine the target character. (2) The mapping from the character to sound streams or stimuli is not intuitive and requires memorization or visual support. (3) It requires a multi-channel audio system, complicating setup and making it unavailable to patients who have hearing impairment in one ear.— Thus, we propose a novel auditory speller BCI protocol for solving these issues, the ASME-speller.

*ASME paradigm:* ASME (for Auditory Stream segregation, Multiclass, ERP) is the paradigm for auditory BCI based on auditory stream segregation. The auditory stream segregation is one of the auditory illusions that alternately presented sounds can be perceived as segregated multiple streams [10]. e.g., when sounds that have different frequencies (A and B) are presented alternately (ABABAB...), they can be perceived as two segregated sound streams (AAA... and BBB ...). The authors proposed an auditory BCI paradigm utilizing auditory stream segregation [11–14]. In this system, the oddball sequence was put into segregated streams and presented simultaneously to the subjects, and the subjects paid attention to the target stimuli in the target stream. The target stream was estimated by detecting ERP responses elicited by the target stimuli. To date, we tested the ASME paradigm with two streams [11, 12], three streams [13], and four streams [14].

*The ASME speller:*

The QWERTY is a keyboard layout widely used in computers and smartphones, and many personal computer



Figure 1: The conceptual diagram of the ASME-speller.

and smartphone users are expected to be familiar with the QWERTY layout. Since all 26 characters are mapped to three-row keys, the entire keyboard layout can be represented with three streams ASME paradigm. Fig. 1 shows the QWERTY layout and the corresponding tone stream on the ASME-speller. Three key rows are assigned to the sound stream, and the top, middle, and bottom rows correspond to the stream, which has high, middle, and low-frequency bands, respectively. Within each stream, each character is presented as a spoken voice. When the user is going to type "T," the user thinks of which row "T" is located in the QWERTY layout. Since the character "T" is in the top row, the user will listen to the corresponding stream (the stream with a high-frequency band) and pay attention to the "T" sound stimuli. Since one stimuli are paid attention to and the others are ignored, this sequence can be considered an oddball, and the target stimuli elicit ERPs, including P300 [11–14]. The target character can be estimated by detecting ERP responses with a machine learning approach. This study tested the ASME-speller paradigm with 15 characters as a pilot study.

## MATERIALS AND METHODS

**Experimental Design:** Two different conditions were conducted. (1) The ASME condition: Each row of the QWERTY layout was spoken by a different person and had a different pitch, so each row could be perceived as a different sound stream. (2) The control condition: all stimuli were spoken by the same person and had the same pitch. In a session, four runs were conducted with changing conditions. In total, two ASME runs and two control runs were conducted. In a run, 15 trials were conducted. Before starting each trial, the target character was shown on the display in front of the subject, and the subject was instructed to pay attention to the target stream and the target character. All subjects were familiar with the QWERTY layout, and no visual support was provided. 225 stimuli (15 targets and 210 nontargets) were delivered in a trial, and the trial length was about 46 seconds.

**Stimuli:** The fifteen characters (E, R, T, I, O, A, S, D, H, L, C, V, B, N, and M) were selected for this study. Each voice stimuli were generated by Amazon Web Services (AWS) Amazon Polly. AWS Amazon Polly is a cloud service that converts text into synthesized spoken audio. The voice stimuli of characters corresponding to the top (E, R, T, I, and O), middle (A, S, D, H, and L), and bottom (C, V, B, N, and M) row on the QWERTY layout were generated with the voice ID of *Ruth* (Female), *Kevin* (Male child) and *Joey* (Male), respectively. The voice IDs were selected as the top, middle, and bottom rows could be perceived as higher, middle, and lower pitch streams.

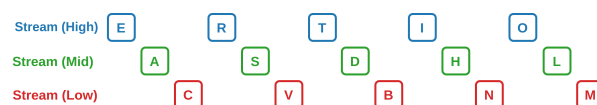


Figure 2: The time chart of presented stimuli block.

To enhance the difference between each stream, the characters corresponding to the top and bottom are shifted in pitch with +2 half-tones and −2 half-tones, respectively. Fig. 2 shows the "block" of the sequence. Each block had 15 stimuli in total. Within each stream, the order of the characters was randomized. For both ASME and control conditions, the stimuli were presented in the order of the characters corresponding to the top, middle, and bottom rows of the QWERTY layout. In a trial, 15 blocks were played. The stimulus onset asynchrony (SOA) was set to 0.2 s. All 15 characters were generated with the voice ID of *Kevin* for the control condition, and no pitch shifting was applied. All other parameters for the control condition were the same as the ASME condition. All sound stimuli were delivered by Fireface 802 (RME, Germany) with headphones (MDR-EX800ST, Sony, Japan).

**Signal Acquisitions:** The following 64-channel (Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT9, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, FT10, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP9, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, and O2) electroencephalogram (EEG) were measured with Ag-AgCl passive electrodes (Easycap, Easycap GmbH, Germany). The vertical and horizontal electrooculogram (EOG) were also measured. All EEG and EOG signals were amplified and recorded with BrainAmp DC and BrainAmp MR plus (Brain Products GmbH, Germany). The reference and the ground electrodes were placed on the right and left ear mastoid, respectively. The signals were recorded at a sampling frequency of 1000 Hz. Subjects sat on a comfortable chair placed in a soundproofing electromagnetic shielded room.

**Subjects:** Six subjects (ages 22 – 27, mean: 24.0) participated in this study. This study protocol was approved by the Review Board on Bioengineering Research Ethics of Shibaura Institute of Technology and was conducted in accordance with the Declaration of Helsinki. Before the experiment, subjects were given information orally and in writing, and written informed consent was obtained from all subjects. No subject had known neurological disorders or hearing problems.

**ERP Analyses:** The EOG artifacts were removed with independent components analysis (ICA). The measured signals were bandpass filtered by 2nd order Butterworth filter in the range of 1–30 Hz, and responses to each stimulus were epoched in the range of −0.1–1.0 s relative to stimulus onset. Then, all epochs were downsampled to 250 Hz. To assess the separability between the responses to the target and nontarget stimuli, signed- $r^2$  values [15] were obtained.

**Binary Classification:** The EOG artifacts were re-

moved using independent components analysis (ICA). The measured signals were bandpass filtered by 2nd order Butterworth filter in the range of 0.1–8Hz, and responses to each stimulus were epoched in the range of 0–1.0 s relative to stimulus onset. Then, all epochs were downsampled to 250 Hz. The mean amplitude in the following ten intervals ([0.0, 0.1; 0.1, 0.2; 0.2, 0.3; 0.3, 0.4; 0.4, 0.5; 0.5, 0.6; 0.6, 0.7; 0.7, 0.8; 0.8, 0.9; 0.9, 1.0] seconds relative to the stimulus onset) were used as the classification feature. The dimension of the feature vector was  $10\text{intervals} \times 64\text{channels} = 640$ . The classification accuracy (AUC: area under the receiver operating characteristic curve) between the responses to the target and nontarget stimuli was obtained by a shrinkage linear discriminant analysis (Shrinkage-LDA) [15] with 4-fold chronological cross-validation. For the binary classification, the chance level was 0.5. The information transfer rate (ITR) was calculated using the equation proposed by Wolpaw et al. [16].

**BCI simulation (target character detection):** In the BCI simulation, the target character of the trial was estimated. For both ASME and control conditions, 30 trials were conducted. The BCI simulation was conducted with chronological 3-fold cross-validation by training data from 20 trials and testing with data from 10 trials. From the training data, the mixing and unmixing matrices were derived using ICA to remove EOG artifacts. The mixing and unmixing matrix was applied to the training data to remove EOG artifacts, and the feature vector was obtained with the same method described in the section "Binary Classification", and Shrinkage-LDA was trained. The classification output  $f(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i + b$  was defined as follows, where  $\mathbf{x}_i$  is a feature vector,  $\mathbf{w}$  is the weight vector obtained by LDA, and  $b$  is a bias. Each feature vector  $\mathbf{x}_i$  had a corresponding class label  $y_i \in \{-1, 1\}$ , and assumed that class label +1 is the target and -1 is nontarget. The LDA was trained as  $f(\mathbf{x}) \geq 0$  if  $\mathbf{x}_i$  was in class +1 and  $f(\mathbf{x}) < 0$  if  $\mathbf{x}_i$  was in class -1. The mixing and unmixing matrix derived using ICA was applied for epoch data in each trial in test data, and the feature vectors were obtained. Then, the classifier output  $f(\mathbf{x}_i)$  for each feature was computed, and the class with the largest mean value of classifier output was estimated as the final classification result. The classification results were evaluated by accuracy. For the BCI simulation, the chance level was 0.067.

**Dynamic Stopping:** To optimize the trial length, the dynamic stopping strategy [17, 18] was also tested for BCI simulation. Dynamic stopping could be triggered after presenting the 75 stimuli in each trial. A one-sided Welch's t-test was applied to the classifier outputs  $f(\mathbf{x})$  of the class, between which the mean value of  $f(\mathbf{x})$  was the largest and second largest. If the difference was significant ( $p < 0.05$ ), the classification procedure was stopped, and the classification result for the trial was determined with the data up to that stimuli.

## RESULTS

Table 1: Binary classification results. The classification accuracy (AUC) for the ASME and control conditions are shown. The chance level was 0.5.

Subject	ASME	control
A	0.72	0.53
B	0.71	0.52
C	0.86	0.63
D	0.76	0.58
E	0.63	0.56
F	0.74	0.61
Average	0.74	0.57

Fig. 3 shows grand averaged ERP responses to the target and nontarget stimuli. In the time range from 0.2 to 0.4 seconds, N2 was observed in the ASME condition. Furthermore, in the time range from 0.4 to 0.8 seconds, P300 was observed. The amplitude of N2 and P300 were larger for the target stimuli than for nontarget stimuli, and the absolute value of the signed- $r^2$  was also larger, which implies it was informative for the machine learning model for classification. In contrast, a clear difference between the responses to the target and the nontarget stimuli was not observed in the control condition. The absolute value of the signed- $r^2$  was small compared to that for the ASME condition; the separability between the response to the target and nontarget stimuli was small compared to that for the ASME condition. Tab. 1 shows the binary classification accuracy (AUC) for the ASME and control conditions. The accuracy for the ASME condition was significantly larger than that for the control condition ( $p = 0.031$ , two-sided Wilcoxon signed-rank test). Fig. 4 shows the result of the BCI simulation (detecting the target character of the trial) without dynamic stopping. The average accuracy was 0.72 (ASME) and 0.31 (control), and the accuracy of the ASME condition was significantly larger ( $p = 0.031$ , two-sided Wilcoxon signed-rank test). The average ITR was 2.79 bits/min (ASME) and 0.726 bits/min (control), and the ITR of the ASME condition was significantly larger ( $p = 0.031$ , two-sided Wilcoxon signed-rank test). Fig. 5 shows the result of the BCI simulation using dynamic stopping. The average accuracy was 0.73 (ASME) and 0.31 (control), and the accuracy of the ASME condition was significantly larger ( $p = 0.031$ , two-sided Wilcoxon signed-rank test). The average ITR was 3.78 bits/min (ASME) and 0.775 bits/min (control), and the ITR of the ASME condition was significantly larger ( $p = 0.031$ , two-sided Wilcoxon signed-rank test). After applying dynamic stopping, the ITR (Information Transfer Rate) was improved without any drop in accuracy. The best ITR was reached at 7.61 bits/min with an accuracy of 0.97 (subject C). The worst ITR was 0.902 bits/min with an accuracy of 0.40 (subject E). Fig. 6 shows the results of BCI simulation using dynamic stopping in the confusion matrix.

## DISCUSSION

The letters in each row of the QWERTY keyboard layout, mapped to three sound streams, were presented as voice

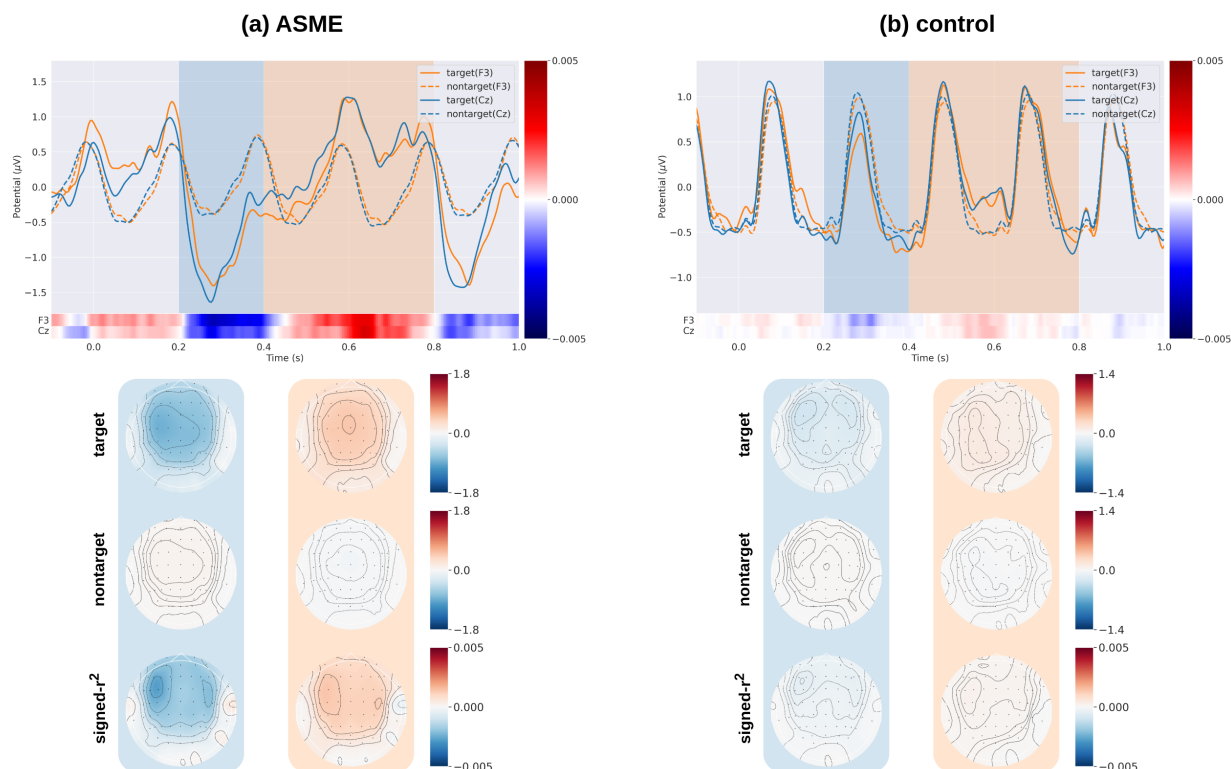


Figure 3: The grand averaged ERP responses to the target (solid line) and nontarget (dashed line) stimuli on electrode Cz (Orange) and F3 (Blue). (a) ASME condition. (b) control condition. Timepoint 0 is the stimulus onset. Stimuli were delivered with SOA of 0.2 s. The colormap below each ERP plot shows the signed- $r^2$  values on each electrode. The topography map shows the responses to the target, nontarget stimuli, and the signed- $r^2$  values from the top in each time range denoted as blue and orange mesh in the ERP plot.

stimuli, and it was shown that it is possible to pay selective attention to a single target letter stimulus. In addition, ERPs such as P300 and N2 were elicited only to the target stimuli by paying attention to them. Furthermore, the target letter could be detected with a machine learning approach with high accuracy. It can be concluded that the ASME-speller can be realized. However, the number of characters in this study was limited to 15. Thus, the speller with 26 letters needs to be tested. By applying the dynamic stopping procedure, the average Information Transfer Rate (ITR) was found to be 3.78 bits/min. Tab. 2 shows ITRs achieved in previous studies. The ITR of this study is superior to other studies except for the work by Schreuder et al. [8]. However, the best ITR was higher than theirs (7.61 bits/min v.s. 7.55 bits/min). ASME-speller has the capability to achieve higher or competitive ITR (Information Transfer Rate) and deliver high performance.

*Necessity of stream segregation:* The ASME-speller was achieved by dividing the sound stimuli into three groups, corresponding to each row of the QWERTY keyboard layout. However, these stimuli can also be delivered with a single stream, and it was not clear whether the sound stimuli needed to be delivered with segregated groups. Therefore, as a control, the condition of delivering all stimuli with a single stream was also tested in this study. Compared to the ASME condition, the amplitude of ERPs was smaller in the control condition, resulting

Table 2: ITRs achieved in previous studies. The ITR of [19] was read from a figure.

Average ITR (bits/min)	Authors
1.54	Furdea et al. [5]
2.0	Klobassa et al. [7]
3.4	Höhne et al. [20]
5.26	Schreuder et al. [8]
about 1.3	Höhne et al. [19]
1.11	Kleih et al. [21]
2.38	Markovinović et al. [22]
3.78	Kojima et al. (this paper)

in low classification results. In this study, the SOA was set to 0.2 s; however, in ASME condition, SOA within the stream was 0.6 s. It is expected that this slower SOA within the stream made the subject find the target stimuli easier and feel less overlap between stimuli. It can be concluded that utilizing the ASME paradigm makes SOA within the stream slower and making easier to find the target stimuli from the sequence.

## CONCLUSION

In this study, the ASME-speller, which detects the users' target letter from 15 characters mapped to three sound streams corresponding to the QWERTY keyboard layout, can be realized as an auditory speller BCI. The achieved ITR was faster than most of the proposed auditory BCI spellers. This study also proved to provide stimuli di-

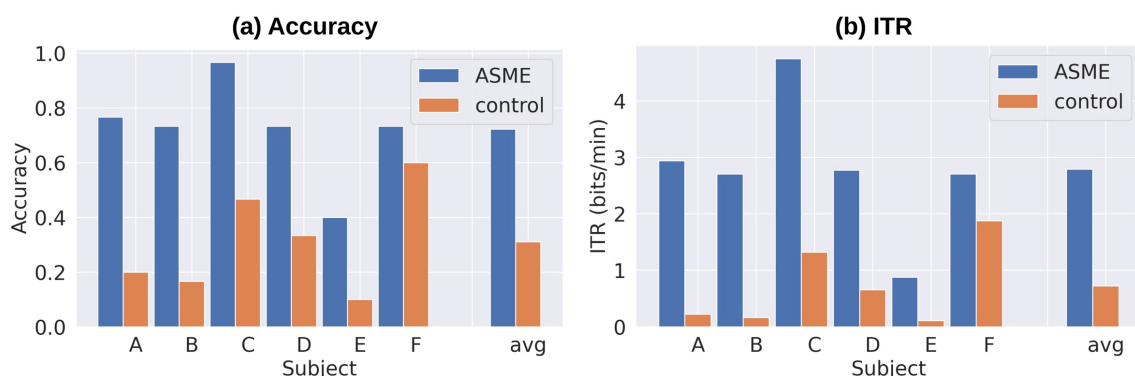


Figure 4: The results of the BCI simulation (estimating the target character of the trial) without dynamic stopping. (a) Accuracy and (b) ITR.

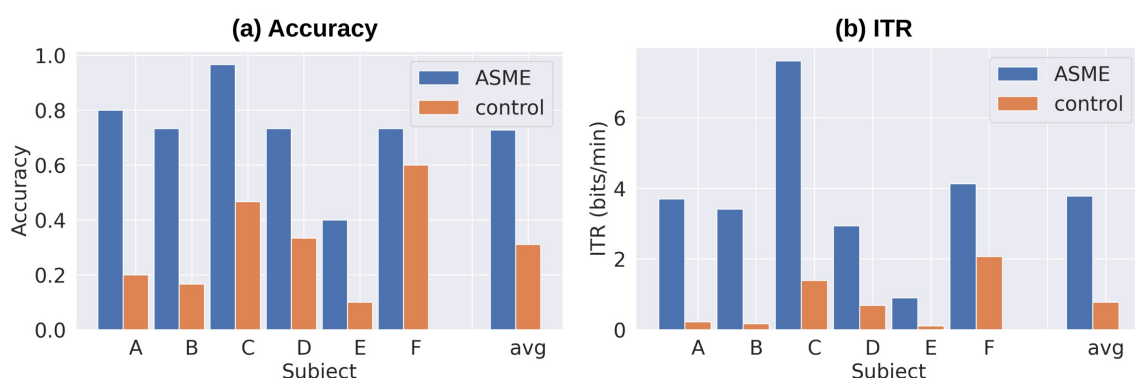


Figure 5: The results of the BCI simulation (estimating the target character of the trial) using dynamic stopping. (a) Accuracy and (b) ITR.

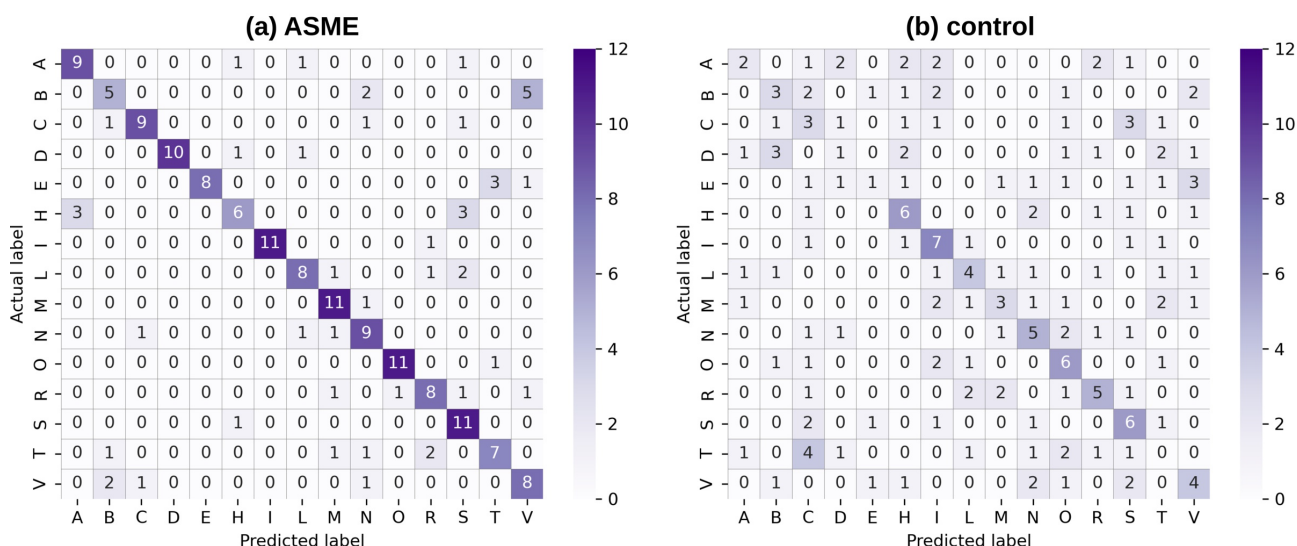


Figure 6: The confusion matrix for conditions (a) ASME and (b) control in the BCI simulation using dynamic stopping.

viding into groups by using auditory stream segregation, drastically improving the ASME-speller's performance. Furthermore, the target letter can be determined with a single trial, and no visual support is required if the users are familiar with the QWERTY layout. Additionally, All sound stimuli can be delivered with a monaural audio channel, solving the issues proposed by the auditory BCI

spellers. This system has the potential to be used by patients who have severe motor impairment or hearing impairment in one ear with high ITR, and it provides users with more choices of auditory BCI spellers.

#### ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP23K11811.

## REFERENCES

- [1] Wolpaw J *et al.* Brain-computer interface technology: A review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*. 2000;8(2):164–173.
- [2] Rao RPN. Brain-computer interfacing: An introduction. Cambridge University Press: Cambridge New York Melbourne New Delhi Singapore (2019).
- [3] Kundu S, Ari S. Brain-Computer Interface Speller System for Alternative Communication: A Review. *IRBM*. 2022;43(4):317–324.
- [4] Choi YJ, Kwon OS, Kim SP. Design of auditory P300-based brain-computer interfaces with a single auditory channel and no visual support. *Cognitive Neurodynamics*. 2023;17(6):1401–1416.
- [5] Furdea A *et al.* An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*. 2009;46(3):617–625.
- [6] Farwell L, Donchin E. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*. 1988;70(6):510–523.
- [7] Klobassa DS *et al.* Toward a high-throughput auditory P300-based brain-computer interface. *Clinical Neurophysiology*. 2009;120(7):1252–1261.
- [8] Schreuder M, Rost T, Tangermann M. Listen, You are Writing! Speeding up Online Spelling with a Dynamic Auditory BCI. *Frontiers in Neuroscience*. 2011;5.
- [9] Schreuder M, Blankertz B, Tangermann M. A New Auditory Multi-Class Brain-Computer Interface Paradigm: Spatial Hearing as an Informative Cue. *PLOS ONE*. 2010;5(4):e9813.
- [10] Bregman AS. Auditory Scene Analysis: The Perceptual Organization of Sound. The MIT Press (1990).
- [11] Kanoh S, Miyamoto Ki, Yoshinobu T. A brain-computer interface (BCI) system based on auditory stream segregation. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Aug. 2008, 642–645.
- [12] Kanoh S, Miyamoto Ki, Yoshinobu T. A Brain-Computer Interface (BCI) System Based on Auditory Stream Segregation. *Journal of Biomechanical Science and Engineering*. 2010;5(1):32–40.
- [13] Kanoh S, Kojima S. Evaluation of auditory BCI system based on stream segregation. In: Proceedings of the 8th Graz Brain-Computer Interface Conference 2019. Graz, 2019.
- [14] Kojima S, Kanoh S. Towards realizing multi-class auditory brain-computer interface paradigm based on stream segregation: A preliminary study. In: 2023 15th Biomedical Engineering International Conference (BMEiCON). IEEE: Tokyo, Japan, Oct. 2023, 1–5.
- [15] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components — A tutorial. *NeuroImage*. 2011;56(2):814–825.
- [16] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*. 2002;113(6):767–791.
- [17] Verschore H, Kindermans PJ, Verstraeten D, Schrauwen B. Dynamic Stopping Improves the Speed and Accuracy of a P300 Speller. In: *Artificial Neural Networks and Machine Learning – ICANN 2012*. Springer: Berlin, Heidelberg, 2012, 661–668.
- [18] Schreuder M, Höhne J, Blankertz B, Haufe S, Dickhaus T, Tangermann M. Optimizing event-related potential based brain-computer interfaces: A systematic evaluation of dynamic stopping methods. *Journal of Neural Engineering*. 2013;10(3):036025.
- [19] Höhne J, Tangermann M. Towards User-Friendly Spelling with an Auditory Brain-Computer Interface: The CharStreamer Paradigm. *PLOS ONE*. 2014;9(6):e98322.
- [20] Höhne J, Schreuder M, Blankertz B, Tangermann M. A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. *Frontiers in Neuroscience*. 2011;5.
- [21] Kleih SC, Herweg A, Kaufmann T, Staiger-Sälzer P, Gerstner N, Kübler A. The WIN-speller: A new intuitive auditory brain-computer interface spelling application. *Frontiers in Neuroscience*. 2015;9.
- [22] Markovinović I, Vrankić M, Vlahinić S, Šverko Z. Design considerations for the auditory brain computer interface speller. *Biomedical Signal Processing and Control*. 2022;75:103546.

# A NEW AUDITORY BRAIN-COMPUTER INTERFACE BASED ON STREAM SEGREGATION UTILIZING ASSR

Shin'ichiro Kanoh<sup>1</sup>, Naoki Mizukami<sup>1</sup>, Simon Kojima<sup>1</sup>

<sup>1</sup>Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan

E-mail: kanoh@shibaura-it.ac.jp

**ABSTRACT:** The authors have proposed an auditory brain-computer interface (BCI) based on stream segregation which detects users' selective attention to one the multiple segregated streams. In this system, several odd-ball sequences with different frequency bands were presented to users. To detect the target stream, this system needed to wait for the arrival of deviant stimuli in odd-ball sequences. In this study, auditory steady-state response (ASSR) was utilized to achieve a higher information transfer rate (ITR) system. Two streams consisting of sinusoidally amplitude-modulated (SAM) tones were presented to subjects, and they were requested to attend to one of the two streams. From the results of the electroencephalogram (EEG) measurement experiment, it was found that the user's selective attention enhanced ASSR corresponding to the modulation frequency of SAM sounds in the target tone stream, and the target stream which subject paid attention to was detected at an average accuracy of 0.77. The best accuracy was 0.92. It was concluded that an auditory BCI based on stream segregation utilizing ASSR is feasible, and it provides high performance and practical BCI options to users.

## INTRODUCTION

Brain-computer interfaces (BCIs) provide their users communication and control channels without using brain's normal output channels of peripheral nerves and muscles [1]. Since Vidal et al. [2] proposed electroencephalogram (EEG) based noninvasive BCI, it has attracted interests by a number of researchers.

Among BCIs, reactive BCI detects a stereotypical brain response elicited by sensory stimuli presented to users [3]. To date, a number of studies on visual-based BCI have been conducted since it is easy to present visual stimulus to users with a precise time resolution. As the visual modality is the most important sense in daily life, users can not do anything else when they use such visual-based BCIs. On the other hand, as auditory BCIs do not rely on visual modality, they might not disturb users' daily activities during operations.

The authors have proposed an auditory BCI based on stream segregation. Stream segregation is one of the illusory phenomena in auditory perception, on which multiple tone sequence with different frequencies which are alternately presented in time tend to be perceived as mul-

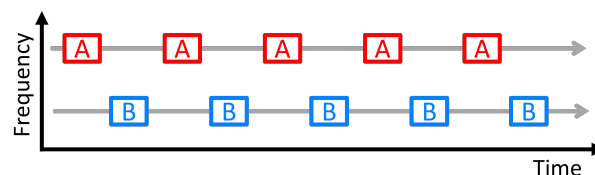


Figure 1: Schematic diagram of auditory stream segregation

multiple segregated sound streams. For example, when two kinds of sounds that have different frequencies (A and B) are presented alternately in time (ABABAB...), such a sound sequence can be perceived as two segregated sound streams (AAA... and BBB...) (Fig. 1).

The authors have proposed an auditory brain-computer interface (BCI) based on stream segregation which detects users' selective attention to one the multiple segregated streams [4, 5]. In these studies, two different odd-ball sequences perceived as segregated streams were presented to one ear of subjects, and they were requested to pay attention to one of the streams. It was shown that the target stream of selective attention could be estimated by detecting ERPs (event-related potentials) including P300 component which was elicited by the deviant stimuli embedded in the target stream. The authors proposed 2-class [4, 5], 3-class [6], and 4-class [7] BCI systems based on auditory stream segregation.

The feature of the auditory BCI based on segregation is that it operates in the frequency domain of the incoming monaural sound. As human auditory system have a rich capability to perceive pitch and melody of the sounds as well as location and movement in three dimensional auditory space, it is expected that such a system could enhance conventional auditory BCIs.

However, since such an auditory BCI detects P300 responses to the deviant stimuli in the oddball sequence perceived as stream, the system needs to wait for the deviant stimuli to detect the target of selective attention. As the frequency of the presentation of the deviant stimuli is low in oddball sequences, the average response time for P300 detection is long and it causes lower information transfer rate (ITR).

Therefore, in this pilot study, an auditory BCI system based on stream segregation without using oddball sequences was proposed and tested. In this proposed system, instead of P300 components, auditory steady-state

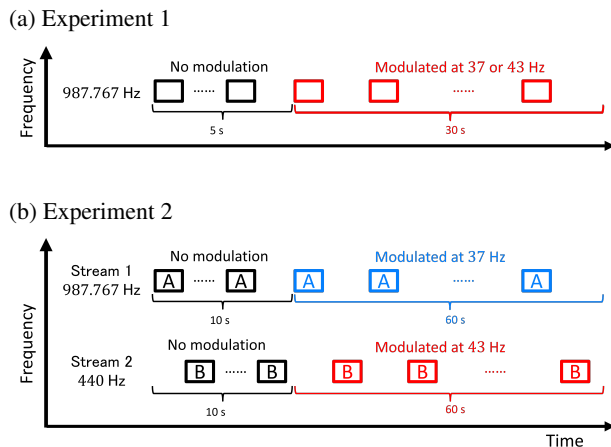


Figure 2: Time chart of presented stimuli used in Experiment 1 (a) and Experiment 2 (b)

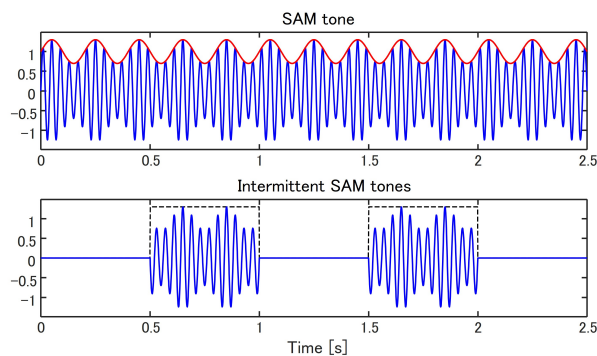


Figure 3: Schematic diagram of SAM tone and intermittent SAM tones. For visualization, the following parameters are used: tone (carrier) frequency 25 Hz, modulation frequency 5 Hz, duration of intermittent SAM tone 0.5 s. Red line shown in the upper figure is an envelope of the amplitude modulation.

response (ASSR) was used to detect user's selective attention to the segregated stream. ASSRs are elicited by temporally modulated auditory stimulation, such as an amplitude-modulated (AM) tone [8].

After starting the presentation of a sustained AM tone, EEG or magnetoencephalogram (MEG) rapidly entrains to the modulation frequency and phase of the stimulus [8]. Lopez et al. [9] reported that ASSR is modulated by selective attention and can be utilized as a BCI paradigm. If ASSR is elicited by segregated AM tone streams and is modulated by selective attention, P300 component can be replaced by ASSR as a new identifier of the target of selective attention on auditory BCI based on selective attention. And it is expected that the target of the users' attention can potentially be detected in a shorter time period than our proposed system [4, 5].

In this paper, the results of two experiments are shown. Experiment 1 is a preliminary experiment to confirm ASSRs elicited by the intermittent tones can be observed and detected. And in Experiment 2, the feasibility of the auditory BCI based on stream segregation utilizing ASSR was investigated.

## EXPERIMENT 1: ASSR TO AMPLITUDE-MODULATED INTERMITTENT TONE SEQUENCE

**Objective:** In the previous studies [4–7], sequence of intermittent short sounds perceived as segregated streams were presented to subjects. To apply ASSR to such an auditory BCI, it is required that ASSR is elicited by a sequence of intermittent short sound. However, it is general to measure ASSR by presenting long lasting AM tones to subjects, and whether ASSRs are evoked using intermittent short sound stimuli has not been evaluated. Hence, in Experiment 1, the ASSRs to amplitude-modulated intermittent sound stimuli were evaluated as a preliminary experiment.

**Methods:** Three males participated in this experiment. The subjects sat on a comfortable chair in a soundproofing electromagnetically shielded room, and 9-channel EEG (F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4) were recorded with passive Ag-AgCl electrodes (Easy-cap, Easycap GmbH, Germany). Reference and ground electrodes were placed on the right and left ear mastoid, respectively. EEG signal was amplified and recorded with a biosignal amplifier (BrainAmp MR plus, Brain Products, Germany) at a sampling frequency of 1000 Hz. Before sampling, bandpass filter (0.1–100 Hz) was applied to the recorded data.

Fig. 2(a) shows the time chart of the tone sequence presented to subjects in Experiment 1. Sequence of tones (duration 180 ms) was presented to subjects, and they were requested to listen to the presented tones. The stimulus onset asynchrony (SOA) was set to 400 ms.

The length of a trial was 35 s. Frequency of each tone was set to 987.767 Hz (musical pitch B5). The first 25 tones (10 s) were pure tone without amplitude modulation, and the following 75 tones (30 s) were sinusoidally amplitude-modulated (SAM) tones (schematic diagram is shown in Fig. 2). Two kinds of modulation frequency (37 Hz or 43 Hz) were tested in different trials.

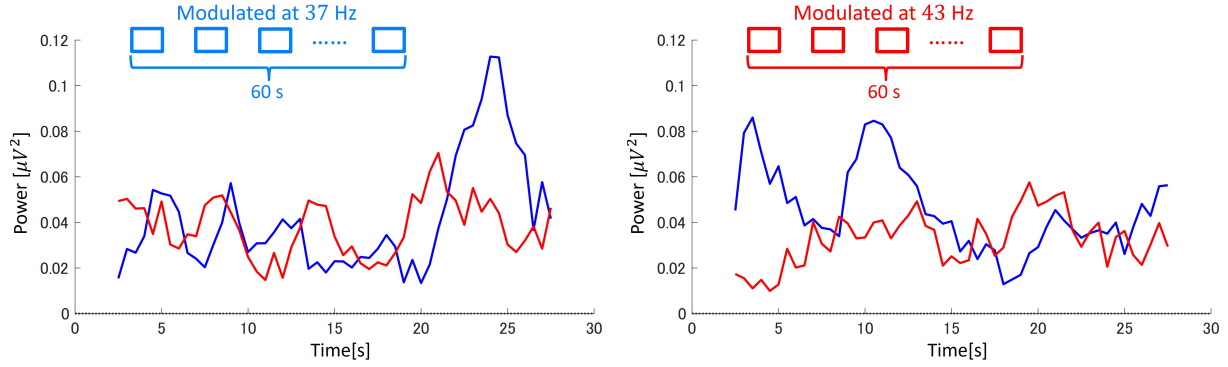
All stimuli were generated by MATLAB (Mathworks, USA). Tone stimuli were presented to subjects' left ear by an audio interface (Fireface 802, RME, Germany) and headphones (HDA200, Sennheiser, Germany).

MATLAB (Mathworks, USA) and EEGLAB [10] were used for analysis. The recorded EEG signal was bandpass filtered at 20–150 Hz.

The data of 30 s during which SAM tones were presented (colored red in Fig. 2(a)) was analyzed by the following way. Additionally, EEG data with removing gaps (EEG data during presentation of each SAM tone was extracted and concatenated, total duration 13.5 s) was also analyzed and results were compared.

EEG signal was segmented into 5 s intervals with an overlap of 4.5 s. Then, canonical correlation analysis (CCA) was applied to the segmented data. The reference signal  $Y(f_i)$  was set as follows.

(a) Using the whole period of EEG data



(b) Using the EEG data with removing gaps

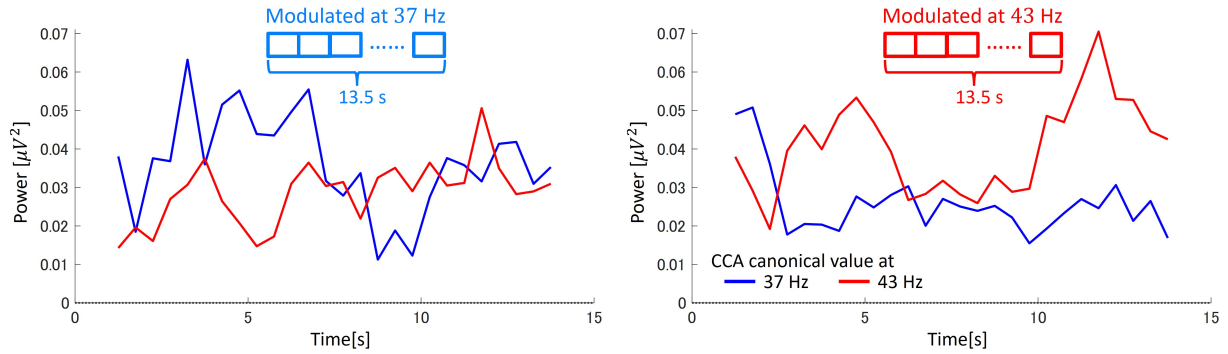


Figure 4: An example of the time courses of the power spectrum value of CCA canonical variables at 43 (red) and 37 Hz (blue) (Subject A). Top row: results of analysis using the whole period of EEG data (30 s). Bottom row: results of analysis of EEG data with removing gaps (13.5 s). The results when this subject was presented SAM tones modulated at 37 Hz (left) and 43 Hz (right) are shown.

$$\mathbf{Y}(f_t) = \begin{bmatrix} \sin(2\pi f_t t) \\ \cos(2\pi f_t t) \\ \sin(4\pi f_t t) \\ \cos(4\pi f_t t) \\ \sin(6\pi f_t t) \\ \cos(6\pi f_t t) \end{bmatrix} \quad t = \frac{1}{f_s}, \frac{2}{f_s}, \dots, \frac{N_s}{f_s} \quad (1)$$

where,  $f_s$  is the sampling frequency,  $N_s$  is the number of samples of the 5 s segmented data, and  $f_t \in \{37, 43\}$  Hz is the target frequency to be detected by CCA, since SAM tones were modulated at 37 or 43 Hz. Fast Fourier transform (FFT) was applied to canonical variable obtained by CCA, and the time course of the power spectrum value at each target frequency  $f_t$  was obtained. If the target frequency to be detected is  $f_d$ , corresponding power spectrum value if FFT power spectrum of canonical variable using reference signal  $\mathbf{Y}(f_t)$  at frequency  $f_t$ .

**Results:** Fig. 4 shows the time courses of the power spectrum value of CCA canonical variables at 37 and 43 Hz of Subject A.

In Fig. 4 (a), the results of analysis using the whole period of data (30 s) are shown. And the responses during the time range when subjects were presented intermittent sounds were discarded are shown in Fig. 4 (b).

When the data from the whole trial was analyzed, there was no clear difference between the power spectrum values of CCA components at two modulation frequencies

(Fig. 4 (a)). However, the CCA power values at corresponding modulation frequency were increased if the segments when the sounds were delivered to this subject were concatenated and analyzed (Fig. 4 (b)). The same tendencies were also observed on the other two subjects, however, the time course of the power spectrum values of CCA components were varied across subjects.

As a result, despite the short length of each SAM tone (180 ms), ASSRs were evoked while the stimuli were played. Thus, it can be concluded that intermittent short tone can also evoke ASSR. It was also confirmed that the power of ASSR corresponding to the modulation frequency of presented tones was increased, and it could be observed clearly if the data was analyzed when the data during the presentation of each intermittent sound was extracted and concatenated.

## EXPERIMENT 2: DETECTION OF SELECTIVE ATTENTION TO SEGREGATED STREAM OF SAM TONES

**Objective:** In this experiment, tone sequence of SAM tones, which were perceived as two segregated tone streams, were presented to subjects, and pattern classification to detect subjects' target of selective attention was estimated offline. This experiment aimed to evaluate the feasibility of the auditory BCI paradigm based on stream segregation utilizing ASSR.

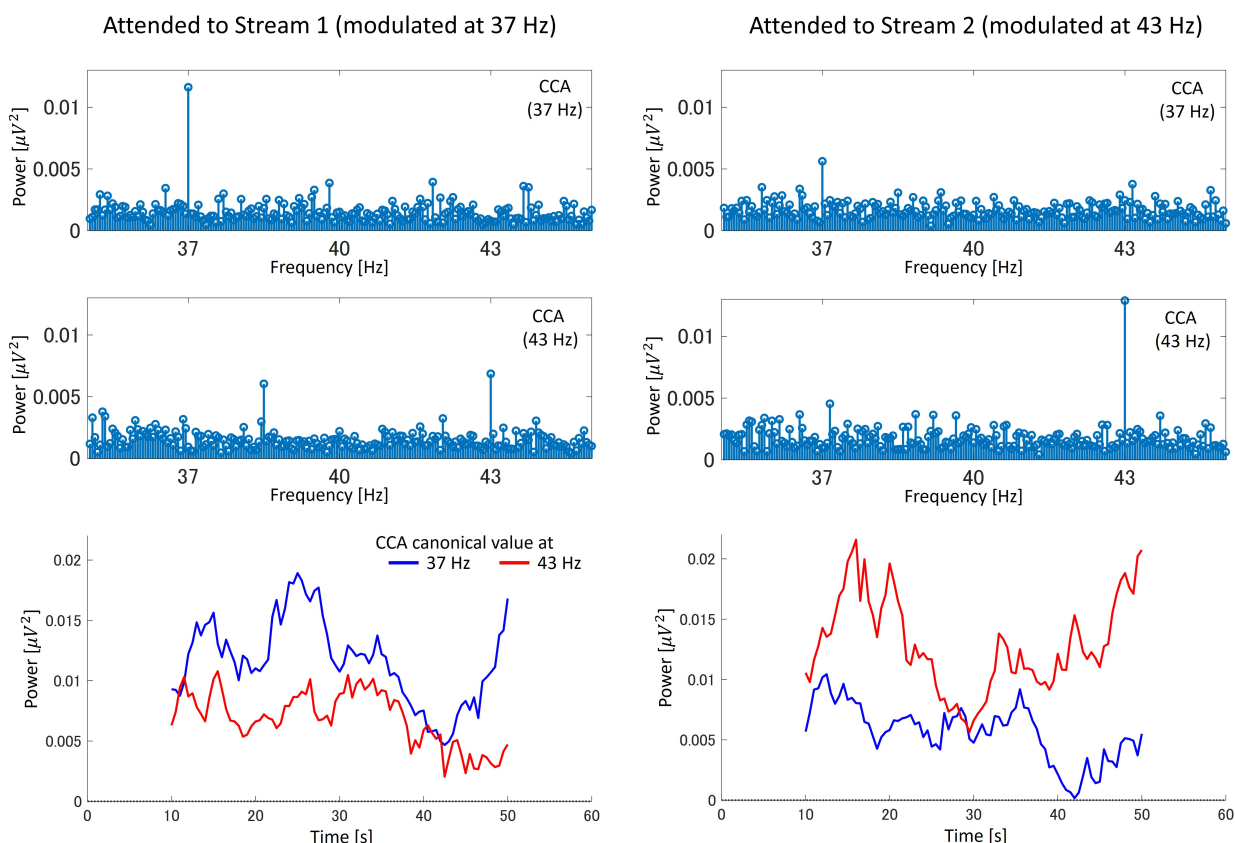


Figure 5: An example of the power spectrum of the CCA canonical variable on 37 and 43 Hz, and time courses of the power spectrum value at 37 (blue) and 43 Hz (red) which a time windows of 20 s (Subject B). Results when the subject pays attention to and Stream 1 (modulation frequency 37 Hz) and Stream 2 (43 Hz) are shown in left and right column, respectively.

**Methods:** Eight male subjects (aged between 21–23) participated in this experiment. The subjects sat on a comfortable chair in a soundproofing electromagnetically shielded room, and 64-channel EEG (Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT9, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, FT10, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP9, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, and O2) were recorded with passive Ag-AgCl electrodes (Easycap, Easycap GmbH, Germany). Other conditions for measurement were the same as in Experiment 1.

Two kinds of SAM tones, 987.767 Hz (musical pitch B5) pure tones amplitude-modulated at 37 Hz and 440 Hz (musical pitch A4) of pure tones amplitude-modulated at 43 Hz were used. The duration of the tones was 180 ms. Fig. 2(b) shows the time chart of the sequence used in this experiment. The stimulus onset asynchrony (SOA) was set to 200 ms. All stimuli were presented to subjects' left ear by an audio interface (Fireface 802, RME, Germany) with headphones (HDA200, Sennheiser, Germany). The length of a trial was 70 s, and the subjects were requested to pay attention to the instructed stream. The unmodulated tones were presented for the first 10 s. MATLAB (Mathworks, USA) and EEGLAB [10] were

used for analyses. The recorded EEG signals were band-pass filtered in a range of 20–150 Hz. The data of 60 s during which SAM tones were presented was used. Based on the result of Experiment 1, EEG data during gap period was removed and concatenated before analysis. To find the optimal length of the window for analysis, four window lengths (5, 10, 15, and 20 s with the overlap of 4.5 s, 9.5 s, 14.5 s, and 19.5 s, respectively) were used for segmentation. Then, CCA was applied to the segmented data. The same reference signals as Experiment 1 were used.

The subjects' target of selective attention was estimated with linear discriminant analysis (LDA). Two features (the power spectrum of CCA canonical variables at 37 and 43 Hz) of each data segment were used for classification. The classification accuracy was evaluated by 5-fold cross-validation.

**Results:** Fig. 5 shows an example of the result from Subject B. The power spectrum of the CCA canonical variable on  $f_t = 37$  and 43 Hz, and time courses of the power spectrum value at 37 (blue) and 43 Hz (red) which a time windows of 20 s are shown. It was shown that the CCA canonical variable at the modulation frequency of the attended stream was larger than that of the unattended stream. On 5 subjects out of 8, the same tendency was observed. However, like in Experiment 1, the time course

Table 1: Classification accuracy for each length of time window.

Subject	5 s	10 s	15 s	20 s
A	0.65	0.64	0.72	0.73
B	0.69	0.72	0.82	0.92
C	0.66	0.70	0.73	0.84
D	0.56	0.65	0.77	0.83
E	0.56	0.63	0.67	0.80
F	0.64	0.62	0.70	0.68
G	0.55	0.58	0.68	0.72
H	0.57	0.54	0.57	0.62
Average	0.61	0.64	0.71	0.77

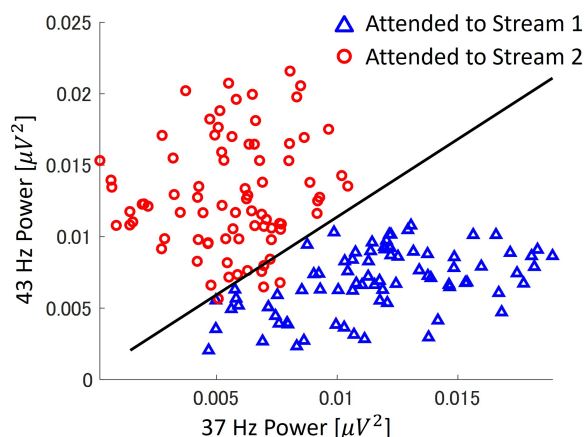


Figure 6: An example of the feature distributions with 20s of the time window (Subject B)

of the power spectrum values of CCA components were varied across subjects.

Table 1 shows the classification accuracy for each length of time window. When 20 s time window was used, the average accuracy reached 0.77. Fig. 6 shows the feature distribution of the data from subject B using 20 s time window. Feature vectors of the two classes were distributed differently in feature space and could be discriminated by linear function.

## DISCUSSION

When subjects paid attention to one of the streams consisting of SAM tones, ASSR at modulation frequency was increased, and the subjects' target of selective attention to the stream could be detected with high accuracy by the machine-learning approach. It was proved that the auditory BCI based on stream segregation utilizing ASSR is feasible. Compared to the previous study which detects P300 component by using oddball sequence as presented tone stimuli, the ASSR-based paradigm doesn't need to wait for the deviant stimuli to be presented. Furthermore, it does not require the precise timing of stimulus onset, while the oddball-based paradigm requires accurate timing of stimulus onset. Therefore, it is expected that an ASSR-based BCI using stream segregation can realize higher ITR with a simple setup.

In this pilot study, classification accuracy was below 0.7 when the length of time window to observe power spec-

trum of CCA canonical variable was 10 s (Table 1). Because it was needed to observe EEG data for the width of time window, power spectrum values could not be calculated at the beginning of the data (on offline analysis in this study, power spectrum values were lacked for half width of the time window at the beginning and at the end of the EEG data. See Figs. 4 and 5).

Two different target frequencies  $f_i$  could be discriminated by relative power spectrum values of ASSR at corresponding the modulation frequencies (37 or 43 Hz) of SAM tones (Fig. 6). It was shown that the values of the feature vector (i.e. power spectrum value of CCA canonical variable at two modulation frequencies) changed over time during attending to the target segregated stream (Figs. 4 and 5), and it was different between subjects. Such phenomenon might reflect the temporal change of the degree of attention on each individual subject.

However, to improve both classification accuracy and ITR, optimization of feature extraction of EEG using CCA is required. This is left for the further study.

## CONCLUSION

An auditory BCI system based on stream segregation of SAM tone sequence was proposed and tested. In this proposed system, instead of P300 components[4–7], ASSR was used to detect user's selective attention to the segregated stream. It was found that the power spectrum of ASSR at the modulation frequency of the target SAM tone stream is increased by selective attention, which is perceived as segregated streams by stream segregation. This result shows that an auditory BCI based on stream segregation utilizing ASSR is feasible.

With this approach, the target stream can be estimated without waiting for special stimuli (e.g., deviant stimuli in oddball sequence), and it can achieve high ITR in principle. Furthermore, as opposed to an ERP-based system, precise timing of stimulus presentation is not required, and it is beneficial for making the system simpler. In principle, it will lead the auditory BCI based on stream segregation to high-performance and more practical and offers users more options for auditory BCIs.

Further investigations on feature extraction and optimization of presenting auditory tone stimuli are needed to improve the performance of the present auditory BCI system.

## ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP23K11811.

## REFERENCES

- [1] Wolpaw J *et al.* Brain-computer interface technology: A review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*. 2000;8(2):164–173.

- [2] Vidal JJ. Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering*. 1973;2(1):157–180.
- [3] Rao RPN. *Brain-Computer Interfacing: An Introduction*. Cambridge University Press: Cambridge (2013).
- [4] Kanoh S, Miyamoto Ki, Yoshinobu T. A brain-computer interface (BCI) system based on auditory stream segregation. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Aug. 2008, 642–645.
- [5] Kanoh S, Miyamoto Ki, Yoshinobu T. A Brain-Computer Interface (BCI) System Based on Auditory Stream Segregation. *Journal of Biomechanical Science and Engineering*. 2010;5(1):32–40.
- [6] Kanoh S, Kojima S. Evaluation of auditory BCI system based on stream segregation. In: *Proceedings of the 8th Graz Brain-Computer Interface Conference 2019*. Graz, 2019.
- [7] Kojima S, Kanoh S. Towards realizing multi-class auditory brain-computer interface paradigm based on stream segregation: A preliminary study. In: 2023 15th Biomedical Engineering International Conference (BMEiCON). IEEE: Tokyo, Japan, Oct. 2023, 1–5.
- [8] O'Donnell BF, Vohs JL, Krishnan GP, Rass O, Hetrick WP, Morzorati SL. Chapter 6 - The auditory steady-state response (ASSR): A translational biomarker for schizophrenia. In: *Supplements to Clinical Neurophysiology*. Elsevier, Jan. 2013, 101–112.
- [9] Lopez MA, Pomares H, Pelayo F, Urquiza J, Perez J. Evidences of cognitive effects over auditory steady-state responses by means of artificial neural networks and its use in brain-computer interfaces. *Neurocomputing*. 2009;72(16):3617–3623.
- [10] Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. 2004;134(1):9–21.

# A NOVEL CHATGPT-DRIVEN COMMUNICATION AID BASED ON CODE-MODULATED VISUAL EVOKED POTENTIALS (CVEP)

Atilla Cantürk, Ivan Volosyak

Faculty of Technology and Bionics, Rhine-Waal University of Applied Sciences, Kleve, Germany

E-mail: [ivan.volosyak@hochschule-rhein-waal.de](mailto:ivan.volosyak@hochschule-rhein-waal.de)

**ABSTRACT:** Brain-computer interface (BCI) systems, including applications based on visual evoked potentials (VEPs), have proven to provide reliable and accurate control. In recent years, communication has remained one of the main application areas of modern BCIs, with a lot of advancements based e.g., on the incorporation of dictionary support and text prediction. This study explores the integration of BCIs with artificial intelligence (AI), specifically focusing on the development and evaluation of an innovative spelling interface powered by the ChatGPT application programming interface (API). Aimed at enhancing communication for individuals with severe motor impairments, this interface combines the precision of code-modulated visual evoked potentials (cVEPs) with the predictive capabilities of AI to offer a more intuitive and efficient user experience. The performance of 13 healthy participants (10 females) was evaluated in an online experiment. The participants successfully completed all spelling tasks using the cVEP BCI with aid from ChatGPT, achieving a mean information transfer rate (ITR) of 33.16 bpm, a mean accuracy of 87.49%, and an average output of 8.74 output characters per minute (OCM) for unique sentence tasks. This was slower than in our previous research using an n-gram model which achieved 18.9 characters per minute.

## INTRODUCTION

A BCI system detects, analyzes and decodes brain activity in real time to provide communication with the external environment, without involving normal output pathways of the human nervous system or muscle activities [1]. Modern BCIs can be used as communication tools for severely impaired people suffering for example from spinal cord injuries, brain stem strokes, amyotrophic lateral sclerosis (ALS), or muscular dystrophies. For the practical use of such BCI applications their accuracy and speed are the most important factors. Visual evoked potentials (VEPs) allow the fastest BCI realisation, between them the code-modulated VEPs (cVEPs), where all stimuli are modulated with different time lags of the same code sequences, typically the m-sequences, yield potentially higher accuracies [2]. Further increase in the communication speed, usually measured in terms of information transfer rate (ITR) is possible by using word prediction methods, particularly based on n-gram

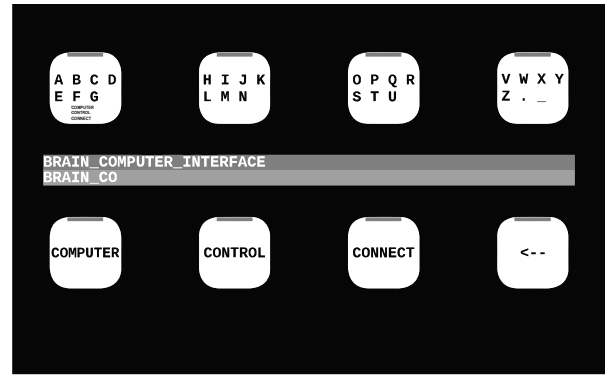
models. In our previous paper [3], c-VEP based BCI system was further extended by several methods for enhanced target identification, including dynamic sliding windows and software-based stimulus synchronization, coupled with an ensemble-based classification. Integrating a dictionary-driven n-gram word prediction model, the system demonstrated improved usability, with significantly better results when the dictionary integration was used. Unfortunately, this software implementation of the signal processing and the dictionary support in form of a single custom-made application limits the transfer of this code into newer BCI applications, necessitating the re-development of the dictionary support. Another issue is that until recently, it was not technologically feasible to accurately reproduce sentences generated using such communication aids, especially concerning the proper verb declension, which requires a complete understanding of the sentence's information content. However, with the advent of new AI-based language systems, it is now possible to bridge this gap and produce stylistically and grammatically correct sentences.

In recent years, AI has garnered significant attention across various domains, revolutionizing the way we interact with technology and transforming traditional workflows. From chatbots for libraries [4] to recommendation system for farms [5], AI-driven solutions have demonstrated remarkable capabilities in understanding and processing human language. The combination of BCIs and AI marks a significant shift in human-computer interaction, especially for individuals with severe motor impairments. Advanced AI models like ChatGPT enhance this synergy, revolutionizing interactions from healthcare to customer service. Its applications, aiding clinical diagnoses to supporting medical education, highlight AI's utility, though ethical and legal considerations accompany it [6]. The introduction of ChatGPT has sparked a robust debate over its potential applications and limitations, underscoring the need for a nuanced exploration of AI's role in healthcare and medical research.

This paper presents an innovative spelling interface that leverages the ChatGPT API, demonstrating the seamless integration of cVEP-based BCIs with AI to create a more intuitive and efficient communication tool. By examining the advantages, limitations, and effects of employing ChatGPT and AI in such interfaces, alongside their practical applications and future prospects in medicine and



(a) "Training" mode, awaiting initiation.



(b) Online "Two-Steps Speller" interface with dictionary hints (top right).

Figure 1: Fig. 1a shows 1<sup>st</sup> selection field highlighted in green and waiting for "F1" key press to start the training session. Once training begins, all text outside of the selection fields will be hidden. In Fig. 1b, API word suggestions appear in selection fields 5 to 7, with dictionary hints also shown in the 1<sup>st</sup> selection field (placed inside the field, under the letter groups).

healthcare, we aim to contribute to the ongoing discourse on the responsible and effective use of AI technologies in enhancing human-computer interaction and healthcare outcomes.

## MATERIALS AND METHODS

**Participants:** 13 participants (10 females) participated in this study; the average age of the subjects was 24.46 years, with a standard deviation (SD) of  $\pm 4.2$ . All participants provided written consent in adherence to the Declaration of Helsinki, and the study received approval from the ethical committee of the medical faculty at the University Duisburg-Essen. The collected data for analysis purposes were stored anonymously, ensuring the confidentiality of the participants. Subjects received compensation for their involvement in our study. Since this was not addressed in the ethical approval, the EEG data cannot be published.

**Hardware:** The computer in use was a Dell Precision Desktop with NVIDIA RTX3070 graphics card that operated on Microsoft Windows 10 (21H2) Education running on an Intel processor i9-10900K (3.70 GHz). For the purpose of presenting the stimuli, a modern display (Asus ROG Swift PG258Q, Full-HD, 240 Hz maximal refresh rate) was used.

An EEG amplifier (g.USBamp medical engineering GmbH, Schiedlberg, Austria) was used, utilizing all 16 signal channels which were placed according to the international system of EEG electrode placement at positions: P7, P3, Pz, P4, P7, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, O9, Iz and O10. Additionally, the reference electrode was positioned at Cz, while the ground electrode was placed at AFz. During the preparation stage, regular abrasive electrolytic electrode gel was used between the electrodes and the scalp to reduce impedances to less than 5 k $\Omega$ .

**GUI:** The graphical user interface (GUI) is illustrated in Fig. 1b. An eight target spelling interface as presented in [3] was utilized. Selecting individual charac-

ters required exactly two steps ("two-steps speller"). The graphical user interface (GUI) was designed with its first row featuring 28 characters, including the 26 letters of the alphabet, an underscore, and a full-stop character, organized into four boxes, each containing seven characters. The second row provided three suggestions generated by the ChatGPT API (dictionary suggestion boxes), along with an option for correction. Utilizing the correction option allowed users to delete the last typed character or word, enhancing the typing experience by integrating both predictive text and error correction functionalities. By selecting a letter group from the first row, the associated characters were presented individually in the "second step". The GUI includes dictionary hints that present the same recommendations as the ChatGPT API at the bottom of the lastly selected letter, facilitating easier selection without the need to divert attention. An illustration of this feature is provided in Fig. 1b (top-left corner selection field). Each selection triggers audio and visual feedback (the selected field briefly enlarges and turns green for correct selections, or turns red for incorrect selections).

**Stimulus Presentation:** The stimuli targeted in the experiment were comprised of eight selection fields (boxes) ( $230 \times 230$  pixel) arranged as  $2 \times 4$  selection field matrix (see Fig. 1). 63 bit m-sequences  $c_i$ ,  $i = 1, \dots, K$  ( $K = 8$  for our case) were assigned to the selection field matrix employing a circular shift of 4 bits ( $c_1$  had no shift,  $c_2$  was shifted by 4 bits to the left,  $c_3$  was shifted by 8 bits, etc.). The codes were allocated to the matrix in a row-wise manner, beginning with the upper left target labeled as  $c_1$ , and subsequent targets were labeled following a row-major sequence. The stimuli linked to the codes switched between "black" (the background color, denoted by "0") and "white" (indicated by "1"). Here,  $c_1$  was defined as

$$c_1 = 10101100110111011010010011100010 \quad (1)$$

$$11110010100011000010000011111110$$

The duration of a stimulus cycle in seconds can be calculated by dividing the code length by the monitor refresh

rate  $r$  in Hz; in this experiment,  $63/60 = 1.05s$  (the used refresh rate was 240 Hz, so the stimulus changed in accordance with the bit sequence, but for every fourth frame). Spatial filters were developed using the information gathered during the recording phase for classification. Canonical correlation analysis (CCA) was used on the training trials in this regard. Further details about used cVEP signal processing methods can be found in [2].

**Training:** During the recording phase, eight stimuli were observed sequentially from 1 to 8 by the participants, as illustrated in Fig. 1a. The recording was divided into six blocks of training, denoted as  $n_b = 6$ . Within each block, every stimulus was focused on once, resulting in a total of  $6 * 8 = 48$  trials. Each trial lasted for 2.1 seconds, during which the code pattern was displayed for two cycles. A visual cue, represented by a green frame, indicated the specific box towards which participants were required to direct their gaze. Following each trial, the subsequent field the user needed to focus on was highlighted, and the flickering paused for one second. After completing each block of eight trials (all eight targets), the software transitioned to the next block of training, with a one-second pause until 48 trials were accomplished.

**ChatGPT API:** The ChatGPT application programming interface (API) works by sending a prompt, typically a piece of text or a question, to the API endpoint. The model then processes this input and generates a response that continues the conversation or provides relevant information based on the context provided in the prompt.

In order to be able to construct API requests, we modified our C++ based software with the help of the documentation provided by OpenAI [7]. We used libcurl to enable our software to be able to communicate with the API and used a JSON parser to make it easier to work with the response from the API. We made adjustments to certain parameters to facilitate the use of the API. For detailed information on the parameters modified and their specific functions, please refer to Table 1.

The word completion algorithm functions by identifying words demarcated by underscore (“\_”) characters. When a user selects a suggested word via the dictionary buttons (see Fig. 1b), the algorithm updates the text by replacing the characters entered after the most recent underscore character with the selected suggestion. When the user selects the correction button after choosing a suggested word, the software restores their manually typed text to its original form before the suggestion was applied.

Once user types a letter, the typed sentence is added into the API request and sent. Once the response is received the contents are extracted and separated into 3 different words and pushed into the dictionary suggestion boxes and also into dictionary hints text, giving the user the chance to type the word recommended by ChatGPT.

**Experimental Protocol:** The experiment was conducted in the BCI-Lab of Rhine-Waal University of Applied Sciences (HSRW). Firstly participants filled a questionnaire with questions regarding their experience with

Table 1: Key ChatGPT API Parameters

<b>Model:</b> “gpt-3.5-turbo-0125”. Chosen for its low cost and high speed.
<b>Instruction:</b> “We are trying to realise the speller. Always return suggestions for just the most likely last word of the query starting with letters of the query, having in mind previous words of the query. Do not return previous words of the query. Always return only three comma-separated words, no further information, no line skips.” Customized for spelling suggestions (52 tokens). This parameter guides the model on how to generate its response depending on the input prompt.
<b>Max tokens:</b> Set to 100. Max tokens parameter represents the maximum number of tokens that can be generated in the chat completion.
<b>Presence_penalty and Frequency_penalty:</b> Set to -0.1 to fine-tune response variability. Both parameters slightly increase the likelihood of repeating information, promoting a less random output.
<b>Top_p:</b> Adjusted to 0.1 for focused response generation. This parameter controls the diversity of the model’s responses by limiting the probability mass considered for sampling the next token. Setting it to 0.1 ensures that the model’s outputs are highly focused and relevant to the given prompt, by choosing the response from the top 10%.

BCI systems. Participants were also asked to provide insights into their experiences with the BCI technology and describe their level of fatigue prior to initiating the study; their answers were recorded. Then, participants were briefed about the procedure and the operation of the speller. Following these explanations, participants engaged in a preliminary test run to accustom themselves to the speller, during which they freely composed a sentence of their choice and got familiar to the use of ChatGPT word recommendations. The threshold, gaze shift, and time window settings were calibrated as necessary during the familiarization. During this study, participants were told to spell the words “BCI”, “KLEVE” then spelled the pangram “THE\_QUICK\_BROWN\_FOX\_JUMPS\_OVER\_THE\_LAZY\_DOG”. After successfully finishing the spelling of the pangram, a unique sentence for each participant was randomly chosen from a pool of sentences that were inspired by news article titles.

After successfully completing the spelling session, participants completed the post-questionnaire containing questions regarding their impressions, opinions and their experience towards the BCI systems. The questions regarding their experience composed of questions regarding the flickering lights and how it affected them and questions regarding the effect of the assistance from the API and the dictionary hints functionality.

Spelling phases concluded automatically upon correct word spelling. On average, each subject’s spelling session (just spelling) lasted 20 to 25 minutes. Resulting accuracy, ITR, and OCM values were recorded for all com-

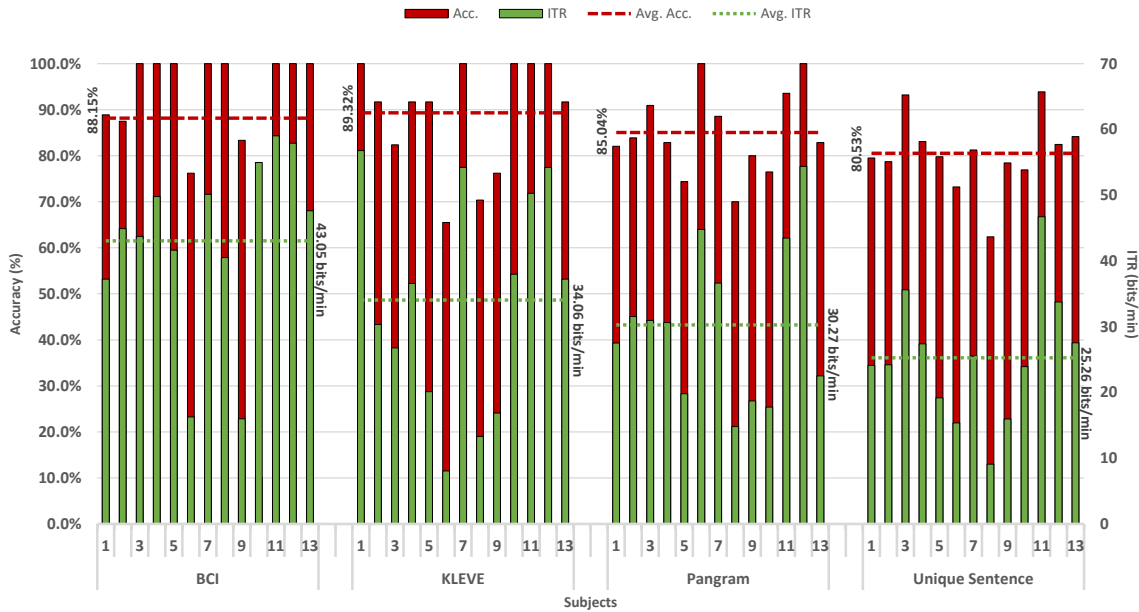


Figure 2: Individual spelling task's accuracy, Information Transfer Rate (ITR) and their average values are presented in sequence. Participants were assigned the following typing tasks: "BCI", "KLEVE", "THE\_QUICK\_BROWN\_FOX\_JUMPS\_OVER\_THE\_LAZY\_DOG" and a unique sentence for each participant.

pleted tasks.

**Evaluation Measures of the BCI Performance:** The BCI system's performance was evaluated using commonly used accuracy (Acc.), Information Transfer Rate (ITR), and in form of the output characters per minute (OCM).

**Accuracy:** The accuracy was calculated by dividing the total number of correct selections (word completions were considered a single command), including user-necessary corrections during speller execution, by the overall commands classified. The resulting accuracy value was displayed as a percentage value on the speller interface.

**ITR:** The Information Transfer Rate (ITR) was calculated in bits per minute (bits/min) using the formula:

$$B = \log_2 N + P \log_2 P + (1 - P) \log_2 \left[ \frac{1 - P}{N - 1} \right], \quad (2)$$

where:

$B$  = information transferred in bits,

$N$  = number of targets (for this study it is equal to 8),

$P$  = classification accuracy.

To obtain the ITR in bits/min,  $B$  is multiplied by the average classification time in minutes. For more information and tools to calculate ITR, visit our webpage: <https://bci-lab.hochschule-rhein-waal.de/en/itr.html>.

**OCM:** The Output Characters per Minute (OCM) measures typing speed by dividing the total number of output characters by the time taken to type them. OCM accounts for error correction time, as participants will require additional time for corrections if mistakes are made.

**Evaluation of the Questionnaire:** A questionnaire was designed to collect participant feedback, with sections dedicated to both pre-experiment and post-experiment questions. These sections are intended to be completed respectively before and after the experiment, focusing on assessing user experience and the improvements ChatGPT has made to user comfort. For further information, refer to Table 2, which outlines these pre- and post-experiment questions.

## RESULTS

The results indicating BCI performance are shown in the Fig. 2 and Fig. 4. Fig 2 illustrates the ITR and accuracy values achieved by participants, along with average values per task. As tasks lengthened, average ITR decreased, and sentence accuracies were lower compared to single-word tasks. Fig 4 displays the output characters per minute (OCM) values and their averages per task, revealing that the average OCM values are lower for the first two tasks compared to the last two. This difference is attributed to ChatGPT's inability to predict these words, likely due to their uncommon usage. However, ChatGPT notably enhanced participants' performance in the final two tasks compared to its performance in previous tasks. This was especially noticeable in the pangram task, where the average Output Characters per Minute (OCM) increased dramatically, more than tripling from 5.72 characters per minute to 17.72 characters per minute, having in mind the total numbers of spelled characters.

Results from the questionnaires indicate that eight out of 13 participants felt more tired after the experiment, while the rest reported no change in their fatigue levels. Four out of 13 considered the flickering disturbing. Majority

Table 2: Used Questionnaires.

<i>Pre-Questionnaire</i>
Have you ever used a BCI system? If yes, please add some information about it.
Do you have a vision prescription? If yes, are you wearing a reading aid now?
How tired do you feel right now? 1: not at all, 6: very much
How many hours did you sleep last night?
<i>Post-Questionnaire</i>
How tired do you feel right now? 1: not at all, 6: very much
Did you find the flickering disturbing? 1: not at all, 6: very much
Was it easy for you to concentrate on the boxes? 1: not at all, 6: very much
Did dictionary hints improve word completion? 1: not at all, 6: very much
Do you prefer the speller with or without ChatGPT aid? With / Without
Would you repeat the experiment? Yes / No / Maybe
Could you use the system daily? Yes / No / Maybe
In your opinion, how long can the system be used without breaks?
What was the unique sentence you had to type?
Do you think the BCI is a reliable control method? Yes / No / Maybe

found concentrating on the boxes easy. Everyone preferred spelling with the aid of ChatGPT and 12 out of 13 found dictionary hints helpful. 12 out of 13 participants reported that they would like to take the experiment again. On average, participants reported that they could use the system for approx. 1.2 hours. Majority of the participants found dictionary hints helpful. Findings related to fatigue levels are presented in Figure 3. Eight participants had no vision prescription, two had prescriptions but opted not to use any corrective wear, and the remainder used their prescribed vision aids.

## DISCUSSION

The general use of digital technologies owned by private companies and located overseas raise many data protection, ethical, safety and security questions. E.g., it is well known that OpenAI has recently removed accounts of hacker groups from China, Russia, North Korea, and Iran. The use of ChatGPT as a language model for the BCI purposes is of course not comparable to this example, but, on the other hand, texts produced by the target group of users with disabilities, a most vulnerable group, need much more privacy and require careful ethical considerations.

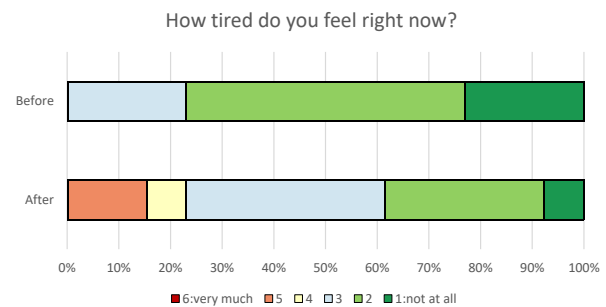


Figure 3: Participant feedback on fatigue levels before and after the experiment, measured on a scale from 1 (not fatigued at all) to 6 (very much fatigued).

We encountered several challenges during the experiments and development phases. One notable issue was ChatGPT's occasional difficulty in accurately following instructions, leading to incorrect recommendations. For instance, in some instances, instead of generating the complete word as instructed, it would only output the missing part of the word, despite clear instructions to do the opposite. We tested many different instructions to minimize this issue, but there might be of course a better instruction set. Using a different ChatGPT version will also likely require a completely different set of queries. We fine-tuned several supplementary parameters (see Tab. 1), to refine and control the API's behavior. We should note that although ChatGPT occasionally failed to adhere strictly to instructions, it sometimes succeeded in enhancing typing speed by suggesting corrections and even predicting the next word before the user began typing it. This behavior became evident to us during some instances when participants were reciting pangrams (see Fig. 4), where the OCM values were the highest. Additionally, it demonstrated the ability to switch languages seamlessly, without needing explicit commands (for instance, recommending German words upon typing "KLEVE"), while still managing to follow instructions to a satisfactory extent. Another challenge encountered was occasional unresponsiveness of the software due to high network traffic impacting the ChatGPT, leading to delays while awaiting responses. For best performance it is recommended to have a good network connection and the API status should be checked (<https://status.openai.com/>). Following an extended period of time spent attempting to type the initial words "BCI" and "KLEVE", two participants opted to withdraw from the experiment (therefore, in total 15 participants were recruited for this study). Some factors that affected the performance include the frequent need for words to be in plural form or to have different endings, which required additional typing for ChatGPT to suggest the appropriate word forms. Another common issue was participants ignoring suggestions and opting for manual typing, resulting in lower characters per minute. When comparing our average OCM with [8], which used an n-gram prediction model, a clear difference is observed. Specifically, for sentences aided by ChatGPT, the average OCM

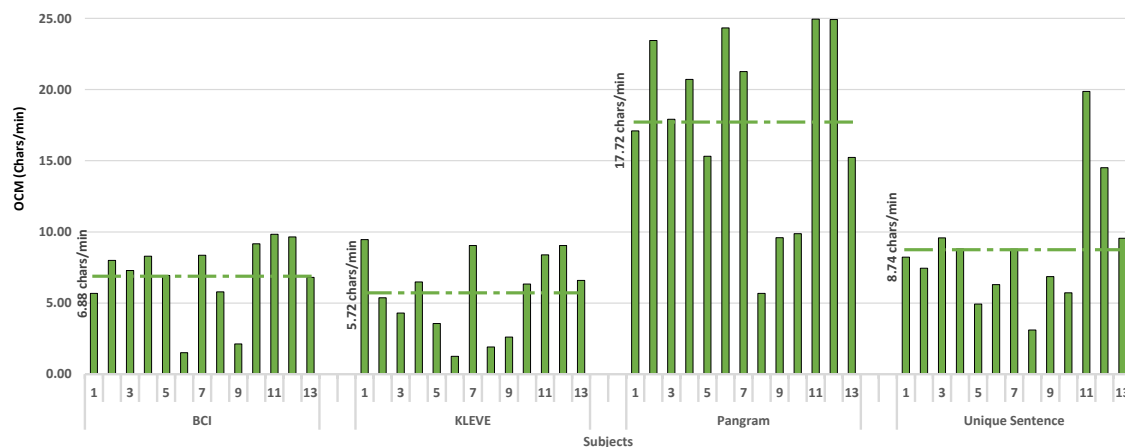


Figure 4: For each task, the Output Characters per Minute (OCM) are represented by bars, and the average values are illustrated with lines. Participants were assigned the following typing tasks: “BCI”, “KLEVE” “THE\_QUICK\_BROWN\_FOX\_JUMPS\_OVER\_THE\_LAZY\_DOG” and a unique sentence for each participant. The average values of the lines can be found above them.

was 8.74 cpm for unique sentences (see Fig. 2), while the n-gram model study showed an average of 18.9 cpm for their unique sentence tasks. This indicates that ChatGPT did perform worse compared to n-gram prediction model, but these differences may also be described by the ChatGPT communication delays. Future steps could involve the implementation of another API that is specifically designed for auto-correction or integrating a locally executed artificial-intelligence(AI) model tailored for automatic text correction, potentially addressing many of the data protection concerns previously mentioned, and also increasing the performance. Additionally, exploring the power consumption and comparing it to the n-gram model.

## CONCLUSION

We successfully developed a cVEP based spelling interface that incorporates the ChatGPT API to assist users with spelling tasks. The integration of ChatGPT expands the software’s functionality, potentially improving communication efficiency. However, it’s noteworthy that its assistance didn’t come close to that of an n-gram model in terms of output characters per minute (OCM) in unique sentence tasks. Despite this limitation, it simplifies future development and reduces processing power requirements for local machines, potentially enhancing the typing experience. Our study underscores the potential of BCI-AI collaboration to enhance communication, autonomy, and quality of life for individuals with physical disabilities, though further research is needed.

## ACKNOWLEDGEMENT

The authors thank the financial support provided by “The Friends of the University Rhine-Waal - Campus Cleve” association. Additionally, we appreciate the student assistants and all participants in this study.

## REFERENCES

- [1] Wolpaw J, Wolpaw EW. Brain-Computer Interfaces: Principles and Practice. OUP USA (2012).
- [2] Volosyak I, Rezeika A, Benda M, Gembler F, Stawicki P. Towards solving of the Illiteracy phenomenon for VEP-based brain-computer interfaces. Biomedical Physics & Engineering Express. 2020;6(3):035034.
- [3] Gembler F, Volosyak I. A Novel Dictionary-Driven Mental Spelling Application Based on Code-Modulated Visual Evoked Potentials. Computers. 2019;8(2):33.
- [4] Lappalainen Y, Narayanan N. Aisha: A Custom AI Library Chatbot Using the ChatGPT API. Journal of Web Librarianship. 2023;17(3):37–58.
- [5] Sharma A, Bhargava M, Khanna AV. AI-Farm: A crop recommendation system. In: 2021 International Conference on Advances in Computing and Communications (ICACC). IEEE, Oct. 2021.
- [6] Dave T, Athaluri SA, Singh S. ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. Frontiers in Artificial Intelligence. 2023;6.
- [7] OpenAI. “OpenAI Documentation.” (2022). [Online]. Available: <https://platform.openai.com/docs/overview> (visited on 02/21/2024).
- [8] Gembler F, Stawicki P, Saboor A, Volosyak I. Dynamic time window mechanism for time synchronous VEP-based BCIs-Performance evaluation with a dictionary-supported BCI speller employing SSVEP and c-VEP. PLOS ONE. 2019;14(6):e0218177.

# A STUDY OF PERFORMANCE VARIABILITY IN DEEP NEURAL NETWORKS FOR MOTOR IMAGERY CLASSIFICATION: TOWARDS A ZERO-CALIBRATION APPROACH

Pasquale Arpaia<sup>1,2,3</sup>, Antonio Esposito<sup>1,2</sup>, Fortuna Galdieri<sup>1,2</sup>, Angela Natalizio<sup>1,4</sup>, Marco Parvis<sup>4</sup>,  
Andrea Pollastro<sup>1,2</sup>

<sup>1</sup> Augmented Reality for Health Monitoring Laboratory (ARHeMLab)

<sup>2</sup>Department of Electrical Engineering and Information Technology (DIETI), Università degli Studi di Napoli Federico II, Naples, Italy

<sup>3</sup>Centro Interdipartimentale di Ricerca in Management Sanitario e Innovazione in Sanità (CIRMIS), Università degli Studi di Napoli Federico II, Naples, Italy

<sup>4</sup>Department of Electronics and Telecommunications (DET), Polytechnic of Turin, Turin, Italy

E-mail: pasquale.arpaia@unina.it

**ABSTRACT:** This study deals with the adoption of deep learning and transfer learning in motor imagery-based brain-computer interfaces to develop a robust system with a zero-calibration approach. Deep neural networks would be also sought to improve the classification accuracies of these interfaces. However, these approaches are affected by inherent variability in their performance, so that dominating uncertainty sources appears crucial. To assess the performance variability of deep neural networks, the effects of parameter initialisation and pre-processing were studied. EEGNet and Sinc-EEGNet were used for this purpose. The results highlight that network's weight initialisation significantly affect the performance. For instance, classification accuracy can improve from 67 %  $\pm$  3 % to 73 %  $\pm$  3 % by just changing the weight initialisation. Meanwhile, EEG pre-processing does not improve the performance, thus it can be avoided to reduce the computational effort. These results pave the way for real-time application scenarios.

**Keyword:** brain-computer interface, motor imagery, deep learning, transfer learning, uncertainty.

## INTRODUCTION

A motor imagery-based Brain-Computer Interface (BCI) measures voluntarily modulated brain signals generated while imagining a movement [1]. Notably, non-invasive and wearable BCIs based on motor imagery have been investigated more and more [2, 3]. These typically exploit electroencephalographic (EEG) signals acquired through electrodes placed on the scalp [4]. During the execution of a motor imagery task, spectral power changes occur in the  $\mu$  (7 Hz to 13 Hz) and  $\beta$  (13 Hz to 30 Hz) bands of the signals recorded over the sensorimotor brain area. In particular, event-related desynchronisation and synchronisation can be observed immediately before and after motor imagery, respectively [5]. Therefore, the interface

attempts to detect the imagined movement through the analysis of temporal, frequency, and spatial features of the acquired signals.

Despite their potential in several fields [3, 6–8], motor imagery-based BCIs still suffer from many limitations. Firstly, the user must learn how to modulate sensorimotor rhythms. Secondly, motor imagery signals are hindered by noise, which can be either baseline neural activity [9] or artefacts. In addition, a large amount of data are needed for an effective training and testing of processing pipelines, typically relying on machine learning [10–12]. Hence, long calibration periods (20–30 minutes) are needed before properly using the BCI system [10]. Moreover, EEG signals suffer from a significant inter- and intra-subject variability [13]. This implies that calibration data should be acquired for each new user and new session of the same user. The highlighted challenges are exacerbated when the BCI system has to discriminate as the number of motor imagery tasks increases [14].

In this framework, research has recently focused on deep learning methods to improve motor imagery classification, especially in multi-class problems [12, 14, 15]. However, deep learning techniques require a bigger amount of data with respect to classical machine learning approaches [14–16].

EEG data are typically acquired in controlled experimental conditions, but the experimental burden make it difficult to obtain large and significant datasets in practice [16]. Therefore, common strategies to compensate for the lack of data are data augmentation and transfer learning (TL) [10, 16, 17]. In particular, TL is based on training a model by relying on the knowledge gained from another pre-trained model. This approach has the advantage of reducing training time [12, 16, 18] or neglecting it in a "zero calibration" scenario, where the EEG data of a subject are classified by a model identified on independent data from other subjects.

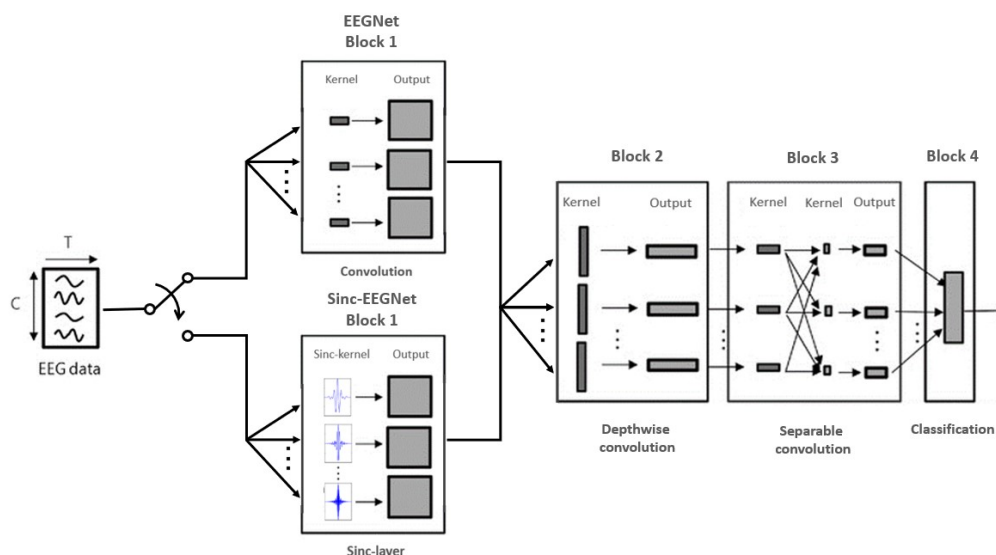


Figure 1: Adopted deep neural network architectures for motor imagery EEG processing, namely EEGNet and Sinc-EEGNet. The only difference between the two is in Block 1, which is a standard convolutional layer for EEGNet and a Sinc-layer for Sinc-EEGNet. The remaining blocks are identical in both architectures.

The combination of TL methods with deep learning-based processing pipelines appears promising in the BCI context [14, 16, 19]. However, it is worth emphasizing that deep neural networks performance is very sensitive to the weight initialisation and data pre-processing: the former affects the training process [20–22], while the latter improves the quality of the data [12, 15, 23, 24]. Unfortunately, in context of TL applications for motor imagery-based BCI, there is a lack of studies investigating those performance variations [15, 21, 22], especially regarding the pre-processing [24].

This paper thus focuses on performance variability due to weight initialisation and pre-processing in the context of TL for motor-imagery BCI towards a zero-calibration approach. Notably, EEGNet [25] and Sinc-EEGNet [26] were investigated. As common pre-processing strategies for EEG signals consist of the use of basic filtering techniques, this paper focuses on the use of band-pass filter [27, 28] and Laplacian filter [29].

Therefore, the remainder of the paper is organised as follows. Sec. MATERIALS AND METHOD describes EEGNet and Sinc-EEGNet architectures as well as the analysis conducted on them by involving TL. Sec. RESULTS presents the exploited dataset and discusses inherent results.

## MATERIALS AND METHODS

The purpose of this Section is to present the architectures used in the study and the analyses performed on them. It is worth remarking that all the analyses were performed under zero calibration. The Section is structured as follows: in Sec. *Architectures*, EEGNet [25] and Sinc-EEGNet [26] are presented (Fig. 1), while the methodology proposed for comparing different setting is detailed

in Sec. *Experimental Setup*.

**Architectures:** EEGNet is one of the most commonly used deep learning architectures in BCI [25]. It is a low-density convolutional neural network designed to robustly extract information from EEGs. It uses both depth-wise and separable convolutions to extract EEG features. The architecture is structured in four blocks. In the first block, two sequential convolutional layers are used as a temporal filter. In the second block, a depth-wise convolutional layer is used. In EEG-specific applications, this type of layer provides a direct way to learn the spatial filters for each temporal filter, allowing for efficient extraction. The third block uses a separable convolution, which reduces the number of parameters to be fitted and explicitly decouples relationships within and between feature maps. Finally, in the classification block, the features are passed directly to a softmax function.

Interestingly, the architecture of EEGNet resembles the steps of the well-known filter bank common spatial pattern algorithm [30], adding flexibility thanks to the end-to-end training procedure of deep learning models. The strengths of this architecture with respects to general-purpose convolutional neural networks include (i) reduced number of trainable parameters due to the use of depthwise and separable convolutions, (ii) applicability to low-dimensional data, and (iii) adaptability across different EEG datasets and tasks [25, 31, 32].

An EEGNet variant called Sinc-EEGNet was also recently proposed [26]. It consists of merging EEGNet [25] with Sinc-Net [33], which is characterised by a convolutional layer having learnable sinc functions as filters. The main strength of Sinc-Net consists in deriving a custom filter bank, specifically tuned for the desired application [33]. Sinc-EEGNet consists of an EEGNet architecture in which the first convolutional layer has been replaced by

a sinc-layer, resulting in a reduced number of trainable parameters [26]. Different ways for combining the two architectures were proposed in literature [34–36]. This work focused on the version proposed in [26]. It faithfully reproduced the original version of EEGNet except for the first block, in which the first traditional convolutional layer was replaced by a sinc layer [33]. The structures of the two architectures are jointly illustrated in Fig. 1 to stress that the only difference resides in the first block. Such architectures are both suitable for TL, with the Sinc-EEGNet variant that is more prone to explainability [26].

Training hyperparameters	Values
n. of training epochs	1000
learning rate	0.0001
batch size	32
optimizer	Adam
early stopping patience	150
weight decay	0.02

Table 1: Training setup for the conducted analyses.

**Experimental Setup:** The analyses performed in this work were based on EEGNet and Sinc-EEGNet with either 4 or 32 filters in the first block. In particular, 4 is the optimal number of filters for EEGNet [25] and 32 is the optimal number of filters for Sinc-EEGNet [26]. The name of the architecture followed by the number of filters is used to refer to the specific architectures (e.g. EEGNet-4 refers to EEGNet with 4 filters in the first convolutional layer). Training hyperparameters were optimized in a previous work [26] and they are recalled for clarity in Tab. 1. All the analyses were done in the zero calibration TL scenario. Notably, the leave-one-subject-out [37] technique was used for this purpose. Once the dataset is selected, all samples relating to a single subject are removed and the model is then trained on the remaining samples. The performance of the model is then evaluated on the independent samples from the left-out subject. Such an offline analysis simulates the straightforward usage of a BCI on a previously unseen subject. Two comparative analyses were carried on by using the above mentioned architectures. The first one consisted of analysing the models’ performances with different weight initialisation. The He initialisation [38] was used in this work. To this aim, the seed for pseudo-random generation of initial parameters was firstly varied from 42 to 56 (15 values). Then, by exploiting the optimal seed for each model, a second analysis step was carried on. This consisted of analysing the models’ performance as the EEG pre-processing strategy was varied. In particular, the following three cases were tested: (i) no pre-processing, (ii) band-pass filter from 4 Hz to 40 Hz, and (iii) Laplacian filter.

For all the analyses, the metric adopted to assess models’ performances was the mean classification accuracy across subjects of the selected dataset and its associated type A uncertainty, i.e. the standard deviation divided by the square-root of the number of averaged accuracies.

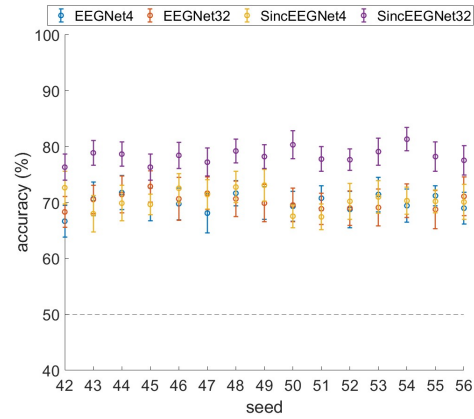


Figure 2: Mean classification accuracies and associated uncertainties obtained as a function of the seed (i.e. weight initialisation) for each architecture. The dotted line refers to the random accuracy.

## RESULTS

This Section presents the results of the analyses. In details, Sec. *Dataset* describes the data and its usage, Sec. *Weight initialisation impact* presents the results of the first analysis step, and Sec. *Pre-processing impact* presents the results associated with different pre-processing techniques.

**Dataset:** the benchmark dataset *BCI competition IV, 2a* was used for the analyses [39]. It includes EEG signals from nine healthy subjects recorded using 22 wet electrodes. The sampling rate was 250 Sa/s. The subjects performed four motor imagery tasks during two sessions recorded on two different days. As the present study adopts an inter-subjective approach, the investigations considered the only first session. Moreover, two classes of motor imagery were used, namely left and right hand motor imagery. Finally, each trial was epoched from 2 s to 6 s, thus including the cue and the motor imagery windows.

**Weight initialisation impact:** Fig. 2 shows the results obtained for each configuration (i.e., EEGNet-4, EEGNet-32, Sinc-EEGNet-4, and Sinc-EEGNet-32) when varying the seed. In particular, each point represents the mean classification accuracy across the nine subjects of the dataset together with its type A uncertainty. The dotted line shows the random accuracy. As previously found [26], Sinc-EEGNet-32 is the most effective configuration architecture, even as the seed varies. This result was confirmed by the Kruskal-Wallis test (see Tab. 3).

Fig. 2 shows the results obtained for each configuration (i.e., EEGNet-4, EEGNet-32, Sinc-EEGNet-4, and Sinc-EEGNet-32) when varying the seed. In particular, each point represents the mean classification accuracy across the nine subjects of the dataset together with its type A uncertainty. The dotted line shows the random accuracy. As previously found [26], Sinc-EEGNet-32 is the most effective configuration architecture, even as the seed varies. This result was confirmed by the Kruskal-Wallis

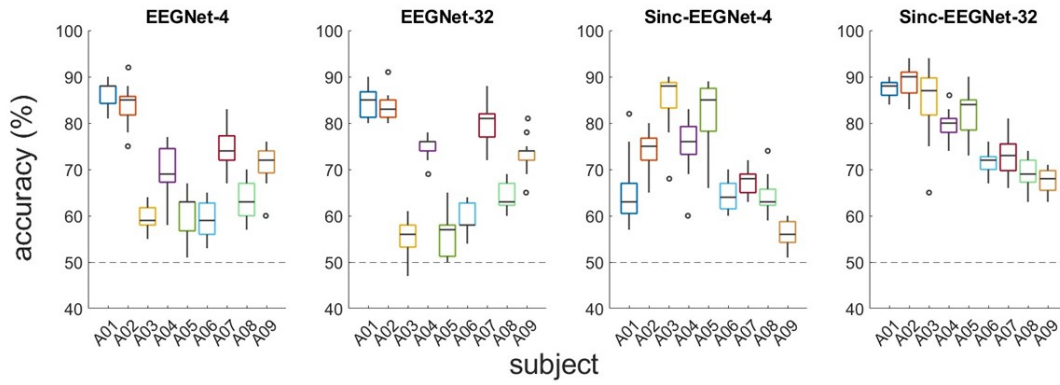


Figure 3: For each configuration, the results obtained for each subject with seed variation are presented in box plots, each containing the 15 associated classification accuracy values for each seed. The dotted line refers to the random accuracy.

	min (%)	max (%)
EEGNet-4	67 ± 4	72 ± 4
EEGNet-32	68 ± 4	73 ± 4
Sinc-EEGNet-4	67 ± 3	73 ± 3
Sinc-EEGNet-32	76 ± 3	81 ± 3

Table 2: Minimum and maximum classification accuracies across seed values and for each configuration. Uncertainties associated with these mean accuracies, estimated by using the standard deviation of the mean, are reported too.

Group A	Group B	p-value
EEGNet-4	EEGNet-32	0.966
EEGNet-4	Sinc-EEGNet-4	0.955
EEGNet-4	Sinc-EEGNet-32	< 0.001
EEGNet-32	Sinc-EEGNet-4	0.100
EEGNet-32	Sinc-EEGNet-32	< 0.001
Sinc-EEGNet-4	Sinc-EEGNet-32	< 0.001

Table 3: Results of post-hoc analysis with Kruskal-Wallis test.

test as extensively reported in Tab. 3.

It is important to stress that, for a specific architecture, the performance can vary significantly with different initialisation. Tab. 2 displays the minimum and maximum results obtained as the seed varies for each configuration. This analysis demonstrates the importance of testing the model's performance with different seeds before effectively using them. For example, in [26] all analyses were performed with a fixed seed of 42, which was found to be the worst case for most of the configurations. However, once the seed was fixed, the accuracy resulted repeatable. Fig. 3 also shows the results obtained for different seeds, subject by subject. Each box plot represents the 15 classification accuracies, while the dotted line indicates the random classification accuracy reference. Although results are consistent among different architectures for some subjects, there is significant variability in others. For instance, subject A09 displays several outliers when EEGNet is employed, resulting in performance differences of over 15 % when just varying the seed. Next, it can be noted that the zero calibration scenario changes the discrimination between "good" and "bad" subjects

with respect to previous literature evidence [40]. For instance, for the notoriously good subject A03, Sinc-EEGNet model leads to a good performance but the EEGNet model does not. In other case, like for A08, performance is relatively low with the proposed approaches, while other literature approaches led to higher performance. In this regard, it is important to recall that the results were obtained by training the models on data from other subjects, which could be associated with a non-compatible probability distribution.

*Pre-processing impact:* Fig. 4 shows the variations in models' performance by varying the pre-processing for each configuration. The networks were initialised with a random seed (seed = 0), but it is worth noting that compatible results were obtained when the best seed from the previous step was selected. Each box plot contains the results obtained for the nine subjects. The dotted line indicates the random classification accuracy. As usually proposed, one type of pre-processing has been used for each experiment [12]. This also facilitates online classification in terms of computational effort. The models performance resulted reduced by applying a band-pass filter. This is in contrast to what is observed by using classical machine techniques, where filtering the data trial by trial is often recommended [15, 24, 30]. The application of the Laplacian filter, instead, led to compatible performance than the "no pre-processing" case. Hence, this evidence suggests that pre-processing can be avoided. It is worth noting that EEGNet was originally proposed with a band-pass filter [25], whereas Sinc-EEGNet was proposed without any kind of pre-processing [26].

## CONCLUSION

Deep learning has attracted more and more attention in the processing of EEG data for motor imagery-based BCIs, and transfer learning promises to improve classification accuracy while reducing the calibration burden. This would disclose a very large use of such BCI technologies in practice. However, literature results are still quite variegated and uncertainty sources are not dominated yet.

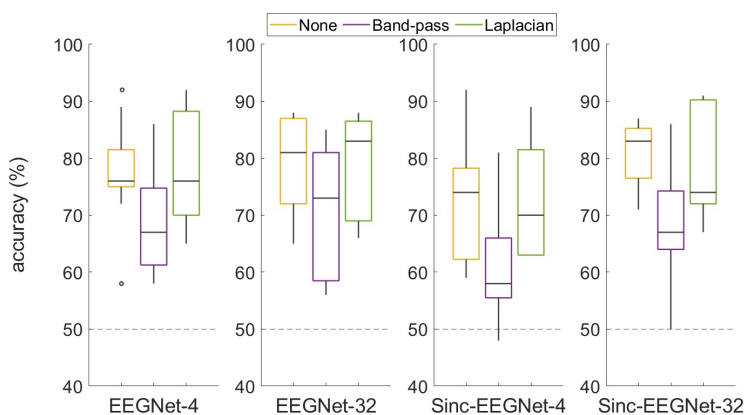


Figure 4: For each configuration, the box plots contains the results obtained using different pre-processing. The dotted line refers to the random accuracy.

To make a further step towards the repeatability and reproducibility of deep neural network results, a comparative analysis of the performance has been proposed for two relevant networks, namely EEGNet and Sinc-EEGNet, when varying fundamental settings. All the analyses were carried out under a zero-calibration approach, hence exploiting the leave-one-subject-out technique on the benchmark dataset *BCI competition IV, 2a* limited to two motor imagery tasks.

It was found that the performance varies significantly with different weights initialisation. Then, by analysing models' performance for varying pre-processing strategies, the performance of the deep network models was unchanged by Laplacian filter and even reduced by band-pass filtering. This finding is in contrast to what is observed by using classical machine learning, where filtering the data is recommended to improve classification results. This outcome suggests the use of pre-trained deep architectures on a new subject without the need for any preliminary data processing. This would reduce the pre-processing time of the data and foster an online classification. Overall, it is worth noting that Sinc-EEGNet-32 resulted the most effective architecture in accordance with previous studies.

This preliminary analysis emphasised the importance of carefully investigating the variability of deep neural networks adopted in BCI. Nonetheless, future works will deal with extending these analyses to more dataset to collect more evidence. Moreover, an experimental plan will be designed for a more comprehensive uncertainty assessment as the main networks settings vary.

## ACKNOWLEDGEMENTS

This work was financially supported by Italian Ministry for Universities and Research (MUR) through the project "AGE-IT - Ageing Well in an ageing society: A novel public-private alliance to generate socioeconomic, biomedical and technological solutions for an inclusive Italian ageing society (Spoke: Ageing and clinical practice)", PNRR PE15, CUP E63C22002050006.

## REFERENCES

- [1] Lotze M, Halsband U. Motor imagery. *Journal of Physiology-paris*. 2006;99(4-6):386–395.
- [2] Lin BS, Pan JS, Chu TY, Lin BS. Development of a wearable motor-imagery-based brain–computer interface. *Journal of medical systems*. 2016;40:1–8.
- [3] Arpaia P *et al*. Paving the way for motor imagery-based tele-rehabilitation through a fully wearable bci system. *Sensors*. 2023;23(13):5836.
- [4] Yadav D, Yadav S, Veer K. A comprehensive assessment of brain computer interfaces: Recent trends and challenges. *Journal of Neuroscience Methods*. 2020;346:108918.
- [5] Pfurtscheller G, Lopes da Silva F. *Handbook of Electroencephalography and Clinical Neurophysiology – Event-related desynchronization*. Elsevier, Amsterdam, Netherlands (1999).
- [6] Paszkiel S. Control based on brain-computer interface technology for video-gaming with virtual reality techniques. *Journal of Automation, Mobile Robotics and Intelligent Systems*. 2016:3–7.
- [7] Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain–computer interfaces for communication and rehabilitation. *Nature Reviews Neurology*. 2016;12(9):513–525.
- [8] Tong Y *et al*. Motor imagery-based rehabilitation: Potential neural correlates and clinical application for functional recovery of motor deficits after stroke. *Aging and disease*. 2017;8(3):364.
- [9] Singh A, Hussain AA, Lal S, Guesgen HW. A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface. *Sensors*. 2021;21(6):2173.
- [10] Huang X *et al*. A review on signal processing approaches to reduce calibration time in eeg-based brain–computer interface. *Frontiers in Neuroscience*. 2021;15:733546.
- [11] Padfield N, Zabalza J, Zhao H, Masero V, Ren J. Eeg-based brain-computer interfaces using

- motor-imagery: Techniques and challenges. *Sensors*. 2019;19(6):1423.
- [12] Arpaia P, Esposito A, Natalizio A, Parvis M. How to successfully classify eeg in motor imagery bci: A metrological analysis of the state of the art. *Journal of Neural Engineering*. 2022;19(3):031002.
- [13] Saha S, Baumert M. Intra-and inter-subject variability in eeg-based sensorimotor brain computer interface: A review. *Frontiers in computational neuroscience*. 2020;13:87.
- [14] Al-Saegh A, Dawwd SA, Abdul-Jabbar JM. Deep learning for motor imagery eeg-based classification: A review. *Biomedical Signal Processing and Control*. 2021;63:102172.
- [15] Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (eeg) classification tasks: A review. *Journal of neural engineering*. 2019;16(3):031001.
- [16] Zhang K *et al.* Application of transfer learning in eeg decoding based on brain-computer interfaces: A review. *Sensors*. 2020;20(21):6321.
- [17] Ko W, Jeon E, Jeong S, Phyo J, Suk HI. A survey on deep learning-based short/zero-calibration approaches for eeg-based brain-computer interfaces. *Frontiers in Human Neuroscience*. 2021;15:643386.
- [18] Apicella A, Isgro F, Pollastro A, Prevete R. On the effects of data normalization for domain adaptation on eeg data. *Engineering Applications of Artificial Intelligence*. 2023;123:106205.
- [19] Lotte F *et al.* A review of classification algorithms for eeg-based brain-computer interfaces: A 10 year update. *Journal of neural engineering*. 2018;15(3):031005.
- [20] Yam JY, Chow TW. A weight initialization method for improving training speed in feedforward neural network. *Neurocomputing*. 2000;30(1-4):219–232.
- [21] Cao W, Wang X, Ming Z, Gao J. A review on neural networks with random weights. *Neurocomputing*. 2018;275:278–287.
- [22] Narkhede MV, Bartakke PP, Sutaone MS. A review on weight initialization strategies for neural networks. *Artificial intelligence review*. 2022;55(1):291–322.
- [23] Pal KK, Sudeep K. Preprocessing for image classification by convolutional neural networks. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTE-ICT). 2016, 1778–1781.
- [24] Alzahab NA *et al.* Hybrid deep learning (hdl)-based brain-computer interface (bci) systems: A systematic review. *Brain sciences*. 2021;11(1):75.
- [25] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. Eegnet: A compact convolutional neural network for eeg-based brain-computer interfaces. *Journal of neural engineering*. 2018;15(5):056013.
- [26] Arpaia P *et al.* Sinc-eegnet for improving performance while reducing calibration of a motor imagery-based bci. In: 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRaine). 2023, 1063–1068.
- [27] Togha MM, Salehi MR, Abiri E. An improved version of local activities estimation to enhance motor imagery classification. *Biomedical Signal Processing and Control*. 2021;66:102485.
- [28] Jeon E, Ko W, Yoon JS, Suk HI. Mutual information-driven subject-invariant and class-relevant deep representation learning in bci. *IEEE Transactions on Neural Networks and Learning Systems*. 2021.
- [29] Cohen MX. Analyzing neural time series data: Theory and practice. MIT press (2014).
- [30] Ang KK, Chin ZY, Wang C, Guan C, Zhang H. Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *Frontiers in neuroscience*. 2012;6:39.
- [31] Özdenizci O, Wang Y, Koike-Akino T, Erdoğan D. Learning invariant representations from eeg via adversarial inference. *IEEE access*. 2020;8:27074–27085.
- [32] Autthasan P *et al.* Min2net: End-to-end multi-task learning for subject-independent motor imagery eeg classification. *IEEE Transactions on Biomedical Engineering*. 2021;69(6):2105–2118.
- [33] Ravanelli M, Bengio Y. Speaker recognition from raw waveform with sincnet. In: 2018 IEEE spoken language technology workshop (SLT). 2018, 1021–1028.
- [34] Shimizu H, Srinivasan R. Improving classification and reconstruction of imagined images from eeg signals. *Plos one*. 2022;17(9):e0274847.
- [35] Bria A, Marrocco C, Tortorella F. Sinc-based convolutional neural networks for eeg-bci-based motor imagery classification. In: International Conference on Pattern Recognition. 2021, 526–535.
- [36] Chen J, Wang D, Yi W, Xu M, Tan X. Filter bank sinc-convolutional network with channel self-attention for high performance motor imagery decoding. *Journal of Neural Engineering*. 2023;20(2):026001.
- [37] Kunjan S *et al.* The necessity of leave one subject out (loso) cross validation for eeg disease diagnosis. In: Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings 14. 2021, 558–567.
- [38] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. 2015, 1026–1034.
- [39] Brunner C, Leeb R, Müller-Putz G, Schlögl A, Pfurtscheller G. Bci competition 2008–graz data set a. Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology. 2008;16:1–6.
- [40] Arpaia P, Esposito A, Moccaldi N, Natalizio A, Parvis M. Online processing for motor imagery-based brain-computer interfaces relying on eeg. In: 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). 2023, 01–06.

# DEEP LEARNING FOR MOTOR IMAGERY-BASED BCIS USING SEEG SIGNALS

Zhichun Fu<sup>1</sup>, Xiaolong Wu<sup>1</sup>, Xin Gao<sup>1</sup>, Dingguo Zhang<sup>1</sup>

<sup>1</sup> Department of Electronic and Electrical Engineering, University of Bath, Bath, United Kingdom

E-mail: d.zhang@bath.ac.uk

**ABSTRACT:** Motor imagery (MI) is the most popular paradigm for brain-computer interfaces (BCIs) based on scalp electroencephalography (EEG), while this paradigm is missing for stereo-electroencephalography (sEEG)-based BCIs. Recently, the first public dataset of sEEG has become available for MI-based BCIs. However, the performance using traditional methods is still inferior. In this study, we employed some state-of-the-art methods based on deep learning to improve the classification accuracy of MI for sEEG-based BCIs. Six different deep learning models were developed, which include Shallow ConvNet, DeepNet, ResNet20, conformer, vision transformer (ViT) and ViT with pre-trained parameters. Among six deep learning models, we achieved an average accuracy of 0.83 in the hand open/closed binary classification task with the conformer model. Compared to the available work, our approach demonstrated a remarkable 16% increase in accuracy.

## INTRODUCTION

Brain-computer interface (BCI) technology serves as a promising solution, enabling direct communication between the human brain and external devices or computer systems. In general, there are two categories of BCI technology, i.e., invasive and non-invasive. Non-invasive BCI relies on capturing brain signals from the scalp in a user-friendly way. Invasive BCI, on the other hand, involves direct implantation into the brain for signal acquisition, which can result in intracranial signals with less noise interference and a higher spatial-temporal resolution. Examples of invasive BCIs include signals such as stereo-electroencephalography (sEEG) and electrocorticography (ECoG) [1], [2].

Current sEEG-based BCIs primarily focus on motor-related decoding, such as various hand gestures, tongue movements, and foot movements [3], [4]. Combrisson et al demonstrated that motor execution, intention movement and rest status can be differentiated by decoding sEEG signals [5]. The authors discovered a relationship between phase, amplitude and PAC during the planning and execution phases of the goal-directed movement. Additionally, they were able to predict continuously changing grasp force through decoding sEEG signals [6].

However, there have been relatively few studies on sEEG decoding of imagined movements using sEEG. Murphy

et al. employed a Support Vector Machine (SVM) to classify the imagined force and rest status in two different grasp configurations, achieving an average accuracy of over 0.6, which was higher than the chance level [7]. When analysing imagined single feature modulation, the alpha band showed a higher modulation level compared to other bands. Ottenhoff et al. demonstrated that non-motor areas contain sufficient information for motor decoding [8]. To avoid the effect of the motor area, they excluded all electrodes originating from the central sulcus and its adjacent area. They used a Riemannian decoder as the classifier, which achieved an average area under curve of 0.68 for imagined movements, with details extracted from the beta band. Individuals with movement disorders or speech impairments often rely on imagery movement as a means of communication. This work aims to enhance the accuracy of sEEG imagined movement decoding by using a deep learning model.

Considering the capability of deep learning models to extract sophisticated features without manual feature extraction, we propose using the same to decode imagined movements. Furthermore, after the advent of the Transformer model, it was demonstrated to be highly effective in sequence-to-sequence tasks due to its attention mechanism. Recent research has shown promising results for deep learning models based on Transformers in reconstructing trajectories of imagined movement [9]. Therefore, the purpose of this study is to evaluate whether a deep learning model can enhance BCI performance for each participant. By utilizing algorithms that have previously been successful in executed movement decoding and regression tasks, they can improve the classification and recognition of imaginary motions with some optimizations. By comparing six different deep learning models with different structures, we have identified a more suitable structure for recognizing imaginary movements which will be valuable for future studies. In summary, our main contributions can be outlined as follows:

- 1) We explore the application of deep learning methods on sEEG motor imagery datasets.
- 2) We demonstrate improvements in recognition results compared to previous studies.

The remainder of the article is organized as follows. In the Methods section, we introduce the deep learning models utilized in this study, along with details regarding the dataset and data preprocessing methods. The Result section presents the experimental findings of the models. Finally, we discuss the outcomes and summarize the

contributions of this article.

## METHODS

To assess the performance of deep learning on MI-based BCIs using sEEG, this study implemented five different state-of-the-art (SOTA) models and an improved model with pre-trained weights. The results were then compared with those obtained from the Riemannian classifier in reference [8].

### Shallow ConvNet

The ConvNet shown in Fig. 1 is a model that utilises temporal convolution and spatial filtering in its initial layers, similar to the bandpass phase in the filter bank common spatial pattern (FBCSP) [10]. The shallow ConvNet's use of a larger kernel size in temporal convolution allows for a broader range of transformations. Additionally, incorporating multiple pooling regions per trial enables the learning of the temporal structure of band power changes, thereby enhancing classification.

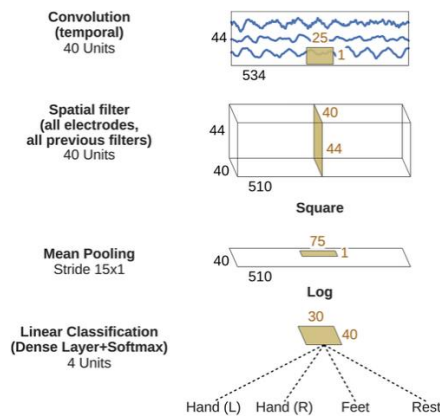


Figure 1: The model structure for shallow ConvNet [10].

### DeepNet

The deepNet model utilized in this study featured a more intricate architecture with a substantial increase in the number of layers compared to the shallow model [10]. The architecture includes temporal convolution, spatial convolution, a fully connected layer and basic blocks which are used to extract the spatial features. Fig. 2 shows the model structure with one basic block. A dropout rate of 0.5 was employed to improve the model's robustness. The process of optimization involves experimenting with different quantities of basic blocks to determine the optimal configuration of the model. For this work, we utilized the deepNet model with two basic blocks.

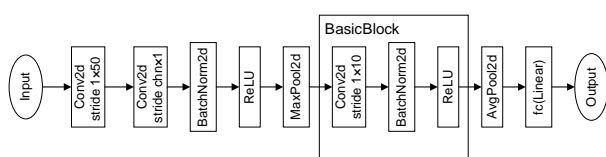


Figure 2: The model structure for DeepNet model with changing depth.

### ResNet20

ResNet architectures have demonstrated success in various computer vision tasks due to their ability to mitigate the vanishing gradient problem and facilitate the training of exceptionally deep networks [11]. The ResNet model, with its residual connections, aims to leverage these advantages to enhance the performance of the imagining motion task. As the ResNet model has been previously used for emotion classification based on EEG image recognition, we incorporated a 20-layer Residual Network (ResNet) into our model, as shown in Fig. 3 [12].

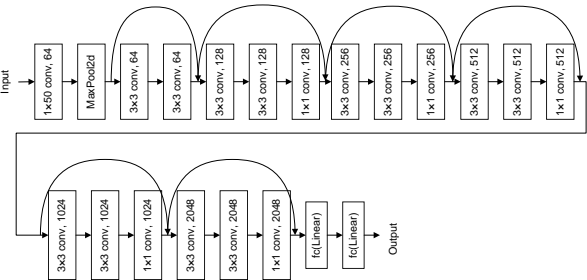


Figure 3: The model structure for Resnet20.

### Conformer

The Conformer model comprises three modules: a convolution module, a self-attention module, and a classifier module, as shown in Fig. 4 [13]. The convolution module uses both spatial and temporal convolutions to capture local spatial and temporal features of EEG signals. This is followed by an average pooling layer to reduce feature dimension and mitigate noise interference. The self-attention module utilizes multi-head attention mechanisms to capture global temporal dependencies of EEG features, complementing the local features learned by the convolution module. The classifier module includes two fully connected layers to output the probability of different EEG categories, such as motor imagery or emotion recognition tasks.

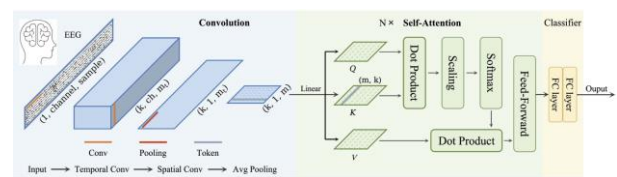


Figure 4: The model structure for conformer [13].

### ViT (Vision Transformer)

ViT is a hybrid model that combines a two-step convolution block with a transformer block, depicted in Fig. 5 [14]. The two-step convolution block is composed of two convolutional layers, one for the temporal dimension and one for the channel (spatial) dimension. This block generates patch embeddings that capture the frequency and spatial information of the sEEG data. The transformer block utilises the ViT architecture, which divides the input into patches and processes them as a sequence using self-attention and multi-layer perceptron. Additionally, it also captures global dependencies and patterns in the patch sequence. The final representation of the input is the hidden state of a special token.

The transformer block was pre-trained on the ImageNet dataset, which contains millions of natural images. This pre-training allows the model to utilise prior knowledge learned from image data and transfer it to the BCI field.

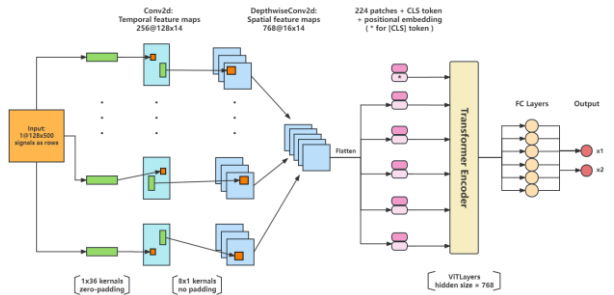


Figure 5: The model structure for ViT [14].

### Dataset

The sEEG dataset was collected from eight subjects performing two motor imagined tasks [8]. Specifically, subjects were instructed to imagine opening and closing their left and right hands, with each action lasting for 3 seconds. Tab. 1 provides additional details about the dataset, including the distribution of contacts for each subject, as well as the number of electrodes in the left and right hemispheres and the presence of electrodes in the motor area. Each participant imagined 30 trials of opening and closed actions for each of their hand, with each action lasting 3 seconds. Consequently, each participant generated a total of 60 movements and 60 rest events throughout the entire experiment.

Table 1: Electrode details of subjects after removing noisy and abnormal signals. Number of contacts also includes the electrodes which are located in the motor areas.

Patient ID	Motor Left	Motor Right	Contacts Left	Contacts Right
1	4	6	37	90
2	0	0	103	24
3	9	0	66	0
4	0	0	54	0
5	6	0	117	0
6	0	0	63	63
7	3	5	67	60
8	0	3	40	75

### Data processing

For each subject, any abnormal signals, including flat signals and signals with abnormal amplitudes, were removed. In brief, the logarithm of the root mean square (LRMS) of each channel's signal was calculated. Then, we normalized these LRMS values, and calculated the corresponding p-values based on a normal distribution assumption. Channel with p-values less than or equal to 0.05 were flagged as having abnormal amplitudes. For the channel where 50Hz frequency band power exceeded two times the interquartile range of the signal, it would be removed as well. The remaining signals underwent

detrending, mean removal, and were subjected to a notch filter at 50Hz, 100Hz, 150Hz, and 200Hz to minimize interference from noise.

Subsequently, the data was downsampled to 500Hz, and each experiment was segmented from -0.5s to 3s. The entire dataset is segmented by selecting fixed window size and stride size and stored as DataLoader formatted data for subsequent input into deep models for training and evaluation. By utilizing a fixed stride size, the optimal window size was identified among 200, 400, 600, 800, 1000 and 1200. With the best window size, best stride size can be found among 20, 50, 100, 200, 300, 400, 500. Based on the performance of all models, 800 and 400 are selected as the window size and stride size respectively. In this work, a learning rate of 0.0001 was employed with a weight decay set to 0.0005. The optimizer used was Adadelta, and the cross-entropy loss was used for loss function calculation. To prevent data leakage from affecting model training and prediction, trials were classified before splitting. 60% of the trials were assigned to the training set, 20% to the validation set, and 20% to the test set. After determining these sets, each trial was further divided into one-second overlapping intervals to simulate data obtained during online experiments. The desktop computer that was used in the tests has the following configuration: 11th Gen Intel i9-11900 16 core CPU, 64 GB of RAM and a NVidia RTX 3080 GPU.

### RESULT

The performance evaluation of six deep learning models was conducted to investigate their effectiveness in the task of imagining motion. Tab. 2 summarizes the performance metrics for each model, including the performance of the original Riemannian decoder. Our results indicate that shallow ConvNet and deep models have a relatively lower performance in imagining motion, with average test accuracy of 0.52 and 0.62, respectively, slightly above the chance level (0.5). Among the other deep learning models, ResNet, Conformer and ViT achieved performance levels of 0.75, 0.83 and 0.71, respectively, demonstrating superior effectiveness in the task. The ViT model with pre-trained parameters achieved an accuracy of 0.76, higher than the ViT model without pre-trained.

Given that the experiment involves movements of both hands, the binary classification only focuses on distinguishing hand movements, neglecting the distinction between left and right hands. Hence, Fig. 6 presents the classification performance of four gestures across six models, considering both the left- and right-hand movements. Additionally, Fig. 7 illustrates the classification performance of the Conformer model on the dataset from Subject 8 (S8).

Table 2: Comparison of classification accuracy results among 6 deep learning mode. The chance level for the classification is 0.5. ViT\_p model refers to ViT model with pretrained data.

Subject	Shallow ConvNet	DeepNet	ResNet20	Conformer	ViT	ViT_p	Ref [8]
S1	0.45	0.55	0.7	1.0	0.8	0.8	0.82
S2	0.4	0.5	0.7	0.8	0.5	0.8	0.7
S3	0.7	0.7	0.8	0.7	0.7	0.7	0.64
S4	0.35	0.65	0.8	0.9	0.6	0.8	0.64
S5	0.4	0.6	0.8	0.8	0.8	0.7	0.58
S6	0.65	0.65	0.8	0.7	0.7	0.8	0.65
S7	0.7	0.6	0.7	0.9	0.7	0.8	0.61
S8	0.5	0.7	0.7	0.8	0.9	0.7	0.7
Avg	0.52	0.62	0.75	0.83	0.71	0.76	0.67

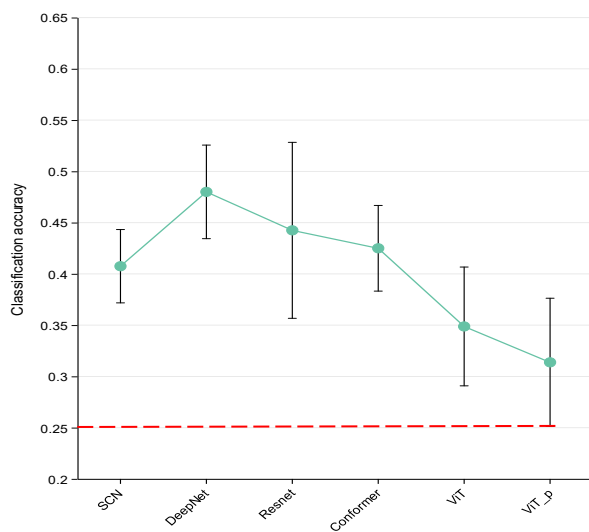


Figure 6: Comparison of 4-gestures classification accuracy results among 6 deep learning mode with red dot line represents for the chance level of 0.25.

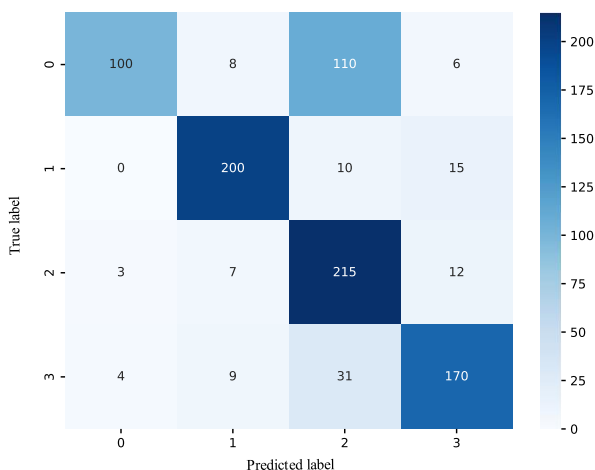


Figure 7: Confusion matrix for 4 gestures classification on S8 with conformer model. Label 0 and label 1 represents the close and open status for left hand respectively. Label 2 and label 3 represents the close and open status for right hand respectively.

## DISCUSSION

For the 2-gesture classification, it suggests the need for sophisticated feature extraction capabilities, which transformer architectures appear to provide, especially when no electrode is located in the motor area. And the result from the ViT with pre-trained data suggests that leveraging pre-training significantly enhances the ViT model's performance in the imagining motion task. Therefore, these results have displayed potential advantages of deep learning models in the imagery motion task. By comparing our findings to previous results, it can be suggested that given limited sEEG dataset, not only the model expressiveness can be improved through data augmentation on dataset itself but also through pretraining on other datasets, such as ImageNet for an EEG regression task [15]. This approach proves effective in improving the classification performance of MI tasks. While transformer-based models may not perform as well as models relying solely on convolutional modules for four-class classification, this can possibly be explained by the electrode distribution. For example, some participants only have electrode implantation on single side of the brain and only some have a limited number of electrodes presented in the motor cortex. Due to contralateral control, the left-hand movements are dominated by the right hemisphere of the brain. As shown in Fig. 7, more than half of the left-hand closing gestures are incorrectly recognized as right-hand closing gestures.

While our study makes valuable contributions, it is essential to acknowledge certain limitations. Due to the limited availability of public sEEG motor imagery datasets, our research focused solely on evaluating the model's performance in classifying two types of gestures within a single dataset. Future investigations should aim for a more comprehensive exploration of task specificity, dataset characteristics, and the impact of model hyperparameters on the ultimate performance. This would allow the exploration of various deep learning architectures, particularly the advantages and limitations of transformer-based models in motor imagery tasks. However, with the varied performance of different

models, especially the enhanced accuracy with pre-trained ViT, we provide more opportunities for further explorations and optimisation.

## CONCLUSION

In this work, we evaluate the performance of sEEG-based imagery motion classification by using multiple deep learning models. By comparing six different deep learning models, we used the conformer model to achieve an accuracy of 0.83 in the binary classification of imagined movements, which is 0.16 higher than the performance of the previous work. This work provides a reference for using deep learning models in BCI imagery movements with sEEG signals.

## ACKNOWLEDGEMENTS

This work is supported by the EPSRC New Horizons Grant of UK (EP/X018342/1).

## REFERENCES

- [1] Z. Xie, O. Schwartz, and A. Prasad, 'Decoding of finger trajectory from ECoG using deep learning', *J. Neural Eng.*, vol. 15, no. 3, p. 036009, Feb. 2018, doi: 10.1088/1741-2552/aa9dbe.
- [2] P. Z. Soroush, C. Herff, S. K. Ries, J. J. Shih, T. Schultz, and D. J. Krusienski, 'The nested hierarchy of overt, mouthed, and imagined speech activity evident in intracranial recordings', *NeuroImage*, vol. 269, p. 119913, Apr. 2023, doi: 10.1016/j.neuroimage.2023.119913.
- [3] G. Li et al., 'Assessing differential representation of hand movements in multiple domains using stereo-electroencephalographic recordings', *NeuroImage*, vol. 250, p. 118969, Apr. 2022, doi: 10.1016/j.neuroimage.2022.118969.
- [4] M. A. Jensen et al., 'A motor association area in the depths of the central sulcus', *Nat. Neurosci.*, vol. 26, no. 7, Art. no. 7, Jul. 2023, doi: 10.1038/s41593-023-01346-z.
- [5] E. Combrisson et al., 'From intentions to actions: Neural oscillations encode motor processes through phase, amplitude and phase-amplitude coupling', *NeuroImage*, vol. 147, pp. 473–487, Feb. 2017, doi: 10.1016/j.neuroimage.2016.11.042.
- [6] X. Wu et al., 'Decoding continuous kinetic information of grasp from stereo-electroencephalographic (sEEG) recordings', *J. Neural Eng.*, vol. 19, no. 2, p. 026047, Apr. 2022, doi: 10.1088/1741-2552/ac65b1.
- [7] B. A. Murphy, J. P. Miller, K. Gunalan, and A. B. Ajiboye, 'Contributions of Subsurface Cortical Modulations to Discrimination of Executed and Imagined Grasp Forces through Stereoelectroencephalography', *PLOS ONE*, vol. 11, no. 3, p. e0150359, Mar. 2016, doi: 10.1371/journal.pone.0150359.
- [8] M. C. Ottenhoff et al., 'Decoding executed and imagined grasping movements from distributed non-motor brain areas using a Riemannian decoder', *Front. Neurosci.*, vol. 17, 2023, Accessed: Nov. 25, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnins.2023.1283491>
- [9] P. Wang, P. Gong, Y. Zhou, X. Wen, and D. Zhang, 'Decoding the Continuous Motion Imagery Trajectories of Upper Limb Skeleton Points for EEG-Based Brain-Computer Interface', *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023, doi: 10.1109/TIM.2022.3224991.
- [10] R. T. Schirrmeister et al., 'Deep learning with convolutional neural networks for EEG decoding and visualization', *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, 2017, doi: 10.1002/hbm.23730.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep Residual Learning for Image Recognition'. *arXiv*, Dec. 10, 2015. Accessed: Dec. 28, 2023. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [12] T. Tian, L. Wang, M. Luo, Y. Sun, and X. Liu, 'ResNet-50 based technique for EEG image characterization due to varying environmental stimuli', *Comput. Methods Programs Biomed.*, vol. 225, p. 107092, Oct. 2022, doi: 10.1016/j.cmpb.2022.107092.
- [13] Y. Song, Q. Zheng, B. Liu, and X. Gao, 'EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization', *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 710–719, 2023, doi: 10.1109/TNSRE.2022.3230250.
- [14] R. Yang and E. Modesitt, 'ViT2EEG: Leveraging Hybrid Pretrained Vision Transformers for EEG Data'. *arXiv*, Aug. 01, 2023. Accessed: Dec. 08, 2023. [Online]. Available: <http://arxiv.org/abs/2308.00454>
- [15] J. Chen, D. Wang, W. Yi, M. Xu, and X. Tan, 'Filter bank sinc-convolutional network with channel self-attention for high performance motor imagery decoding', *J. Neural Eng.*, vol. 20, no. 2, p. 026001, Mar. 2023, doi: 10.1088/1741-2552/acbb2c.

# HIGH-PERFORMANCE NEURAL DECODING OF 14 DUTCH KEYWORDS

Elena C. Offenberg<sup>1</sup>, Julia Berezutskaya<sup>1</sup>, Zachary V. Freudenburg<sup>1</sup> and Nick Ramsey<sup>1</sup>

<sup>1</sup> Brain Center, University Medical Center Utrecht, Utrecht, Netherlands

E-mail: e.c.offenberg@umcutrecht.nl

**ABSTRACT:** Brain-computer interfaces (BCIs) can help people with locked-in syndrome to communicate. While continuous speech decoding can be used in everyday communication, navigating a computer menu or interacting with external devices may be easier and more reliable using a small set of distinct command keywords.

In this preliminary study, two able-bodied epilepsy patients, temporarily implanted with high-density electrocorticography (ECoG) electrodes, spoke 14 potential keywords out loud in Dutch. With optimized Support Vector Machines (SVMs), the maximal decoding accuracy reached was a median of 93.3% for 50 repetitions per word (practical chance level 9.6%). We also identified that a minimum of 30 repetitions was needed to achieve this result, and determined that the most relevant electrodes for decoding were on the ventral sensorimotor cortex, close to the central sulcus.

## INTRODUCTION

People with a neurodegenerative disease, such as amyotrophic lateral sclerosis (ALS), can over time lose their ability to communicate verbally. Such a loss of communication has a direct negative impact on their perceived quality of life [1]. Fortunately, this effect can be assuaged: using communication devices has been shown to have a positive impact on quality of life and mood in dysarthric people with ALS [2].

Recently, there have been advancements in large-vocabulary decoding using brain-computer interfaces (BCIs) [3,4]. However, these large language models have not yet been shown to work in a home-use scenario. For daily communication or for controlling the home environment, a more robust and stable solution may be preferred. A very robust solution, using high gamma signals for one-dimensional cursor control, has been utilized successfully over several years [5]. However, such one-dimensional control limits the agency of the user. The next step allowing for more complex computer control could be the use of a limited set of keywords. For example, using six-keyword navigation with commands “up”, “down”, “left”, “right”, “enter” and “back” decoded from ECoG signals has recently demonstrated reliable control of computer menus and environment at high accuracy (median accuracy of 90.59%) over several months [6]. However, scaling up the individual word decoding from

ECoG has proven rather challenging due to inter-subject variability, limited amounts of data and inherent limitations of decoding from brain signals. Further recent studies of keyword decoding from brain signals achieved 47.1% accuracy decoding 50 words in a paralyzed person with anarthria [7], 74.1% accuracy in a person with vocal paralysis [8], and 92%–100% decoding accuracy of 12 words in able-bodied participants [9].

One very relevant question both for researchers working with limited time and for BCI patients wanting to utilize their assistive devices with as little delay as possible is how much data is necessary to reach an acceptable decoding accuracy. Current decoding attempts are often limited by the small amounts of data researchers are able to collect with transient ECoG recordings. Therefore, determining the minimum amount of data necessary for satisfying decoding accuracies is one of our main goals.

In this study, we acquired high-density ECoG data from two subjects who spoke 14 Dutch words out loud, with S1 repeating each word 50 times. Using this data, we investigated three specific questions: 1. What is the highest accuracy of decoding 14 individual words? 2. How much data is needed to reach this accuracy, and 3. Which cortical areas are relevant for decoding?

We found that an accuracy of 93.3% (chance level 9.6%) could be reached with optimized SVMs with only 30 repetitions per word and that the electrodes contributing to the decoding performance the most were located on the postcentral gyrus of the ventral sensorimotor cortex, close to the central sulcus.

## MATERIALS AND METHODS

Two human subjects S1 and S2 (1 male, 1 female, 26 and 46 years old, respectively) with medication-resistant epilepsy were implanted with 32-electrode high density (HD-)ECoG grids with platinum-iridium electrodes, a 4 mm inter-electrode distance and 1 mm exposed diameter. The grids were located on the left hemisphere covering the ventral sensorimotor cortex. Participant S1 had a previous tissue resection in the left temporal cortex. The study was approved by the Medical Ethical Committee of the University Medical Center Utrecht in accordance with the Declaration of Helsinki (2013). The subjects gave written informed consent to participate in research tasks.

The 14 words were candidates for navigational words in Dutch, namely (in alphabetical order): "beneden" (down), "boven" (up), "kiezen" (choose), "links" (left), "noord" (north), "omhoog" (upwards), "omlaag" (downwards), "oost" (east), "rechts" (right), "selecteer" (select), "terug" (back), "verwijder" ("remove"), "west" (west) and "zuid" (south). During each repetition of the task, the 14 words and "-" (for rest trials, referred to as "rest" hereafter) were shown 5 times each in random order and read aloud by the participant. For S1, the task was repeated 10 times over the course of four days, resulting in a total number of  $10 \times 5 = 50$  repetitions per word. For S2, one run was recorded, resulting in 5 repetitions of each word. HD-ECOG data was recorded using a Micromed system at 2048 Hz. Simultaneously, microphone data was recorded to determine the voice onset times. Voice onsets were defined manually as the first moment during which speech could be audibly perceived. In S1, one trial of the word "terug" was excluded from further analysis, since the subject did not say it during this trial.

Data pre-processing consisted of notch filtering of line noise (50 Hz) and its harmonics, common average re-referencing, and high frequency band (HFB) component extraction (70-170 Hz) using a Morlet wavelet decomposition in 1 Hz frequency bins, implemented via MNE-Python [10]. The high-frequency components were then averaged across frequencies, log-transformed and downsampled to 100 Hz.

The HFB signals of each run were "re-calibrated", i.e. normalized individually per run using the mean and standard deviation of a 2-second rest period prior to the beginning of the task. The signals were then concatenated across runs and split into trials of 0 to 1 second after each voice onset time. Extending the trial length and including signals from before the voice onset, namely from 0.5 seconds before to 2 seconds after, did not change the accuracy results and was therefore not further pursued.

Due to the limited sample size, theoretical chance levels and practical chance levels differ [11]. Therefore, we used a binomial cumulative distribution to derive statistical significance thresholds for the obtained accuracies [11]. The practical chance levels were set at  $p < 10^{-3}$  for the given sample sizes.

One repetition of every word was used as validation, and another as part of the test set, resulting in as many folds as there were repetitions per word. The flattened trial data (vectorized electrodes x time-points) was used to train an optimized Support Vector Machine (SVM) with a linear kernel in a one-vs-one approach with leave-one-group-out nested cross-validation. In the inner loop, the SVM regularization parameter was optimized using an automatic hyperparameter selection library Optuna [12], while the outer loop was necessary for cross-validating the classification results.

Since SVMs do not inherently provide probability estimates, the class membership probability estimates for the SVM were calculated with Scikit-learn [13],

which uses Platt scaling and five-fold cross-validation.

The electrode weights were determined by the L2 norm of the respective coefficients in the trained SVM. As per calculation of the L2 norm, we summed over the time dimension. The electrodes with the largest absolute classifier weights have the biggest impact on the classification.

## RESULTS

### Accuracy and Misclassified Trials

In S1, when trained and tested using all 50 repetitions per word or rest trial, the SVM reached a median accuracy of  $93.3 \pm 6.7\%$  across folds. For S2 with 5 repetitions, a median accuracy of  $73.3 \pm 6.7\%$  across folds was reached. The practical chance levels were 9.6% and 17.3% for S1 and S2, respectively - thus, both results were well above chance.

Not performing the re-calibration, which normalized the HFB data per run for S1, did not change the decoding accuracy.

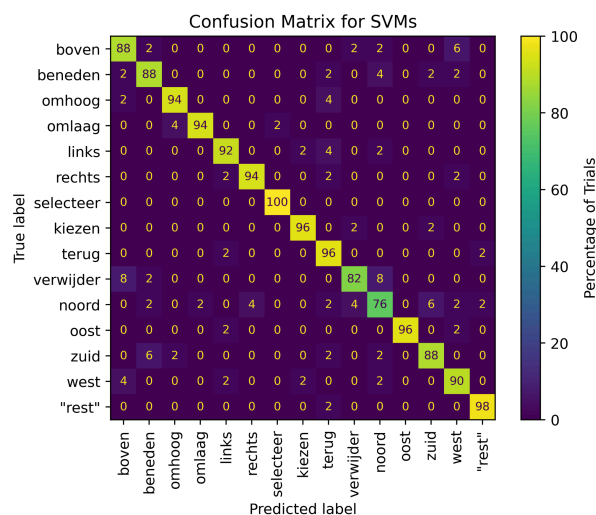


Figure 1: Normalized confusion matrix for S1 for 14 words and rest, with 50 repetitions for every word but "terug" (49 repetitions).

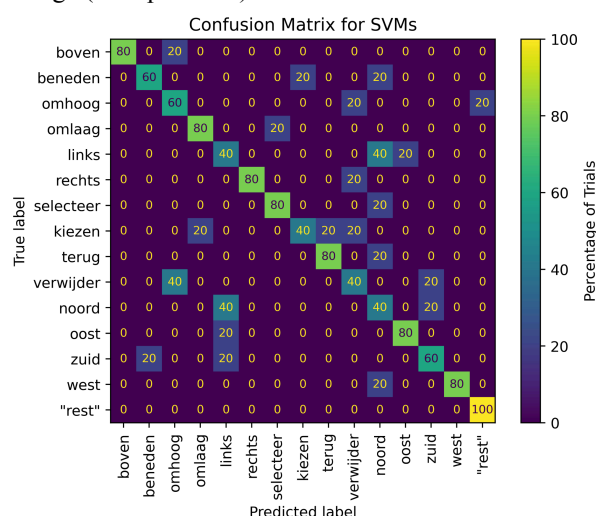


Figure 2: Normalized confusion matrix for S2 for 14 words and rest, with 5 repetitions for every word.

As can be seen in Figures 1 and 2, the prediction accuracy varied between different words. Although for S1, the highest accuracy was a 3-syllable word, “selecteer”, there was no significant correlation between decoding accuracy and length of the words (Pearson coefficient of 0.16,  $p$ -value 0.90).

To see how uncertain the trained SVM for S1 was about its predictions, we visualized the calculated class probabilities for each trial, sorted by words (Figure 3). For trials classified correctly (in gray), the probability often peaked at the near-maximum for the target class and was quite low for all non-target classes. For the misclassified trials (in red), there were often several probability peaks that included the target class.

Quantitatively, across all misclassified trials, the target class was assigned the second or third highest probability in 78% of cases, showing that even in trials with incorrect predictions, correct patterns were still being picked up by the SVM. In total, in 98.3% of trials, the correct class was among the top three predictions.

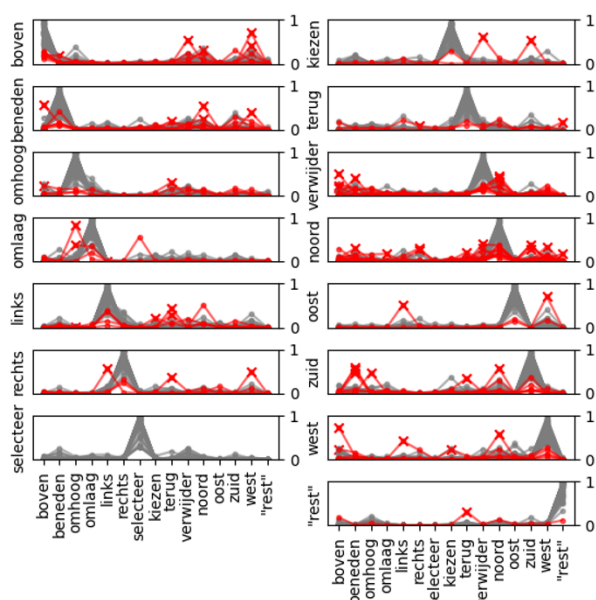


Figure 3: The probability distributions of the SVM predictions for each word and rest-trials for S1. The target words are on the y axis, the predicted words on the x axis. Correctly classified trials are plotted in gray, the misclassified trials are plotted in red, and the incorrect predictions are marked with red crosses.

### Amount of Data Necessary

How much data is necessary in order to reach an acceptable decoding accuracy? For S1, 50 repetitions per word were recorded across 10 runs over the course of several days (5 word repetitions per run). When trained with the data from successive runs cumulatively, the decoding accuracy increased (Figure 4). This analysis uses mean accuracy values instead of medians since the mean provides a smoother statistic over the number of repetitions.

A mean accuracy of  $68.0 \pm 9.8\%$  for S1 and  $66.6 \pm 11.2\%$  for S2 was reached after the first run. A similar

result was achieved for the calculations based on median values.

Notably for S1, even after the decoding accuracy reached the ceiling, the variance in performance decreased as more trials were added, suggesting a further stabilization of the decoding performance.

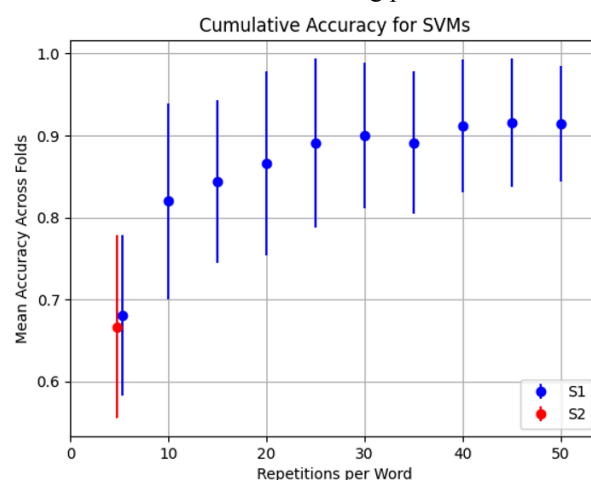


Figure 4: Cumulative accuracy for SVMs as a function of number of word repetitions.

### Electrodes Relevant for Decoding

Not all of the 32 electrodes in each grid contributed to the decoding performance in the same proportion.

In Figures 5 and 6, the positions of the grids for S1 and S2 on the left hemisphere are shown. In addition, we visualize the normalized absolute SVM weights to highlight electrodes most relevant for the decoding performance. For both subjects, the most relevant electrodes were located close to the central sulcus on the ventral sensorimotor cortex, an area associated with the cortical control of articulation [14]. Importantly, electrodes with highest SVM weights were located on the postcentral gyrus.

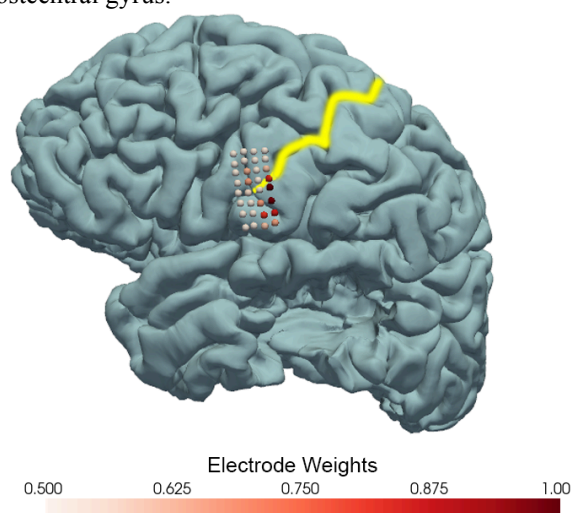


Figure 5: The electrode grid of S1, with darker colors corresponding to higher normalized SVM weights of the electrode. The central sulcus is highlighted in yellow.

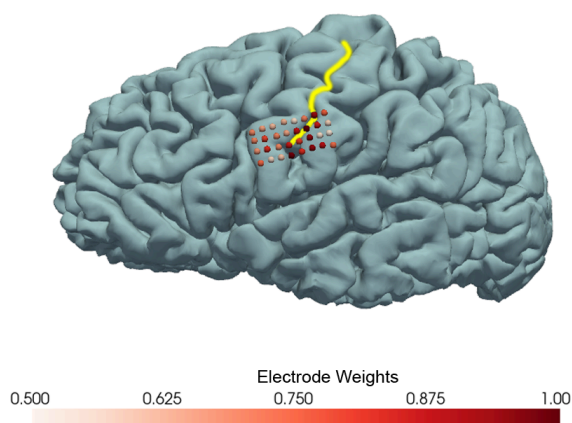


Figure 6: The electrode grid of S2, with darker colors corresponding to higher normalized SVM weights of the electrode. The central sulcus is highlighted in yellow.

## DISCUSSION

In this study, we set out to answer three questions concerning decoding of individual words from HD-ECoG, namely: what is the highest decoding accuracy, what is the minimal amount of data necessary and what are the locations of the most informative electrodes.

We were able to reach a median accuracy of 93.3% for decoding 14 words with 50 repetitions per word. Upon closer inspection, we saw that in over 98% of all trials, the correct words were among the top three SVM predictions. This result could be used to improve predictive performance even further, for example by giving the user a list of top-ranking alternatives or combining top predictions with statistics of previous use to enable quick corrections by the user in case of misclassification.

For controlling external devices, an accuracy of command identification of at least 90% has been determined as the acceptable threshold in a survey among ALS-patients [15]. This threshold was reached after 6 runs, corresponding to 30 repetitions per word, see Figure 4.

For both subjects, we obtained a mean accuracy of over 66% with only five repetitions per word, which is well above chance. While this accuracy does not yet approach the threshold desired for long-term use, a model trained on only five repetitions could already be used to provide immediate feedback to the user while continuing to update the decoder in the background. In our experience, participants find tasks with feedback more engaging, leading to a higher quantity of data collection. The concrete influence of such early feedback on both motivation and performance could be further investigated in a future work.

The most relevant electrodes in both subjects were located on the dorsal part of the ventral sensorimotor cortex, with the highest weights found closest to the central sulcus. This mirrors the results of another

ECoG-study for keyword decoding [6]. Interestingly, most contributing electrodes seemed to be located on the postcentral gyrus – the somatosensory area of the brain. Since our subjects were able-bodied people, one might attribute this result to the sensory feedback from mouth movements. However, motor decoding from the somatosensory cortex has previously been shown both for amputees [16] and people with paralysis due to ALS [17], suggesting that there is information about movement in somatosensory areas even in the absence of direct sensory feedback. This is in line with work on an efference copy of voluntary movements in the somatosensory cortex [18, 19]. It remains to be seen whether comparable decoding performance from the sensorimotor cortex can be achieved in locked-in individuals using a BCI.

One important difference between our study subjects and future locked-in users is that the voice onset times will not be available as ground truth for training the classifiers. As an alternative, different methods of extracting activity onset directly from brain signals have already been proposed and used [4,6].

## CONCLUSION

In the present study, we achieved a high accuracy of decoding 14 individual words from HD-ECoG brain activity recorded from the ventral sensorimotor cortex of two able-bodied subjects. For a subject with 50 repetitions per word, 30 repetitions per word were sufficient to reach a decoding accuracy of over 90%, and the most informative electrodes for both patients were located in the ventral postcentral gyrus.

## ACKNOWLEDGEMENTS

This publication is part of the project Dutch Brain Interface Initiative (DBI2) with project number 024.005.022 of the research programme Gravitation, which is financed by the Dutch Ministry of Education, Culture, and Science (OCW) via the Dutch Research Council (NWO).

In addition, this project is funded by the European Union's HORIZON-EIC-2021-PATHFINDER CHALLENGES program under grant agreement No 101070939 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00198. Finally, the authors were supported by the National Institutes of Health under award number UH3NS114439 (NINDS).

## REFERENCES

- [1] Felgoise SH, Zaccheo V, Duff J, Simmons Z. Verbal communication impacts quality of life in patients with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2016;17(3-4):179-183
- [2] Körner S, Siniawski M, Kollwe K, Rath KJ, Krampfl K, Zap, A et al. Speech therapy and

- communication device: Impact on quality of life and mood in patients with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2013;14(1):20-25.
- [3] Willett FR, Kunz EM, Fan C, Avansino DT, Wilson GH, Choi EY et al. A high-performance speech neuroprosthesis. *Nature*. 2023;620(7976):1031-1036.
- [4] Metzger SL, Littlejohn KT, Silva AB, Moses DA, Seaton MP, Wang R et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*. 2023;620(7976):1037-1046.
- [5] Pels EGM, Aarnoutse EJ, Leinders S, Freudenburg ZV, Branco MP, van der Vijgh BH et al. Stability of a chronic implanted brain-computer interface in late-stage amyotrophic lateral sclerosis. *Clinical Neurophysiology*. 2019;130(10):1798-1803.
- [6] Luo S, Angrick M, Coogan C, Candrea DN, Wyse-Sookoo K, Shah S, et al. Stable Decoding from a Speech BCI Enables Control for an Individual with ALS without Recalibration for 3 Months. *Advanced Science*. 2023;10(35):2304853.
- [7] Moses DA, Metzger SL, Liu JR, Anumanchipalli GK, Makin JG, Sun PF et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine*. 2021;385(3):217-227.
- [8] Metzger SL, Liu JR, Moses DA, Dougherty ME, Seaton MP, Littlejohn KT et al. Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis. *Nat Commun*. 2022;13(1):6510.
- [9] Berezutskaya J, Freudenburg ZV, Vansteensel MJ, Aarnoutse EJ, Ramsey NF, Van Gerven MAJ. Direct speech reconstruction from sensorimotor brain activity with optimized deep learning models. *J Neural Eng*. Published online July 19, 2023.
- [10] Gramfort A, Luessi M, Larson E, Engeman DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-Python. *Front Neurosci*. 2013;7.
- [11] Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*. 2015;250:126-136.
- [12] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM; 2019:2623-2631.
- [13] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al. Scikit-learn: Machine Learning in Python. Published online 2012. 12(85):2825-2830, 2011.
- [14] Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. *Nature*. 2013;495(7441):327-332.
- [15] Huggins JE, Wren PA, Gruis KL. What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*. 2011;12(5):318-324.
- [16] Bruurmijn MLCM, Pereboom IPL, Vansteensel MJ, Raemaekers MAH, Ramsey NF. Preservation of hand movement representation in the sensorimotor areas of amputees. *Brain*. 2017;140(12):3166-3178.
- [17] Leinders S, Vansteensel MJ, Piantoni G, et al. Using fMRI to localize target regions for implanted brain-computer interfaces in locked-in syndrome. *Clinical Neurophysiology*. 2023;155:1-15.
- [18] Umeda T, Isa T, Nishimura Y. The somatosensory cortex receives information about motor output. *Sci Adv*. 2019;5(7):eaaw5388.
- [19] Christensen MS, Lundbye-Jensen J, Geertsens SS, Petersen TH, Paulson OB, Nielsen JB. Premotor cortex modulates somatosensory cortex during voluntary movements without proprioceptive feedback. *Nat Neurosci*. 2007;10(4):417-419.

# PROCESSING OF INCONGRUENT INFORMATION CAN BE DECODED FROM SINGLE-TRIAL EEG: AN AR-STUDY

M. Wimmer<sup>1,2</sup>, A. Pepicelli<sup>3</sup>, B. Volmer<sup>3</sup>, N. ElSayed<sup>1</sup>, A. Cunningham<sup>3</sup>,  
B.H. Thomas<sup>3</sup>, E.E. Veas<sup>1,4</sup>, G.R. Müller-Putz<sup>2,5</sup>

<sup>1</sup>Know-Center GmbH, Graz, Austria

<sup>2</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>3</sup>Wearable Computer Lab, University of South Australia, Adelaide, SA, Australia

<sup>4</sup>Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria

<sup>5</sup>BioTechMed Graz, Graz, Austria

E-mail: eveas@know-center.at, gernot.mueller@tugraz.at

**ABSTRACT:** Augmented reality (AR) allows users to display additional digital information about their physical environment. We present an interactive AR study, in which participants manipulated a Rubik's cube which served as a physical referent for presented digital information showing the current status of the cube. In 30% of the instances, the presented information did not match its status. We recorded the electroencephalographic data of 19 participants to study their responses to incongruent stimuli and assessed if they could be classified on a single-trial level. We found that the processing of incongruent data in AR elicits both N400 and P600 components. Further, we could classify them in 15 out of 19 participants with accuracies above chance. These results contribute to the design of brain-computer interfaces, as the decoding of such correlates could inform the system about the current mental context of the user.

## INTRODUCTION

Augmented reality (AR) allows the integration of virtual content into the real world [1]. With the increasing number of head-mounted displays (HMDs) and other personal electronic devices that can create AR visualizations, the technology has become a widely available tool with manifold applications [2]. One goal of situated AR visualizations is to communicate information about physical objects to users or assist them in specific tasks [3]. To give two practical examples, this information could guide users through procedural tasks via visual cues [4] or support them, e.g., in their purchasing decisions by displaying relevant details directly next to products [5].

Efforts have been made to study users' electroencephalographic (EEG) responses to the presentation of data, e.g., showing anomalous information. A particularly prominent component of the event-related potential (ERP) caused by incongruent stimuli is the N400. It is a negative deflection relative to congruent stimuli that peaks approximately 400 ms after stimulus onset in centro-parietal areas of the scalp [6]. The N400 component has been found

in response to numerous types of conflicting stimuli, such as incongruent words in sentences [7], incongruent solutions of simple mathematical problems, [8], pictures [9], or gestures [10]. Hence, Kutas et al. [7] described the N400 as "an electrophysiological sign of the 'reprocessing' of semantically anomalous information". This is relevant in the context of brain-computer interfaces (BCIs) [11], as decoding incongruent stimuli could allow systems to infer information about the user's mental context without making it explicit.

However, only a few studies attempted to decode semantic incongruencies on a single-trial level. Geuze et al. [12] found an N400 effect in a word association task using related and unrelated word pairs and could decode them with accuracies between 54% and 67%. In a similar task, Dijkstra et al. [13] studied the responses to multiple consecutive word stimuli after presenting a target word. They reported similar neural responses and achieved a classification accuracy of 59.5%. Both works used an  $L_2$  regularized logistic regression algorithm for classification. Finally, Tanaka et al. [14] presented semantically correct and incorrect sentences and found both N400 and P600 components. Using a multilayer perceptron, they could correctly identify them in up to 59.5% of the instances. Interested readers are referred to [15] for an overview of N400 for BCIs.

All three above-mentioned papers explored EEG correlates of semantic anomalies in language processing. In this work, we studied the neural responses to incongruent information in an AR scenario. In particular, we investigated the following two research questions (RQs):

**RQ1)** Can we find EEG responses to the presentation of incongruent information using AR?

**RQ2)** Can we use these responses to discriminate congruent and incongruent information on a single-trial level?

For this, we designed an interactive paradigm, in which users were visually instructed on how to manipulate a physical Rubik's cube and presented situated information related to the cube, which could either match the users' expectations or not.

## MATERIALS AND METHODS

**Participants:** Twenty healthy volunteers (20 to 45 years old, 28.2 years on average, 14 male and six female) participated in the study. The study was conducted in accordance with the Declaration of Helsinki (1975) and approved by the ethics committee of the University of South Australia. All participants gave their written consent before conducting the experiment and received vouchers worth 40 AUD as compensation.

**AR HMD:** AR visualizations were presented using a HoloLens 2 (Microsoft, Redmond, WA, USA) and designed in Unity 2021.1.3<sup>1</sup>. We used the HMD to record gaze signals at 30 Hz during the data presentations.

**EEG recordings:** EEG signals were acquired using a BrainAmp amplifier (Brain Products, Munich, Germany) at 500 Hz. We positioned 32 electrodes according to the international 10-10 system at AFz, F3, F1, Fz, F2, F4, FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P3, P1, Pz, P2, and P4. Reference and ground electrodes were placed at Fpz and the right mastoid, respectively. EEG signals, gaze data, and markers from the experimental paradigm were synchronized utilizing the lab streaming layer (LSL) protocol<sup>2</sup>.

**Experimental setup:** Participants sat at a table such that they could easily reach a tricolor (red, blue, white) Rubik's cube (see Fig. 1). A camera (Canon EOS 200D II, Tokyo, Japan) pointed at the cube to detect the nine colors of its top surface (camera not visible in Fig. 1). Color detection was performed based on the Qbr Rubik's cube solver<sup>3</sup>, implemented in Python and OpenCV.



Figure 1: **Experimental setup.** A participant wearing an EEG cap and an HMD. The tricolor Rubik's cube is in its starting position.

**Experimental procedure:** First, participants took an Ishihara test<sup>4</sup> to assess their color vision [16]. After successful completion, participants performed one training run consisting of ten trials to familiarize themselves with the experimental paradigm. The following experiment consisted of 6 runs of 33 trials (23 congruent, 10 incon-

gruent). The order of the trials was randomized. Between runs, the participants could take short breaks of usually one to five minutes to avoid fatigue.

**Experimental paradigm:** The timings of one trial of the paradigm are shown in Fig. 2. Each trial started with the presentation of one or two visual cues indicating which one or two cube manipulations the participants should perform. Manipulations included rotating a specific row or column of the cube, or the whole cube in a given direction. Depending on the number of manipulations, the visual cues were presented for 1 or 2 seconds (s). Thereafter, participants were instructed to take the cube, perform the indicated manipulations, and return it to its starting position. At this point, participants should count the number of red, blue, and white squares on the cube's top surface. For example, in Fig. 2, after performing the manipulations the count would be three red, two blue, and four white squares. The participants indicated that they knew the correct count by pressing a physical button on the keyboard. This triggered the presentation of a fixation cross and a frame on the left side of the cube. After 1.3 to 1.7 s (randomized), a congruent or incongruent count was presented inside the frame for .75 s following the order red-blue-white. Incongruent answers deviated strongly from congruent answers, i.e., 1-1-1 or 0-8-0, which are impossible counts per se. Participants were instructed to fixate their gaze on the cross and to avoid gaze shifts during the data presentation as much as possible. Each data presentation was followed by a break of .75 s before a new trial was introduced with a countdown from two to zero (1.5 s).

**EEG data preprocessing:** The data processing and analysis were performed offline using Matlab R2022a (The MathWorks, MA, USA) incorporating the EEGLAB toolbox (v2022.0) [17].

EEG data were filtered between 1 and 25 Hz using a zero-lag Butterworth filter of order 4. We applied two notch filters at 30 Hz and 50 Hz to suppress noise from the HMD and the power line (zero-lag, Butterworth, second order). After resampling the signals to 125 Hz, we applied the extended Infomax algorithm [18] to perform independent component analysis (ICA) [19]. Based on visual inspection, we rejected components corresponding to eye movements or blinks. Thereafter, we segmented the data into epochs of 1.5 s ([-.5, 1] s relative to the stimulus onset).

We rejected contaminated epochs through visual inspection and based on amplitude ( $\pm 35 \mu V$ ), kurtosis, and joint probability (both 5 times the standard deviation) [20], similar to [21]. To avoid the influence of possible residual eye-related artifacts, we rejected trials with excessive eye movements. For that, we computed the variance of the eye movements of each trial and removed epochs with a z-score outside  $\pm 3$ . On average, we rejected 9% of the trials per participant.

One participant could not identify the incongruent stimuli and was subsequently removed from the analysis.

**Classification:** We performed stimulus-locked classification with two classes (congruent and incongruent). As

<sup>1</sup><https://unity.com>

<sup>2</sup><https://github.com/scn/labstreaminglayer>

<sup>3</sup><https://github.com/kkoomen/qbr>

<sup>4</sup><https://www.colorblindnesstest.org/ishihara-test/>

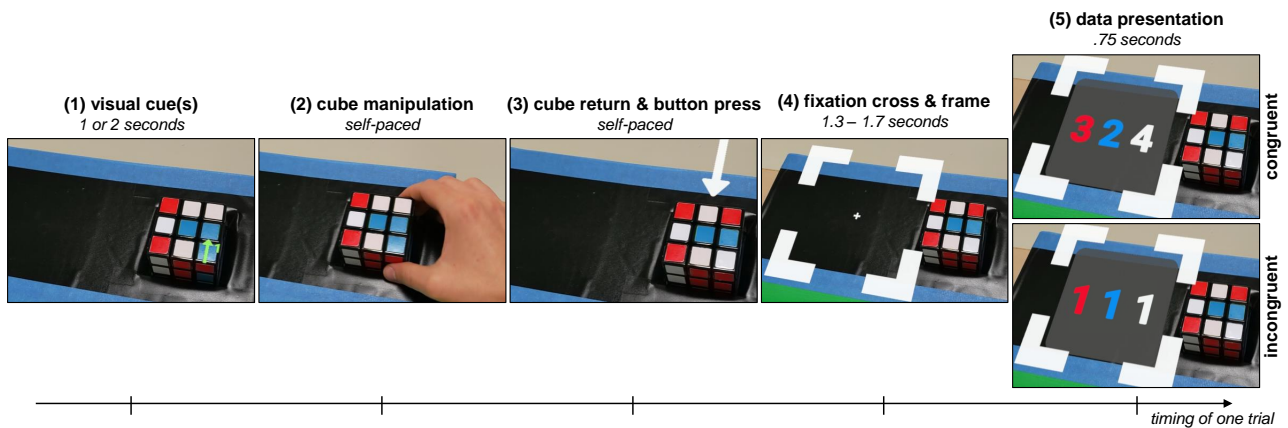


Figure 2: **Timing of one trial.** (1) Each trial started with the presentation of one or two visual cues (green arrow). In the given example, the participant was instructed to rotate the right column upwards. (2) After performing the manipulation, participants returned the cube to its starting position and counted the colors on the top surface. In the example, the count is 3 red, 2 blue, and 4 white. (3) As soon as the participants knew the correct count, they pressed a physical button. (4) Next, a fixation cross and a frame for the following data presentation were presented for 1.3 to 1.7 s. (5) Finally, a correct (top) or incorrect (bottom) count was presented for .75 s.

features, for each channel, we computed the average amplitude of overlapping windows of 152 ms in steps of 32 ms, where the first window started at 252 ms after the stimulus and the last window after 844 ms (e.g., 252-404 ms, 284-436 ms, etc.) [22]. These features were z-scored and used to train a shrinkage linear discriminant analysis (sLDA) classifier [23]. For every participant, we trained and tested the personalized classifier ten times in a 5-fold cross-validation approach on balanced datasets by choosing a random sample of congruent trials for each of the ten iterations.

We report the true positive rates (TPR) as the fraction of incongruent trials that were correctly classified. Analogously, the true negative rate (TNR) is the fraction of correctly classified congruent trials. The accuracy is the mean of the TPR and the TNR. TPR, TNR, and accuracy are calculated from the average of the 50 folds.

**Statistics:** We performed Wilcoxon signed-rank tests for a sample-wise comparison of the ERPs following congruent and incongruent stimuli (Fig. 3). To correct for multiple comparisons, we applied the false discovery rate (FDR) procedure ( $\alpha = .05$ ). To assess if the classification accuracies are significantly above chance [24], we computed the 95% confidence interval of the chance level using a cumulative binomial distribution [25]. We calculated the significance level (SL) for each participant individually (Tab. 1).

## RESULTS

Figure 3 shows the grand average EEG results, i.e., the mean of the 19 participant averages. The grand average ERPs at CPz are depicted for both classes (mean  $\pm$  2 times the standard error of the mean (SEM)) as well as their difference computed by subtracting congruent from incongruent (Fig. 3a). This difference signal has a negative peak with a maximum amplitude of  $-2.53 \mu\text{V}$  at  $t = .47$  s and a positive peak with a maximum of

$1.82 \mu\text{V}$  at  $t = .68$  s, relative to the stimulus onset. Congruent and incongruent responses differ significantly ( $p < .05$ ) in the intervals [.40, .52] s, [.64, .72] s, and [.76, .77] s. For the first two, we show the mean topographical distributions of the intervals, revealing mainly centro-parietal responses (Fig. 3b).

Table 1 summarizes the classification accuracies for each participant, including the TPR, TNR, and SL. On average, 63.3% (TPR = 62.3%, TNR = 64.3%) of the trials were correctly classified, which exceeds the average SL by 5.3%. In 15 out of 19 participants (79%), incongruent stimuli could be distinguished from congruent ones above chance, five achieved accuracies of 70% or higher.

Table 1: **Classification results.** Participants with accuracies exceeding the SL are marked with “\*”.

Participant	TPR	TNR	Accuracy	SL
	%	%	%	%
P1*	76.0	76.8	76.4	58.0
P2	51.6	56.8	54.2	57.6
P3*	59.6	60.6	60.1	57.6
P4*	60.0	56.1	58.1	58.0
P5	47.6	52.4	50.0	58.2
P6*	63.8	64.4	64.1	58.0
P7	55.3	54.1	54.7	58.0
P8*	56.5	60.3	58.4	57.5
P9*	67.1	72.9	70.0	58.2
P10*	59.1	61.8	60.4	57.6
P11*	57.5	63.9	60.7	58.3
P12*	73.5	77.3	75.4	58.3
P13	57.5	55.3	56.4	58.0
P14*	63.3	63.6	63.5	58.0
P15*	72.5	70.3	71.4	58.0
P16*	59.8	61.8	60.8	58.2
P17*	63.4	63.8	63.6	58.0
P18*	56.8	61.6	59.2	57.8
P19*	82.4	88.0	85.2	58.2
Average	62.3	64.3	63.3	58.0

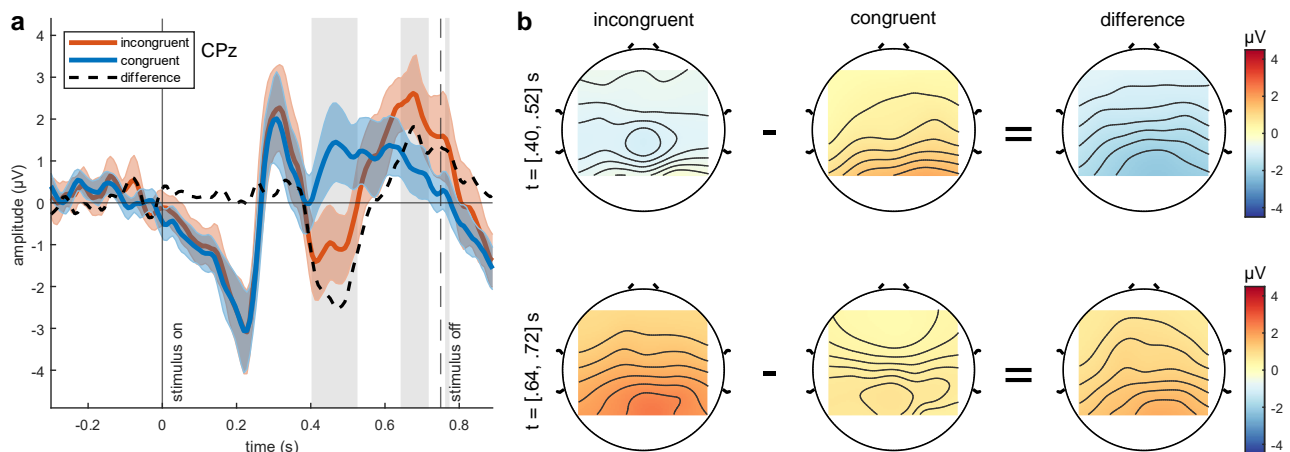


Figure 3: **Grand average EEG results.** (a) ERPs at CPz following the presentation of congruent (blue) and incongruent (orange) stimuli, and their difference *incongruent* – *congruent* (dashed black). Shaded areas indicate  $\pm 2 \cdot SEM$ . Significant differences ( $p < .05$ ) are highlighted in gray. (b) Average topographical distributions of the significant intervals  $[.40, .52] s$  (top row) and  $[.64, .72] s$  (bottom row) for incongruent (left column), congruent (middle column), and their difference (right column).

## DISCUSSION

We developed an interactive AR task incorporating situated information related to a Rubik's cube. In situ data presentation is an application area for AR technologies that offers complementary information about the user's physical environment [26]. The information is usually derived from available data and may not match the expectations of the users, given their mental context in their current situations.

Subsequently, to answer **RQ1**, we studied the participants' neurophysiological responses following incongruent stimuli and found centro-parietal N400 and P600 effects. The N400 has previously been linked to the processing of incongruent information. Since the current experiment yielded very similar patterns, we conclude that participants perceived the erroneous counts as incongruent with their semantic context, i.e., the Rubik's cube. For instance, both the N400 and P600 components have been reported for arithmetic incongruencies, found after simple multiplication errors, e.g., " $7 \cdot 8 = 54$ " [8]. Judged by its scalp distribution, the authors hypothesized that the positive peak belongs to the family of P300 effects and subsequently reflects the participants' surprise following improbable stimuli, which is likely to be the case in our work too. Similarities in time course, topography, and polarity have already been suggested earlier [27]. This aligns with Coulsen et al. [28], whose experiments revealed the sensitivity of the P600 amplitude to the frequency of improbable stimuli, similar to the P3b, a sub-component of the P300. The increased latency was explained by differences in the stimulus complexity [29]. However, counterevidence was provided when different neural generators were found to play crucial roles in the modulation of the P600 and the P300 [30]. The debate on whether the P600 is a form of a P3b is still ongoing, we refer to Leckey and Federmeier [31] for an overview.

For **RQ2**, we studied the feasibility of distinguishing congruent and incongruent responses on a single-trial basis. Using personalized classifiers, i.e., trained and tested with data of the same participant, we could decode incongruent stimuli with accuracies between 50% and 85%. Given the relevance of the N400 and P600 in neuroscience research, these components have not yet been granted much attention in the BCI community. To our best knowledge, only three works have attempted time-locked classification of incongruent information using linguistic stimuli, achieving accuracies of nearly 60% [12–14]. Our classification results are in a similar range, exceeding their reported accuracies slightly by about 4%. However, we provide first evidence that decoding the processing of incongruent in situ information in AR is feasible.

## CONCLUSION

In this work, we demonstrated that neural correlates of the processing of incongruent information can be measured in AR scenarios. These correlates are consistent with the existing literature that focuses on monitor-based tasks. Further, we showed that the classification of incongruent trials on a single-trial level above chance is possible for most participants. This can be relevant for the design of BCIs since these correlates could allow to infer active mental concepts of users.

## ACKNOWLEDGEMENTS

The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth, and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

## REFERENCES

- [1] Azuma RT. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments*. 1997;6(4):355–385.
- [2] Marriott K *et al.* *Immersive analytics*. Springer (2018).
- [3] Kalkofen D, Sandor C, White S, Schmalstieg D. Visualization techniques for augmented reality. In: *Handbook of Augmented Reality*. Springer, 2011, 65–98.
- [4] Volmer B, Liu JS, Matthews B, Bornkessel-Schlesewsky I, Feiner S, Thomas BH. Multi-level precues for guiding tasks within and between workspaces in spatial augmented reality. *IEEE Transactions on Visualization and Computer Graphics*. 2023;29(11):4449–4459.
- [5] ElSayed N, Thomas B, Marriott K, Piantadosi J, Smith R. Situated analytics. In: *Proc. BDVA*. 2015, 1–8.
- [6] Kutas M, Federmeier KD. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*. 2011;62:621–647.
- [7] Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*. 1980;207(4427):203–205.
- [8] Niedeggen M, Rösler F, Jost K. Processing of incongruous mental calculation problems: Evidence for an arithmetic N400 effect. *Psychophysiology*. 1999;36(3):307–324.
- [9] Barrett SE, Rugg MD. Event-related potentials and the semantic matching of pictures. *Brain and Cognition*. 1990;14(2):201–212.
- [10] Wu YC, Coulson S. Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*. 2005;42(6):654–667.
- [11] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*. 2002;113(6):767–791.
- [12] Geuze J, Gerven MA van, Farquhar J, Desain P. Detecting semantic priming at the single-trial level. *PloS One*. 2013;8(4):e60377.
- [13] Dijkstra K, Farquhar J, Desain P. Electrophysiological responses of relatedness to consecutive word stimuli in relation to an actively recollected target word. *Scientific Reports*. 2019;9(1):14514.
- [14] Tanaka H, Watanabe H, Maki H, Sakriani S, Nakamura S. Electroencephalogram-based single-trial detection of language expectation violations in listening to speech. *Frontiers in Computational Neuroscience*. 2019;13:15.
- [15] Dijkstra K, Farquhar J, Desain P. The N400 for brain computer interfacing: Complexities and opportunities. *Journal of Neural Engineering*. 2020;17(2):022001.
- [16] Birch J, McKeever LM. Survey of the accuracy of new pseudoisochromatic plates. *Ophthalmic and Physiological Optics*. 1993;13(1):35–40.
- [17] Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. 2004;134(1):9–21.
- [18] Lee TW, Girolami M, Sejnowski TJ. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*. 1999;11(2):417–441.
- [19] Makeig S, Bell A, Jung TP, Sejnowski TJ. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*. 1995;8.
- [20] Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*. 2007;34(4):1443–1449.
- [21] Wimmer M, Weidinger N, Veas E, Müller-Putz GR. Multimodal decoding of error processing in a virtual reality flight simulation. *Scientific Reports*. 2024;14(1):9221.
- [22] Yasemin M, Cruz A, Nunes UJ, Pires G. Single trial detection of error-related potentials in brain–machine interfaces: A survey and comparison of methods. *Journal of Neural Engineering*. 2023;20(1):016015.
- [23] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*. 2011;56(2):814–825.
- [24] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: A closer look on BCI results. *International Journal of Bioelectromagnetism*. 2008;10:52–55.
- [25] Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*. 2015;250:126–136.
- [26] Veas E, Grasset R, Ferencik I, Grünwald T, Schmalstieg D. Mobile augmented reality for environmental monitoring. *Personal and Ubiquitous Computing*. 2013;17:1515–1531.
- [27] Osterhout L, Holcomb PJ. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*. 1992;31(6):785–806.
- [28] Coulson S, King JW, Kutas M. Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*. 1998;13(1):21–58.
- [29] Kutas M, McCarthy G, Donchin E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*. 1977;197(4305):792–795.
- [30] Frisch S, Kotz SA, Von Cramon DY, Friederici AD. Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology*. 2003;114(2):336–340.
- [31] Leckey M, Federmeier KD. The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*. 2020;57(7):e13351.

# AUTO-ADAPTATION OF ECOG-BASED MOTOR BCI USING NEURAL RESPONSE DECODER: A CROSS-PATIENT STUDY

H. Lafaye de Micheaux<sup>1</sup>, F. Martel<sup>1</sup>, F. Sauter-Starace<sup>1</sup>, G. Charvet<sup>1</sup>, H. Lorach<sup>2,3,4</sup>,  
T. Aksenova<sup>1</sup>

<sup>1</sup> Univ. Grenoble Alpes, CEA, LETI, Cinatec, Grenoble, France

<sup>2</sup> NeuroX Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL),  
Geneva, Switzerland

<sup>3</sup> Department of Clinical Neuroscience, Lausanne University Hospital (CHUV) and University of  
Lausanne (UNIL), Lausanne, Switzerland

<sup>4</sup> NeuroRestore, Defitech Center for Interventional Neurotherapies, EPFL/CHUV/UNIL, Lausanne,  
Switzerland

E-mail: [hugo.lafayedemicheaux@cea.fr](mailto:hugo.lafayedemicheaux@cea.fr) and [felix.martel@cea.fr](mailto:felix.martel@cea.fr) and [tetiana.aksenova@cea.fr](mailto:tetiana.aksenova@cea.fr)

**ABSTRACT:** Motor imagery brain-computer interfaces (BCIs) face challenges in practical application, notably in decoder training. Traditionally, decoders are trained in a supervised manner. This approach requires labeled data and restricts users to predefined actions during the training period. Moreover, regular decoder updates are needed. To address these issues, the auto-adaptive BCI (aBCI) infers training labels directly from brain signals using a neural response (NR) decoder, eliminating the need for supervised sessions. This study investigates the performance and replicability of the aBCI and explores labeling strategies using electrocorticography data from three spinal cord injured patients across diverse paradigms. Results demonstrate that aBCI can be used to significantly increase decoding performance above chance level in all three patients. Performance depended on patients and labeling strategy. The labeling strategy, focusing on correct neural responses (CNR), demonstrates significantly improved performance compared to correct/error neural responses (CENR) labeling strategy. Despite limitations of pseudo-online simulation, our findings underscore the aBCI's promise in advancing BCI technology.

## INTRODUCTION

Motor brain computer interfaces (BCIs) come with a number of recognized limitations that hinder their practical use in everyday situations. Many of these limitations relate to the need for training of BCI decoders. Traditionally, motor control (MC) decoders in BCIs are trained using supervised learning. In such a framework, access to the neural data along with the labels is required. Labels are derived from the user's intention. Consequently, during MC decoder training, BCI users are constrained to perform predefined actions under the supervision of researchers or the dedicated environment [1] [2]. In addition, the MC decoder must be regularly updated due to the degradation of performance over time.

Facing these limitations, the use of neural responses (NR) to BCI task performance for unsupervised updating of BCI decoders has been explored [3] [4] [5]. The majority of studies use event-related NR, namely event-locked error-related potentials, e.g. [6]. Most studies focus on NR in brain areas outside the sensorimotor cortex using electroencephalography (EEG)-based BCIs [7]. A limited number of studies investigates NR within the sensorimotor cortex. Invasive electrocorticography (ECoG)-based [8] [9] or microelectrodes array-based BCIs [10] reveal detectable NR following discrete erroneous events in a sensorimotor cortex.

Continuous in time NR (in contrast to event-locked NR) is explored by Rouanne et al. [11] [12], demonstrating detectability of such NR in the sensorimotor cortex using ECoG recording device. For complex BCI auto-adaptation, access to continuous in time NR is powerful as it would provide performance assessments at each time point, whereas event-locked NR would have to extrapolate performance around measured points. On the bases of such continuous in time NR, an auto-adaptive BCI (aBCI) framework with the objective of training the MC decoder during the free use of ECoG-based motor BCI is proposed [11] [12]. The core idea is to infer the training labels directly from the brain signals rather than from the environment, thus removing the necessity for training sessions. Within this aBCI framework, the user can update the MC decoder at will, enabling greater user autonomy in determining their actions. This first proof-of-concept study demonstrated in offline simulation that aBCI can be used to train in an unsupervised manner a MC decoder from scratch, eliminating the necessity for precise label assignment. However, to evaluate the aBCI framework, Rouanne et al. worked on data from a single patient. In order to build a robust and replicable aBCI framework, several questions are still to be addressed. In this paper, we explore the replicability of the aBCI framework [11] [12] with three patients. In addition, we

compared two aBCI labeling strategies to improve aBCI performance, and make the aBCI framework more versatile and compatible across different paradigms.

## MATERIALS AND METHODS

*Experimental recordings:* To investigate the replicability, we tested the aBCI on datasets from three patients implanted with two ECoG-recording WIMAGINE implants, one on each hemisphere, on the sensorimotor cortex [1] [13]. Two of these patients, referred as BCI001 and BCI002, are involved in the “BCI and Tetraplegia” clinical trial at CEA/Clinatoc (NCT02550522). Data for the third patient (BSI001) was collected in the STIMO-BSI clinical trial (NCT04632290). Both clinical trials focus on recording and decoding motor intentions with different effectors respectively. Consequently, experimental paradigms slightly differed between BCI and BSI patients.

For the BCI001 and BCI002 patients, we used the dataset collected during the Runner paradigm experiments (Fig. 1A). Runner represents a binary classification test where the BCI user controls a human avatar to either walk or stand still. BCI001 dataset spans a period of 5 months, from September 2019 to January 2020, comprising 13 half-day sessions for a total of 142 minutes of recordings. BCI002 dataset spans a period of 12 months, from November 2019 to October 2020, comprising 34 half-day sessions for a total of 653 minutes of recordings. For the BSI001 patient, we used a dataset collected during the Gait paradigm (Fig. 1B). In this paradigm, the patient used the BCI system to modulate electrical stimulation of the spinal cord enabling walking. A 3-class decoder (left/right hip flexion and rest) was used to decode the intention to perform each independent step and modulate the amplitude of stimulation according to the decoder prediction [13]. This dataset spans a period of 4 months, from September 2022 to January 2023, comprising 19 sessions for a total of 518 minutes of recordings.



Figure 1: Experimental paradigms of datasets included to the study. (A) Runner paradigm, binary classification of human avatar to either walk or stand still. (B) Gait paradigm, 3-class classification of left/right hip flexion and resting to control spinal cord stimulator.

During real time BCI experiments, time–frequency information was extracted for each of the 64 electrodes used [1] from each 1s-long epoch (spaced by 0.1s, 90% overlap), using continuous complex wavelet transform (Morlet) with 15 central frequencies 10 Hz apart from 10 to 150 Hz for patients BCI001 and BCI002. 0.2s-long

epoch (spaced by 0.1s, 50% overlap) with 24 central frequencies (2, 5:5:100, 125, 150, 200 Hz) were used in BSI001 patient sessions. Recursive Exponentially Weighted Markov-Switching multi-Linear Model (REW MSLM) was employed as MC decoder as in [2].

The aBCI framework have been evaluated across these three labeled datasets, shortly noted Runner BCI001, Runner BCI002 and Gait BSI001.

*aBCI framework description:* The overview of the aBCI framework [12] is given in Fig. 2. In the aBCI framework, the labels for the MC decoder training are not acquired through traditional training paradigm employed in supervised learning. Instead, they are estimated thanks to the auto-adaptive module. This module consists in a neural response (NR) decoder, also known as task performance decoder or satisfaction decoder. Its role is to interpret from the input features how well the effector’s actions match the user’s intentions. In other words, the NR decoder predicts from the brain signals whether the user is satisfied or dissatisfied with the action decoded by the MC decoder. The NR decoder is trained in a supervised manner. The MC decoder is then trained / updated in real time in an unsupervised manner, relying on the labels estimated by the NR decoder during the free use of the BCI.

In the current system, the same feature space described above is used by both decoders, which are trained using the REW MSLM algorithm [2].

*aBCI labeling strategy:* The process of automatic labeling of the training data for the MC decoder update is not a straightforward task. Indeed, the estimated labels are derived from the output of the NR decoder, noted  $\hat{y}_{NR}$ , which have not a perfect accuracy. Therefore, the derived labels cannot be expected to be perfect either. To limit this imperfection, the epochs with high level of uncertainty on the task performance estimation from the NR decoder are not labeled, and thus, discarded from the MC decoder update dataset. In this study, we compare two discarding strategies, resulted in two labeling strategies.

The first labeling strategy (Fig. 3A), proposed in [12], considers correct and error neural responses (CENR). It relies on the use of two thresholds,  $th_{corr}$  and  $th_{err}$ , for the classification of epochs respectively as correct and erroneous. Epochs are considered correct when  $\hat{y}_{NR} > th_{corr}$  and erroneous when  $\hat{y}_{NR} < th_{err}$ . Epochs for which  $th_{err} < \hat{y}_{NR} < th_{corr}$  are unlabeled and so not included in the MC training / update dataset. To evaluate the thresholds, the output of the NR decoder  $\hat{y}_{NR}$  is modeled as a mixture of two Gaussians,  $\mathcal{N}(\mu_{corr}, \sigma_{corr}^2)$  for the correct class and  $\mathcal{N}(\mu_{err}, \sigma_{err}^2)$  for the error class. The parameters of the two Gaussians are estimated on the training data for each class. Then, the thresholds are defined as  $th_{corr} = \mu_{corr} + a \sigma_{corr}$  and  $th_{err} = \mu_{err} - a \sigma_{err}$ , where  $a$  is a hyper-parameter to balance accuracy and data inclusion. Similarly to [12],

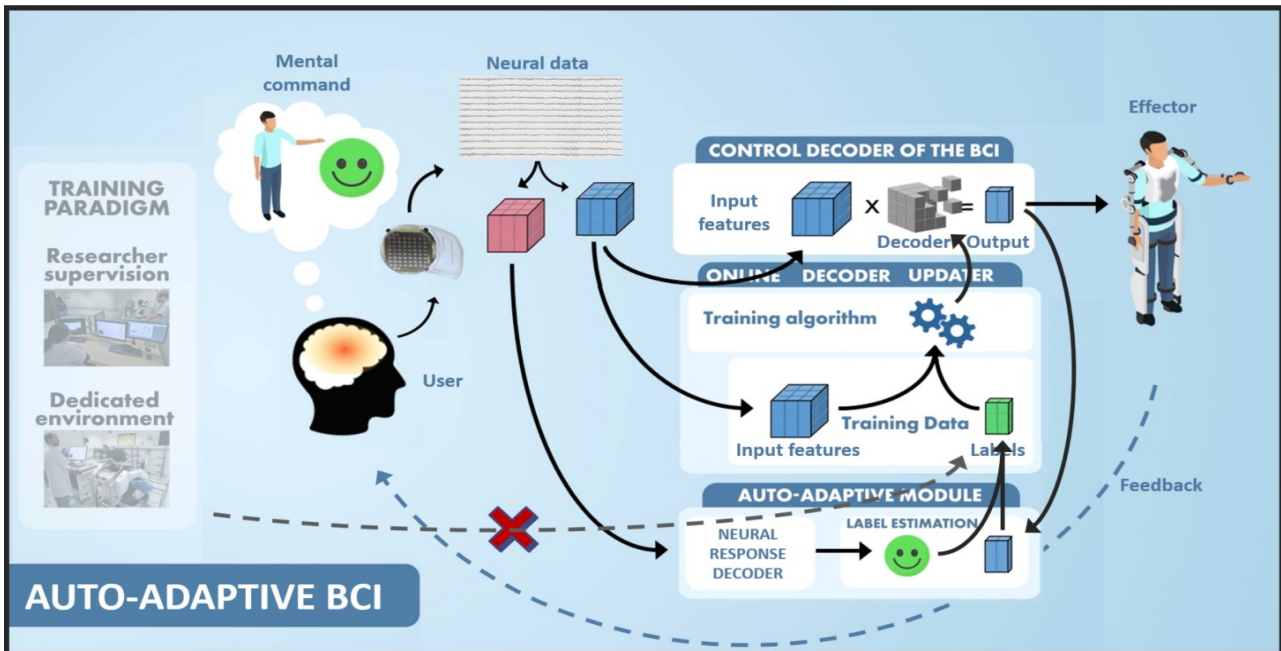


Figure 2: Diagram of the aBCI framework. An extra auto-adaptive module is added to the classic BCI framework, which is usually composed of a (motor) control (MC) decoder and an updater. The aBCI module includes a neural response (NR) decoder aiming at detecting continuous in time NRs to task performance and estimating the labels to update the MC decoder, instead of using the ones supplied by the researcher’s supervision or the dedicated environment as it is commonly done in a classic BCI framework.

we set  $\alpha = 1$  in this study. Finally, the fixed thresholds are used during the update of the MC decoder. When the epoch is estimated correct, the true MC label is the most probable MC output, while when it is estimated erroneous, the true MC label is the second most probable MC output.

The second labeling strategy (Fig. 3B), that we propose, is more restrictive and focuses exclusively on correct neural responses (CNR). It relies on the use of only one threshold for the classification of epochs as correct when  $\hat{y}_{NR} > th_{corr}$  and unlabeled when  $\hat{y}_{NR} < th_{corr}$ . The rest of the conditions and parameters employed with the CNR labeling strategy were consistent with those from the CENR one. We have headed for the CNR labeling strategy to make the aBCI framework more generic and adapted to multiclass classification, regression problems or combinations, where wrong decoded motor actions are very hard to relabel.

*Pseudo-online simulation:* We conducted a pseudo-online simulation to evaluate the performance of the aBCI framework, aiming to replicate conditions closely resembling online uses. To achieve this, we divided each dataset into three non-overlapped splits containing approximately the same number of recording sessions. The first split was dedicated to train the NR decoder. The second split was allocated to train the MC decoder from scratch within the aBCI framework, i.e. without knowledge of the real labels for the MC decoder. The third split was reserved for evaluating the performance of the newly trained MC decoder. The training data for the NR decoder were labeled according to the decoded MC outputs obtained during the online experiment: an epoch

with a decoded MC output being consistent with the desired MC output was labeled correct, while it was labeled error when inconsistent. We chose to train the MC decoders from scratch, meaning that no prior training was required, using solely the aBCI framework to highlight its capacity in training MC decoders.

In a typical online use, the neural data corresponding to the second split would be gathered during free use of the BCI. However, in our simulation study, we utilized pre-existing labeled datasets. The MC decoder training process was emulated in a pseudo-online fashion, where neural data was iteratively fed into the algorithm to mimic online acquisition. Labels are continuously estimated and training of the MC decoder were conducted using the aBCI framework every fifteen seconds, corresponding to the acquisition of labeled data. Notably, the newly updated MC decoder did not influence BCI actions, as the datasets were pre-recorded.

*Performance evaluation:* Cross-validation with the three splits by permuting their roles, which leads to six performance measures, was used to evaluate aBCI performances. For the Runner paradigm (binary classification), the performance was evaluated using the AUC of the ROC curve of the MC decoder. For the Gait paradigm (3-class classification), the performance was evaluated using a generalized version of the AUC of the ROC curve for multi-class classification [14]. The final performance of the aBCI for each paradigm was assessed with the mean AUC of the ROC curves over each test split. For a comparative evaluation, the MC decoder trained from scratch using the aBCI framework was compared to MC decoders trained in two other ways. The

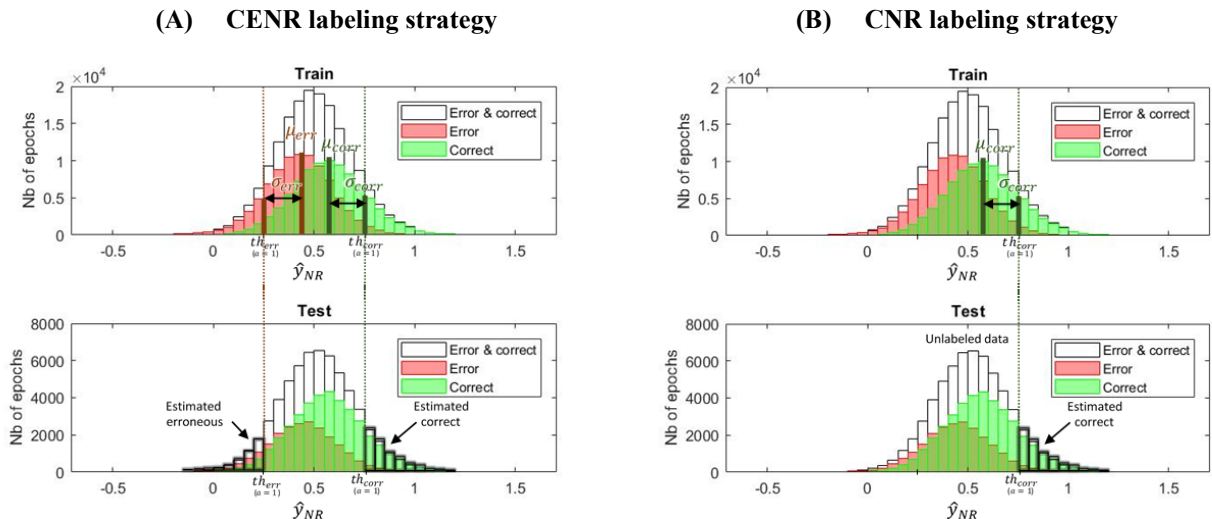


Figure 3: Two aBCI labeling strategies, one (A) focusing on correct/error neural responses (CENR) and the other (B) on correct neural responses (CNR). These histogram examples show the outputs of the neural response (NR) decoder,  $\hat{y}_{NR}$ , on one training fold (top) and its associated test set (bottom). The thresholds for the inclusion of epochs in the training set of the aBCI-based motor control (MC) decoder are based on a tradeoff parameter  $\alpha$  and the means and standard deviations of the Gaussians fitted to the correct and error class (for the CENR labeling strategy) or only the correct class (for the CNR labeling strategy) on the training set.

first one was a supervised training of the MC decoder using the true labels of each epoch from the recorded dataset. The second one was a MC decoder trained following the aBCI framework, but with random outputs of the NR decoder (chance level). For each dataset, the three training strategies were tested against each other through two-sided Wilcoxon Mann Whitney tests.

## RESULTS

Fig. 4 shows the mean AUC of the ROC curves for the MC decoding of the three compared training methods (random aBCI training / aBCI training / supervised training), across the three examined datasets (Runner BCI001 / Runner BCI002 / Gait BSI001) and using both labeling strategies (CENR / CNR).

*Replicability study across patients:* First, we could remark performance variations among patients, especially looking at the supervised trainings with mean AUC of the ROC going from 0.650 to 0.894. The aBCI decoding performances follow a similar trend. Second, we could also note that the aBCI decoding performances consistently exceed chance levels (50%, whatever the number of classes) and almost always in the significant manner (p-values < 0.05), except for the CENR labeling strategy on the Runner BCI002 (p-value > 0.05).

*Comparison of labeling strategies:* First, one should note that the aBCI decoding performances for the CENR labeling strategy always fall short of the gold standard performances achieved through supervised training, sometimes very significantly as for the Runner BSI001 and the Gait BSI001 (p-values < 0.01). Second, a direct comparison of aBCI decoding performances using both labeling strategies reveals a discernible improvement

when exclusively using correct neural responses. Notably, this enhancement is particularly pronounced in the case of the Runner BCI001 dataset, with the AUC of the ROC increasing from 0.637 to 0.819. Although not displayed on the figures, a p-value of 0.0022 for this dataset means significance when comparing the results of both labeling strategies through a two-sided Wilcoxon Mann Whitney test. No significant difference were observed when comparing CENR and CNR for the other datasets. Third, CENR presents high AUC variabilities in terms of standard deviation compared to low AUC variabilities for CNR.

## DISCUSSION

*Replicability across patients:* We showed over three patients that the aBCI control decoding performances are significantly higher than the random auto-adaptive trainings. This result demonstrates the potential of the aBCI framework for replication across different patients and its capacity to train / update MC decoders. aBCI performance is lower than supervised BCI, with essential cross-patient differences observed for the CENR labeling strategy (-29%, -12% and -15%, respectively). However, the cross-patient results are rather consistent for the CNR labeling strategy with smaller differences, down to -8%, -0.3% and -9% for the three patients respectively.

*Improvements using CNR labeling strategy:* The CNR labeling approach resulted in an improvement of 29%, 13%, and 8% in AUC compared to the CENR labeling approach, also reducing drastically AUC variabilities as indicated by lower standard deviations. These improvements brought aBCI decoding performance closer to supervised ones in terms of mean AUC and standard deviation of AUC. Several reasons may explain

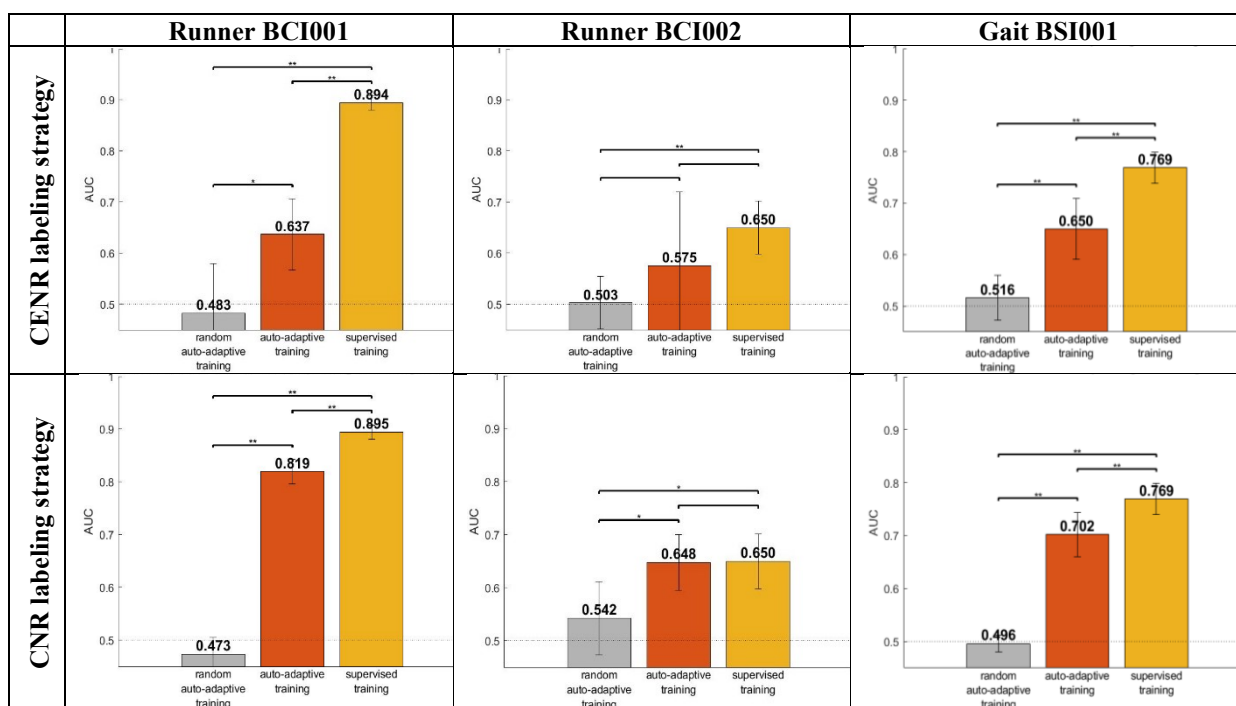


Figure 4: Motor control (MC) decoding performances for three MC training methods, on three datasets (as columns) and using two aBCI labeling strategies (as rows). Performances are given in terms of mean AUC of the ROC curves of the MC decoders trained using the aBCI (in red) compared to MC decoders trained using supervised learning (in yellow) or using random outputs of the neural response decoders (in gray). CENR stands for correct/error neural response and CNR stands for correct neural response. Stars denote significant differences between training methods (two-sided Wilcoxon Mann Whitney test, \* p-value < 0.05, \*\* p-value < 0.01).

this performance improvement. First, the distribution of class ‘correct’ is possibly better evaluated as it is better presented in the recordings. Therefore, the model of class ‘correct’ may have better generalization ability compared to the model of class ‘error’. Second, in case of detection of class ‘error’ by the NR decoder, supplementary relabeling is needed: the second most probable class is used as label in the CENR labeling strategy. In case of more than two classes in the MC decoder, it may increase the probability of erroneous labeling.

In addition, we suggest that using the CNR labeling strategy, the aBCI becomes more versatile and compatible across different paradigms, including classification, regression problems and combinations. On the other hand, the CNR labeling strategy is more selective and keeps less data for the model update compared to the CENR labeling strategy. It may result in a slower MC decoder adaptation.

*Limitations and perspectives:* By modifying the aBCI labeling strategy to the only use of correct neural responses, we get rid of the uncertainty on relabeling error ones but we also reduce the quantity of data used for updating the MC decoder. This reduction of data could be dramatic for cases with lots of error neural responses. The labeling strategies should be further explored.

According to the results of this study, on simple BCI paradigms with three patients, the aBCI framework

seems highly promising. However, it remains essential to validate this approach on more complex datasets, featuring additional degrees of freedom and a combination of discrete and continuous tasks, through classification and regression. Such paradigms will be tested with our aBCI framework in the near future.

A significant limitation of the study lays in the pseudo-online simulation rather than actual online use. While pseudo-online simulation studies allow for greater parameter exploration, they may not fully capture the variability of online experiments, even if it was designed to closely mimic the online use. In the near future, we will test the online version of the proposed aBCI framework.

On another hand, a more in-depth cross-paradigm and cross-patient study of features extraction should be conducted in terms of frequency and spatial characterization, for NR decoders. Indeed, the NR decoder is of critical importance in the aBCI framework and features extraction have not been optimized yet. For now, extracted features are the same as for the MC decoder. Therefore, studying other feature extraction methods would allow better interpretation of the aBCI performance results.

## CONCLUSION

The aBCI framework addresses critical limitations associated with traditional BCIs, especially the need for

supervised retraining sessions by allowing MC decoders to be updated during the free use of the BCI. This innovation not only offers greater user autonomy but also the potential for more natural and intuitive control.

In the continuation of the initial work of Rouanne et al. [12], the present paper provides valuable insights into the replicability and performance of the aBCI framework. Our investigation into using data from multiple patients and diverse paradigms with varying number of degrees of freedom, demonstrates the framework's adaptability. However, the variation of performance observed across patients and paradigms, highlights the need for further research to enhance the framework's robustness and generalizability.

Furthermore, we delved into refining the labeling strategy for training the MC decoder, emphasizing the use of correct neural responses exclusively. This approach yielded significant improvements in aBCI control decoding performance, showcasing the potential of this labeling strategy for future development.

In summary, the aBCI framework represents a promising avenue for advancing BCI technology, offering the potential for greater user autonomy and more natural control. Current and further exploration of the framework's capabilities and optimization strategies will undoubtedly contribute to its continued development for real-world applicability.

## ACKNOWLEDGEMENTS

This study was supported by Translational Medical Research Award 2021 from the Leenaards Foundation, the Swiss National Science Foundation, the Lead Agency Program with the French National Research Agency (Think2Move SNF-32003BE-205563, ANR-21-CE19-0038), the European Commission (ERC-2019-PoC Braingait 875660, EIC 2021-TransitionChallenges-01-01 ReverseParalysis 101057450, Horizon-EIC-2021-Pathfinderchallenges-01-02 NEMO-BMI 101070891), Fonds Clinattec (WIMAGINE implant development) and Institut Carnot Leti.

## REFERENCES

- [1] Benabid AL, Costecalde T, Eliseyev A, Charvet G, Verney A, Karakas S, et al. An exoskeleton controlled by an epidural wireless brain-machine interface in a tetraplegic patient: A proof-of-concept demonstration. *The Lancet Neurology*. 2019;18(12):1112–1122. DOI: [10.1016/S1474-4422\(19\)30321-7](https://doi.org/10.1016/S1474-4422(19)30321-7)
- [2] Moly A, Costecalde T, Martel F, Martin M, Larzabel C, Karakas S, et al. An adaptive closed-loop ECoG decoder for long-term and stable bimanual control of an exoskeleton by a tetraplegic. *Journal of Neural Engineering*. 2022;19(2):026021. DOI: [10.1088/1741-2552/ac59a0](https://doi.org/10.1088/1741-2552/ac59a0)
- [3] Buttfield A, Ferrez PW, Millan JR. Towards a robust BCI: error potentials and online learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2006;14(2):164–168. DOI: [10.1109/TNSRE.2006.875555](https://doi.org/10.1109/TNSRE.2006.875555)
- [4] Gürel T, Mehning C. Unsupervised adaptation of brain-machine interface decoders. *Frontiers in Neuroscience*. 2012;6:164. DOI: [10.3389/fnins.2012.00164](https://doi.org/10.3389/fnins.2012.00164)
- [5] Vortmann LM, Urban T, Putze F. Machine learning from mistakes: self-improving attention classifier using error-related potentials. In *IEEE International Conference on Systems, Man, and Cybernetics*. 2023:4770-4777. DOI: [10.1109/SMC53992.2023.10393901](https://doi.org/10.1109/SMC53992.2023.10393901)
- [6] Chavarriaga R, Sobolewski A, Millán JDR. Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Frontiers in Neuroscience*. 2014;8:208. DOI: [10.3389/fnins.2014.00208](https://doi.org/10.3389/fnins.2014.00208)
- [7] Omedes J, Iturrate I, Minguez J, Montesano L. Analysis and asynchronous detection of gradually unfolding errors during monitoring tasks. *Journal of Neural Engineering*. 2015 ;12(5):056001. DOI: [10.1088/1741-2560/12/5/056001](https://doi.org/10.1088/1741-2560/12/5/056001)
- [8] Milekovic T, Ball T, Schulze-Bonhage A, Aertsen A, Mehning C. Detection of error related neuronal responses recorded by electrocorticography in humans during continuous movements. *PLOS ONE*. 2013;8(2):e55235. DOI: [10.1371/journal.pone.0055235](https://doi.org/10.1371/journal.pone.0055235)
- [9] Wilson NR, Sarma D, Wander JD, Weaver KE, Ojemann JG, Rao RP. Cortical topography of error-related high-frequency potentials during erroneous control in a continuous control brain-computer interface. *Frontiers in Neuroscience*. 2019;13:502. DOI: [10.3389/fnins.2019.00502](https://doi.org/10.3389/fnins.2019.00502)
- [10] Even-Chen N, Stavisky SD, Kao JC, Ryu SI, Shenoy KV. Augmenting intracortical brain-machine interface with neurally driven error detectors. *Journal of Neural Engineering*. 2017;14(6):066007. DOI: [10.1088/1741-2552/aa8dc1](https://doi.org/10.1088/1741-2552/aa8dc1)
- [11] Rouanne V. Adaptation of discrete and continuous intracranial Brain-Computer Interfaces using neural correlates of task performance decoded continuously from the sensorimotor cortex of a tetraplegic. PhD thesis. Université Grenoble Alpes [2020–2022]. 2022. Available: [tel-03917556](https://tel-03917556)
- [12] Rouanne V, Costecalde T, Benabid AL, Aksenova T. Unsupervised adaptation of an ECoG based brain-computer interface using neural correlates of task performance. *Scientific Reports*. 2022;12:21316. DOI: [10.1038/s41598-022-25049-w](https://doi.org/10.1038/s41598-022-25049-w)
- [13] Lorach H, Galvez A, Spagnolo V, Martel F, Karakas S, Interling N, et al. Walking naturally after spinal cord injury using a brain-spine interface. *Nature*. 2023;618:126–133. DOI: [10.1038/s41586-023-06094-5](https://doi.org/10.1038/s41586-023-06094-5)
- [14] Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*. 2001;45:171–186. DOI: [10.1023/A:1010920819831](https://doi.org/10.1023/A:1010920819831)

# CORRECTING TRAJECTORY-DECODING ERRORS VIA CORTICAL SUBSTRATES OF CONTINUOUS ERRONEOUS FEEDBACK PROCESSING

H.S. Pulferer<sup>1</sup>, K. Kostoglou<sup>1</sup>, N. Srisrisawang<sup>1</sup>, G.R. Müller-Putz<sup>1,2</sup>

<sup>1</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>2</sup>BioTechMed-Graz, Graz, Austria

E-mail: gernot.mueller@tugraz.at

**ABSTRACT:** Decades of research thoroughly established various neural correlates of processing *discrete errors*, i.e., events that may be classified as either *correct* or *wrong*. However, despite many successful demonstrations of brain-computer interfaces (BCIs) utilizing these discrete correlates, a range of everyday tasks (e.g., car driving) requires fine-tuned feedback control that already transgresses such coarse distinction. Following up on recent research in the field of *continuous* erroneous feedback processing, we propose the regression of *continuous feedback-target deviations* from the electroencephalogram (EEG). Within thirty pre-recorded sessions of data in ten participants, employing a 2D target-tracking task that offered online feedback, we thus utilized a convolutional neural network to infer ongoing feedback-target deviations and correct the feedback's position accordingly in an offline evaluation. The presented correction approach significantly improved correlations between feedback and target kinematics - a first indication that *continuous* error-related cortical activity can be utilized in BCIs as well.

## INTRODUCTION

Over the last thirty years, a large body of research extensively documented the cortical response to discrete erroneous stimuli [1, 2]. In this context, the error-related negativity (ERN) and error positivity (Pe) – frontocentral and centroparietal deflections in the scalp potentials, respectively – quickly emerged as two key markers for error processing and error awareness [3, 4]. Arising approximately 100ms (ERN) and 300-500ms (Pe) after an erroneous stimulus, the sequence of these two potentials was termed the *error-related potential* (ErrP) and met increasing interest as a control signal within the field of brain-computer interfaces (BCIs) [5–7]. However, while various approaches emerged to utilize the ErrP as a control signal - seeking, e.g., to prevent the execution of erroneous commands altogether, or recalibrate the interface in response to errors [5] - a number of issues surrounding the discrete correlates to error processing persist.

For one, the neurophysiology of the ErrP proved notably sensitive to a range of factors. Advanced age [8], lowered levels of attention [9], or reduced attributed significance to an error [10] reportedly alter the measured po-

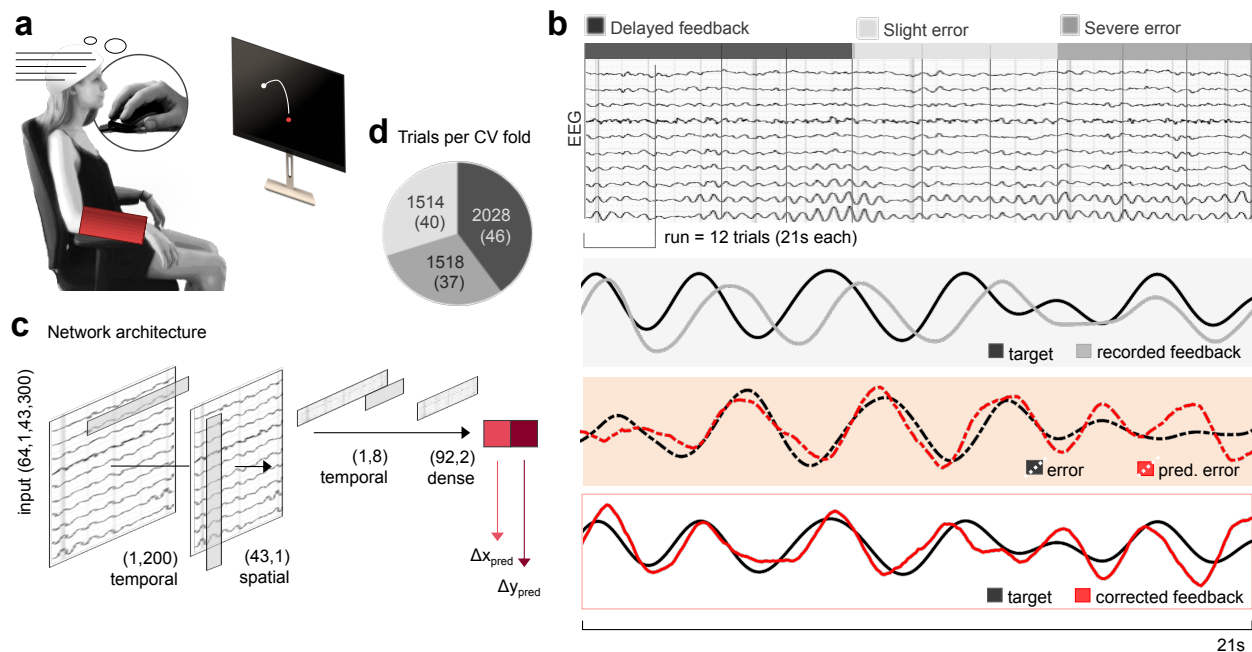
tentials, leading to corresponding difficulties in BCI operation [11]. Furthermore – and arguably the biggest limitation to utilizing ErrPs for BCI applications – the presence or absence of the ErrP intrinsically encodes binary information only. However, various tasks such as car driving or moving a cursor on-screen demand minute adjustments depending on the perceived discrepancy between intention and outcome rather than a coarse distinction into *error* or *no error*.

While some literature already attempted an expansion to continuous error processing within BCIs, this predominantly encompassed the presentation of discrete stimuli in a continuous paradigm [12–15]. Recent work by our group first reported the occurrence of cortical modulations with continuous feedback-target deviations [16], however, the usefulness of this neural substrate within BCIs remains to be established.

In this work, we thus aimed to close the gap in knowledge and answer two major questions. First, is it possible to regress target-feedback discrepancies from neural markers of continuous erroneous feedback processing within the electroencephalogram (EEG)? And second, can these inferences be used to subsequently correct the initial feedback and alleviate the feedback-target mismatch? Using thirty sessions of previously recorded data of an online target-tracking task with different feedback conditions, we trained a convolutional neural network (CNN) to infer the discrepancy (error signal) between feedback and target position in two spatial dimensions from the EEG. We then adjusted the recorded feedback trajectories as decoded and presented during the online measurements by the inferred error signal to obtain *corrected feedback trajectories*. Compared to the recorded feedback trajectories, correlations with the target trajectories significantly improved for the corrected feedback trajectories, indicating notable merit to using neural signatures of continuous erroneous feedback processing for automated correction of interface-related errors within a BCI.

## MATERIALS AND METHODS

**Dataset:** The prerecorded dataset [17] consisted of the 60-channel EEG (10-10 electrode system) and 4-channel electrooculogram (EOG) of ten able-bodied,



**Figure 1:** (a) Experimental setup. Participants attempted cursor-like movement of their strapped dominant arm to trace a moving object on-screen (snake, white) with real-time feedback (red dot). (b) Feedback conditions and approach. Target and feedback kinematic (grey box), as well as EEG, were previously recorded within the three different feedback conditions *Delayed feedback*, *Slight error*, and *Severe error*. Distances in  $x$  and  $y$  between recorded target and feedback trajectories are calculated as error signals (red box) and predicted from a convolutional neural network (c). Corrected feedback trajectories are obtained from initial feedback and decoded error signals (red framed box). (d) Average number and standard deviation of available trials per cross-validation fold for each participant and session (training: 4 folds, validation: 1 fold, testing: 1 fold). Exemplary trajectories for the correction approach in (b) are taken from session 1 of participant P2 ( $y$  coordinate, *Slight error* condition).

right-handed participants in total of 30 sessions, sampled at 200Hz. The dataset was chosen due to previously unveiled cortical modulations with the ongoing target-feedback deviations [16], indeed suggesting error-related brain activity elicited within the employed task.

**Paradigm:** During each session of this previously recorded online study, participants *attempted* cursor-like movement as if wielding a computer mouse to track a moving target on-screen (*snake*). An encasing around their dominant arm (see Fig.1a) limited overt movement. Utilizing a combination of partial least squares regression and an unscented Kalman filter (PLSUKF) [18], estimates for the snake's trajectory were decoded in real time from the EEG and delivered within different feedback conditions in the form of a feedback dot on-screen. Each participant underwent three separate sessions of measurements within the time span of a week. The multiple sessions (each employing the identical paradigm conditions) were initially designed to evaluate session-to-session differences in performance; as no significant changes were found in the initial work [17], we disregarded the session information and pooled all data to a total of 30 sessions for the current work.

**Feedback conditions:** Each session commenced with four calibration runs, followed by three 50% and three 100% EEG-decoded online feedback runs (see Fig.1b). One run comprised 12 trials of 23s length, respectively, during which participants tracked the moving snake on-screen. The first and last second in each run were omitted from further analysis to further minimize movement-

related artifacts as the participants became aware of the start and end of each run, leading to 21s of data per run. Within the calibration runs, EEG data to fit the PLSUKF online-decoder was recorded; as such, no EEG-decoded trajectory information was available yet to display as feedback. To accustom the participants to the additional visual input of the feedback dot from the beginning nonetheless, *fake feedback* in the form of a slightly delayed snake was presented during the calibration runs. As feedback dot and snake largely coincided throughout the calibration runs, leading to minimal discrepancy between target and feedback, calibration runs are henceforth termed *Delayed feedback*.

After fitting the PLSUKF decoder with the calibration data, the measurement proceeded with *online* EEG-decoded feedback. To transition smoothly between fake and online-decoded feedback, three intermediate 50% EEG-decoded feedback runs were introduced, wherein the arithmetic mean between actual (snake) and EEG-decoded target positions was displayed. Due to the EEG-decoded information, the discrepancy between target and feedback increased notably with respect to the calibration runs. We thus term the 50% EEG-decoded feedback runs *Slight error* in the following.

In the final three runs, 100% EEG-decoded feedback was displayed. In contrast to the mixed information shown during the 50% EEG-decoded feedback runs, participants were now faced with considerable discrepancy between target and feedback due to limitations in decoding. We thus term the 100% EEG-decoded feedback runs *Severe*

error in the following.

**Data processing:** Both EEG and EOG data were band-pass filtered between 0.2-10Hz (10th order Butterworth). Subsequently, noisy channels of the EEG as identified during the online experiment were spherically interpolated from neighboring channels. Eye artifacts were similarly attenuated utilizing the stored correction matrices of the SGEYESUB algorithm [19] from the online measurement, after which the EOG channels were removed. Persistent eye artifacts as well as muscle artifacts at peripheral EEG channels were identified via independent component analysis (ICA) and removed. To further eliminate any eye- or muscle-related influence, the outermost EEG channels (i.e., positions AF7-AF8, F7/8, FT7/8, T7/8, TP7/8, P7/8, PO7/8) were excluded from our analysis scheme, leading to a remaining number of 43 EEG channels.

The recorded x and y coordinates of the target (snake), as well as of the displayed feedback dot, were smoothed using a Savitzky-Golay filter (second order polynomials, 21 sample window  $\sim 100\text{ms}$ ). The distance vector  $\vec{\Delta}(t)$  between feedback (fb) and target (tg) position in each time point  $t$ , i.e., the *error signal* in two dimensions, was then derived via:

$$\vec{\Delta}(t) = \vec{X}(t)_{fb} - \vec{X}(t)_{tg} \in \mathbb{R}^2. \quad (1)$$

**Neural network architecture:** The aim of this work was to utilize the continuous error-related brain activity representing the ongoing target-feedback discrepancy to *correct* the initially recorded EEG-decoded predictions for the target position. To this end, we modified EEGNet [20] and changed the output layer to simultaneously regress the x and y coordinates of the error signal ( $\vec{\Delta}$ ) from 300-sample windows of EEG data (i.e., 1.5s). Importantly, several other architectures, such as Deep ConvNet [21] or EEG-TCNet [22], could have been employed as well; as the current study however mainly corresponds to a proof of concept, benchmarking has not been undertaken in the scope of this work.

In detail, the utilized network consisted of three layers. Within the first layer, the temporal dimension of the input frames was zero-padded (75 samples at both edges) and temporally convolved in 16 filters to extract temporal features from the EEG (kernel size (1,200)). The large kernel size corresponding to a 1s window was chosen to enable the model to learn from frequency information down to 1Hz, as previous findings revealed contributions of predominantly the delta band for the used dataset [16]. In the second layer, a consecutive depthwise convolution (kernel size (43,1)) extracted spatial features in 8 filters by condensing the information of all considered channels to one single value, followed by subsequent temporal average pooling (kernel size (1,4)). In the final third layer, a second temporal convolution in 4 filters was employed (kernel size (1,16)), followed by average pooling (kernel size (1,2)). Each mentioned convolutional layer was followed by batch normalization to accelerate convergence [23], as well as dropout to impede an overfit on the train-

ing data (rates of 0.25, 0.35, and 0.45 for 1st, 2nd and 3rd layer). For activation functions in each neuron, exponential linear units were used [24]. Finally, the outputs of the third layer were flattened and passed through a fully connected layer (92 input features, 2 output features), returning the 2D prediction  $\vec{\Delta}_{pred}$  of the error signal. For each batch, the predictions were smoothed with a Savitzky-Golay filter to alleviate noise (2nd order polynomials, 21 samples). All network models were implemented, trained and evaluated using PyTorch.

**Network training procedure and regression performance:** For each participant, session, and condition, we sliced the EEG data into windows containing 300 samples with a stride of 15 samples (i.e., 1.5s windows, sampled every 75ms). We retained the last sixth of the resulting windows for testing in a causal fashion, while the optimum model - i.e., the model maximizing the validation correlation (mean of both x and y coordinate) between actual ( $\vec{\Delta}$ ) and predicted ( $\vec{\Delta}_{pred}$ ) error signal - was found via 5-fold cross-validation on the remaining data. For each fold, we trained a model for 30 epochs using a batch size of 64 and a learning rate of  $1e-4$ . Notably, predicting the error signal for 64 consecutive 1.5s-windows at a stride of 75ms per batch lead to  $64 \cdot 0.075\text{s} = 4.8\text{s}$ -long trajectories of predictions, for which both the correlation and the RMSE values with the ground truth were evaluated. To optimize the model's parameters with respect to the mean squared error loss, we used the Adam optimization algorithm [25]. The overall performance of each regression model was then assessed via Pearson's correlation coefficient, as well as the root mean square error (RMSE) between actual ( $\vec{\Delta}$ ) and predicted ( $\vec{\Delta}_{pred}$ ) error signal within each batch. The overall testing performance then corresponded to the average across all batches within the testing set.

To additionally analyze each EEG channel's contribution to the regression performance, we iteratively set one of the 43 channels within the testing data to zero (i.e., we simulated one dead channel at a time) and reevaluated the regression performance. The absolute difference between the optimum model's testing performance and the dead-channel performance for each removed channel then provided an estimate for the specific channel's importance.

**Trajectory correction and correction performance:** Utilizing the prediction  $\vec{\Delta}_{pred}$  of the error signal acquired via the neural network (Fig.1c), we obtained *corrected feedback trajectories* (i.e., new target predictions<sup>1</sup>) via Equ.(1) as:

$$\vec{X}_{fb,corr} = \vec{X}_{fb} - \vec{\Delta}_{pred} = \vec{X}_{tg,pred} \in \mathbb{R}^2. \quad (2)$$

The overall merit of our correction approach was judged by comparing Pearson's correlation coefficients and the RMSE values between a) target and recorded feedback

<sup>1</sup> Note that from the relation in Equ.(2), we indeed gain *target predictions*. However, to keep consistent terminology (the recorded feedback  $\vec{X}_{fb}$  depicted during the measurement corresponded to target predictions as well, if from the PLSUKF regressor), we term these new target predictions *corrected feedback*.

(i.e.,  $\vec{X}_{tg}$  and  $\vec{X}_{fb}$ ) and b) target and corrected feedback (i.e.,  $\vec{X}_{tg}$  and  $\vec{X}_{fb,corr}$ ).

**Chance level estimation:** To establish whether the network models truly discern cortical activity rather than noise, we estimated individual chance levels for the regression performance in each participant, session, and condition. To this end, we randomly shuffled the target information (i.e., the error signal  $\vec{\Delta}$ ) across batches of testing data, effectively breaking any causal relation between brain activity and corresponding target-feedback discrepancy without changing the frequency content of  $\vec{\Delta}$  (equivalently, the input signals could have been shuffled as well, leading however to the same results at a higher computational cost). The previously trained optimum model for the corresponding participant, session, and condition was then evaluated on the shuffled data. This approach was repeated for 100 times, yielding a chance distribution of correlations and RMSE values between actual ( $\vec{\Delta}$ ) and predicted ( $\vec{\Delta}_{pred,shuff}$ ) error signals. The chance levels for the correlations and RMSE values within each regression model were then identified as the 95th and 5th percentiles of the corresponding chance distributions, respectively.

We further investigated the nature of possible improvements in correlations and RMSE values due to our approach in Equ.(2). Specifically, improvements in correlation or RMSE values between target and corrected feedback (i.e.,  $\vec{X}_{tg}$  and  $\vec{X}_{fb,corr}$ ) compared to those between target and recorded feedback (i.e.,  $\vec{X}_{tg}$  and  $\vec{X}_{fb}$ ) may merely correspond to spurious fluctuations in performance due to the addition of the smoothed (low-frequency) predictions to the recorded feedback. To eradicate this concern, we once more exploited the previously outlined shuffling approach. We randomly shuffled the pairs of recorded target and feedback information across batches, breaking all causal relations with the EEG input while retaining the temporal structure within each batch. The optimum model then predicted the error signals  $\vec{\Delta}_{pred}$ , which in turn served to obtain randomly corrected feedback trajectories according to Equ.(2). The shuffling was repeated for 100 times, yielding chance distributions for the correlation and RMSE values between target ( $\vec{X}_{tg}$ ) and randomly corrected feedback ( $\vec{X}_{fb,corr,shuff}$ ) trajectories. The chance levels for the correlations and RMSE values within the correction approach for each participant, session, and condition were then found as the 95th and 5th percentiles of the chance distributions, respectively.

Table (1) summarizes all utilized evaluation approaches.

## RESULTS

**Regression performance:** Individual CNN models were trained for each participant, session and condition; the overall regression performance for each feedback condition is outlined in Fig.2(a-b). The single dots correspond to the 30 sessions (pooling all three sessions per participant in the 10 participants); mean and median

Table 1: Summary of evaluation approaches

Assessment	Correlation/RMSE between
Regression performance	$\vec{\Delta}, \vec{\Delta}_{pred}$
Regression chance level	$\vec{\Delta}, \vec{\Delta}_{pred,shuff}$ (p95 for correlation, p5 for RMSE)
Correction performance	$\vec{X}_{tg}, \vec{X}_{fb,corr}$ compared to $\vec{X}_{tg}, \vec{X}_{fb}$
Correction chance level	$\vec{X}_{tg}, \vec{X}_{fb,corr,shuff}$ (p95 for correlation, p5 for RMSE)

are displayed as dashed and solid lines within the boxes, respectively, solid black lines denote the corresponding chance levels.

On average, we obtained mean correlations across both coordinates of 0.36, 0.32 and 0.23 between actual ( $\vec{\Delta}$ ) and predicted ( $\vec{\Delta}_{pred}$ ) error signals for Delayed feedback, Slight error and Severe error conditions, each of which ranges above the corresponding chance level (approximately 0.15, 0.17 and 0.17). Similarly, the observed mean RMSE values across both coordinates of 56px, 106px and 187px for Delayed feedback, Slight error and Severe error condition cut beneath their corresponding chance levels (64px, 115px and 193px; see Fig.2b).

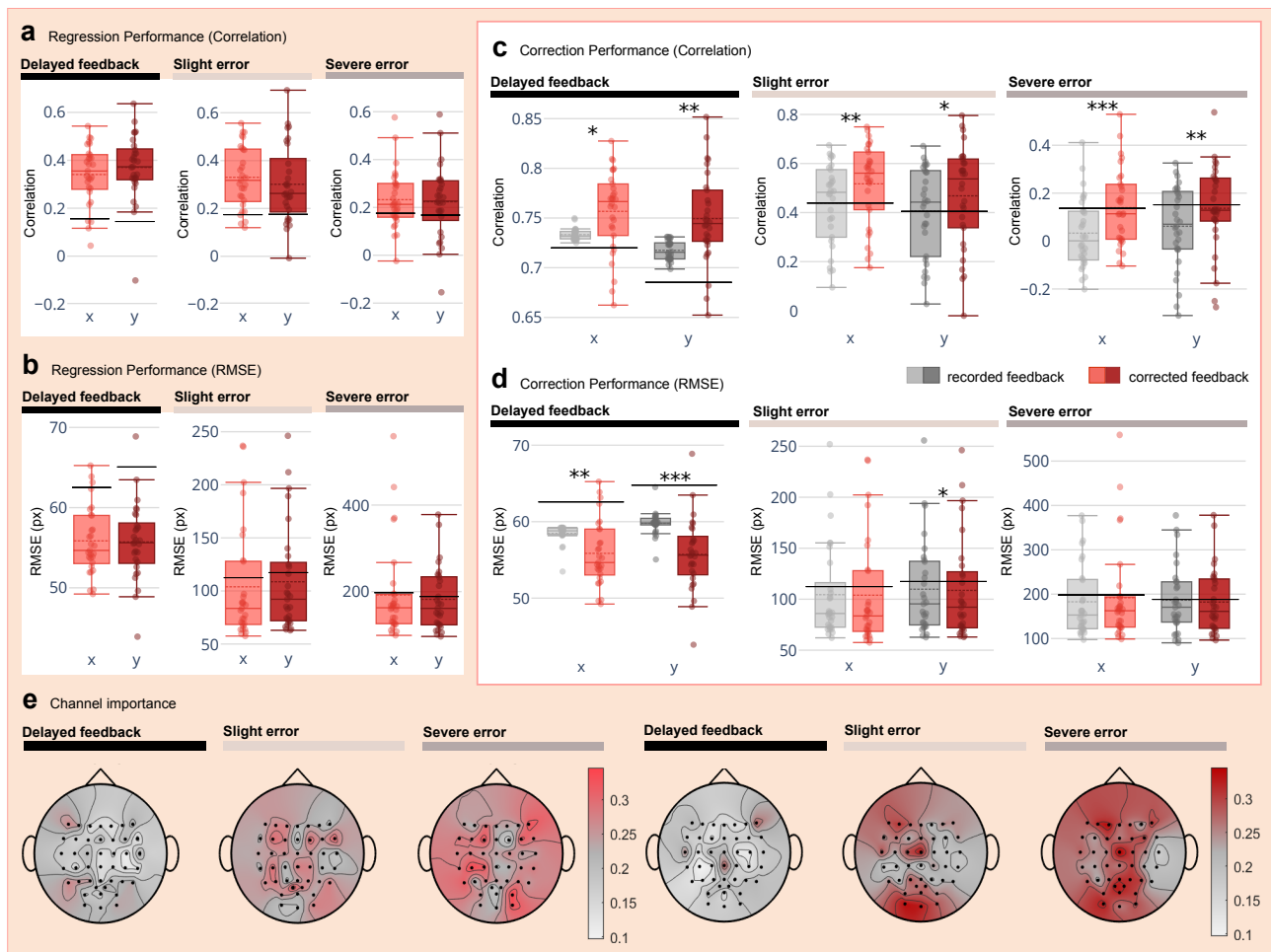
Notably, the regression performance worsened for both metrics with increasing discrepancy between target and feedback (note the different scales). In this context, our feature analysis unveiled growing scalp regions of importance across feedback conditions, displaying increasing central and parietooccipital engagement as the feedback deviates from the target (see Fig.2(e)). Left (bright) and right (dark) topographical maps correspond to the most relevant EEG channels for predicting x and y coordinate of the error signal  $\vec{\Delta}$ , respectively.

**Correction performance:** The mean correction performance due to our approach is displayed in the white panel of Fig.2. Light and dark gray results correspond to the x and y coordinate results pertaining to the recorded feedback ( $\vec{X}_{fb}$ ); light and dark red display the x and y coordinate results for the corrected feedback ( $\vec{X}_{fb,corr}$ ).

A right-tailed paired Wilcoxon signed rank test, Bonferroni-corrected for six tests (three conditions in two coordinates), revealed significantly higher correlations with the target trajectories ( $\vec{X}_{tg}$ ) for our corrected feedback trajectories ( $\vec{X}_{fb,corr}$ ) compared to the recorded ones ( $\vec{X}_{fb}$ ) (see Fig.2(c); significance levels of 0.05, 0.01, and 0.001 are marked as \*, \*\*, and \*\*\*).

Overall, we observed an average improvement in correlation across both coordinates of approximately 0.03, 0.07, and 0.08 for Delayed feedback, Slight error, and Severe error conditions due to the feedback correction; however, only the corrections for the first two conditions range on average above chance. In terms of RMSE values, significant differences arose only for the Delayed feedback condition and one coordinate within the Slight error condition, even though all mean values fell beneath the chance levels (see Fig.2(d)).

## DISCUSSION AND CONCLUSION



**Figure 2:** (a-b) Correlations and RMSE values between actual and decoded error signal as obtained by the CNN-regression. Bright and dark red color indicate results in the x and y coordinate, respectively. (c-d) Correlations and RMSE values for the correction approach. Measures between target and recorded feedback trajectories are shown in gray, measures between target and corrected feedback trajectories in red. Bright and dark colors indicate results in the x and y coordinate. For all box plots, single session means are depicted as dots, dashed and solid lines denote mean and median of the distributions, respectively. Black horizontal lines indicate the corresponding chance levels. (e) Averaged normalized channel importance maps for the error signal regression as obtained by setting one channel at a time to zero and evaluating the resulting drop in performance. Left and right topographical maps correspond to the importance in predicting the x and y coordinate of  $\Delta$ , respectively.

Within the offline analysis of 30 sessions of EEG recordings, we present a first attempt at correcting previously inferred feedback trajectories during a target tracking task by utilizing markers for continuous erroneous feedback processing within the brain.

Using an adaption of the well-known neural network architecture of EEGNet, we presented evidence for the successful inference of continuous feedback-target deviations (*error signals*) from the EEG for the first time. Achieving mean correlations with the actual error signals of between 0.23 and 0.36 for the investigated conditions (Fig2(a)), the obtained predictions proved to range above chance level, indicating that the regression of this type of error-related information from the EEG is indeed feasible.

Interestingly, the regression performance peaked within the only minor deviations during the Delayed feedback condition and dropped steadily across conditions with increasing feedback-target deviation. However, previous neurophysiological findings for the used dataset indicated an opposing effect, e.g., increasingly prominent cortical

modulations with increasing absolute distance from the target [16]. While the obtained channel importance maps affirm the previous findings with the emergence of increasing central and centrooccipital relevance with increasing error severity and indeed indicate that the models learned from error-related features, our overall results might suggest that the error signal's x and y components are not sufficiently encoded within the brain. Future approaches will have to clarify whether better performance could be achieved by taking the modulus, i.e., the Euclidean distance between feedback and target, into account; possible implementations could for example switch to radial coordinates or add the error signal's modulus to the network's training procedure. Despite the moderate correlations for the regression itself, the correction approach nonetheless proved to be of merit. We observed significant improvements in correlations with the target trajectories for our corrected feedback compared to the initially recorded feedback trajectories across all feedback conditions. A chance level evaluation certified these improvements as better

than random for all but the Severe error condition. However, for this condition, initial correlations between recorded target and feedback trajectories already failed to meet the chance level, which could be improved considerably due to our approach nonetheless.

In summary, we conclude that the use of continuous error-related brain activity can significantly improve the performance of a BCI and that further work in this field will be of great value for future implementations.

## ACKNOWLEDGMENTS

The authors thank the extended Graz BCI team, especially Adyasha Dash, for valuable comments. Research was supported by the NTU-TUG Joint Ph.D. Program.

## REFERENCES

- [1] Falkenstein, Hohnsbein, Hoormann, Blanke. Error processing in choice reaction tasks with focused and crossmodal divided attention. an ERP study. I Conference on EventRelated Potentials of the Brain. May 23-June 3, Noordwijk, the Netherlands. 1989.
- [2] Gehring WJ. The error-related negativity: Evidence for a neural mechanism for error-related processing. University of Illinois at Urbana-Champaign (1992).
- [3] Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GP, Kok A. Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*. 2001;38(5):752–760.
- [4] Holroyd CB, Nieuwenhuis S, Mars RB, Coles MGH. Anterior cingulate cortex, selection for action, and error processing. In: *Cognitive neuroscience of attention* (pp. The Guilford Press, xiii: New York, NY, US, 2004, 219–231.
- [5] Chavarriaga R, Sobolewski A, Millán JDR. Errare machinale est: The use of error-related potentials in brain-machine interfaces. *Front. Neurosci.*. 2014;8:208.
- [6] Müller-Putz GR *et al.* Feel your reach: An EEG-based framework to continuously detect goal-directed movements and error processing to gate kinesthetic feedback informed artificial arm control. *Frontiers in Human Neuroscience*. 2022;110.
- [7] Wimmer M, Weidinger N, Veas E, Müller-Putz GR. Multimodal decoding of error processing in a virtual reality flight simulation. *Sci. Rep.*. 2024;14(1):1–14.
- [8] Falkenstein M, Hoormann J, Hohnsbein J. Changes of error-related ERPs with age. *Exp. Brain Res.*. 2001;138(2):258–262.
- [9] Yeung N, Botvinick MM, Cohen JD. The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychol. Rev.*. 2004;111(4):931–959.
- [10] Hajcak G, Moser JS, Yeung N, Simons RF. On the ERN and the significance of errors. *Psychophysiology*. 2005;42(2):151–160.
- [11] Chavarriaga R, Millán JDR. Learning from EEG error-related potentials in noninvasive brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.*. 2010;18(4):381–388.
- [12] Zhang H, Chavarriaga R, Khaliliardali Z, Gheorghe L, Iturrate I, Millán JdR. EEG-based decoding of error-related brain activity in a real-world driving task. *J. Neural Eng.*. 2015;12(6):066028.
- [13] Lopes-Dias C, Sburlea AI, Müller-Putz GR. On-line asynchronous decoding of error-related potentials during the continuous control of a robot. *Sci. Rep.*. 2019;9(1):17596.
- [14] Wimmer, Weidinger, ElSayed, Müller-Putz, Veas. EEG-Based error detection can challenge human reaction time in a VR navigation task. In: 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Oct. 2023, 970–979.
- [15] Mondini V, Sburlea AI, Müller-Putz GR. Towards unlocking motor control in spinal cord injured by applying an online EEG-based framework to decode motor intention, trajectory and error processing. *Sci. Rep.*. 2024;14(1):4714.
- [16] Pulferer HS, Kostoglou K, Müller-Putz GR. Getting off track: Cortical feedback processing network modulated by continuous error signal during target-feedback mismatch. *Neuroimage*. 2023;274:120144.
- [17] Pulferer HS, Ásgeirsdóttir B, Mondini V, Sburlea AI, Müller-Putz GR. Continuous 2D trajectory decoding from attempted movement: Across-session performance in able-bodied and feasibility in a spinal cord injured participant. *J. Neural Eng.*. 2022.
- [18] Mondini V, Kobler RJ, Sburlea AI, Müller-Putz GR. Continuous low-frequency EEG decoding of arm movement for closed-loop, natural control of a robotic arm. *J. Neural Eng.*. 2020;17(4):046031.
- [19] Kobler RJ, Sburlea AI, Lopes-Dias C, Schwarz A, Hirata M, Müller-Putz GR. Corneo-retinal-dipole and eyelid-related eye artifacts can be corrected offline and online in electroencephalographic and magnetoencephalographic signals. *Neuroimage*. 2020;218:117000.
- [20] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.*. 2018;15(5):056013.
- [21] Schirrmeister RT *et al.* Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.*. 2017;38(11):5391–5420.
- [22] Ingolfsson TM, Hersche M, Wang X, Kobayashi N, Cavigelli L, Benini L. EEG-TCNet: An accurate temporal convolutional network for embedded Motor-Imagery Brain-Machine interfaces. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, Oct. 2020, 2958–2965.
- [23] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 2015.
- [24] Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). 2015.
- [25] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014.

## PREDICTORS OF ECOG-BCI PERFORMANCES ACROSS SUBJECTS AND SESSIONS DERIVED FROM IDLE STATE CHARACTERISTICS

L. Struber<sup>1</sup>, F. Martel<sup>1</sup>, S. Karakas<sup>1</sup>, V. Juillard<sup>1</sup>, A. Bellicha<sup>1</sup>, F. Sauter<sup>1</sup>, S. Chabardès<sup>1,2</sup>, H. Lorach<sup>3,4</sup>, G. Charvet<sup>1</sup>, T. Aksenova<sup>1</sup>

<sup>1</sup> Univ. Grenoble Alpes, CEA, LETI, Clnatec, Grenoble, France

<sup>2</sup> Univ. Grenoble Alpes, Grenoble University Hospital, Grenoble, France

<sup>3</sup> NeuroRestore, Defitech Center for Interventional Neurotherapies, EPFL/CHUV/UNIL, Lausanne, Switzerland

<sup>4</sup> NeuroX Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Geneva, Switzerland

E-mail: [lucas.struber@cea.fr](mailto:lucas.struber@cea.fr)

**ABSTRACT:** Comprehension of performance variabilities across subjects and sessions is crucial for real life brain-computer-interfaces (BCI) applications. This study compared three subjects that underwent implantation of minimally invasive WIMAGINE ECoG recording implants. Three training strategies to discern best achievable performance, session drift, and variability were evaluated offline using datasets recorded during real-time closed-loop BCI experiments. Results revealed distinct BCI profiles across patients, consistent with qualitative observations made during online training. These performances were correlated with two indicators computed in feature space during idle periods of BCI sessions: Euclidean distance between the current session and the session of model creation in a low-dimensional UMAP embedding, and intrinsic dimension. Between sessions distances demonstrated statistically significant correlation with models' performances, then recalibration need may be potentially anticipated from the characteristics of idle state periods. Additionally, the intrinsic dimension was significantly correlated to subjects' overall BCI capabilities. The results are consistent with pre-implantation MEG-BCI experiments, which could make it useful for patient selection.

### INTRODUCTION

Brain-computer-interface (BCI) technology has shown promising advances in the past years, in terms of rehabilitative potential, performances and usability [1], [2], [3]. Despite the progress, there are still challenges to overcome before BCI use in day-to-day scenarios. In particular, the need to regularly train / update decoders poses a significant obstacle to translate BCI into real-life applications [4]. Minimally invasive electrocorticography (ECoG) based BCI, provides a much higher signal stability than electroencephalography (EEG) based BCIs, or than highly invasive

Microelectrode Array (MEA) based BCIs [5]. ECoG-BCIs showed their ability to properly decode brain activity without recalibration for several months [6], [7]. However, these studies included only one subject who was intensively trained to control the BCI. It is well known, however, that BCI control performances and motor imagery capabilities can vary significantly across subjects, 15-30% of patients even being described as BCI-illiterate or inefficient [8]. Furthermore, although studies [6], [7] showed globally stable performances over time, inter-sessions variability remained significant.

While the community widely acknowledges issues of inter- and intra-subject variability, characterization of good or poor BCI performance is still not well established. Several studies investigated potential neurophysiological EEG-based predictors of inter-subjects BCI performances variability, associating the frontal theta rhythm (4-8Hz) [9] and the amplitude of the motor cortex mu band peak in the power spectrum [10], [11] during a relax condition to the ability of the subject to control a BCI. At the subject level, it has been shown that quality of motor imagery within a session (assessed by classification of left and right motor imagery) was correlated with gamma power during the task [12]. While these studies presented promising results for patient selection, it also shows that a multitude of currently unknown brain processes most likely affects BCI performance, and may vary across experimental paradigms. Furthermore, none of these studies investigated the neurophysiological markers of session-to-session variability of subjects' performances. Recently, some studies explored transfer learning approaches to compensate for this drift over sessions in EEG [13] and MEA [14]. However, these methods perform systematic domain adaptation and model retraining which requires labelled data and computation time. In an online perspective with patients chronically implanted with ECoG recording implants, in which it is possible to keep the same decoder functioning for several

sessions, establishing predictors that can be rapidly estimated on an idle state period could help determine if the decoder needs to be recalibrated and predict how well a previous decoder would fit the incoming data. Implementing predictors of day's performance of a subject using a former model is also crucial to develop better session-to-session variability compensation techniques, and to elaborate more effective and personalized training procedures.

In this study, we propose to compare BCI performances across sessions of three patients implanted with chronic ECoG implants, and relate them to data-driven characteristics extracted from idle state. We hypothesize that idle state signals recorded in motor and sensorimotor cortices are informative both on inter- and intra-subjects performances variability, and in particular that idle state characteristics can explain this variability. Interestingly, we relate these long-term ECoG-BCI results to MEG-BCI sessions that subjects performed before implantation, speculating that individual long-term performances was somehow predictable.

## MATERIALS AND METHODS

*Subjects:* Three subjects who underwent bilateral implantation of chronic wireless WIMAGINE implants on the motor and sensorimotor cortices were included in this study. Subjects 1 and 2 (S1 and S2) were respectively 28 and 29 years-old males (at the time of surgery), with traumatic sensorimotor tetraplegia which were included within the 'BCI and Tetraplegia' clinical trial (clinicaltrials.gov, [NCT02550522](https://clinicaltrials.gov/ct2/show/study?term=NCT02550522)) and implanted over the upper limb region of the cortex [15]. Subject 3 (S3) was a 38-year-old male who had sustained an incomplete cervical (C5/C6) spinal cord injury and was included within the 'STIMO-BSI' clinical trial (clinicaltrials.gov, [NCT04632290](https://clinicaltrials.gov/ct2/show/study?term=NCT04632290)) [1]. He was implanted more centrally to approach the legs motor regions.

*Online experiments:* During online BCI-sessions, the three subjects were trained to control different effectors. In the data considered in this study, S1 was controlling an avatar in a 3D virtual environment over eight continuous degrees of freedom (right and left hand 3D translations, and right and left wrist rotations) using motor imagery of both hands fingers. S2 controlled a virtual keypad in four directions, each of them being associated to a discrete state of the controller (up, down, left, right) using motor imagery of shoulders, legs and both hands. As for S3, he controlled directly his own legs independently through an epidural stimulator of the spinal cord allowing two discrete stimulation patterns (left leg and right leg) using direct motor imagery. For the three subjects, in addition to the controlled degrees of freedom, the decoders were trained to discern an idle class, corresponding to the periods of recordings in which the patient was relaxing. These periods were used to implement idle state indicators that are described below. ECoG was sampled at 585Hz. For each subject each BCI-session lasted approximately 2h, but only parts of data were labelled (rest of data was online testing). Prior to

implantation, the three subjects performed a ~1h single magneto-encephalography (MEG) BCI session, sampled at 1kHz, in which they controlled a runner avatar through motor imagery of walking (2-states brainswitch control).

*Offline dataset:* In order to obtain comparable results, the online ECoG dataset of each subject was narrowed to three discrete states: idle for every subject and motor imagery of right and left hand for subjects 1 and 2 and right and left hip for subject 3. Since one subject had only one functioning implant due to an electronic dysfunction, only data from left implant were kept for all subjects, leading to 32 electrodes per subject. For S1 and S2 recorded electrodes were distributed following a checkerboard-pattern over the implant and for S3 more central electrodes were favored to get a better coverage of leg motor area [1]. Datasets were also balanced in terms of number of sessions and number of samples per states in each session. Finally, this led to a dataset comprising 15 sessions of 1800 labelled motor imagery samples per subject (one each 0.1s; 600 per state), acquired over 6, 10 and 5 months for subjects 1,2 and 3 respectively.

*Feature extraction:* Feature extraction procedure is described in details in [6]. After interpolation of missing points in the raw ECoG data, 1s-long epochs of neural signals with a 100ms sliding step, were mapped to the temporal frequency space using a complex continuous wavelet transform (CCWT) (Morlet) with a frequency range from 10 to 150 Hz (10 Hz step). The absolute value of the CCWT coefficients was then decimated along the temporal modality to obtain a 10-timepoints description of the epoch for each frequency band and each channel, resulting in the temporal-frequency-spatial neural feature tensor  $X_t \in \mathbb{R}^{10 \times 15 \times 32}$ . Same features were used during online experiments and offline analyses, except for subject 3 for which 0.2s-long epoch and 24 frequency bins were used in online experiments leading to  $X_t \in \mathbb{R}^{2 \times 24 \times 64}$  (offline features were recomputed to match the other two subjects). Similar features were extracted from MEG experiments, with feature tensor  $X_t \in \mathbb{R}^{10 \times 9 \times 105}$  (10 temporal steps, 9 frequency steps distributed between 1 and 40Hz, and 105 MEG channels).

*BCI performances evaluation strategies and criteria:* To assess subjects' performances across sessions, 3 classes classification models were trained offline for each subject. As previously explained, BCI performances were assessed offline on equivalent datasets to have a fair comparison between subjects that performed different online experiments. Similar to online experiments, Hidden Markov Models (HMM) combining emission and transition probabilities were trained and used for classification in a pseudo-online manner [6]. Emission probability was computed using REW-NPLS with one-hot encoded class labels, post-processed by softmax function [16]. Transition probability matrix was estimated by counting the number of transition in the training set. The class prior was established to ensure equal probability distribution among classes. In order to evaluate general performances but also their session-to-session variability, three training strategies were carried

out:

- Within-session training: models were trained and tested on the same session, with a 5-folds cross-validation scheme.
- Session-1 training: models were trained only with data of session 1 and tested on every next sessions.
- Session-to-session training: models were trained each session, and tested on the following session.

Subjects' performances were assessed with the accuracy of classification (total number of correct predictions divided by the total number of samples). They were also evaluated in prior MEG experiments with two states classification models using within-session training strategy. As the different states were balanced, the chance level was of 1/3 for all classifications.

**Idle state variability evaluation:** All idle state samples of the three patients were projected together in an unsupervised manner from the feature space into a low dimensional (2D) space using Uniform Manifold Approximation & Projection for dimension reduction (UMAP [17]). The centroid of each session for each patient was identified, and idle state variability was evaluated through the Euclidean distance between each session's centroid and the centroid of session 1, and between each pair of consecutive sessions' centroids (Fig. 2). Measure of distances in a 2D UMAP embedding was chosen to investigate variability over sessions because we showed in a previous study on a comparable dataset that the projected patterns were remarkably stable over time [18].

**Idle state dimensionality:** Intrinsic dimension (ID) was computed in the feature space for idle state samples of each session. ID can be defined as the minimum number of parameters needed to describe the data with a minimal loss of information. This was done using two widely used classical estimators, MLE and DANCo [19]. ID was also assessed in MEG experiments to evaluate to what extent the dimensionality is predictable before implantation.

**Correlations:** To investigate if idle state characteristics could explain inter-sessions and inter-subjects variability, linear regressions were estimated between BCI performances, both distances in the UMAP embedding as well as the ID. Specifically, we investigated relationships between 1) within-sessions training performances and idle state ID indicators, 2) session-1 training performances and distances between each session's centroid and the centroid of session 1, and 3) session-to-session training performances and distances between each pair of consecutive sessions' centroids. Goodness of fits were estimated using the Pearson correlation coefficient  $R$ .

**Statistical tests:** Differences in performances, distances and ID between subjects (mean across sessions) were assessed using one-way analysis of variance (ANOVA), followed by post-hoc Tukey's honestly significant difference tests when ANOVA were significant. Statistical significance threshold was set to  $p < 0.05$ .

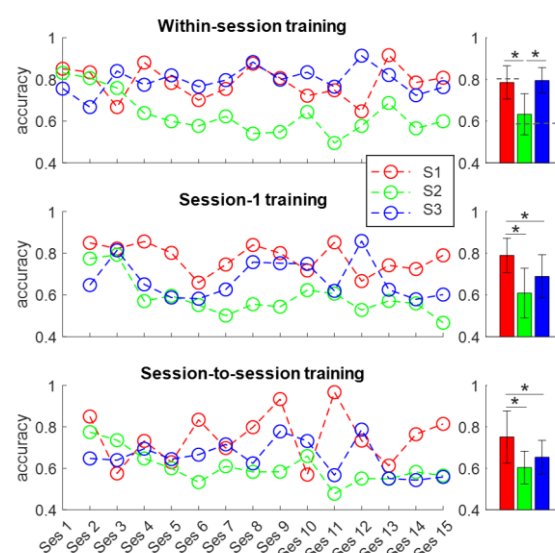


Figure 1: Models' decoding accuracies over sessions (left) and average across sessions (right) with the three training strategies. Bar graph are presented as mean  $\pm$  standard dev. \* reports significant differences. For comparison purposes, grey dotted lines on the top bar graph represents models' decoding accuracies in MEG experiments.

## RESULTS

**BCI performances:** BCI performances (models accuracy on test sets) of the three subjects with the three training strategies are presented in Fig. 1. S1 and S3 had significantly better performances than S2 in the within-session training ( $p < 0.001$ ; S1:  $0.79 \pm 0.08$ ; S2:  $0.63 \pm 0.10$ ; S3:  $0.79 \pm 0.06$ ). When model was trained on day 1 only, performances of S1 were significantly better than S2 and S3 ( $p < 0.001$  and  $p = 0.02$  respectively; S1:  $0.79 \pm 0.08$ ; S2:  $0.61 \pm 0.12$ ; S3:  $0.69 \pm 0.10$ ). Only performances of S3 dropped in this scenario compared to within-session training. For session-to-session training, S1 performances were significantly better than S2 and S3 ( $p < 0.001$  and  $p = 0.03$  respectively; S1:  $0.75 \pm 0.13$ ; S2:  $0.60 \pm 0.08$ ; S3:  $0.65 \pm 0.08$ ). Again, performances of S3 dropped particularly with this training strategy compared to within-session training. Regarding prior to implantation MEG experiment, BCI performances were better for S1 in comparison to S2 and S3 (S1: 0.80; S2: 0.58; S3: 0.59).

**Idle state variability:** Distances between centroids of idle state features projected into the 2D UMAP embedding are presented in Fig. 2. Whether comparing distance to first session or session-to-session distances, it appeared that idle state features of S1 remained more stable over time, with a smaller average distance and a smaller variability of distances. Distance to session 1 was significantly lower for S1 than for S2 and S3 ( $p = 0.01$  and  $p = 0.02$  respectively; S1:  $0.73 \pm 0.41$ ; S2:  $1.99 \pm 1.78$ ; S3:  $1.33 \pm 1.13$ ), while session-to-session distance was significantly lower for S1 in comparison to S3 only ( $p < 0.01$ ; S1:  $1.11 \pm 0.58$ ; S2:  $1.49 \pm 1.07$ ; S3:  $1.95 \pm 1.38$ ).

**Idle state dimensionality:** ID was globally stable over ECoG sessions for the three subjects with a gradation between them (Fig. 3). Whether computed with MLE or

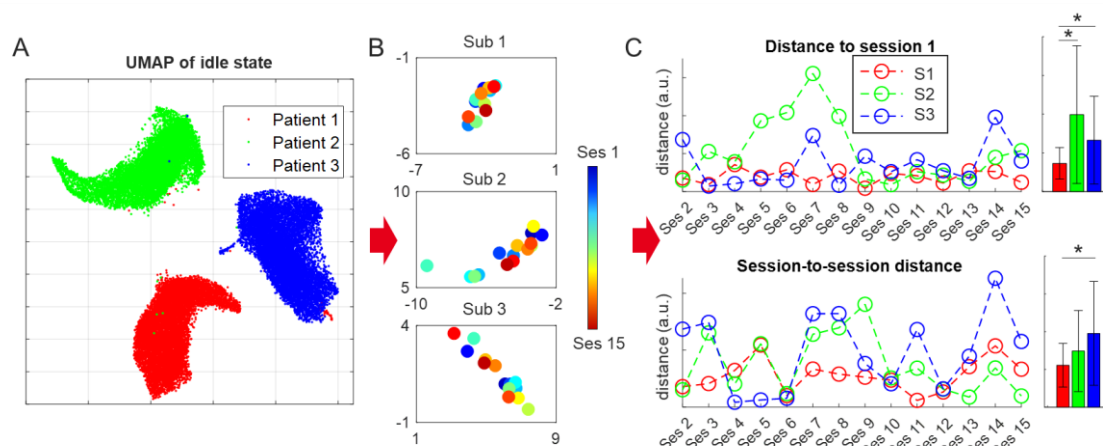


Figure 2: (A) Unsupervised UMAP of all idle state features of the three patients in a 2D-space; (B) Representation of the centroids of each BCI-ECOG session for each subject separately; (C) Euclidean distance between centroid of session 1 and centroids of the following sessions (top) and distance between each session and the previous one (bottom) in the UMAP embedding. Average across sessions is presented as bar graphs on the right (mean  $\pm$  standard dev). \* reports significant differences).

DANCo estimator, ID was significantly lower for S1 compared to S2 and S3, and significantly lower for S2 than for S3 ( $p < 0.001$  for all cases; MLE-ID: S1:  $15.3 \pm 1.8$ ; S2:  $31.3 \pm 4.8$ ; S3:  $24.7 \pm 2.9$ ; DANCo-ID: S1:  $19.2 \pm 4.5$ ; S2:  $53.7 \pm 11.4$ ; S3:  $39.9 \pm 7.4$ ). ID was also evaluated during prior to implantation BCI-MEG experiments, and seemed to follow a similar pattern between subjects, especially with DANCo estimator (MLE-ID: S1: 34.0; S2: 36.3; S3: 36.8; DANCo-ID: S1: 27.2; S2: 43.3; S3: 37.2).

**Correlations:** When pulling subjects' together, significant correlations between cross-sessions models' accuracies and variability of idle state between sessions as well as between subjects' performances and intrinsic dimensions were found (cf. Fig. 4 – Pearson correlation coefficient and associated p-values are indicated in the figure). These correlations remained insignificant for individual subjects.

## DISCUSSION

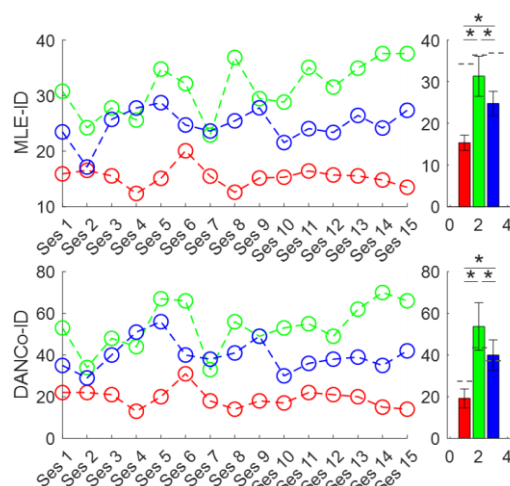
The main objective of this study was to investigate if markers of inter and intra-subjects' BCI performances could be unraveled from idle state. To do so, we compared different indicators of idle state brain signals between sessions and patients, and related them to the BCI performances. BCI performances were assessed offline, with three training strategies to disentangle subjects' best achievable performance (within-session training), drift over time (session-1 training) and session-to-session variability (session-to-session training). Note that for the purpose of this study, subjects' best achievable performance were assessed with relatively small training sets (5 folds of 1440 features, i.e. 2.4 min of data) and do not reflect the best performances they could achieve with a longer training.

The three training strategies allowed us to observe that the patients presented distinct BCI profiles. S1 showed high BCI capabilities with no drift and only low variability over sessions. S2 performances were lower, but we did not observe strong drift or variability between

sessions either. Regarding S3 high BCI performances were observed, with an important drop of performances when the model of a previous session was used, indicating a drift and/or variability over sessions. Interestingly, these observations are consistent with what was noticed in online experiments. Indeed, S1 was able to control accurately up to 8 continuous degrees of freedom without recalibration of the model for up to 6 months [6], while S2 controlled models with 5 discrete states with fluctuant performances even in the same session. As for S3, he was somewhere in between, and controlled with good performances 7 continuous states [1], although regular model recalibration (approximately every 2 weeks) was necessary.

This more frequent need for recalibration was also coherent with more fluctuations of idle state, measured as distances between centroid features of sessions in a low-dimensional projection. Indeed, S2 and S3 presented higher variability in idle state features over sessions. Furthermore, the performances in a session using a decoding model of another session was significantly negatively correlated with the distance between these sessions. Thus, measuring the distance with model's calibration session through an idle state recording acquired prior to the BCI experiment could be a good predictor of session expected performances, and of the need to recalibrate the model. Although this has to be confirmed on more sessions, we believe to have obtained here a session-to-session predictor of subjects' performances, in contrast to previous studies that investigated only inter-subjects predictors of performance [10], [11], [20]. In addition, this distance is a data-driven indicator that is not based on neurophysiological hypotheses, and then could be adapted to other recording techniques.

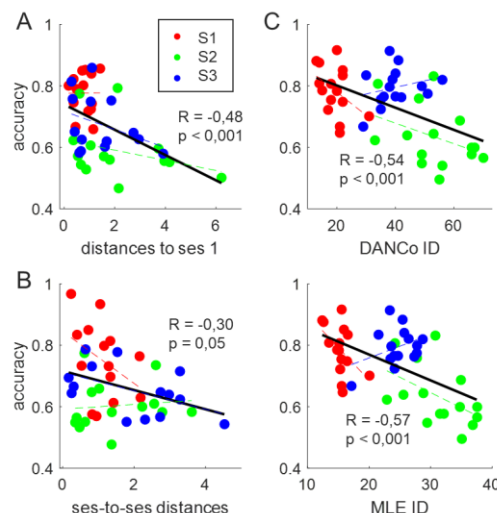
Regarding ID of the idle state in features space, we observed a clear gradation between subjects, with a lower ID for S1, higher for S2, and relatively stable across sessions. Crosschecking this result with subjects'



**Figure 3:** ID of idle state features over sessions (left) and average across sessions (right) with MLE (top) and DANCo (bottom) estimators. Bar graph are presented as mean  $\pm$  standard dev. \* reports significant differences. For comparison purposes, grey dotted lines on the bar graphs represents ID of idle state features in MEG experiments.

performances indicated that ID could be a good global predictor of subject BCI capabilities: the lower ID, the more decodable and stable brain signals. This was confirmed by correlation between model's accuracy and ID. This is in agreement with previous results that reported the same relationship in images dataset [21].

Although predicting the global long-term performances of implanted patients through ECoG experiments could be of interest, it would be much more valuable to assess it before implantation. With this in mind, we examined prior to implantation MEG-BCI experiments that were performed by the three subjects. This session was the first BCI experiment of patients and was performed to assess their adhesion. While S1 and S2 presented similar MEG-BCI performances than in ECoG-BCI, S3 presented much lower performances (in comparison to within-session training, which is the same strategy than in MEG). This tends to indicate that, if high MEG-BCI performances would ensure high ECoG-BCI performances, lower MEG-BCI performances does not necessarily leads to lower ECoG-BCI performances. This is not surprising as the patients certainly have a different BCI learning potential, which cannot be estimated within a single session. Thus, MEG-BCI sessions seems to be an important step for subjects prior to the implantation to assess their BCI "compatibility" (in addition to their adhesion), but we strongly suggest to perform more than one BCI sessions (ideally, enough to observe a learning curve). Interestingly ID of idle state during these MEG-BCI experiments was also estimated and showed the same gradation between patients than in ECoG-BCI experiments (especially with DANCo estimator). Since ID seems to remain relatively stable and discriminating across subjects after implantation, it could be a strong predictor of subjects' global BCI long-term performances. Although this need to confirmed on more patients, this could be an important finding as it could help identifying BCI-inefficient patient for whom an



**Figure 4:** Correlations analyses: (A) between models' accuracy (session-1 training strategy) and distance to session 1 in idle state UMAP embedding, (B) between models' accuracy (session-to-session training strategy) and session-to-session distances in idle state UMAP embedding and (C) models' accuracy (within-session training strategy) and idle state features ID (DANCo and MLE).

implantation would be an unnecessary risk in addition to a waste of time and resources. A simple assessment of subject's brain signals complexity in MEG (or EEG) in a relax state could participate to reveal these patients before implantation.

## CONCLUSION

To our knowledge, this work is the first study investigating session-to-session BCI performance predictors on implanted ECoG patients. Even though this was based on offline analysis, on patients that used different motor imagery, effectors and online decoding models, the conclusions drawn here on narrowed comparable datasets are reflecting our experimental observations done during the online experiments. Based on idle state features variability over sessions, we first uncovered a predictor of the performances of a previous decoding model on current session. Then, analyzing the dimensionality of brain signals in idle state, we revealed a more "long-term" indicator, which predicted the global BCI-capabilities of the patients. Furthermore, although this must be confirmed on more subject, the latter followed the same pattern between patients in MEG-BCI experiments performed before implantation. These results are of particular importance on one hand to anticipate the need of model recalibration in ECoG-BCI training experiments, and on the other hand for selecting patients to be implanted with BCI neuroprosthesis.

## ACKNOWLEDGMENTS

This study was Supported by Translational Medical Research Award 2021 from the Leenaards Foundation, the Swiss National Science Foundation, the Lead Agency Program with the French National Research Agency

(Think2Move SNF-32003BE-205563, ANR-21-CE19-0038), the European Commission (ERC-2019-PoC Braingait 875660, EIC 2021-TransitionChallenges-01-01 ReverseParalysis 101057450, Horizon-EIC-2021-Pathfinderchallenges-01-02 NEMO-BMI 101070891), Fonds Clinatéc (WIMAGINE implant development, BCI4Grasp) and Institut Carnot Leti.

## REFERENCES

- [1] H. Lorach *et al.*, “Walking naturally after spinal cord injury using a brain–spine interface,” *Nature*, vol. 618, no. 7963, pp. 126–133, Jun. 2023, doi: 10.1038/s41586-023-06094-5.
- [2] S. L. Metzger *et al.*, “A high-performance neuroprosthesis for speech decoding and avatar control,” *Nature*, vol. 620, no. 7976, pp. 1037–1046, Aug. 2023, doi: 10.1038/s41586-023-06443-4.
- [3] A. Biasiucci *et al.*, “Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke,” *Nat. Commun.*, vol. 9, no. 1, p. 2421, Jun. 2018, doi: 10.1038/s41467-018-04673-z.
- [4] D. J. Krusienski *et al.*, “Critical issues in state-of-the-art brain–computer interface signal processing,” *J. Neural Eng.*, vol. 8, no. 2, p. 025002, Apr. 2011, doi: 10.1088/1741-2560/8/2/025002.
- [5] C. Larzabal *et al.*, “Long-term stability of the chronic epidural wireless recorder WIMAGINE in tetraplegic patients,” *J. Neural Eng.*, vol. 18, no. 5, p. 056026, Oct. 2021, doi: 10.1088/1741-2552/ac2003.
- [6] A. Moly *et al.*, “An adaptive closed-loop ECoG decoder for long-term and stable bimanual control of an exoskeleton by a tetraplegic,” *J. Neural Eng.*, vol. 19, no. 2, p. 026021, Apr. 2022, doi: 10.1088/1741-2552/ac59a0.
- [7] S. Luo *et al.*, “Stable Decoding from a Speech BCI Enables Control for an Individual with ALS without Recalibration for 3 Months,” *Adv. Sci.*, vol. 10, no. 35, p. 2304853, Dec. 2023, doi: 10.1002/advs.202304853.
- [8] C. Vidaurre and B. Blankertz, “Towards a Cure for BCI Illiteracy,” *Brain Topogr.*, vol. 23, no. 2, pp. 194–198, Jun. 2010, doi: 10.1007/s10548-009-0121-6.
- [9] J.-H. Kang, J. Youn, S.-H. Kim, and J. Kim, “Effects of Frontal Theta Rhythms in a Prior Resting State on the Subsequent Motor Imagery Brain-Computer Interface Performance,” *Front. Neurosci.*, vol. 15, p. 663101, Aug. 2021, doi: 10.3389/fnins.2021.663101.
- [10] E. Tzdaka, C. Benaroch, C. Jeunet, and F. Lotte, “Assessing The Relevance Of Neurophysiological Patterns To Predict Motor Imagery-based BCI Users’ Performance,” in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Toronto, ON, Canada: IEEE, Oct. 2020, pp. 2490–2495. doi: 10.1109/SMC42975.2020.9283307.
- [11] B. Blankertz *et al.*, “Neurophysiological predictor of SMR-based BCI performance,” *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, Jul. 2010, doi: 10.1016/j.neuroimage.2010.03.022.
- [12] M. Grosse-Wentrup, B. Schölkopf, and J. Hill, “Causal influence of gamma oscillations on the sensorimotor rhythm,” *NeuroImage*, vol. 56, no. 2, pp. 837–842, May 2011, doi: 10.1016/j.neuroimage.2010.04.265.
- [13] A. Bleuzé, J. Mattout, and M. Congedo, “Tangent space alignment: Transfer learning for Brain-Computer Interface,” *Front. Hum. Neurosci.*, vol. 16, p. 1049985, Dec. 2022, doi: 10.3389/fnhum.2022.1049985.
- [14] A. D. Degenhart *et al.*, “Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity,” *Nat. Biomed. Eng.*, vol. 4, no. 7, pp. 672–685, Apr. 2020, doi: 10.1038/s41551-020-0542-9.
- [15] A. L. Benabid *et al.*, “An exoskeleton controlled by an epidural wireless brain–machine interface in a tetraplegic patient: a proof-of-concept demonstration,” *Lancet Neurol.*, vol. 18, no. 12, pp. 1112–1122, Dec. 2019, doi: 10.1016/S1474-4422(19)30321-7.
- [16] A. Eliseyev *et al.*, “Recursive Exponentially Weighted N-way Partial Least Squares Regression with Recursive-Validation of Hyper-Parameters in Brain-Computer Interface Applications,” *Sci. Rep.*, vol. 7, no. 1, p. 16281, Nov. 2017, doi: 10.1038/s41598-017-16579-9.
- [17] L. McInnes, J. Healy, and J. Melville, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction,” arXiv, Sep. 17, 2020. Accessed: Feb. 16, 2024. [Online]. Available: <http://arxiv.org/abs/1802.03426>
- [18] M. Śliwowski, M. Martin, A. Souloumiac, P. Blanchart, and T. Aksenova, “Impact of dataset size and long-term ECoG-based BCI usage on deep learning decoders performance,” *Front. Hum. Neurosci.*, vol. 17, 2023, Accessed: Aug. 28, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnhum.2023.1111645>
- [19] P. Campadelli, E. Casiraghi, C. Ceruti, and A. Rozza, “Intrinsic Dimension Estimation: Relevant Techniques and a Benchmark Framework,” *Math. Probl. Eng.*, vol. 2015, pp. 1–21, 2015, doi: 10.1155/2015/759567.
- [20] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, “High Theta and Low Alpha Powers May Be Indicative of BCI-Illiteracy in Motor Imagery,” *PLoS ONE*, vol. 8, no. 11, p. e80886, Nov. 2013, doi: 10.1371/journal.pone.0080886.
- [21] P. Pope, C. Zhu, A. Abdelkader, M. Goldblum, and T. Goldstein, “The Intrinsic Dimension of Images and its Impact on Learning,” *ICLR 2021*, 2021.

# UNCERTAINTY QUANTIFICATION FOR CROSS-SUBJECT MOTOR IMAGERY CLASSIFICATION

Prithviraj Manivannan\*, Ivo Pascal de Jong\*, Matias Valdenegro-Toro, Andreea Ioana Sburlea

Department of Artificial Intelligence, Bernoulli Institute, University of Groningen, The Netherlands

E-mail: ivo.de.jong@rug.nl

**ABSTRACT:** Uncertainty Quantification aims to determine when the prediction from a Machine Learning model is likely to be wrong. Computer Vision research has explored methods for determining epistemic uncertainty (also known as model uncertainty), which should correspond with generalisation error. These methods theoretically allow to predict misclassifications due to inter-subject variability. We applied a variety of Uncertainty Quantification methods to predict misclassifications for a Motor Imagery Brain Computer Interface. Deep Ensembles performed best, both in terms of classification performance and cross-subject Uncertainty Quantification performance. However, we found that standard CNNs with Softmax output performed better than some of the more advanced methods.

## INTRODUCTION

Machine Learning systems for Brain Computer Interfaces (BCI) are normally optimised to their predictive accuracy. The availability of public datasets and benchmarking systems allow for faster progress in this direction. However, for successful BCI systems there are more aspects that need to be explored.

This study explores the options of Uncertainty Quantification (UQ) for Machine Learning models [1] as applied to non-invasive Motor Imagery BCIs. Uncertainty Quantification aims to estimate how likely a prediction from a Machine Learning model is to be correct. For this two types of uncertainty are commonly considered.

*Two types of Uncertainty:* Aleatoric uncertainty (also referred to as data uncertainty) is the uncertainty inherent in the data. This cannot be reduced by better models, only by better EEG recordings or better paradigms. Noisy EEG recordings or extracted features that are poorly correlated to the to-be-predicted classes introduce aleatoric uncertainty.

Epistemic uncertainty (also referred to as model uncertainty) is the uncertainty in the model. This kind of uncertainty can be reduced by collecting more training samples that are similar to what the model is being evaluated on. In BCI contexts this uncertainty can come from limited amounts of training data [2], but also from between-subject variability [3].

While there is some Motor Imagery BCI research dedicated to UQ [2–5], it is worth noting that simple methods of estimating aleatoric uncertainty are often readily available. For example, Neural Networks used for classification generally use Softmax or Sigmoid activation functions for the output, which also gives a crude estimate of aleatoric uncertainty.

This study, like most research on modelling epistemic uncertainty is mostly done in the domain of Deep Learning.

*Using Uncertainty for Rejection:* UQ is often considered as a method for improving interpretability of predictions from a Machine Learning model [6]. There, the goal is to have a precise and well calibrated prediction of the class probability. This means that a prediction with 90% certainty should be correct 90% of the time. This results in methods aimed at addressing overconfidence of Neural Networks [7].

However, for BCIs there is often no time for human interpretation of the classification. Instead, the system should automatically deal with certain and uncertain predictions. Typically this means "rejecting" the uncertain predictions and abstaining from sending a control command to the device. We focus on this rejection case, as it aligns with how BCIs are implemented in practice, and highlight that it comes with different methods and metrics.

*Research Aim:* This paper investigates whether UQ methods that account for epistemic uncertainty can identify wrong predictions in cross-subject classification. This expands on previous work [3, 5] by exploring a larger variety of UQ methods and by applying a leave-one-subject-out cross validation paradigm to get a more realistic estimate of model performance.

We investigate whether available UQ methods for CNNs that account for epistemic uncertainty are actually able to reject the uncertain predictions when applied cross-subject better than the crude methods readily available.

Previous work has shown success with rejection methods [5], but a comparison with simple baseline methods such as Softmax is missing. Moreover, by using different measures of uncertainty we can see how much aleatoric and epistemic uncertainty contribute to the total uncertainty. This disentangling of uncertainties has not been applied to BCIs before [8]. Lastly, we cover a wider range of UQ methods and explain how they have different underlying assumptions.

\*These authors contributed equally to this work

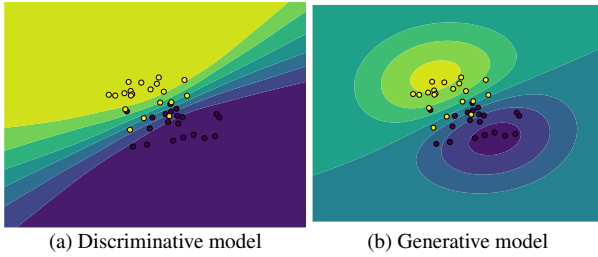


Figure 1: An illustration of a discriminative and a generative model. The yellow and purple dots indicate the training samples of two different classes. The background indicates the prediction. The green color indicates uncertainty.

*Background on uncertain models:* Following [9] we consider two assumptions for how epistemic uncertainty may be modelled. [9] calls these two assumptions *discriminative* and *generative* models.

Discriminative models learn a boundary that optimally separates the classes. Samples that are far away from this boundary are considered "certain", whereas samples that are close to this boundary are considered uncertain. When the model uncertainty is considered, these methods consider multiple decision boundaries that are all valid with the training data. When samples fall between different decision boundaries this is considered epistemic uncertainty. Figure 1a shows what this looks like in a 2D feature space. This could be the band power following 2 CSP filters, but a similar concept can also be applied at a higher dimensional space for Neural Networks. In contrast, generative models learn the distribution of each class. A sample that matches the distribution of the training data is considered "certain", whereas a sample that is far away from the training data is considered "uncertain". Figure 1b visualises this concept.

Both approaches have similar behaviour under aleatoric uncertainty. This is seen in the parts where the two classes overlap. However, they exhibit very different behaviour under epistemic uncertainty. Since it is not known which of these underlying assumptions is most suitable it is important to consider models from either family.

*Bayesian Neural Networks:* Bayesian Neural Networks (BNNs) fall under the category of discriminative models. Standard Neural Networks learn a single optimal vector  $\theta$  of the parameters learned on the training data  $D$ . They then do classification according to the Softmax function to capture aleatoric uncertainty. BNNs instead consider a weight distribution  $p(\theta|D)$ . This captures all possible weights for the Neural Networks, based on how well they fit the data. Inference is then made according to the predictive posterior distribution:

$$p(y = c|x) = \int \underbrace{p(y = c|x, \theta)}_{\text{aleatoric}} \underbrace{p(\theta|D)}_{\text{epistemic}} d\theta. \quad (1)$$

Truly Bayesian Neural Networks are computationally infeasible, so instead various methods to approximate it

have been proposed [1, 8]. We will be considering MC-Dropout [10], MC-DropConnect [11], Deep Ensembles [12] and Flipout [13].

While they have differences in approximation quality, implementation complexity, and computational cost, they all rely on BNN fundamentals.

*Deterministic Uncertainty Quantification (DUQ):* DUQ [14] uses a different approach to Uncertainty Quantification in Neural Networks. DUQ uses a standard Neural Network as a feature extractor, and then learns a centroid for each class. Samples that are far away from the centroids are deemed uncertain, whereas samples that are close to a centroid are deemed certain.

This different underlying assumption of how uncertainty should arise is inspired by generative models, though DUQ is not actually a generative model. A true generative model models the distribution of the training samples directly, whereas DUQ only models class centroids. Still, this makes it fundamentally different from the discriminative BNNs, and may therefore give different results than the BNN approach. It also means that aleatoric and epistemic uncertainty cannot be clearly distinguished, but they are both included in the predicted uncertainty.

## METHODS

*Dataset:* We used the public Motor Imagery dataset: BCI Competition IV, dataset 2a [15]. This dataset contains 22 channel EEG and 3 monopolar EOG channel recordings of 9 subjects performing one of 4 different motor imagery tasks— left hand (class 1), right hand (class 2), both feet (class 3) and tongue (class 4). The sampling rate was 250Hz and the dataset comes pre-applied with a 50Hz notch filter and a bandpass filter of 0.5Hz to 100Hz.

The Braindecode [16] and MNE [17] Python libraries were used to load and pre-process the data.

The training setup (shown for a single subject in figure 2) was designed to allow the observation of aleatoric uncertainty and the combination of aleatoric and epistemic uncertainty. This allows the impact of epistemic uncertainty to be observed in isolation.

We used leave-one-subject-out cross-validation with a slight variation. Normally leave-one-subject-out involves splitting  $N - 1$  subjects into a training set and leaving the last subject as the out-of-population (cross-subject) set. Our variation to this procedure is as follows: 10% of the data from each training set subject is used as a within-population test set\*. This within-population dataset allows for an observation with minimal epistemic uncertainty, and comparing it to the cross-subject set allows us to isolate the impact of cross-subject generalisation.

*Preprocessing:* Some EEG pipelines employ extensive signal processing and feature extraction in order to operate with ML algorithms. However, it is often unclear what value each processing step introduces, and various

\*The remaining training data was in turn split into 90% train and 10% validation for hyperparameter optimisation

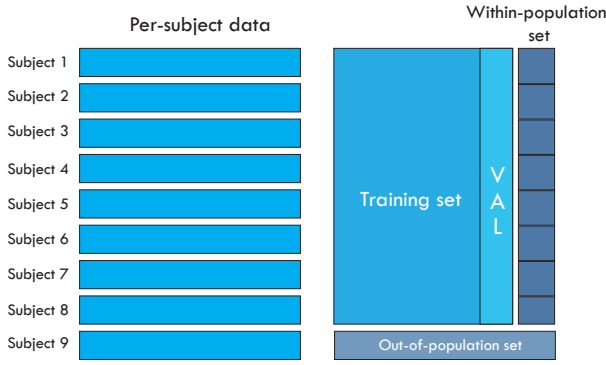


Figure 2: Training setup for a single subject. One subject is excluded and used as an out-of-population set while the other 10% of the data from each subject is separated into a within-population set. The data of the remaining subjects are concatenated and split 90-10 into a training and validation set. This procedure is repeated for every subject.

researchers and labs use different pipelines. The use of CNNs (and DL methods in general) in EEG is promising because of their ability to automatically extract features from raw data and perform classifications, with minimal preprocessing required [18, 19].

Hence the following preprocessing steps are very minimal. It consists of: dropping the EOG channels, converting the EEG signals from volts to microvolts ( $\mu V$ ), applying an exponential moving standardisation with parameters described by [16] and epoching from 0.5 seconds before the trial cue at  $t = 2s$  to end of the trial at  $t = 6s$  (for a total trial window of 4.5 seconds). Creating epochs as such leads to a single trial being a matrix  $(C, T)$  with  $C = 22$  being the number of channels and  $T = 1125$  being the number of timestamps.

**Model Architecture:** We used Keras [20] to implement the Shallow ConvNet CNN [16], and the Keras Uncertainty library [21] to implement the UQ adaptations. \* Although all UQ methods followed the same Shallow ConvNet architecture, minor differences existed in the implementation of the UQ layers. Two standard models regularised with Dropout and DropConnect were used as baselines.

MC-Dropout and MC-DropConnect and their standard counterparts both had only a single UQ layer. In MC-Dropout this layer was positioned before the dense classification layer with a drop rate of 0.2. In MC-DropConnect it was positioned after the second convolutional layer with a drop rate of 0.1. A grid search was done on a single subject (due to computational complexity) to decide this configuration. Normal Dropout and DropConnect sets the value of a node or weight to 0 during training. The equivalent UQ versions retain this during testing, resulting in slightly different predictions each forward pass, thereby representing epistemic uncertainty. The Ensemble model simply consisted of 10 standard Shallow ConvNet CNNs, regularised with dropout identical to the dropout baseline model. Disagreement between

these 10 models represents epistemic uncertainty.

Flipout changes the final dense classification layer to a standard dense layer using ReLU activation with 10 units, following which are two flipout layers. Both flipout layers use a prior  $P(\theta) = \mathcal{N}(0, 1.0^2) + \pi \mathcal{N}(0, 2.5^2)$  with  $\pi = 0.1$ . Additionally, the first flipout layer had 10 units. Both sets of parameters were determined using a grid search.

MC-Dropout, MC-Dropconnect and Flipout are stochastic during inference. Therefore, a number of forward passes  $T$  needs to be selected.  $T$  was chosen to be 50 as it has been found to be point where the improvement in accuracy stabilises [3].

DUQ changes the final layer of the Shallow ConvNet CNN to a dense layer with 100 units using a ReLU activation, following which is an RBF classification layer with a length scale of 0.4 with trainable centroids of dimension 100. These parameters were found using a grid search. Additionally, compared to the categorical cross-entropy loss used by the other methods, DUQ utilizes binary cross-entropy.

Other hyperparameters follow common practice in Deep Learning literature. Specifically we set the learning rate ( $1 \times 10^{-4}$ ), loss function (categorical cross entropy), and optimiser (Adam).

**Uncertainty Measures:** The BNN-based methods rely on  $T$  forward passes from a stochastic model. Each forward pass predicts class probabilities  $p_c$ , resulting in a distribution over probabilities. To this we can apply various Uncertainty Measures to measure either aleatoric uncertainty, epistemic uncertainty or the total uncertainty [8].

The total uncertainty is based on the mean of the predicted probability for each class and is measured by the Predictive Entropy:

$$\mathbb{H}_{\text{pred}}(p) = -\sum_c \bar{p}_c \log \bar{p}_c. \quad (2)$$

The Expected Entropy first determines the uncertainty of each forward pass, and then takes the average over those uncertainties.

$$\mathbb{H}_{\mathbb{E}}(p) = -T^{-1} \sum_t \sum_c p_{ct} \log p_{ct} \quad (3)$$

In this approach, Expected Entropy takes the "average uncertainty" of each individual model. As such, it only corresponds to aleatoric uncertainty [22].

Lastly, subtracting the aleatoric uncertainty from the total uncertainty results in the remaining epistemic uncertainty. This measure is referred to as Mutual Information [23]:

$$\mathbb{I}(p) \approx \mathbb{H}_{\text{pred}}(p) - \mathbb{H}_{\mathbb{E}}(p) \quad (4)$$

Predictive Entropy and Expected entropy may be applied to a standard Neural Network, but they will result in the same prediction. This approximation for Mutual Information cannot be applied to standard Neural Networks.

\* All code is available at <https://github.com/p-manivannan/UQ-Motor-Imagery>

Because DUQ does not follow the same discriminative assumptions for uncertainty, these measures of uncertainty do not apply. Instead, it gives a single uncertainty measure that responds to both aleatoric and epistemic uncertainty.

## RESULTS

Classification accuracy for each method is given in table 1. It can be seen that performance is higher within-population than out-of-population, with Ensembles outperforming all other methods for both groups. The performance of the Ensemble is in-line with benchmarks for out-of-population and within-population accuracies [19] while the other methods are slightly underperforming.

To find out whether UQ can improve performance, uncertainty estimation was treated as a binary classification task, where the aim was to classify wrong predictions as uncertain. Therefore, the Area Under the ROC curve (AUROC) is considered as a performance metric [24]. Note that this can never approach 100 as the uncertain samples are "guessed", which will be correct 25% of the time. These would be labelled as false positives in this framework.

This metric is chosen in place of common metrics like Expected Calibration Error [7], because our goal is to detect misclassifications, whereas ECE aims to detect overconfidence or underconfidence.

The uncertainty AUROC scores for each method and each uncertainty measure on the within-population set is given in table 2a. This table shows that Mutual Information (which corresponds only to epistemic uncertainty) performs the worst. Predictive Entropy and Expected Entropy perform similarly, suggesting that the modelling of epistemic uncertainty is not beneficial to the uncertainty estimation. It also shows that DUQ has the worst uncertainty estimation, and that most discriminative models show similar performance.

Table 2b shows the performance of uncertainty estimation on the out-of-population dataset. The performance of uncertainty estimation is consistently lower here than on the within-population set. Mutual Information, which represents epistemic uncertainty, still does not offer better uncertainty estimation. This suggests that none of the available models are able to fully account for the epistemic uncertainty introduced by cross-subject classification. We again see that DUQ has noticeably worse UQ performance.

It can be seen that the quality of uncertainty estimation is worse cross-population than within-population. This behaviour is inevitable for measures of aleatoric uncertainty, but measures of epistemic uncertainty should be more robust to this [25].

When predictive entropy is disentangled into aleatoric and epistemic uncertainty, it can be seen that epistemic uncertainty based thresholding is consistently slightly worse than aleatoric uncertainty based thresholding. This suggests either that aleatoric uncertainty is more preva-

Table 1: Mean accuracy per subject for each method. Within-population accuracy is higher overall than cross population accuracy, with ensembles outperforming other methods in both categories. Standard DropConnect performs noticeably worse, but most methods perform similar to Standard Dropout.

Method	Within pop. Acc%	Cross pop. Acc %
Dropout	68.98 $\pm$ 2.73	55.54 $\pm$ 7.95
MC-Dropout	69.00 $\pm$ 2.73	55.56 $\pm$ 7.94
DropConnect	66.67 $\pm$ 2.23	53.51 $\pm$ 11.67
MC-DropConnect	69.27 $\pm$ 1.34	54.96 $\pm$ 9.76
Flipout	69.90 $\pm$ 2.55	54.99 $\pm$ 8.67
Ensembles	<b>73.05 <math>\pm</math> 2.22</b>	<b>59.05 <math>\pm</math> 8.11</b>
DUQ	70.47 $\pm$ 2.93	55.42 $\pm$ 9.16

lent than epistemic uncertainty, or that epistemic uncertainty is not captured well by the models. Since the accuracy does go down when moving to cross-population, it is clear that there must be an increase in epistemic uncertainty which the models are not accounting for.

It can be seen that no BNN method is substantially better than another at uncertainty quantification. Only DUQ performs substantially worse than other methods, performing even lower than standard neural networks.

## DISCUSSION

Surprisingly, we find that the specific UQ methods designed to observe epistemic uncertainty are not able give better uncertainty estimations than a similar Neural Network with Softmax activation. It is still possible for all methods to reject some of the uncertain samples to increase accuracy, but this is trivial.

A possible reason for this is that since aleatoric uncertainty seems more prevalent, the ability of these UQ methods to take into account epistemic uncertainty does not help, hence explaining how standard models are able to achieve comparable performance. However, it is clear that the decrease in accuracy should be attributable to an increase in epistemic uncertainty. This could be caused by how these methods model uncertainty, but the results show that the discriminative models and DUQ suffer the same problems.

*Relation to background:* Our findings contradict the expectation that cross-subject classification should introduce epistemic uncertainty, and that therefore BNNs should perform better.

Epistemic uncertainty should arise when a model is tested on data that is different from the data it was trained on. In this case, the cross-subject testing samples are different from the data that the model is trained on, but the models capturing epistemic uncertainty were not able to offer better uncertainty estimates.

It is difficult to attribute this to problems with a specific approximation of BNNs, as a variety of approximations show this effect consistently. We also cannot attribute this to flaws in the discriminative model as shown in Figure 1a, because this problem is consistent even when using DUQ which has a fundamentally different assumption of uncertainty.

Table 2: Uncertainty AUROC scores for each method both within-population and out-of-population. Predictive Entropy and Expected Entropy both perform equally well for all BNN models. At the same time Mutual Information performs noticeably worse, and shows more difference for the different models. The uncertainty of all models and all uncertainty measures is consistently worse when moving out-of-population.

(a) Within-population

Method	Predictive Entropy (Ale+Epi)	Expected Entropy (Ale)	Mutual Information (Epi)
Standard Dropout	$76.07 \pm 2.918$	$76.07 \pm 2.918$	-
MC-Dropout	$76.07 \pm 2.927$	$76.06 \pm 2.927$	$74.24 \pm 3.398$
Standard DropConnect	$75.44 \pm 2.691$	$75.44 \pm 2.691$	-
MC-DropConnect	$75.3 \pm 3.303$	$75.29 \pm 3.297$	$73.33 \pm 2.835$
Flipout	$75.56 \pm 2.461$	$76.70 \pm 2.460$	$70.56 \pm 2.488$
Ensembles	$76.92 \pm 2.868$	$76.66 \pm 3.046$	$70.02 \pm 2.064$
DUQ	$73.19 \pm 2.379$	-	-

(b) Out-of-population

Method	Predictive Entropy (Ale+Epi)	Expected Entropy (Ale)	Mutual Information (Epi)
Standard Dropout	$67.46 \pm 4.646$	$67.46 \pm 4.646$	-
MC-Dropout	$67.43 \pm 4.611$	$67.43 \pm 4.611$	$66.6 \pm 4.164$
Standard DropConnect	$68.23 \pm 4.532$	$68.23 \pm 4.532$	-
MC-DropConnect	$68.48 \pm 4.625$	$68.48 \pm 4.626$	$66.82 \pm 5.311$
Flipout	$67.79 \pm 5.156$	$67.79 \pm 5.152$	$63.95 \pm 4.024$
Ensembles	$67.39 \pm 5.446$	$67.29 \pm 5.564$	$63.86 \pm 4.354$
DUQ	$65.30 \pm 4.01$	-	-

The previous studies in this direction [3, 5] show more positive findings for approximations of BNNs, but by considering an equivalent CNN and using Softmax as a baseline we were able to that those results can also be achieved with simpler methods.

*Limitations:* Our study also only focuses on the use of uncertainty for rejecting difficult samples, and does not actively look at the absolute epistemic uncertainty. It may be that the epistemic uncertainty did increase for cross-subject samples, but if this happens uniformly for a given subject we are not able to capture it. This does not affect the validity of the findings, but does make it harder to know why these Bayesian Neural Networks are not performing well.

There may also be limitations underlying how Predictive Uncertainty is disentangled into aleatoric and epistemic uncertainty. The proposed approach follows a line of existing work [22, 23], but there is also a line of work that assumes an entirely different formulation for disentangling uncertainty [26, 27]. There, the BNNs have two outputs. One for predicting the prediction, and one for the variance. The mean of the variances is then the aleatoric uncertainty, and the variance of the predictions is then epistemic uncertainty. This approach explicitly models aleatoric and epistemic as part of the model, which may give more favourable results.

*Directions for future research:* We showed that UQ did not work to reject the cross-subject samples with the most epistemic uncertainty. However, it may still be usable for deciding whether or not to make a prediction under noisy EEG, or for identifying a model well suited for a certain subject, or even for detecting off-task thoughts.

## CONCLUSION

Available Deep Learning methods that capture and disentangle epistemic uncertainty are not able to improve the robustness of within-subject nor cross-subject Motor Imagery BCIs in the context of a benchmark dataset. However, there are other contexts in BCIs where epistemic uncertainty may be expected. Off-task thoughts, rare artifacts, or insufficient training data can all introduce epistemic uncertainty, and the methods demonstrated here may be able to improve robustness in those cases. This has not yet been investigated.

We want to emphasise the need to study the behaviour and uses of uncertainty estimates from non-Deep Learning models. Classical Machine Learning models for classification often come with an adaptation to return class probabilities, but the behaviour of these may vary substantially. Assessing their ability to reject segments of EEG that are likely to be false positives may allow for more robust BCI systems. The robustness promised by good UQ may be a step towards making BCIs more usable outside of the lab.

## REFERENCES

- [1] M. Abdar *et al.*, “A review of uncertainty quantification in deep learning: Techniques, applications and challenges,” *Information Fusion*, vol. 76, pp. 243–297, 2021.
- [2] T. Duan, Z. Wang, S. Liu, Y. Yin, and S. N. Srihari, “Uncer: A framework for uncertainty estimation and reduction in neural decoding of eeg signals,” *Neurocomputing*, vol. 538, p. 126 210, 2023.

- [3] D. Milanés-Hermosilla *et al.*, “Monte carlo dropout for uncertainty estimation and motor imagery classification,” *Sensors*, vol. 21, no. 21, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/21/7241>.
- [4] E. I. Chetkin, S. L. Shishkin, and B. L. Kozyrskiy, “Bayesian opportunities for brain-computer interfaces: Enhancement of the existing classification algorithms and out-of-domain detection,” *Algorithms*, vol. 16, no. 9, p. 429, 2023.
- [5] D. Milanés-Hermosilla *et al.*, “Robust motor imagery tasks classification approach using bayesian neural network,” *Sensors*, vol. 23, no. 2, p. 703, 2023.
- [6] A. Campbell, L. Qendro, P. Liò, and C. Mascolo, “Robust and efficient uncertainty aware biosignal classification via early exit ensembles,” in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 3998–4002.
- [7] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On calibration of modern neural networks,” in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup and Y. W. Teh, Eds., ser. Proceedings of Machine Learning Research, vol. 70, PMLR, 2017, pp. 1321–1330.
- [8] I. P. de Jong, A. I. Sburlea, and M. Valdenegro-Toro, “Uncertainty quantification in machine learning for biosignal applications—a review,” *arXiv preprint arXiv:2312.09454*, 2023.
- [9] S. J. Prince, *Computer vision: models, learning, and inference*. Cambridge University Press, 2012.
- [10] Y. Gal and Z. Ghahramani, *Dropout as a bayesian approximation: Representing model uncertainty in deep learning*, 2016. arXiv: 1506 . 02142 [stat.ML].
- [11] A. Mobiny, H. V. Nguyen, S. Moulik, N. Garg, and C. C. Wu, *Dropconnect is effective in modeling uncertainty of bayesian deep networks*, 2019. arXiv: 1906.04569 [cs.LG].
- [12] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and scalable predictive uncertainty estimation using deep ensembles,” *Advances in neural information processing systems*, vol. 30, 2017.
- [13] Y. Wen, P. Vicol, J. Ba, D. Tran, and R. Grosse, “Flipout: Efficient pseudo-independent weight perturbations on mini-batches,” *arXiv preprint arXiv:1803.04386*, 2018.
- [14] J. Van Amersfoort, L. Smith, Y. W. Teh, and Y. Gal, “Uncertainty estimation using a single deep deterministic neural network,” in *International conference on machine learning*, PMLR, 2020, pp. 9690–9700.
- [15] M. Tangermann *et al.*, “Review of the bci competition iv,” *Frontiers in Neuroscience*, vol. 6, 2012.
- [16] R. T. Schirrmeister *et al.*, “Deep learning with convolutional neural networks for eeg decoding and visualization,” *Human Brain Mapping*, 2017. [Online]. Available: <http://dx.doi.org/10.1002/hbm.23730>.
- [17] A. Gramfort *et al.*, “MEG and EEG data analysis with MNE-Python,” *Frontiers in Neuroscience*, vol. 7, p. 267, 2013.
- [18] N. Tibrewal, N. Leeuwis, and M. Alimardani, “The promise of deep learning for bcis: Classification of motor imagery eeg using convolutional neural network,” *bioRxiv*, 2021. eprint: <https://www.biorxiv.org/content/early/2021/06/18/2021.06.18.448960.full.pdf>.
- [19] A. Zancanaro, G. Cisotto, J. R. Paulo, G. Pires, and U. J. Nunes, “Cnn-based approaches for cross-subject classification in motor imagery: From the state-of-the-art to dynamicnet,” in *2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2021, pp. 1–7.
- [20] F. Chollet *et al.*, *Keras*, <https://keras.io>, 2015.
- [21] M. Valdenegro, *Keras-uncertainty*, <https://github.com/mvaldenegro/keras-uncertainty>, 2023.
- [22] J. Mukhoti, A. Kirsch, J. van Amersfoort, P. H. Torr, and Y. Gal, “Deterministic neural networks with inductive biases capture epistemic and aleatoric uncertainty,” *arXiv preprint arXiv:2102.11582*, vol. 2, 2021.
- [23] L. Smith and Y. Gal, *Understanding measures of uncertainty for adversarial example detection*, 2018. arXiv: 1803.08533 [stat.ML].
- [24] X. Huang, J. Yang, L. Li, H. Deng, B. Ni, and Y. Xu, “Evaluating and boosting uncertainty quantification in classification,” *arXiv preprint arXiv:1909.06030*, 2019.
- [25] J. Mukhoti, A. Kirsch, J. van Amersfoort, P. H. Torr, and Y. Gal, “Deep deterministic uncertainty: A simple baseline,” *arXiv preprint arXiv:2102.11582*, 2021.
- [26] M. Valdenegro-Toro and D. S. Mori, “A deeper look into aleatoric and epistemic uncertainty disentanglement,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 2022, pp. 1508–1516.
- [27] A. Kendall and Y. Gal, “What uncertainties do we need in bayesian deep learning for computer vision?” *Advances in neural information processing systems*, vol. 30, 2017.

# TOWARDS RIEMANNIAN EEG CLASSIFIERS TO DETECT AWAKE AND ANESTHETIZED STATES USING MEDIAN NERVE STIMULATION

Valérie Marissens Cueva<sup>1,2</sup>, Sébastien Rimbart<sup>1</sup>, Ana Maria Cebolla Alvarez<sup>3</sup>, Mathieu Petieau<sup>3</sup>, Viktoriya Vitkova<sup>3</sup>, Iraj Hashemi<sup>3</sup>, Guy Cheron<sup>3</sup>, Claude Meistelman<sup>4</sup>, Philippe Guerci<sup>4</sup>, Denis Schmartz<sup>5</sup>, Seyed Javad Bidgoli<sup>6</sup>, Laurent Bougrain<sup>2</sup>, Fabien Lotte<sup>1</sup>

<sup>1</sup>Inria Center at Univ. Bordeaux / LaBRI, Talence, France; <sup>2</sup>Université de Lorraine, CNRS, LORIA, Nancy, France; <sup>3</sup>Laboratory of Neurophysiology and Movement Biomechanics, Université Libre de Bruxelles, Bruxelles, Belgium; <sup>4</sup>CHRU de Nancy, Vandoeuvre-lès-Nancy, France; <sup>5</sup>Hôpital Universitaire de Bruxelles - Université Libre de Bruxelles, Bruxelles, Belgium; <sup>6</sup>CHU Brugmann, Bruxelles, Belgium. E-mail: valerie.marissens@loria.fr

**ABSTRACT:** Among all the operations carried out under general anesthesia worldwide, some patients have had the terrible experience of Accidental Awareness during General Anesthesia (AAGA), an unexpected awakening during the surgical procedure. The inability to predict and prevent AAGA before its occurrence using only conventional measures, such as clinical signs, leads to the use of brain activity monitors. Given AAGA patients' first reflex to move, impeded by neuromuscular-blocking agents, we propose using a new Brain Computer Interface with Median Nerve Stimulation (MNS) to detect their movement intentions, specifically in the context of general anesthesia. Indeed, MNS induces movement-related EEG patterns, improving the detection of such intentions. In this article, we compared MNS effects on the motor cortex before and during surgery under general anesthesia. Then, a Riemannian Minimum Distance to the Mean classifier achieved 97% test balanced accuracy in distinguishing awake and anesthetized states. Additionally, we observed how the classifier's response evolves with anesthesia depth, in terms of distance to the awake class centroid. This distance appears to track the patients' awareness level during surgery. This holds promises for developing a future one-class classifier using only awake EEG data, as anesthesia EEG data are usually unavailable for classifier training, to detect AAGA.

## INTRODUCTION

Accidental Awareness during General Anesthesia (AAGA) is an unexpected awakening during surgery that can be a truly traumatic experience for the patients. It occurs in about 1% of high-risk interventions [1], although the incidence remains controversial as it is a subjective experience that may be underestimated in the absence of the necessary questionnaires to follow the patients [2]. During an AAGA, the patient may experience pain, and recall events related to the surgery, which can lead to potentially devastating psychological sequelae, such as Post-Traumatic Stress Disorder (PTSD)

[3]. The risk of AAGA is higher with Total Intravenous Anesthesia (TIVA), such as propofol, in comparison to volatile-based anesthesia [4]. In addition, the use of Neuromuscular-Blocking Agents (NMBAs) further increases this risk [2]. The first reaction to noxious stimulation when anesthesia depth is insufficient is the patient's movement, which may act as a potential detector of AAGA, but this response is suppressed by NMBAs, which paralyze the patient [5]. Besides, traditional clinical signs like hypertension, tachycardia and lacrimation are also unreliable indicators of anesthesia depth [5]. As a result, electroencephalography (EEG) has been used to monitor the depth of anesthesia [6], but awareness may still occur with current monitors [7, 8].

Since the first reflex of a patient experiencing AAGA is to attempt to move to prevent what is happening [2], using a Brain-Computer Interface (BCI) based on motor imagery could be relevant [9, 10]. Indeed, the power variations within the mu and beta frequency bands, called Event-Related Desynchronizations (ERD) and Event-Related Synchronizations (ERS) could be useful markers for detecting whether the patient is experiencing AAGA [11]. Interestingly, Median Nerve Stimulation (MNS) is a painless stimulation of the median nerve that generates a similar ERD/ERS pattern to the one induced by an intention to move (MI) [12]. Combining both MI and MNS has a significant impact on the patterns generated by the MNS, resulting in a better classification accuracy in MI detection [9, 13]. Also, MNS intrinsically provides a trigger to know when to analyze the signal [14], which leads to better classification (+18%) results than those obtained with asynchronous BCI [9]. The originality of this BCI paradigm is to exploit this MNS induced phenomenon to accurately detect the patient's motor intention during an AAGA. In preliminary results, we have shown that propofol sedation (at 0.5 µg/ml and 1 µg/ml) has no negative impact on the ability of an MNS-based BCI to detect movement intention. Concretely, at relatively low propofol concentrations, ERD/ERS patterns are still present in the sensorimotor cortex [15, 16]. However, it appears that

high doses of propofol strongly affect the oscillatory activity generated by the MNS [17], which makes it difficult to detect MI under anesthesia.

Current BCIs require a subject-specific calibration due to large between-subject variabilities. However, before a surgery, patients' EEG examples of MNS during anesthesia are not available. In the long term, we thus need to develop new machine learning tools that can detect MI under anesthesia without any EEG example from this patient. With this future objective in mind, in this paper, we first propose to use a Riemannian Minimum Distance to the Mean (MDM) classifier [18] to differentiate the EEG activities induced by the two types of stimulations: preoperative MNS when the patient is awake (MNS-awake), from intraoperative MNS when the patient is under different stages of general anesthesia (MNS-anesthesia). Our findings suggest that an MDM classifier is indeed capable of distinguishing between these two classes with high accuracy. Additionally, the distance to the MNS-awake class centroid (which does not require data under anesthesia) varies according to the concentration of propofol throughout the surgery. These results are promising for the future development of a one class MNS based-BCI that detects AAGA, as EEG covariance matrices appear to contain information related to the patients' level of awareness throughout the surgery under propofol.

## MATERIALS AND METHODS

**Participants:** 13 volunteers (7 females;  $50 \pm 7.39$  years old) were enrolled for surgery at the CHU Brugmann, Brussels, Belgium, and accepted to participate in this protocol. This study was approved by the ethical committee of the CHU Brugmann (CE 2021/225) and was registered at EUDRACT (2021-006457-56). The study protocol [14] was also registered on ClinicalTrials.gov (NCT05272202) and follows the principles of the Declaration of Helsinki and the Medical Research Involving Human Subjects Act [19]. Subjects 1, 2, and 5 were excluded either due to technical issues or because the surgery was canceled, resulting in only 10 subjects being included. 2 of them (subjects 10 and 11) were stimulated on the right median nerve, and the remaining 8 on the left. This was due to difficulties in placing the electrodes on the left median nerve and limited time in order not to delay the surgical intervention.

**Protocol:** The patients were equipped with a TMSi 64-channel EEG cap covering the entire scalp, and the signals collected through an eego mylab system (ANT Neuro) at 4096 Hz. MNSs consisted of square electrical pulses of 0.2 milliseconds of duration, and were generated by Micromed device SD Ltm Stim Energy and delivered through a pair of grass gold cup electrodes (cathode [-] placed proximally) to the right/left median nerve at the wrist [12]. The intensities were adjusted to elicit visible small thumb twitches, below 15 mA [14]. First, a preoperative EEG recording session (approximately 1 hour) was conducted by stimulating the median nerve of the awake patient during 1 or 2 runs. This was

followed by a second intraoperative session recorded during the entire surgery under general anesthesia; the duration and number of runs depended on the length of the surgery (from 3 to 10 runs, depending on the subject). One run consisted of 150 stimulations, spaced by 3 to 4 seconds. The anesthesia protocol was left at the anesthesiologist's discretion, except for the loss of consciousness, which was achieved using propofol with a Target-Controlled Infusion (TCI) pump with the Schnider pharmacokinetic model designed to predict propofol concentration at the effect-site [20]. If necessary, NMBA agents were used (patients n°4, 6, 7, 8, 11, 13). Data collection of the target propofol concentration administered to the patients was directly recorded alongside the EEG signals. After the surgery, sedation was discontinued, allowing the patient to recover and to be monitored afterward in the post-anesthesia care unit.

**Time-frequency EEG analyses:** Time-frequency analyses to identify the differences in MNS patterns between preoperative and intraoperative sessions were performed using the EEGLAB toolbox [21] and MATLAB R2023a (The MathWorks Inc). EEG signals have very low amplitudes and are thus susceptible to external interference. For example, the use of electrocautery during a surgery produces visible noise in the EEG signals [22]. In order to clean these electrocautery-related artifacts, all trials were visualized and those affected were rejected. Then, EEG signals were downsampled to 128 Hz and epoched into 4.5 s windows (1.5 s before and 3 s after MNS).

Because Event Related Spectral Perturbation (ERSP) time-frequency analyses are conducted at the group level, it is essential for the subjects to have homogeneity, meaning they should be stimulated in the same hand. Therefore, ERSP time-frequency analyses were averaged only across the 8 participants where the left median nerve was stimulated, the electrode of interest being C4 (Fig. 1).

**Classification:** Offline BCI performances for MNS-awake vs. MNS-anesthesia classification were analyzed to determine if the MNS pattern might be used to track the patient awareness level throughout surgery. Classifications were performed using the MNE [23], Scikit-learn [24] and pyRiemann [25] packages in Python 3.10. EEG signals, with trials affected by electrocautery-related artifacts rejected, were downsampled to 128 Hz, band-pass filtered (8-30 Hz), and epochs for MNS-awake and MNS-anesthesia were extracted from 250 to 1000 ms after the stimulation. Epochs do not start at time 0 ms to avoid MNS-induced electrical artifacts. All 64 electrodes were used for the classification, and since the algorithm calibration is subject-specific, all 10 subjects were included. A Riemannian MDM [18, 26] was used, as Riemannian classifiers are currently the state-of-the-art in EEG-based BCIs [27, 28]. Each EEG trial  $\mathbf{X}$  is represented by a covariance matrix  $\mathbf{P} \in \mathbb{R}^{n \times n}$ , with  $n$  the number of electrodes,  $s$  the number of sampled time points in each trial and the superscript  $T$  as matrix transposition:

$$\mathbf{P} = \frac{\mathbf{X}\mathbf{X}^T}{s-1} \quad (1)$$

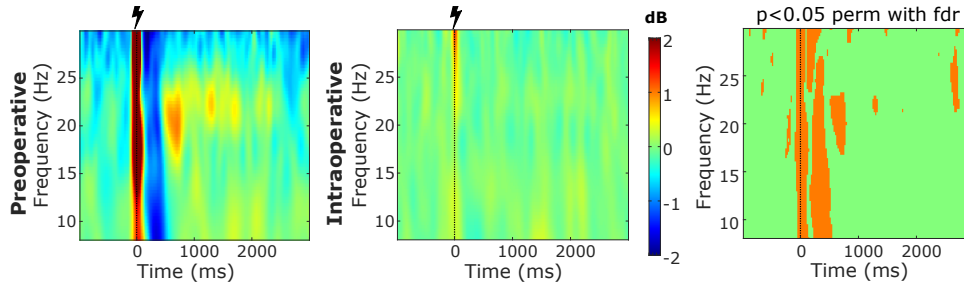


Figure 1: Grand average Event-Related Spectral Perturbation (ERSP) time-frequency analysis across 8 subjects for both conditions: preoperative (left figure) and intraoperative (central figure) median nerve stimulation, for electrode C4. The flash icon indicates the beginning of the stimulation. Statistical differences at a significance level of 0.05 are shown on the right figure. Red color corresponds to a strong ERS and blue to a strong ERD.

Such matrices are symmetric-positive definite matrices, that can be manipulated using Riemannian geometry, and compared using a dedicated Riemannian geodesic distance  $\delta_R$ :

$$\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \|\log(\mathbf{P}_1^{-1/2} \mathbf{P}_2 \mathbf{P}_1^{-1/2})\|_F \quad (2)$$

where  $\|\mathbf{A}\|_F$  is the Frobenius norm of a matrix  $\mathbf{A}$ . For algorithm training, the MDM algorithm first estimates the mean covariance matrix  $\mathbf{P}_G^k$  for each class  $k$  (here, MNS-awake or MNS-anesthesia). This is performed using the covariance matrices of the  $I \geq 0$  training EEG signals from class  $k$ , as follows:

$$\mathbf{P}_G^k = \arg \min_{\mathbf{P}^k} \sum_{i=1}^I \delta_R^2(\mathbf{P}^k, \mathbf{P}_i^k) \quad (3)$$

These two mean covariance matrices, one for each class, can then be used as class centroids. For algorithm testing, covariance matrices are also estimated for each trial. Thus, the class  $\hat{k}$  of an unseen EEG covariance matrix  $\mathbf{P}$  is determined based on the nearest centroid's class:

$$\hat{k} = \arg \min_k \delta_R(\mathbf{P}, \mathbf{P}_G^k) \quad (4)$$

The number of training and testing trials differs among subjects, depending on the quantity of clean trials recorded preoperatively or during their surgery. The classifier was trained with the first half of preoperative trials for the MNS-awake class (65 to 148 trials), and the first half of deep-anesthesia intraoperative trials for the MNS-anesthesia class (113 to 535 trials). It was then tested on the remaining trials (279 to 1192 trials). Standard cross-validation was not used to remain realistic and more similar to an actual online use, where no future data would be available for cross-validation. Employing this technique might lead to an overestimation of the accuracy [29].

## RESULTS

*Impact of general anesthesia on ERD/ERS induced by MNS:* During the preoperative condition (i.e., before the general anesthesia), the ERSP analysis (Fig. 1) revealed the usual EEG pattern associated with median nerve stimulation [9, 30]. In particular, immediately after the MNS, a powerful ERS appears between 0 and 250 ms, in the whole 8 to 30 Hz frequency band. For the remainder of this article, this very first ERS will be referred to as the post-stimulation rebound (PSR; see [9, 30]) and could be due partially to an electrical artifact. The PSR is followed

by an ERD period of approximately 500 ms in both the alpha and beta frequency bands (8-30 Hz). Finally, a post-movement beta rebound (PMBR) occurs in the beta frequency band (18-23 Hz) about 500 ms after stimulation and lasts 1 s. According to a permutation test comparing the two surgical conditions, for  $p \leq 0.05$  with a correction for false discovery rate for multiple comparisons, both ERD and PMBR seem to disappear significantly during the intraoperative condition.

*MNS-awake vs MNS-anesthesia classification:* The MDM algorithm is able to correctly distinguish the preoperative MNS pattern (MNS-awake) from the intraoperative pattern (MNS-anesthesia), with an average test set balanced accuracy of 97% (Fig. 2). The test set balanced accuracy for each subject is never below 85%, which is not surprising given the clear difference in the MNS pattern between these two sessions (Fig. 1).

*Distances to the MNS-awake class centroid:* The distances to the MNS-awake class centroid of the test trials of subjects S6, S8 and S9 over the whole experiment are shown in Fig. 3. When a trial is classified as MNS-awake by the MDM classifier, a bar is presented below it, beneath the graph. The corresponding test set balanced accuracies of subjects 6, 8 and 9 are 96%, 98% and 86%, respectively, with the latter being the lowest accuracy among the 10 subjects. For S6 and S8, the MDM classifier correctly identified the patients as awake during the preoperative session and at the beginning of the intraoperative session. For S6, this prediction persisted until the propofol concentration reached approximately 1.5  $\mu\text{g/ml}$ , and for S8 until it reached 3  $\mu\text{g/ml}$ . At the end of the intraoperative session, S8 was also identified as awake for a few trials. As for S9, the preoperative session was accurately labeled. Towards the end of the surgery, with propofol below 1.5  $\mu\text{g/ml}$ , the subject was identified as awake, which is consistent with reality, as the patient was already responding. However, the classifier associated some trials that were part of the induction (with propofol concentration between 2 and 8  $\mu\text{g/ml}$ ), as part of the MNS-awake class. The evolution of the distance is congruent with anesthetic concentration evolution, suggesting that covariance matrices might indeed reflect the progression of the MNS pattern throughout the surgery.

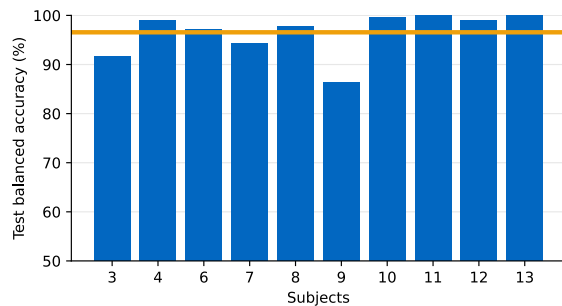


Figure 2: Riemannian MDM test set balanced accuracy of the 10 subjects, for MNS-awake vs MNS-anesthesia classification, where the EEG signals were filtered in the  $\mu$ +beta band (8-30 Hz), and all 64 electrodes were used. The yellow line indicates the average test set balanced accuracy (97%).

## DISCUSSION

According to these results, a Riemannian MDM is indeed capable of distinguishing the MNS pattern of an awake patient from the MNS pattern when the same patient is under propofol. This distinction varies throughout the surgery, according to the propofol concentration. The classifier accurately identifies the patient as awake during the preoperative session, at the beginning of the induction and at the end of the emergence phases. Furthermore, trials of deep anesthesia are farther away from the MNS-awake class centroid than trials corresponding to induction or emergence. We will suggest a few hypotheses that may explain the changes in MNS patterns between the two sessions and how to make the algorithm more robust. Additionally, we will discuss the possibility of evolving this algorithm into a one-class model, as well as future analyses to be explored.

**ERD/ERS differences between preoperative and intra-operative conditions:** As mentioned, sensorimotor modulations after stimulation are strongly modulated by general anesthesia. First, the amplitude of the PSR (i.e., ERS following the MNS) completely disappears in the presence of propofol. This is also the case for the post-MNS ERD and for the PMBR (Fig. 1). Our previous results had shown that with light propofol sedation (concentrations below 1.5  $\mu\text{g/ml}$ ), ERD/ERS patterns were still present in the sensorimotor cortex [16]. However, at deepest concentrations, propofol decreases excitatory inputs from the thalamus to the cortex [6], leading to a decrease in the metabolic activity of the central nervous system, thus making the ERD/ERS disappear after MNS.

In addition to the effects of general anesthesia, other factors could explain the modulations of ERD/ERS. For example, this could be attributed to a change in skin conductance, for instance due to stress. Skin conductance increases in response to stress [31], as is the case right before the intervention, and decreases during the surgery when the patient is no longer conscious. This remains a hypothesis to be further analyzed. It might also result from the change in environment between the two sessions, with the preoperative session conducted in a different room than the operating one because of the high cost

of the operating rooms. Also, the position of the patient's arm was not exactly the same in the two sessions.

**Robustness of the classification:** The MDM classifier is indeed capable of correctly identifying the periods when the patient is awake (preoperative session, beginning of induction and end of emergence). It is important to note that the MNS-anesthesia class was trained with trials under deep anesthesia, which are not easily obtained under clinical conditions. Moreover, for certain subjects such as S9, it mistakenly classifies some trials where the patient is at a propofol concentration of 2 to 6  $\mu\text{g/ml}$  as MNS-awake. During this period, some external elements occurred that could explain why these covariance matrices were closer to the centroid of the MNS-awake class rather than to the one associated with MNS-anesthesia. For example, the patient raised an arm, MNS electrodes were repositioned, the patient was intubated, and the medical team adjusted the patient's position on the operating table. Even though the biggest artifacts were rejected from the EEG, the signals remained considerably noisy. Furthermore, there is minimal difference between S9 and the other subjects, except for its age of 75, compared to an average of 50. Age-related variations in the effects of anesthesia on the EEG have already been shown to impact the effectiveness of EEG-based monitors [32]. Additionally, S9 had notably fewer training trials (77 for MNS-awake and 137 for MNS-anesthesia) compared to S6 and S8, which had 2 to 4 times more trials, making them the subjects with the highest number of training trials. Despite this, Spearman's correlation revealed no significant relationship between the number of training trials and accuracies. Thus, while S9's lower accuracy compared to S6 and S8 may be due to its fewer training trials, this explanation cannot be generalized across all subjects. Burst suppression, a high-voltage activity alternating with isoelectric flat EEG [33], is related to deep levels of general anesthesia [6, 34]. Thus, current depth of anesthesia monitors have a burst suppression sub-variable to avoid a paradoxical increase [5]. In our algorithm, we have not yet taken into account this paradoxical increase in amplitude.

**One-class classifier:** In a real-world application, the MNS-anesthesia class trained with deep anesthesia trials will not be available to calibrate the BCI. Therefore, a one-class approach [35] calibrated only with MNS-awake trials will be required. Given that the distance to class MNS-awake appears to evolve logically in relation to the anesthesia concentration evolution, one could imagine a one-class method that only takes into account preoperative data to compute an MNS-awake centroid using a Riemannian distance, as in the MDM algorithm. A threshold would then be defined, beyond which the trials of the testing set no longer match, meaning the MNS pattern corresponds to an anesthetized patient. We will explore this approach in future works.

**Future analyses:** To detect the propofol concentration at which the MNS patterns are no longer visible, we will further analyze an average time-frequency across subjects

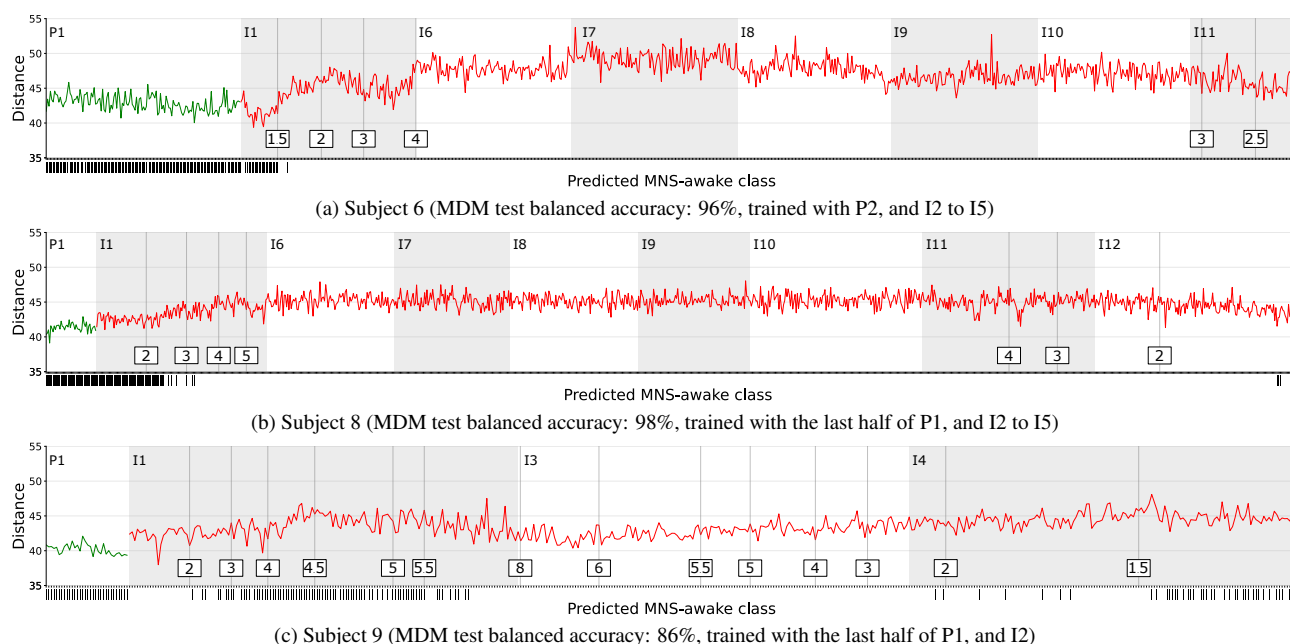


Figure 3: Distances of each test trial to the MNS-awake centroid, for subjects 6, 8 and 9. The real labels are presented by colors: green for the preoperative session (MNS-awake) and red for the intraoperative session (MNS-anesthesia). Below, a bar indicates when the classifier predicted MNS-awake. When propofol reaches the target concentration, a gray line with a label indicating the corresponding concentration (in  $\mu\text{g/ml}$ ) is displayed. Each run is differentiated by alternating background colors (white, gray), and labeled as preoperative (P) or intraoperative (I) runs.

for each concentration. This will allow us to observe the pattern evolution. An extension of this MNS-based BCI will also be explored by integrating other EEG features, still visible under deep levels of propofol. Some of these features might be the signal entropy, already used in some monitors [36], functional connectivity or somatosensory evoked potentials. Further analyses will also aim to identify the specific frequency band and the most relevant electrodes to correctly detect the changes in the MNS pattern when a patient is anesthetized. In order to validate this MNS-based BCI, a protocol will be carefully conceived to try to simulate an AAGA and observe if the BCI will indeed be able to detect it.

## CONCLUSION

In this paper, we evaluated the feasibility of a new MNS-based BCI to detect intraoperative awareness during general anesthesia by tracking the EEG pattern in the motor cortex associated to MNS during the surgery. We compared this MNS pattern when the subject is awake (MNS-awake) with the MNS pattern when the same patient is undergoing a surgical procedure at different concentrations of propofol (MNS-anesthesia). The two patterns are indeed very different from each other, more particularly, the ERD and PMBR present after the stimulation seem to disappear at deeper concentrations of propofol. A Riemannian MDM was used to differentiate the MNS-awake and MNS-anesthesia classes. The average test balanced accuracy was of 97%, which was expected considering how different the two patterns are. The evolution of the classifier was further analyzed, by tracking the distance between the centroid of the MNS-awake class and covariance matrices of other trials throughout the surgery. This

distance is greater during the maintenance phase, under deeper concentrations of anesthesia, compared to the induction or emergence phases. This indicates that covariance matrices associated with the MNS pattern seem to evolve consistently with the patient's level of awareness. Hence, a one-class approach only based on MNS-awake trials, utilizing this distance to detect when the patient is awake during the surgery might be developed. Such a one-class method will be necessary, as the MNS-anesthesia trials will not be available to calibrate the BCI.

## ACKNOWLEDGMENTS

The authors thank the Brugmann Foundation, the National Funds of Research of Belgium, the French National Research Agency (ANR-22-CE19-0016; project BCI4IA) and all the participants of the study.

## REFERENCES

- [1] Tasbihgou SR, Vogels MF, Absalom AR. Accidental awareness during general anaesthesia – a narrative review. *Anaesthesia*. 2018;73(1):112–122.
- [2] Pandit J *et al.* 5th national audit project (nap5) on accidental awareness during general anaesthesia: Summary of main findings and risk factors. *Br. J. Anaesth.* 2014;113(4):549–559.
- [3] Leslie K, Chan MTV, Myles PS, Forbes A, McCulloch TJ. Posttraumatic stress disorder in aware patients from the b-aware trial. *Anesth. Analg.* 2010;110(3):823–828.
- [4] Nimmo AF *et al.* Guidelines for the safe practice of total intravenous anaesthesia (tiva). *Anaesthesia*. 2019;74(2):211–224.

- [5] Bruhn J, Myles PS, Sneyd R, Struys MMRF. Depth of anaesthesia monitoring: what's available, what's validated and what's next? *Br. J. Anaesth.* 2006;97(1):85–94.
- [6] Purdon PL, Sampson A, Pavone KJ, Brown EN. Clinical Electroencephalography for Anesthesiologists: Part I: Background and Basic Signatures. *Anesthesiology*. 2015;123(4):937–960.
- [7] Myles P, Leslie K, Mcneil J, Forbes A, Chan M. Bispectral index monitoring to prevent awareness during anaesthesia: The b-aware randomised controlled trial. *Lancet*. 2004;363:1757–63.
- [8] Sebel P *et al.* The incidence of awareness during anesthesia: A multicenter united states study. *Anesth. Analg.* 2004;99:833–9, table of contents.
- [9] Rimbart S, Riff P, Gayraud N, Schmartz D, Bougrain L. Median nerve stimulation based bci: A new approach to detect intraoperative awareness during general anesthesia. *Front. Neurosci.* 2019;13.
- [10] Blokland Y *et al.* Decoding motor responses from the EEG during altered states of consciousness induced by propofol. *J. Neural Eng.* 2016;13(2):026014.
- [11] Pfurtscheller G, Lopes da Silva F. Event-related eeg/meg synchronization and desynchronization: Basic principles. *Clin. Neurophys.* 1999;110(11):1842–1857.
- [12] Schnitzler A, Salenius S, Salmelin R, Jousmäki V, Hari R. Involvement of primary motor cortex in motor imagery: A neuromagnetic study. *NeuroImage*. 1997;6(3):201–208.
- [13] Avilov O, Rimbart S, Popov A, Bougrain L. Optimizing motor intention detection with deep learning: Towards management of intraoperative awareness. *IEEE Trans. Biomed. Eng.* 2021;68(10):3087–3097.
- [14] Rimbart S *et al.* Detection of motor cerebral activity after median nerve stimulation during general anesthesia (stim-motana): Protocol for a prospective interventional study. *JMIR Res. Protoc.* 2023;12:e43870.
- [15] Guerci P, Schmartz D, Losser MR, Bougrain L, Meistelman C, Rimbart S. Effet d'une sédation légère au propofol sur les oscillations cérébrales du cortex moteur. In: *SFAR*. 2021.
- [16] Rimbart S, Guerci P, Gayraud N, Meistelman C, Bougrain L. Innovative bci based on motor cortex activity to detect accidental awareness during general anesthesia. In: *IEEE SMC*. 2019, 3931–3936.
- [17] Rimbart S *et al.* Detection of Cerebral Electroencephalographic Patterns After Median Nerve Stimulation During Propofol-Induced General Anesthesia : a Prospective Interventional Cohort Study. *Anesth. Ann. Mtg. Poster. ASA*, 2023.
- [18] Congedo M, Coelho Rodrigues PL, Jutten C. The Riemannian Minimum Distance to Means Field Classifier. In: *8th Intern. BCI Conf. Graz, Austria, Sep.* 2019.
- [19] Association WM. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*. 2013;310(20):2191–2194.
- [20] Schnider T *et al.* The Influence of Method of Administration and Covariates on the Pharmacokinetics of Propofol in Adult Volunteers. *Anesthesiology*. 1998;88(5):1170–1182.
- [21] Delorme A, Makeig S. Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J Neurosci. Meth.* 2004;134(1):9–21.
- [22] Hight D, Kaiser H, Sleight J, Avidan M. Correction to: An updated introduction to electroencephalogram-based brain monitoring during intended general anesthesia. *Can. J. Anesth.* 2021;68.
- [23] Gramfort A *et al.* Meg and eeg data analysis with mne-python. *Front. Neurosci.* 2013;7:267.
- [24] Pedregosa F *et al.* Scikit-learn: Machine learning in Python. *JMLR*. 2011;12:2825–2830.
- [25] Barachant A *et al.* *Pyriemann/pyriemann: V0.5*. Version v0.5. 2023. [Online]. Available: <https://doi.org/10.5281/zenodo.8059038>.
- [26] Barachant A, Bonnet S, Congedo M, Jutten C. Multiclass Brain-Computer Interface Classification by Riemannian Geometry. *IEEE Trans. Biomed. Eng.* 2012;59(4):920–928.
- [27] Chowdhury A, Andreu-Perez J. Clinical brain-computer interface challenge 2020 (cbcic at wcci2020): Overview, methods and results. *IEEE Trans. Med. Robot. Bionics*. 2021;3(3):661–670.
- [28] Roy R *et al.* Retrospective on the first passive brain-computer interface competition on cross-session workload estimation. *Front. Neuroergon.* 2022;3.
- [29] Thomas E, Dyson M, Clerc M. An analysis of performance evaluation for motor-imagery based bci. *J. Neural Eng.* 2013;10(3):031001.
- [30] Rimbart S, Marissens Cueva V, Bougrain L, Lotte F. Prediction of Motor-Imagery-BCI performance using Median Nerve Stimulation. In: *12th IEEE Intern. Winter Conf. on BCI*. 2024.
- [31] Jacobs SC *et al.* Use of skin conductance changes during mental stress testing as an index of autonomic arousal in cardiovascular research. *Am. Heart J.* 1994;128(6):1170–1177.
- [32] Purdon P *et al.* The ageing brain: Age-dependent changes in the electroencephalogram during propofol and sevoflurane general anaesthesia. *Br. J. Anaesth.* 2015;115:i46–i57.
- [33] Pawar N, Barreto Chang OL. Burst suppression during general anesthesia and postoperative outcomes: Mini review. *Front. in Syst. Neurosci.* 2022;15.
- [34] Ching S, Purdon PL, Vijayan S, Kopell NJ, Brown EN. A neurophysiological-metabolic model for burst suppression. *PNAS*. 2012;109(8):3095–3100.
- [35] Khan SS, Madden MG. One-class classification: Taxonomy of study and review of techniques. *KER*. 2014;29(3):345–374.
- [36] Viertiö-Oja H *et al.* Description of the entropy<sup>TM</sup> algorithm as applied in the datex-ohmeda s/5<sup>TM</sup> entropy module. *Acta Anaesthesiol. Scand.* 2004;48(2):154–161.

# NEURONAL AVALANCHES FOR EEG-BASED MOTOR IMAGERY BCI

C. Mannino<sup>1</sup>, P. Sorrentino<sup>2,3</sup>, M. Chavez<sup>1</sup>, M.-C. Corsi<sup>1</sup>

<sup>1</sup>Sorbonne Université, Institut du Cerveau – Paris Brain Institute -ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

<sup>2</sup>Institut de Neurosciences des Systèmes, Aix-Marseille Université, 13005 Marseille, France

<sup>3</sup>University of Sassari, Department of Biomedical Sciences, Viale San Pietro, 07100, Sassari, Italy

E-mail: marie-constance.corsi@inria.fr

## ABSTRACT:

Current features used in motor imagery-based Brain-Computer Interfaces (BCI) rely on local measurements that miss the interactions among brain areas. Such interactions can manifest as bursts of activations, called neuronal avalanches. To track their spreading, we used the avalanche transition matrix (ATM), which contains the probability that an avalanche would consecutively recruit any two brain regions. Here, we proposed to use ATMs as a potential alternative feature. We compared the classification performance resulting from ATMs to a benchmark model based on Common Spatial Patterns.

In both sensor- and source-spaces, our pipeline yielded an improvement of the classification performance associated with reduced inter-subject variability. A correspondence between the selected features with the elements of the ATMs that showed a significant condition effect led to higher classification performance, which speaks to the interpretability of our findings.

In conclusion, working in the sensor space provides enough spatial resolution to classify. However the source space is crucial to precisely assess the involvement of individual regions.

## INTRODUCTION

Neuroscientists have been exploring and researching Brain-Computer Interface (BCI) since the 70s as a way to restore communication and motor capabilities for severely disabled people., such as patients affected by amyotrophic lateral sclerosis, stroke, or spinal cord injury [1].

In non-invasive BCI, Event Related Desynchronization/Synchronization, Event Related Potentials, and Steady State Evoked Potentials are the most informative brain activity patterns for communication and control applications to design electroencephalography (EEG)-based BCI. One of the main drawbacks of the current systems lies in the high inter/intra-subject variability, notably in terms of performance. Indeed, multiple studies reported that 15%–30% of the subjects fail in controlling a BCI device. This is a phenomenon referred to as the “BCI inefficiency” [2]. Among the potential causes are the selected data features. Indeed, relying mostly on local measurements might not effectively capture brain functioning, as some information is encoded in the interactions between areas [3].

To overcome these limitations and to take advantage of the EEG time-resolution, in a recent work, we proposed

to use a metric that captures the dynamic nature (i.e. changing in space and time) of the brain activities: the neuronal avalanches. Neuronal avalanches are characterised by the propagation of cascading bursts of activity [4]. Previous studies show that their spreading preferentially across the white-matter bundles [5] and that neuronal cascades are a major determinant of spontaneous fluctuations in brain dynamics at rest [6]. Furthermore, in our previous work we showed that neuronal avalanches, estimated from source-reconstructed data, spread differently according to the task performed by the user, demonstrating the potential relevance of neuronal avalanches as an alternative feature for detecting the subjects' intent [7].

Here, we investigated to which extent this framework would be compatible with a BCI experiment. For this purpose, instead of working in the source domain that requires additional data (e.g. individual magnetic resonance imaging) and computational resources, we tested the performance of neuronal avalanches directly in the sensor domain. Indeed, the methodological validity of sensor space measures is especially relevant for online studies in a clinical setting due to time and economic constraints. We hypothesised that despite a reduction of the spatial resolution, using the neuronal avalanches in the sensor space could help in classification performance, and that the selected features could be neurophysiologically interpretable and relevant.

## MATERIALS AND METHODS

### *Participants*

The research was conducted in accordance with the Declaration of Helsinki. A written informed consent was obtained from subjects after explanation of the study, which was approved by the ethical committee CPP-IDF-VI of Paris. All participants received financial compensation at the end of their participation. Twenty healthy subjects ( $27.5 \pm 4.0$  years old, 12 men), with no medical or psychological disorder, were recruited.

### *Experimental protocol*

The dataset used in our study originates from Corsi et al. [8] and involves a BCI task structured around a two-target box task. Participants were required to adjust their brain's alpha and/or beta activity levels to control a cursor's vertical movement, aiming to reach a vertical bar, referred to as a target displayed on the screen. Achieving the upper target necessitated the subjects to engage in continuous motor imagery (MI) of right-hand

grasping. Conversely, reaching the lower target required the subjects to remain in a resting state. Each session comprised 32 trials, evenly and randomly split between the up and down targets, correlating with the MI and Rest conditions, respectively. For a complete description of the protocol, the reader can refer to [8].

#### *EEG data acquisition & pre-processing*

EEG data were captured using a 74-channel EEG system equipped with Ag/AgCl passive sensors (Easycap, Germany), arranged according to the 10-10 standard montage. Reference electrodes were placed on the mastoids, with the ground electrode on the left scapula. Recordings took place in a magnetically shielded room, utilising a 0.01-300Hz bandwidth and sampling at 1kHz. Two channels (namely T9 and T10) were identified as bad and rejected based on the amplitude of the signals recorded, with a threshold of three standard deviations. For a complete description of the pre-processing steps, please refer to [8].

#### *Neuronal Avalanches extraction*

The neuronal avalanches analysis consists of identifying large signal excursions beyond a given threshold. The cascades are captured by clustering these discrete supra-threshold events based on temporal proximity, thus, defining neuronal avalanches as periods of collective spatio-temporal organization. Each signal was z-scored (over time), and set to 1 when above a threshold, and to 0 otherwise. An avalanche was defined as starting when at least one channel is above threshold (referred here as active channel), and as finishing when all channels were inactive [4,5,6]. For each avalanche, we estimated a transition matrix  $A$ , called Avalanche Transition Matrix (ATM), structured with channels in rows and columns, and the  $ij^{\text{th}}$  element of matrix  $A$  defined as the probability that the electrode  $j$  would be active at time  $t+1$ , given the electrode  $i$  was active at time  $t$ . For each subject, we obtained a transition matrix over all avalanches for each condition (MI and Rest conditions).

#### *Classification Analysis*

To explore the applicability of the ATM method in the context of a BCI training, we performed a subject-specific analysis.

The classification step was done using a Support Vector Machine (SVM). To assess the extent to which the ATMs might be considered as an alternative feature for BCIs, we compared our approach (ATM+SVM) to a framework that relies on spatial filters, namely Common Spatial Patterns (CSP+SVM) [9, 10].

For each approach (namely ATM+SVM or CSP+SVM), we classified different tasks at the individual level. To evaluate the classification performance, we divided the dataset to include 80% of the trials in the train split and 20% of the trials in the test split. The classification scores for all pipelines were evaluated with an accuracy measurement using a random permutation cross-validator. To assess the robustness of our framework, we also tested a different number of re-shuffling & splitting iterations (5/25/50/75).

For each subject, the CSP method decomposes signals using spatial filters, and then selects the  $n$  modes that capture the highest inter-class variance. Here, we selected eight spatial modes and returned the average power of each.

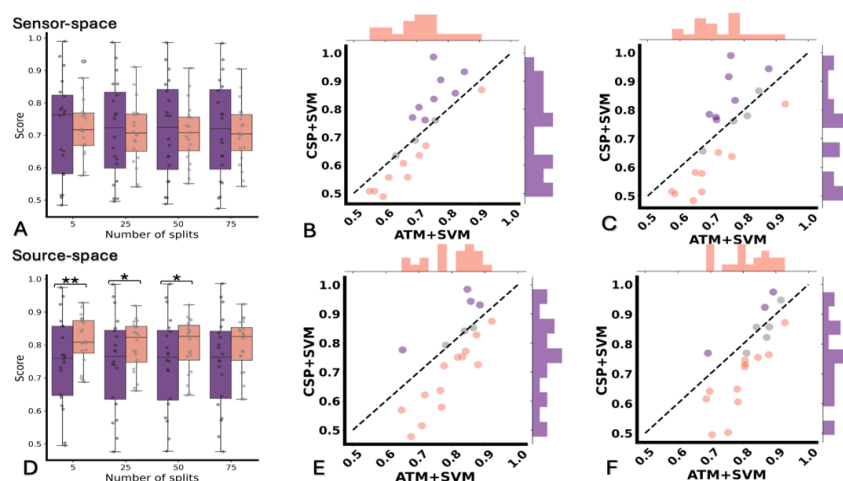
As for the ATMs, to consider the subjects' specificity, we optimised two parameters, namely: the threshold applied to the z-scored signals (ranging from 1.0 to 4.0), and the minimal duration of the considered avalanches (ranging from 2 to 8) [11]. Inside the ATM pipeline the choice of the best decoding parameters relied on a posteriori classification accuracy performance rate.

Finally, we individually compared the classification performance obtained with the CSP+SVM and with the ATM+SVM approaches, respectively. We run t-tests under the null hypothesis that, for a given subject, CSP+SVM and ATM+SVM would not yield statistically significant differences in classification. We repeated the comparison for all the subjects and corrected these statistical comparisons for multiple comparisons across subjects using the False Discovery Rate (FDR) [12]. Such an analysis has been performed across 25, 50, and 75 splits. However, given its poor statistical power, it is not possible to apply a statistical test over 5 splits classification. Therefore, to evaluate whether the difference between the two pipelines could be considered as significant, we calculated the averaged classification performance across splits using both CSP+SVM and ATM+SVM and we determined the difference for each subject. Ultimately, we compared the absolute value of the difference with our predefined threshold, considering the classification performance not statistically different if the magnitude of the difference between the two methods was less than a threshold. We established the threshold at an arbitrary value of 0.05. As a sanity check, we performed this analysis in the source space, as we did in [7]. For a complete description of the source-reconstruction steps, the reader can refer to [8].

In this work, we used preprocessed signals that were bandpass filtered between 3 and 40Hz. To investigate the potential effect induced by the choice of the frequency band, we performed the same analysis in the  $\mu$  band (8 – 13 Hz) and in the beta band (13 – 30 Hz) [not shown]. We performed a one way ANOVA ( $df = 2$ ) among these three-frequency bands under the null hypothesis ( $H_0$ ) that these groups came from the same population. For both approaches (CSP+SVM and ATM+SVM), no frequency band effect on the classification performance was observed ( $p\text{-value} > 0.05$ ). Therefore, in the next sections, we will report the results were obtained within the 3-40 Hz band.

#### *Decoding: Features importance analysis*

To investigate the interpretability of the classification performance, we examined the relative importance of the features derived from the absolute values of the classification coefficients in the model. To better understand the features importance across subjects, we carried out a quantitative reliability analysis across the cohort to identify the repetition of the selected features in



**Figure 1: Classification performance.** (A/D) Effect of splits tested on ATM+SVM and CSP+SVM at group level in sensor-space (A) and source-space (D). (B/E) Individual level classification performance in sensor-space (B) and source-space (E) using 50 random splits. (D/F) Individual level classification performance in sensor-space (D) and source-space (F) using 5 random splits. Color coded: in salmon, ATM+SVM pipeline; in violet, CSP+SVM pipeline. Statistical difference between CSP + SVM & ATM + SVM: \*  $pval < 0.05$ , \*\*  $pval < 0.01$ .

at least half of the subjects. To investigate features importance from a nodal point of view we set as a threshold the median value across channels and subjects, then we evaluated which nodes were over threshold in the majority of the subjects. We computed this analysis over the entire dataset (20 subjects) but also independently on two different subgroups: on the 10 most responsive subjects according to ATM classification performance and on the 10 least responsive subjects respectively. All these investigations were also performed in the source space.

#### Encoding: Quantification and statistical analysis

To identify the edges (i.e. functional links) that are more likely to be recruited during a hand motor imagery task as compared to resting state, for each participant, we calculated the variance in the probability of perturbations traversing a specific edge between resting state and MI task. To assess the statistical significance, we randomized the labels of individual avalanches for each person. This shuffling was repeated 10,000 times to generate a distribution of differences for each edge under the null hypothesis that the transition matrices revealed no distinction between the two conditions. We then determined statistical significance for each edge against this null distribution, applying Benjamini-Hochberg correction for multiple comparisons across edges. This process yielded a matrix for each subject, highlighting  $S_{ij}$  values (here referred to as edges) with statistically significant differences between conditions. We assessed the consistency of these matrices across individuals, concentrating on edges consistently implicated in the task. Then, we performed a node-wise analysis to identify the nodes over which significant differences were clustered. These nodes were referred to as “task-specific” areas.

## RESULTS

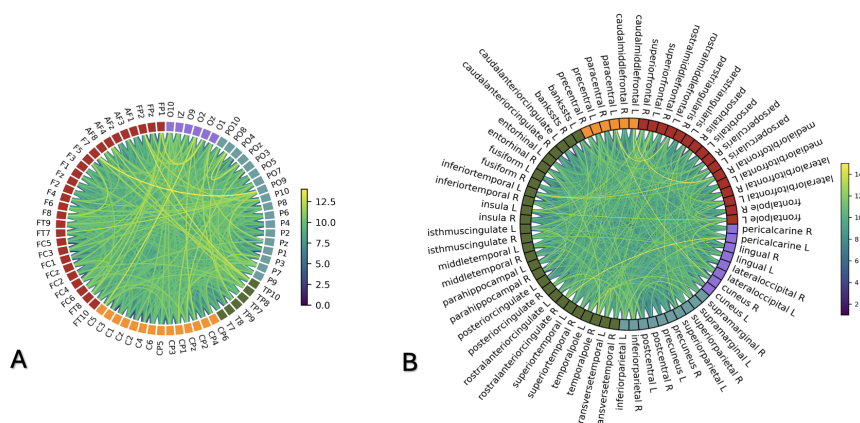
### Classification performance

Working on the entire dataset of 20 subjects, as a

standard configuration, we used 50 random splits.

At the group-level, the classification performance in the sensor space, between CSP+SVM and ATM+SVM is similar (t-test,  $pval > 0.05$ ). Nevertheless, we observed a larger inter-subject variability with CSP+SVM (71% $\pm$ 15%) as compared to ATM+SVM (71%  $\pm$  9%). In the source-space, ATM+SVM (80% $\pm$ 8%) led to a statistical improvement of the classification performance as compared to CSP+SVM (75% $\pm$ 14%) (t-test,  $pval < 0.05$ ) such as a reduced inter-subject variability. At the individual level, in the sensor-space ATM+SVM yielded a statistically better classification accuracy than CSP+SVM for 9 subjects. In 8 subjects, CSPs yielded better accuracy than ATMs. In 3 subjects, there was not any statistically significant difference between the two approaches (Fig. 1B). In the source-space, ATM+SVM yielded significantly higher classification accuracy than CSP+SVM for 13 subjects, while the opposite was true for 4 subjects. For the remaining 3 subjects, there was not any statistically significant difference between the decoding performances of the two approaches (Fig. 1E).

To investigate the possibility to reduce the computational time to get closer to a configuration more compatible with the online requirements, we investigated the accuracy performance across different random splits configurations (5, 25, 50, 75) both at the individual and at the group level. As shown in Fig. 1A & D the performance was robust across splits for both CSP+SVM and ATM+SVM pipelines (one-way ANOVA  $p > 0.05$ ), and we observed a higher accuracy score for most of the subjects with 5 splits both in sensor and source space. Based on the observations made on the inter-subject variability, we validated the significant difference of the variance of these two populations via the F-test ( $pval < 0.05$ ). The statistical difference between the two pipelines was achieved both in the sensor and in the source space.



**Figure 2: Features selection.**

(A) Edges-wise, valid at group level in sensor-space

(B) Edges-wise, valid at group level in source-space

When considering 5 splits, at the group level, in the sensor space, no significant difference was observed between the two pipelines (t-test,  $p$ -value  $> 0.05$ ) but CSP+SVM (72%  $\pm$  15.55%) showed a larger inter-subject as compared to ATM + SVM (73%  $\pm$  9.14%). In the source space, ATM + SVM (81%  $\pm$  7.5%) led to a statistical improvement of the performance (t-test,  $p$ -value  $< 0.01$ ) as compared to CSP + SVM (75%  $\pm$  14%) and a significant reduction of inter-subject variability. At the individual level: in the sensor-space (Fig. 1C), with 5 splits, ATM+SVM yielded a statistically better classification accuracy than CSP+SVM for 9 subjects. In 7 subjects, CSPs yielded better accuracy than ATMs and in four subjects, there was not any significant difference between the two approaches. However, CSP+SVM pipeline led to a larger number of subjects with a performance below the chance level, set to 58% here [13] (6 subjects) than with ATM+SVM (1 subject). In the source space, ATM+SVM showed an improved performance in 12 subjects, while the opposite was true for 3 subjects with CSP+SVM (Fig. 1F).

From now on, unless specified otherwise, the chosen configuration will involve 5 splits to closely mimic a real-time setup, and the subsequent sections will deal with ATM data only.

#### Sensor and source space selected features

To investigate the interpretability of the decoding performance, we estimated the weights attributed to each feature. A preliminary probabilistic analysis showed that most of the selected features presented a lower feature importance and that only a few were notably higher, suggesting that only a reduced number of features were relevant. When considering the features selected in at

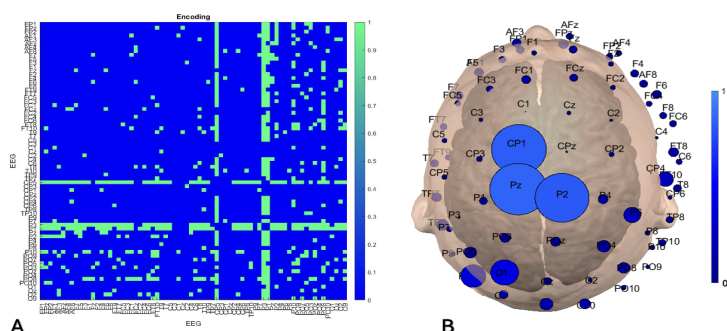
least half of the cohort, an edge involving left central electrodes (C5) and occipital electrodes (O2) was obtained in 13 subjects (Fig. 2A). We observed a predominant involvement of left central electrodes connected to occipital electrodes, between left and right central electrodes connected to parietal electrodes. Similar observations were possible in the source-space. Looking for a recurrent path across most of the subjects, see Fig 2B, in 15 subjects, most of the connections involved left paracentral, rostral anterior cingulate cortex, caudal middle frontal gyrus and medial lateral orbito-frontal regions.

These interactions correspond to edge clusters that were task-dependent and consistent across subjects in encoding investigation shown in our previous paper [7]. Moreover, the features with higher weight often involved the left paracentral and the precentral areas.

To get a more synthetic vision of these results, we performed a similar analysis at the nodal level, confirming the results previously obtained. To increase the statistical validity of such observations, in this part, we worked with the 50-split configuration. In the sensor-space, the electrodes with highest features' importance were C5 and P8. Nevertheless, it is possible to observe a general activation in electrodes over the bilateral motor cortex, and the bilateral parietal lobe. In the source-space, the most frequently selected brain regions were the right paracentral area, the left frontal pole and the right rostral anterior cingulate.

#### Encoding-Decoding Match in sensor-space

To investigate the neurophysiological validity of the selected features, we compared them with the results obtained with an encoding framework. To achieve this,



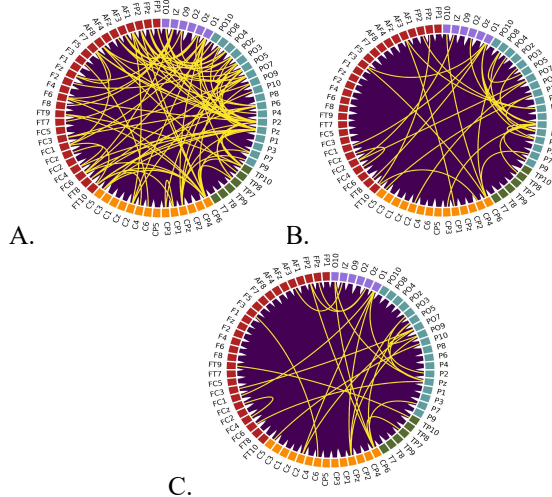
**Figure 3: Encoding analysis in sensor-space.**

(A) Encoding reliably different edges cluster at group level in sensor-space

(B) Encoding at nodal and group level in sensor-space

we examined differences between the two experimental conditions in the probabilities of perturbations propagating across two brain regions. Our results show that there is a set of links over CP and P electrodes (CP5, P1, P2 edges-wise and CP1, Pz, P2 at nodal level), in which the difference between two conditions (MI and rest) was consistently significant across most of the subjects ( $p < 0.0001$ , BH corrected) (Fig. 3A/B).

Following the features' importance analysis described in the previous section, we performed a quantitative reliability analysis to consider only the edge-wise selected in at least half of the subjects (Fig. 2A). The final goal of this analysis was the comparison between reliably different edges selected in the encoding phase and the features selected in most subjects with the larger attributed weight. Indeed, we observed a higher level of match score with the ten subjects with a highest classification performance (37%, see Fig. 4A) as compared to the ten subjects with the lowest classification performance (6%, see Fig. 4B) but also to the entire dataset (9%, see Fig. 4C).



**Figure. 4: Edges matches between encoding and decoding:** (A) Results obtained from the ten subjects with the highest classification performance; (B) Results obtained from the ten subjects with the lowest classification performance; (C) Results obtained from the entire dataset.

## DISCUSSION

In the sensor space as well as in the source space, the classification of ATMs led to an improvement of the decoding performance with respect to the benchmark (namely the spatial filter-based approach) in most of the subjects robustly across the different number of tested random splits. Importantly, in both source and sensor domains, we observed a reduced intra and inter-subject variability with ATM+SVM as compared to CSP+SVM. These findings suggest that the use of our approach could be a tool to reduce the BCI inefficiency phenomenon. Beyond the classification performance, we also investigated the interpretability of our findings through the study of the selected features. ATMs present a straightforward interpretability as opposed to CSPs, which operate on large-scale components of the signal that are not as readily interpretable. Indeed, it is possible

to study and to identify the selected features at the subject level but a quantitative analysis at the group level is not applicable because of the difficulty to identify a common precise pattern across different subjects and different selected features. At the individual level, the information captured by the two types of feature extraction (namely CSPs and ATMs) are complementary, as seen in Fig. 5. ATMs is based on edge-wise representations and focus on strong coherent interactions that intermittently occur on the large-scale whereas CSP features, that embedded pipelines based on techniques that assume stationarity, rely on local measurements (mostly frequency band power features and time-point features) disregarding the propagation of brain dynamics at consecutive time instants.

To further study the meaning of the features selected with ATMs, we adopted an encoding framework identified here as a set of functional connections (i.e., edges) that consistently exhibited a higher likelihood of dynamic recruitment during a hand motor imagery task as compared to the resting state at the group level. This straightforward approach, validated on the entire dataset, allowed us to reliably extract functional information specific to the task execution at the individual level, an observation not achievable through traditional functional metrics (namely power spectra and phase-locking value) [7]. Therefore, from a theoretical standpoint, our study establishes the foundation for exploring neuronal avalanche metrics as a novel functional connectivity measure for investigating changes during motor tasks based not on the functional activation between two brain areas at the same time but on consecutive activations.

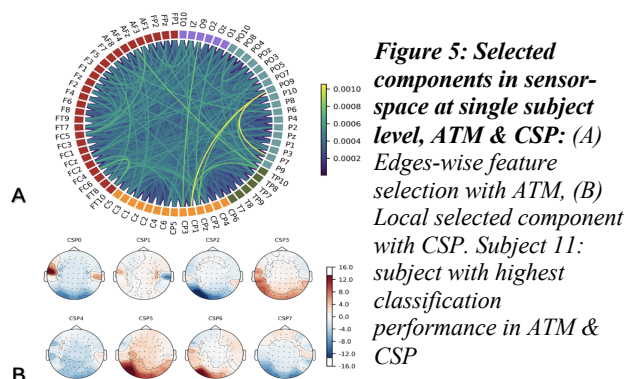
In the sensor space, the electrodes that showed a higher feature importance, identified through the decoding framework, were located over the same brain areas defined as “reliably edges” in the encoding framework. Moreover, we noticed that an increased match between the selected features and the edges-clusters led to an improvement of the classification performance. This finding suggests a possible way to apply a dimensionality reduction in the features used in the decoding step, to improve the classification performance. An ongoing work consists in investigating the key-parameters of the neuronal avalanches to be tuned and the associated features selection approaches to assure that the most relevant information will be considered for the classification step. Considering such approaches will improve the performance, but they will also reduce the computational time as well; two key-aspects of the feasibility of our pipeline in real-world scenarios.

In our work, to emphasise this possible future development, we dealt with epochs of 5s from which 25ms and 27ms (respectively for ATM + SVM and CSP+SVM) were required to extract the features and to perform the classification. Such computational time estimations are in line with current real time settings that rely on similar time windows and propose an update of the provided feedback every 28 ms. Future work will consist in identifying strategies to extract neuronal

avalanches, and therefore ATMs, in shorter time windows to make our framework completely compatible with online settings.

To further investigate the physiological meaning of our findings we compared the results respectively obtained in the sensor and in the source space. The most frequently selected features involved central electrodes (C-CP) in the sensor-space, and the paracentral area in the source-reconstructed data, implying the motor-area. Moreover, our results showed that other networks were involved in a motor-imagery task, through the selection of electrodes above parietal and occipital areas. The parietal lobe is structurally divided into inferior parietal lobe, superior parietal lobe, and precuneus [14]; its principal functions are the perception of the body, the integration of somatosensory information (e.g. touch, pain, pressure and temperature), visuospatial processing and coordination of movement. As such, the parietal activation is in line with our observations [7], because the subjects were instructed to perform a kinesthetic motor imagery task, that involves imagining movements as well as sensing the touch caused by the grasped object, and because coordinating hand, arm, and eye motions is required to perform our task. A similar role of precuneus in coordination of motor behaviour is achieved by anterior cingulate cortex [13] and its involvement has come to light in source-reconstructed data [7]. The occipital lobe [15] is primarily responsible for visual processing. Its recurrent activation and connection with central electrodes usually happens during a kinesthetic task, and when a visual stimulation is proposed as it was during our experiments.

Moreover, mainly in the source-space, we observed the involvement of the caudal portion of the middle frontal gyrus and of the medial-orbital frontal area. Within the caudal portion of the middle frontal gyrus, at the intersection with the precentral gyrus, is the frontal eye fields (Brodmann area 8). The frontal eye fields control saccadic eye movements, rapid, conjugate eye movements that allow the central vision to scan numerous details within a scene or image, same meaning of orbital regions involvement [16], instead medial-orbital frontal region reflects the allocation of attentional resources, which are typically engaged in cognitive/motor tasks [7]. Such findings demonstrate the neurophysiological validity of the selected features.



**Figure 5: Selected components in sensor-space at single subject level, ATM & CSP: (A) Edges-wise feature selection with ATM, (B) Local selected component with CSP. Subject 11: subject with highest classification performance in ATM & CSP**

## CONCLUSION

Our results suggest that the integration of periodic and aperiodic features would be a straightforward way to capture functionally relevant processes, in turn, to apply them to the design of BCIs and to improve task classification. The good performance of the ATMs on the EEG data in the sensor space is relevant to translate our methodology to real-world scenarios. Until now, we tested this new feature only during a hand motor imagery BCI task. Future work will consist in considering a wider range of BCI paradigms for communication and movements recovering applications.

## REFERENCES

- [1] Alwi Alkaff, *et al.* Applications of Brain Computer Interface in Present Healthcare Setting. in *Artificial Intelligence* vol. 0 (IntechOpen, 2024).
- [2] Allison, B. Z. & Neuper, C. Could Anyone Use a BCI? in *Brain-Computer Interfaces* (eds. Tan, D. S. & Nijholt, A.) 35–54 (2010).
- [3] Lotte, F. *et al.* A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update. *J. Neural Eng.* (2018)
- [4] Arviv, O. *et al.* Neuronal avalanches and time-frequency representations in stimulus-evoked activity. *Sci. Rep.* **9**, 13319 (2019).
- [5] Sorrentino, P. *et al.* The structural connectome constrains fast brain dynamics. *eLife* **10**, e67400 (2021).
- [6] Rabuffo, G., *et al.* Neuronal Cascades Shape Whole-Brain Functional Dynamics at Rest. *eNeuro* **8**, ENEURO.0283-21.2021 (2021).
- [7] Corsi, M.-C. *et al.* Measuring brain critical dynamics to inform Brain-Computer Interfaces. *iScience* **27**, 108734 (2024)
- [8] Corsi, M.-C. *et al.* Functional disconnection of associative cortical areas predicts performance during BCI training. *NeuroImage* **209**, 116500 (2020).
- [9] Koles, Z.J., *et al.* Spatial patterns underlying population differences in the background EEG. *Brain Topogr.* **2**, 275–284 (1990).
- [10] Blankertz, B., *et al.* Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Process. Mag.* **25**, 41–56 (2008).
- [11] Shriki, O., *et al.* Neuronal Avalanches in the Resting MEG of the Human Brain. *Journal of Neuroscience*, **33** (16) 7079-7090.
- [12] Benjamini, Y., and Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
- [13] Müller-Putz, *et al.* G. Better than random? A closer look on BCI results. *International Journal of Bioelectromagnetism* (2008).
- [14] Wenderoth, N., *et al.* The role of anterior cingulate cortex and precuneus in the coordination of motor behaviour. *Eur. J. Neurosci.* **22**, 235–246 (2005).
- [15] Kwon, S., *et al.* Neuropsychological Activations and Networks While Performing Visual and Kinesthetic Motor Imagery, *Brain Sci.* **2023**,13,983.
- [16] Schall J.D., Frontal Eye Fields, *Encyclopedia of Neuroscience*. Pages 367-374 (2009).

# LOCALIZING NEURAL SOURCES IN THE CERVICAL SPINAL CORD

Markus E. Oberndorfer<sup>1</sup>, Gernot R. Müller-Putz<sup>1,2</sup>

<sup>1</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>2</sup>BioTechMed Graz, Graz, Austria

E-mail: gernot.mueller@tugraz.at

**ABSTRACT:** Mapping neural activity along the spinal cord is a task that is hardly researched compared to human brain mapping. By identifying neural sources in the spinal cord and detecting unique activity patterns associated with various motor tasks or specific sensory input, it becomes possible to establish a baseline for healthy individuals. This could be utilized to classify spinal cord injuries or monitor changes in the spinal cord. This study demonstrates the effective application of an innovative approach to localizing the spinal sources of spinal cord potentials (SCPs) using the finite element method (FEM) to solve the forward problem and an abstraction of the sLORETA algorithm to identify the neural sources, which were induced by functional electrical stimulation (FES) on the forearms of healthy individuals.

## INTRODUCTION

The field of human brain mapping, once limited to static classifications like Brodmann areas [1] based on structural composition, has now evolved into a dynamic process that allows researchers to gain more knowledge about the signal processing inside the brain. This dynamic process involves functional brain imaging, which is a set of imaging methods, e.g., functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and electroencephalography (EEG) that can be applied to analyze real-time neural activity [2].

In the early work by Pfurtscheller et al. [3] brain patterns of hand movement imagination were used to allow a tetraplegic person to control an electrically driven hand orthosis. Additionally, this was one first work where multichannel EEG was projected onto the surface of a brain model in a BCI context. Since then, the field of functional brain imaging has evolved significantly. Different software packages, e.g., Brainstorm [4], EEGLab [5] and MNE [6] are currently available to visualize EEG data on a head model but also have large signal processing pipelines incorporated. Furthermore, those software packages allow to identify the neural sources in the brain by utilizing forward and inverse calculations. Forward computations involve modeling the propagation of electromagnetic fields from the neural sources within the brain to the scalp electrodes, taking into account the conductivity properties of the head tissues [7]. As for the inverse computations, there is a multitude of different al-

gorithms available, which try to identify the location and magnitude of the neural sources for given EEG measurements [8]. These advancements have facilitated groundbreaking research in many fields, such as allowing researchers to identify conversely modulated gamma frequency bands in central sensorimotor areas during the human gait cycle [9] or proving that two different types of neural networks are active during rhythmic finger movements [10].

With this experience gained in human brain mapping, the way is paved to explore the mapping of sensory stimulations of the peripheral nervous system and motor tasks involving the limbs to the spinal cord. So far, there are hardly any publications that tackle the spinal cord mapping problem. In Stroman et al. [11], the neural activity changes in the lumbar spinal cord due to locally applied low temperatures on the skin were analyzed. The neural activity was recorded with fMRI. In Nierula et al. [12], a comprehensive recording of spinal cord somatosensory evoked potentials (SEPs) was performed in order to assess the functional architecture of somatosensory processing. A first attempt of sources localization of neural sources inside the spinal cord was done by Moffitt and Grill [13]. With their inverse model, they aimed to create an initial framework that can be used to obtain a more detailed map of the neuroanatomy of the spinal cord such that intraspinal, microstimulating electrodes are placed more effectively. However, they only simulated the signals using a basic cylinder structure, intended to mimic a simplified model of the spinal cord. Additionally, given that they were simulating invasive measurements of the spinal cord, they selected a relatively high artificial signal-to-noise ratio (SNR) compared to measurements typically obtained from the neck's skin. Currently, there is no work that uses non-invasively measured SCPs to find correlating neural sources. Therefore, the approach to spinal cord mapping herein is to identify neural sources in the spinal cord that are responsible for discernible potential changes following a predetermined stimulus.

Neural sources in the brain are believed to arise from synchronized synaptic activity [14]. Similar effects lead to measurable potential changes in the spinal cord, e.g., dorsal root potential (DRP), dorsal root reflex (DRR), primary afferent depolarization (PAD) [15]. These similarities provide additional support for choosing sLORETA as the preferred method for source localization in this initial

attempt.

Our objective was to demonstrate the presence of movement-related neural activity in the spinal cord through non-invasive recordings of spinal cord potentials. In particular, we aimed to identify neural responses elicited by afferent stimuli from functional electrical stimulation (FES) induced wrist movements.

## MATERIALS AND METHODS

**Data:** We based our study on the data from Wimmer et al. [16]. Eight healthy participants underwent FES of the forearm, leading to a wrist extension. Simultaneously, a 16-channel EEG as well as 16-channel SCP recordings were performed (see Fig. 1). The FES-induced wrist movement was performed with a Microstim 8 stimulator (Krauth+Timmermann, Hamburg, Germany) and a stimulation frequency of  $f_S = 35\text{Hz}$  (pulse width of  $300\text{ }\mu\text{s}$ ). Only the raw 16-channel SCP recordings were considered for further analysis in this work. After processing the data with a high-pass ( $f_H = 0.5\text{Hz}$ ), low-pass ( $f_L = 60\text{Hz}$ ) and notch ( $f_{N1} = 35\text{Hz}$ ,  $f_{N2} = 50\text{Hz}$ ) filter, the signals were split into trials and averaged (for more detail see [16]).

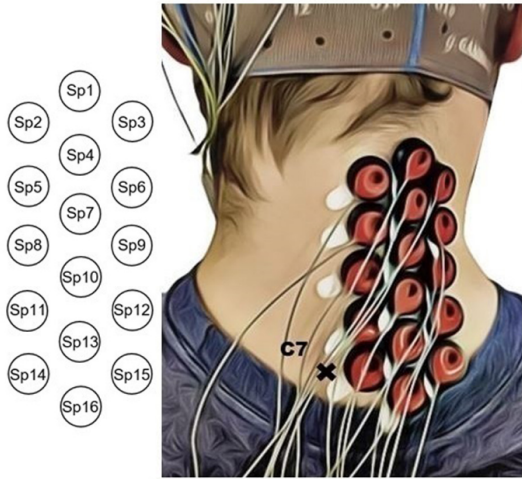


Figure 1: Electrode setup used in Wimmer et al. [16] to measure the Spinal Cord Potentials (SCPs). Image taken with permission from [16].

**Geometrical model:** To create the geometry for this model, the NGSolve/NETGEN-package in Python was used. For simplicity, only the following structures were added to the model: spinal cord, vertebrae, intervertebral discs, trachea, oesophagus and neck.

**Forward problem:** A Volume-Conduction model was used to describe the potential field distribution. To mathematically formulate this, Poisson's Equation was incorporated into the model:

$$\sigma \cdot (-\Delta V) = I_m \quad (1)$$

For some given current source density  $I_m$  and conductivity  $\sigma$ , the resulting potential  $V$  was calculated with the finite element method (FEM) [17]. The forward problem

was implemented with the NGSolve/NETGEN-package in Python.

For simplicity, the conductivity of all materials was assumed to be homogeneous. Several publications [18–22] were taken into consideration to determine the conductivity values. The conductivity values are listed in Tab. 1.

Table 1: List of all materials with their respective conductivity.

Material	Conductivity $\sigma$ [S/m]
Spine	0.22
Vertebrae	0.0014
Disc	0.008
Trachea	0.015
Oesophagus	0.015
Muscle (Neck)	0.01

**Inverse problem:** sLORETA [23] is widely acknowledged as a commonly utilized technique in EEG source localization, thereby validating its adoption in our proposed approach. Moreover, sLORETA was the preferred inverse algorithm method in [13]. In this study, we utilized a simplified version of the sLORETA algorithm. Derived from the original algorithm [23], the simplified version incorporates two key modifications: a cylinder-like geometry and the exclusion of deep sources. The number of electrodes is defined as  $N_E = 16$  and the number of dipoles assumed to occur in the cervical part of the spinal cord are defined as  $N_V$ .

Starting with the following equality

$$\Phi = \mathbf{K}\mathbf{J} + c\mathbf{1} \quad , \quad (2)$$

in which  $\Phi \in \mathbb{R}^{N_E \times 1}$  represents the SCP measurements,  $\mathbf{K} \in \mathbb{R}^{N_E \times (3N_V)}$  and  $\mathbf{J} \in \mathbb{R}^{(3N_V) \times 1}$  are the leadfield matrix and the solution vector, respectively. The term  $3N_V$  is due to the dimensionality of the model. While the positions of the dipoles remain fixed, it's important to consider the x-, y-, and z-directions individually for each electric dipole, as they can assume any orientation in space. The  $c$  variable introduces noise to the model. As customary, the forward problem is employed to compute the leadfield matrix  $\mathbf{K}$ . To obtain  $\mathbf{J}$ , the functional  $F$  has to be minimized with respect to  $\mathbf{J}$  and  $c$

$$F = \|\Phi - \mathbf{K}\mathbf{J} - c\mathbf{1}\|^2 + \alpha \|\mathbf{J}\|^2 \quad , \quad (3)$$

in which  $\alpha$  is a regularization parameter. The solution to this optimization problem is

$$\hat{\mathbf{J}} = \mathbf{T}\Phi \quad , \quad (4)$$

in which  $\hat{\mathbf{J}}$  is the solution vector containing all the neural sources' location, direction and magnitude.  $\mathbf{T}$  is a pseudo-inverse of the leadfield matrix  $\mathbf{K}$ , calculated as

$$\mathbf{T} = \mathbf{K}^T \mathbf{H} [\mathbf{H} \mathbf{K} \mathbf{K}^T \mathbf{H} + \alpha \mathbf{H}]^+ \quad , \quad (5)$$

in which  $\mathbf{H}$  is the centering matrix.

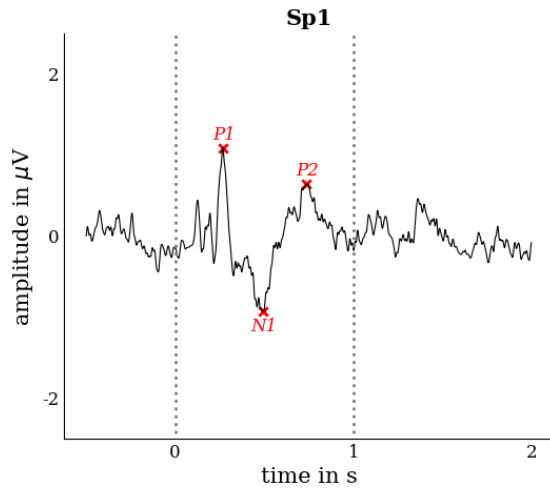


Figure 2: Average SCP obtained from one channel (Sp1) placed along the spine (Fig. 1). Three distinct peaks P1, N1, P2 are marked, as they are typical for potentials in the spinal cord.

More information regarding the calculations can be found in the original paper [23].

**Optimization:** From an optimization point of view, there are two hyperparameters that can be tuned. The first one is the regularization parameter  $\alpha = [0.005, 0.01, 0.05, 0.1, 0.2, 0.35, 0.5]$  in Eq. 3. The second one is the number of dipoles  $N_V = [21, 42, 63, 84, 105]$ , which defines the size of the leadfield matrix  $\mathbf{K}$ . We computed the error functional  $F$  in Equation 3 for each combination of hyperparameters. Subsequently, we selected the hyperparameter combination that resulted in the lowest error for the final calculation.

## RESULTS

After processing the data, the SCP signals exhibited a typical triphasic spike [24]. One of the 16 channels is shown representatively in Fig. 2. The characteristics of the triphasic spikes, calculated from all channels placed along the spine, are listed in Tab. 2.

Table 2: Amplitude and latency for the waveform points. Mean and standard deviation (SD) are calculated from the  $n=16$  channels.

Waveform-point	Latency		Amplitude	
	Mean [ms]	SD [ms]	Mean [μV]	SD [μV]
P1 (n=16)	0.264	0.003	1.108	0.103
N1 (n=16)	0.487	0.004	-0.944	0.077
P2 (n=16)	0.739	0.005	0.708	0.097

The simplified neck geometry is visualized in Fig. 3. It captures only the most important features of a human neck that are necessary for the forward problem to deliver meaningful results. To show the functionality of the forward model, an axially oriented example dipole was inserted into the spinal cord (Fig. 4). A visualization of the solution vector  $\hat{\mathbf{j}}$  is shown in Fig. 5 as red arrows. The vector length is normalized to the cylinder diameter

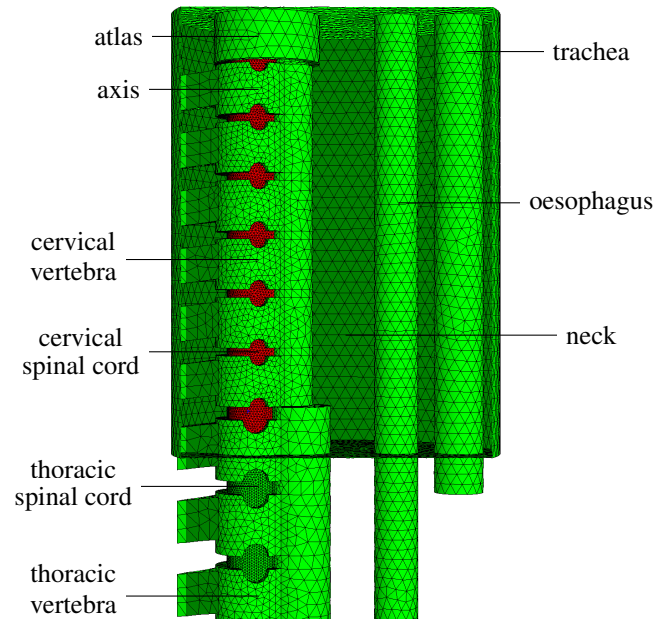


Figure 3: Neck part of the model dissected to reveal the spine, the trachea and the oesophagus with the cervical part of the spinal cord marked in red. The model incorporates a different structure for atlas and axis as well as different sizes for the vertebral bodies.

for each of the three solutions. The gray cylinder represents the neck.

The behaviour of the error functional  $F$  with respect to the hyperparameters  $\alpha$  and  $N_V$  is shown in Fig. 6 for the first peak P1.

## DISCUSSION

In this study, we successfully employed well-established methods for the forward and inverse computations to accurately localize neural sources within the spinal cord.

Based on the results of the forward problem, the potential field propagates primarily through the nervous tissue, indicating that the forward model works properly.

The nerves in the forearm are part of the Plexus Brachialis, which is a composition of spinal nerves C5-C8 and Th1 [25]. Since the large dipoles of P1 are located in the lower section of the cervical spinal cord, this result coincides with the anatomical structure of the nervous system. This is still partially true for the neural sources seen in points N1 and P2. Therefore, the results demonstrate the possibility of localizing the neural sources in the spinal cord. The influence of the hyperparameters (Fig. 6) on the error of the model indicates that  $\alpha = 0.1$  offers the lowest error and that  $\alpha$  has a more significant impact on the error than the number of dipoles  $N_V$ . Similar results were found for the N1 and P2 wave. Since the model does not capture much of the complexity the human body has to offer, these hyperparameter results should only be considered as initial guesses for future,



Figure 4: Visualization of the potential field distribution inside the model for an axially oriented electric dipole located in the spinal cord.

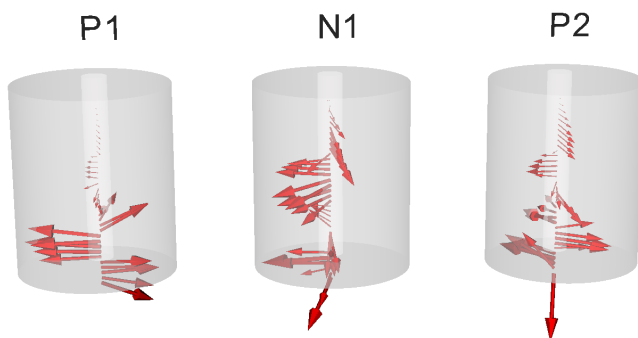


Figure 5: Solution of the source localization algorithm. The calculated dipoles are shown as red arrows originating from the spinal cord. The solution is shown for every wave P1, N1 and P2. The gray cylinder represents the neck.

hyperparameter search for: P1

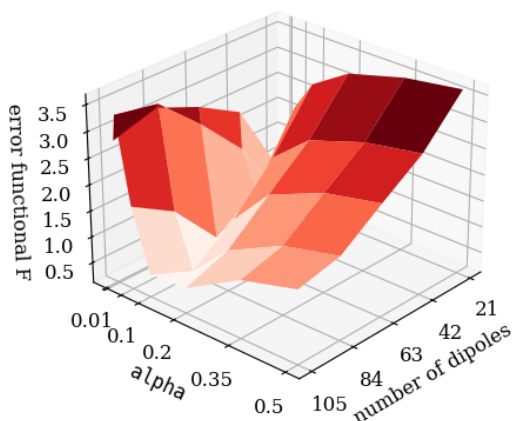


Figure 6: Behaviour of the error with respect to the two hyperparameters  $\alpha$  and  $N_V$ .

more complex, models.

In this paper, only three points in time were compared. For actual mapping purposes, it would be appropriate to analyze the complete time domain to possibly obtain more differences in the neural activity induced by different stimuli.

Assuming that this mapping was conducted across numerous healthy individuals, it would establish a reference dataset that could be applied in various contexts and applications. For example, classifying the spinal cord injury level, monitoring disease progression through tracking changes in the spinal cord neural activity (e.g., in amyotrophic lateral sclerosis (ALS)) or changes during rehabilitation of spinal cord injured individuals.

Additionally, one could try to analyze the characteristics of the triphasic spike (latency and magnitude, Tab. 2) to distinguish between different motor tasks or stimuli. If more SCP recordings were available with a different stimulus, the characteristics might be slightly different and therefore, could be used for mapping.

This framework does come with a few limitations. First, the mathematical and geometrical model is very simple and does not capture the complexity of the human physiology and anatomy, respectively. Especially, when it comes to large vessels like the Arteria carotis, which can distort the electric field. Further, the conductivity of biological tissue is anisotropic and is different for every human. Second, the measurements were taken from only one side of the neck and therefore, introduce a bias to the resulting sources. It is assumed that this bias is expressed in less precise directions of the dipoles. Third, since the stimulation of the forearm was performed with a 1s long 35Hz biphasic current pulse, the resulting SCPs are a summations of several consecutive stimuli, not just a single stimulus.

These considerations emphasize the importance of further refining both models and measurements in future research.

## CONCLUSION

In this work, the feasibility of localizing electric dipoles in the cervical spinal cord based on recorded spinal cord potentials was demonstrated. By resolving the complicated challenges associated with the source localization problem, the findings reveal the potential for reasonably precise spatial identification within this neural region. These findings conclude that a spinal cord mapping of certain motor tasks or sensory stimuli is possible.

## REFERENCES

- [1] Strotzer M. One century of brain mapping using brodmann areas. *Clinical Neuroradiology*. 2009;19(3):179.
- [2] Raichle ME. A brief history of human brain mapping. *Trends in neurosciences*. 2009;32(2):118–126.

- [3] Pfurtscheller G, Guger C, Müller G, Krausz G, Neuper C. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience letters*. 2000;292(3):211–214.
- [4] Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. Brainstorm: A user-friendly application for meg/eeg analysis. *Computational intelligence and neuroscience*. 2011;2011:1–13.
- [5] Delorme A, Makeig S. Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*. 2004;134(1):9–21.
- [6] Gramfort A *et al.* Mne software for processing meg and eeg data. *neuroimage*. 2014;86:446–460.
- [7] Hallez H *et al.* Review on solving the forward problem in eeg source analysis. *Journal of neuroengineering and rehabilitation*. 2007;4(1):1–29.
- [8] Grech R *et al.* Review on solving the inverse problem in eeg source analysis. *Journal of neuroengineering and rehabilitation*. 2008;5(1):1–33.
- [9] Seeber M, Scherer R, Wagner J, Solis-Escalante T, Müller-Putz GR. High and low gamma eeg oscillations in central sensorimotor areas are conversely modulated during the human gait cycle. *Neuroimage*. 2015;112:318–326.
- [10] Seeber M, Scherer R, Müller-Putz GR. Eeg oscillations are modulated in different behavior-related networks during rhythmic finger movements. *Journal of Neuroscience*. 2016;36(46):11671–11681.
- [11] Stroman PW, Tomanek B, Krause V, Frankenstein UN, Maliszka KL. Mapping of neuronal function in the healthy and injured human spinal cord with spinal fmri. *Neuroimage*. 2002;17(4):1854–1860.
- [12] Nierula B *et al.* Non-invasive multi-channel electrophysiology of the human spinal cord—assessing somatosensory processing from periphery to cortex. *bioRxiv*. 2022:2022–12.
- [13] Moffitt MA, Grill WM. Electrical localization of neural activity in the dorsal horn of the spinal cord: A modeling study. *Annals of biomedical engineering*. 2004;32:1694–1709.
- [14] Jackson AF, Bolger DJ. The neurophysiological bases of eeg and eeg measurement: A review for the rest of us. *Psychophysiology*. 2014;51(11):1061–1071.
- [15] Giuliano LMP, Nunes KF, Manzano GM. The potentials recorded around the spinal cord: Different sides of the same dice. *Revista Neurociências*. 2013;21(3):449–454.
- [16] Wimmer M, Kostoglou K, Müller-Putz GR. Measuring spinal cord potentials and cortico-spinal interactions after wrist movements induced by neuromuscular electrical stimulation. *Frontiers in Human Neuroscience*. 2022;16:858873.
- [17] Zienkiewicz OC, Taylor RL, Zhu JZ. The finite element method: Its basis and fundamentals. Elsevier (2005).
- [18] Vanrumste B. Eeg dipole source analysis in a realistic head model. Ph.D. dissertation. Ghent University. 2002.
- [19] Jackson A, Travascio F, Gu W. Effect of mechanical loading on electrical conductivity in human intervertebral disk. *Journal of biomechanical engineering*. 2009;131:054505.
- [20] Lee JH *et al.* In vivo electrical conductivity measurement of muscle, cartilage, and peripheral nerve around knee joint using mr-electrical properties tomography. *Scientific Reports*. 2022;12(1):73.
- [21] Haueisen J *et al.* The influence of brain tissue anisotropy on human eeg and meg. *Neuroimage*. 2002;15(1):159–166.
- [22] Huiskamp G, Vroeijenstijn M, Dijk R van, Wieneke G, Huffelen AC van. The need for correct realistic geometry in the inverse eeg problem. *IEEE Transactions on Biomedical Engineering*. 1999;46(11):1281–1287.
- [23] Pascual-Marqui RD *et al.* Standardized low-resolution brain electromagnetic tomography (sloreta): Technical details. *Methods Find Exp Clin Pharmacol*. 2002;24(Suppl D):5–12.
- [24] Yates BJ, Thompson FJ, Mickle PJ. Origin and properties of spinal cord field potentials. *Neurosurgery*. 1982;11(3):439–450.
- [25] Orebaugh SL, Williams BA, *et al.* Brachial plexus anatomy: Normal and variant. *The Scientific World Journal*. 2009;9:300–312.

## A SMALL STEP TOWARDS THE DETECTION OF MENTAL FATIGUE INDUCED BY BCI-HMD TRAINING

A. Polyanskaya<sup>1,2</sup>, R. Rosipal<sup>1</sup>, G. Sobolová<sup>1</sup>, Z. Rošťáková<sup>1</sup>, N. Porubcová<sup>1</sup>

<sup>1</sup>Institute of Measurement, Slovak Academy of Sciences, Bratislava, Slovakia

<sup>2</sup>Middle European Interdisciplinary Programme in Cognitive Science, University of Vienna, Vienna, Austria

E-mail: arina.polyanskaya@univie.ac.at

**ABSTRACT:** We used a proprietary constructed brain-computer interface system with a head-mounted display for motor neurorehabilitation training of a subject after a stroke. This study analyzes quantitative EEG (qEEG) changes during resting state periods before and after the neurorehabilitation training. Eyes closed and eyes open resting state EEG collected during 13 training sessions is analyzed to determine qEEG changes indicating mental state changes like increased mental fatigue, tiredness, or sleepiness. We decomposed the EEG spectrum into oscillatory and fractal parts, allowing us to investigate changes in the oscillatory component of qEEG separately. We observed increased post-training oscillatory EEG amplitudes in slow frequency bands (delta and theta) and decreased in faster alpha to beta bands. A shift to a slower frequency of the dominant alpha frequency was also observed in the post-training resting state EEG. Compared with existing literature, these changes indicate increased mental fatigue and sleepiness.

### INTRODUCTION

A growing body of evidence suggests that integrated technologies of brain-computer interfaces (BCI) and virtual reality (VR) environments provide a flexible platform for neurorehabilitation therapies, including significant post-stroke motor recovery and cognitive-behavioral therapy. If VR scenarios are realized through head-mounted displays (HMDs), a compact BCI-HMD system that is exceptionally flexible and rich for implementing various scenarios and tasks can be constructed. As some studies have shown, BCI-based neurorehabilitation therapies are effective in improving the motor abilities of stroke survivors [2, 3].

One of the challenges of BCI is its decreased performance over time, making it unreliable for long-term use [4]. The oscillations of psychological states can cause such inconsistencies. While performing BCI tasks, mental states such as level of frustration, mental fatigue, and attention may shift, therefore influencing the outcomes of a BCI session [5]. This is a concern because stroke survivors exhibit a higher prevalence of fatigue; it is often severe and frequent even long after stroke [6]. Therefore, it is important to consider mental states, especially

fatigue, for stroke patients while performing rehabilitation procedures. Previous studies concentrated on subjective fatigue measurements, such as the Visual Analogue Scale (VAS) or other qualitative reports; however, EEG has been the more reliable predictor due to its temporal precision [2, 7, 8].

Fatigue is a decreased ability to initiate or sustain voluntary actions, including difficulties with alertness, mental performance, and reduced efficiency. It is gradual and cumulative and applies to psychological and physical activity. Mental fatigue is concerned explicitly with reduced or impaired cognitive functions that are believed to be caused by prolonged cognitive activity. Mental fatigue can also influence physical performance.

There have been different approaches to defining mental fatigue with EEG, including detecting an increase in the ratio of slow wave to fast wave as the fatigue progresses [9]. Additionally, particular areas of the brain and frequencies that indicate the increase in fatigue have been determined. Some of the findings seem to contradict each other. For example, a study by Eoh and colleagues (2005), and a study by Stern found a decrease in the alpha band as drowsiness increased [11, 12]. However, a later study by Jap and colleagues found that alpha waves increased in the occipital lobe as fatigue progressed [13]. On the other hand, Eoh and Jap outlined that the beta band decreases with the progression of fatigue in temporal and frontal areas, and the (theta+alpha)/beta ratio increases [11, 13]. Both studies looked more into the drivers' fatigue and sleepiness, possibly influencing the outcomes. Trejo and colleagues (2015) came to the same conclusions as Jap and colleagues, stating that an increase in parietal alpha and frontal theta was present, along with a shift in alpha frequency to the lower alpha band [13, 14]. According to the meta-study, the increase of theta waves in the frontal, central, and posterior regions is associated with fatigue, additionally, the rise of alpha in central and posterior frequency serves as a biomarker of fatigue [15].

Other researchers have previously investigated fatigue in BCI rehabilitation. In 2010, a paper was published by Prasad and colleagues who researched BCI used for upper-limb recovery [2]. The Visual Analogue Scale

(VAS) was used to examine fatigue. Some participants exhibited increased mental fatigue, but the details were unclear as it did not allow for temporally precise results. The study by Foong and colleagues (2020) also concentrated on the neurorehabilitation of upper limbs for stroke survivors through the use of BCI technologies [16]. To examine mental fatigue, the researchers extracted 3-second data before each trial to correlate it with the subject's performance, and they concentrated on the shifts of amplitude of beta waves across different brain areas separately (frontal, central, parietal-occipital). The findings showed a significant positive correlation of beta power with accuracy in frontal and central brain regions, which suggested that mental fatigue in BCI tasks was associated with the performance outcomes [16]. However, the shortcoming of this approach is that it only considers beta bands without looking into theta, delta, and alpha, which were associated with mental fatigue in the previous literature. On the contrary, a study by Talukdar and colleagues found that there is a significant increase of spectral power in the range of 0.1-12 Hz but no significant findings in the beta band after performing the MI task on BCI [17]. With that in mind and to our knowledge, no studies have examined the effects of BCI-HMD systems on mental fatigue after performing MI tasks.

In this study, we used a proprietary constructed real-time BCI-HMD system for motor rehabilitation of the upper limbs of subjects after stroke [18]. We performed a series of 13 training sessions (days) on a subject with post-stroke motor impairment of the left upper limb. Part of the training process is the collection of EEG data during resting state periods preceding and following the training itself. This paper focuses on quantitative EEG analysis (qEEG) of changes the training can induce on the resting state eyes closed (EC) and eyes open (EO) EEG, or passive BCI. We focused on the oscillatory part of the EEG spectrum in the frequency range from 2.5 Hz to 18.0 Hz. Significant post-training changes were observed in the EC condition, indicating EEG slowing. Changes in the EO condition were sporadic and restricted to faster alpha and beta EEG frequencies. A shift in the dominant alpha frequency to the lower alpha band was also observed. These changes align with changes associated with increased mental fatigue, as reported by Trejo and colleagues [14].

## MATERIALS AND METHODS

In the study, the previously developed and described BCI-HMD system was used [18]. Its architecture is depicted in Fig. 1, representing the standard BCI design consisting of

- signal acquisition,
- signal processing and classification and
- environment control.

Publicly available OpenVibe<sup>1</sup> software for BCI and real-time neuroscience interconnect three major blocks of the

<sup>1</sup><http://openvibe.inria.fr>

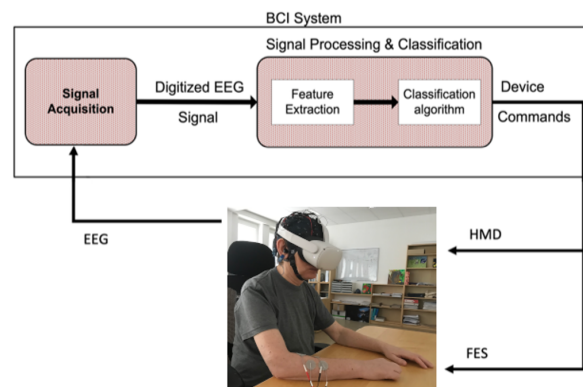


Figure 1: The architecture of the brain-computer interface with a head-mounted display (BCI-HMD) and the functional electrical stimulation (FES) element.

BCI-HMD architecture. An autonomous Oculus Quest 2 (Meta Platforms, Inc.) headset with a fast processor, a new-generation graphics card, and 256 GB of internal storage is used as the HMD. The neurorehabilitation system is also enriched with the functional electrical stimulation (FES) component applied to selected muscles. This is done through the programmable two-channel externally controlled Microstim FES device (Medel GmbH).

The main element of BCI training is the task of motor imagery (MI), during which the subject imagines the movement without including movement of the limb. During this effort, required changes in brain activity are recorded using an EEG on the subject's scalp. If these changes are detected successfully, the requested visualization in the VR environment will begin. Such a visualization could exhibit an avatar's hand grasping a cup on the table (Fig. 2). In this study, we used three randomly varying tasks: grasping a cup, a cube, and turning a key in a lock. Because our goal is not the description and analysis of the BCI training itself but the analysis of brain activity changes during the resting state before and after the training, we do not describe the design of the BCI-HMD further and refer the reader to [18].



Figure 2: An example of a virtual environment (a cup) with an object grip animation.

*Experimental Protocol:* Each training day (session) started with two minutes of the resting state block with eyes closed (EC) followed by two minutes of the resting state block with eyes open (EO). The same EC and EO resting state EEG was recorded after each BCI-HMD training session. During the EO condition, the subject fixated his eyes on a small cross on the wall in front of him.

A self-made questionnaire was provided at the beginning of each session and after completing resting state EC and EO EEG recordings. The subject was asked to answer the question "*Do you feel tired?*" on a seven-level scale (1 - absent feeling, 7 - extreme feeling). In addition, after each block of BCI-HMD training trials, six questions focused on the subjective evaluation of cybersickness and tiredness were applied. However, in this paper, EEG data recorded during the BCI-HMD training and their connection to subjective assessment of cybersickness and mental fatigue development during the training itself are not analyzed.

The BCI-MHD experiment consists of a series of trials in which the subject is instructed to imagine a movement of their avatar hand in VR mentally. In this study, each session consisted of three blocks with ten MI trials in each block.

**Participant:** The subject of this study is an 86-year-old male who has experienced left-sided hemiparesis due to a stroke in the basal ganglia region on the right side with residual upper limb weakness (acral part). After the stroke, he also suffers from fatigue syndrome with Parkinson's syndrome. In the time before the stroke, the subject was a healthy, active athlete with no presenting psychiatric diagnosis.

The subject participated in 13 days of BCI training from May 11, 2023, to June 22, 2023. The sessions occurred at intervals of 1 to 6 days.

**EEG Signal Acquisition and Processing:** Wireless g.Nautilus PRO FLEXIBLE FDA-cleared and CE-certified recording system was used for EEG data acquisition. The current experiment included 11 Ag/AgCl wet electrodes attached to a fabric cap following a 10-20 international system. The electrodes included on the right hemisphere were FC4, C2, C4, C6, and CP4; the electrodes on the left hemisphere included FC3, C1, C3, C5, CP3, O1, as well as a linked-ears reference and one ground electrode AFz.

For resting state conditions (EC/EO), we performed an initial analysis with the sampling rate set to 250 Hz. EEG data processing consisted of multiple steps applied in Brain Vision Analyzer 2.3.0<sup>2</sup> (BVA) with templates and expert supervision. First automatic artifact detection step with criteria of maximal allowed voltage set at 50  $\mu$ V/ms, maximal absolute amplitude set at 70  $\mu$ V, lowest allowed activity in intervals of 100 ms set to 0.5  $\mu$ V, and maximally allowed difference of voltages in intervals of 20 ms was to 70  $\mu$ V was applied. EEG traces with detected artifact segments were then visually inspected by a trained expert, and artifact markers were edited.

Artifact markers were exported from the BVA software, and further analysis was carried out in the MATLAB<sup>3</sup> software.

EEG band amplitudes were analyzed for the oscillatory part of the frequency spectrum. The decomposition of the total frequency spectrum into fractal (representing back-

ground EEG) and oscillatory components was done using the irregular-resampling auto-spectral analysis (IRASA) [19]. IRASA decomposes the amplitude spectrum of each segment into a fractal (scale-free) and an oscillatory part. Different mechanisms may generate EEG oscillatory and fractal components, so it is essential to estimate them separately, mainly when the measurement focuses on localized narrow-band oscillatory rhythms, as is the case here. The oscillatory part of the amplitude spectrum was obtained by subtracting the fractal part from the total spectrum estimate. Negative values of the oscillatory spectrum were set to zero. In the study, we focused on the oscillatory part of the spectrum to measure band amplitudes from 2.5 Hz to 18.0 Hz. Both spectrum parts were computed with a resolution of 0.4883 Hz. The analyzed EEG endpoints were standardized quantitative qEEG measures in the following ranges: delta (2.5-4Hz), theta (4-8 Hz), alpha1 (8-10 Hz), alpha2 (10-12 Hz), beta1 (12-15 Hz), beta2 (15-18 Hz), alpha individual (6.4-9.8Hz). Additionally, ASI (alpha slow wave index defined as a ratio of (alpha1 + alpha2)/(delta + theta)), TBR (theta/(beta1 + beta2) ratio), and BAR (beta1/(alpha1 + alpha2) ratio) derived measures were included.

Using a paired t-test, we analyzed differences between the above-defined measures computed from EEG traces recorded before and after the training. This testing was done separately for each electrode in eyes closed and eyes open condition.

In the subjective evaluation of fatigue, the days were separated by the answers to the tiredness question. The first group included sessions in which increased tiredness was reported after the BCI training compared to the pre-training response (sessions 1, 2, 8, 11, and 12). The remaining eight sessions created the second group, including the days of the same reported fatigue or decreased fatigue after the training session. Then, the percent of change (PC)

$$PC = (post - pre) / (post + pre)$$

between post- and pre-training endpoint values were computed for each endpoint and electrode. The PC of each of the two groups was computed by averaging all sessions corresponding to that group.

## RESULTS

**Eyes Closed:** Fig. 3 shows a summary of significant before and after BCI-HMD training amplitude band differences. A significant after-training delta increase can be observed at the C2 and CP4 electrodes. Theta significantly increased at both fronto-central FC3 and FC4 EEG channels. Alpha1 post-training decreased at the O1 and CP4 electrodes. Alpha2 post-training decreased at the C1, C2, and FC4 EEG electrodes, but increased at the O1 electrode. Significant changes in the beta range were sporadic and limited to the beta1 post-training increase at O1.

<sup>2</sup><https://www.brainproducts.com/solutions/analyzer>

<sup>3</sup><https://www.mathworks.com>

The derived ASI, TBR, and BAR measures mimicked the observed changes in amplitude bands and indicated overall EEG slowing. Increasing slower delta and theta frequency bands and decreased alpha bands are reflected by a significantly broader post-training ASI decrease in the left hemisphere (FC3, C1, C3) and the right hemisphere (FC4, C2, C4, C6). TBR increases at the FC4 electrode and reflects an observed significant increase of theta at the same electrode site. A significant post-training increase in BAR was limited to the O1 electrode.

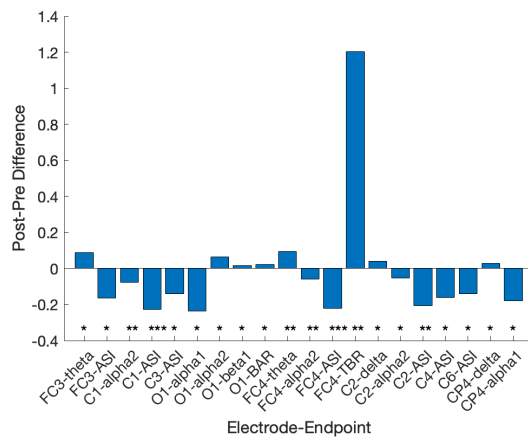


Figure 3: Bar chart showing significant post-training and pre-training amplitude band differences for the eyes closed (EC) condition. The statistical significance of the differences was tested using a paired t-test. Results are shown by EEG channel and amplitude band. Positive values indicate a post-training increase and negative values decrease. Values represent amplitude per Hz differences ( $\mu\text{V}/\text{Hz}$ ). For ASI, TBR, and BAR, the values represent a ratio. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

Fig. 4 shows oscillatory spectrum differences between post-training and pre-training spectrum averages calculated over 13 sessions. Pre-training and post-training oscillatory spectrum overlay is depicted in Fig. 5. As can be seen, the theta band significantly increased in frontal and central areas (more prevalent on the left side). There is also a decrease of alpha1 in central, frontal, and parietal electrodes, mainly on the right hemisphere.

When analyzing post- versus pre-training endpoint values grouped according to the subjective evaluation of tiredness, there was a 22.24% (at electrode C2) and 19.74% (at electrode CP4) PC increase in the delta band in the increased fatigue group. In the group where no increase in fatigue was subjectively indicated, the delta PC was 12.69% (electrode C2) and 7.43% (electrode CP4). Although limited to a single endpoint, these findings are consistent with some previous research [8, 20].

*Eyes Open:* Significant post-training changes in the EO condition were sporadic and restricted to faster EEG frequencies (alpha, beta) than the resting state EC condition. A summary of significant before and after BCI-HMD training amplitude band differences is shown in Fig. 6. A post-training increase was observed in alpha1 at CP4, but a more systematic decrease of alpha2 was

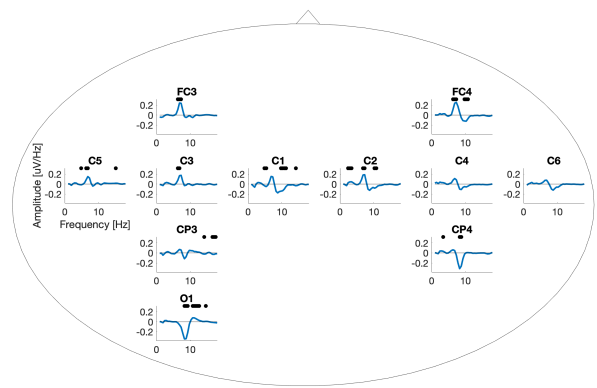


Figure 4: Eyes closed (EC) oscillatory spectrum difference computed as a difference between post-training and pre-training oscillatory spectrum averages calculated over 13 sessions. Positive values indicate a post-training increase and negative values decrease. Dots indicate frequencies where significant differences were observed using a paired t-test ( $p < 0.05$ ).

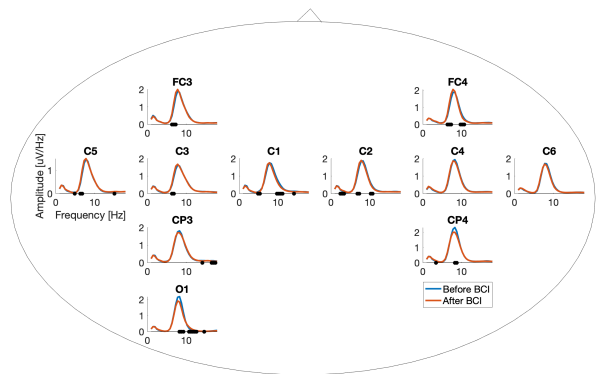


Figure 5: Eyes closed (EC) spectrum overlay of pre-training and post-training oscillatory spectrum averages calculated over 13 sessions. Dots indicate frequencies where significant differences were observed using a paired t-test ( $p < 0.05$ )

observed in the right hemisphere (C4, C6, CP4). TBR increased at CP3. This significant change in TBR is driven by post-training theta increase and beta decrease. The change can be observed in Fig. 7 and Fig. 8.

Similarly to Fig. 4, Fig. 7 shows significant EO changes depicted as oscillatory spectrum differences between post- and pre-training spectrum averages calculated over 13 separate sessions. Pre-training and post-training oscillatory spectrum overlay is depicted in Fig. 8. Around 9 to 10 Hz, a rapid negative to positive post-training change can be observed. This frequency span defines the subject's alpha frequency range, and the rapid change represents a shift from the dominant alpha to the lower alpha band.

No consistent results were found when subjectively reported fatigue levels collected before and after the training sorted endpoint values.

## DISCUSSION

The study assessed mental fatigue using EEG collected

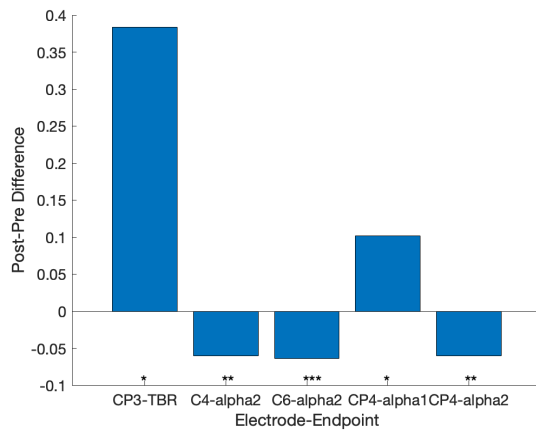


Figure 6: Bar chart showing significant post-training and pre-training amplitude band differences for the eyes open condition (EO). See Fig. 3 for description.

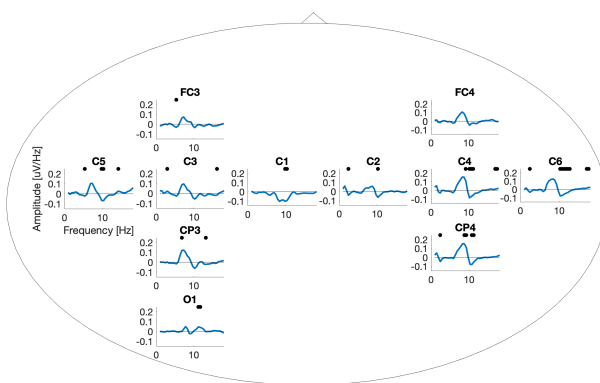


Figure 7: Eyes open (EO) oscillatory spectrum difference computed as a difference between post-training and pre-training oscillatory spectrum averages calculated over 13 sessions. Dots indicate frequencies where significant differences were observed using a paired t-test ( $p < 0.05$ ).

from resting state EC and EO periods recorded before and after the BCI-HMD training. As previous research has shown, mental fatigue is associated with the increase of theta band power in the frontal area and alpha in the parietal area, as well as with a shift in alpha frequency to the lower alpha band [14]. These changes could be seen as a result of completing a monotonous task consisting of solving simple mathematical problems and lasting up to three hours. These findings were also restricted to the EO condition because subjects solved the task displayed on a screen [14]. In the current study, we observed a significant post-training theta band increase at the frontocentral spatial region, and this was true for the EC condition. This theta increase is also consistent with the findings by Jap, Barwick, and DeGennaro [7, 8, 13].

The alpha effect was less clear. Considering alpha1 and alpha2 frequency sub-bands, we observed an increased EO alpha1 in the right central-parietal region but a decrease of the same alpha1 in the EC condition. Alpha2 decreased in both EC and EO conditions, which was true mainly for the central and central-parietal regions, except

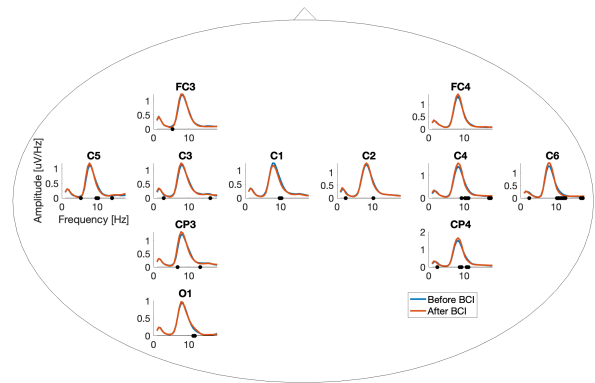


Figure 8: Eyes open (EO) spectrum overlay of pre-training and post-training oscillatory spectrum averages calculated over 13 sessions. Dots indicate frequencies where significant differences were observed using a paired t-test ( $p < 0.05$ ).

for the O1 electrode in the EC condition, alpha2 post-training increased.

The described study analyzed only one participant; therefore, no more robust conclusions can be made from the findings, and broader research is necessary. Instead, the presented study serves as an introduction to the problem of induced mental fatigue in BCI-HMD MI training of stroke patients. The study doesn't address other essential elements that need to be controlled for, such as circadian rhythms, caffeine intake, the quality of sleep, and other phenomena affecting mental fatigue. Therefore, it remains an open question whether the reported post-training differences could be explained by mental fatigue or other possible factors associated with using the BCI-HMD system, such as visual fatigue, lack of motivation or interest, frustration, sleepiness, or even dizziness. The provided questionnaire considered some of those questions, and the obtained subjective scores will be analyzed in future research. Including these elements in further research may lead to a more precise separation of mental fatigue from sleepiness, lack of engagement, or other similar yet different mental phenomena.

## CONCLUSION

The research findings could be utilized to compose customized machine-learning algorithms for motor rehabilitation of post-stroke patients using the BCI-HMD environment. Considering mental fatigue in training sessions could increase rehabilitation outcomes; more specifically, it could suggest improvements in task design and data analysis.

## ACKNOWLEDGEMENTS

This research was supported by the CHIST-ERA (grant 20-BCI-004, ReHaB), Slovak Research and Development Agency (APVV-21-0105, TInVR), and VEGA (grant 2/0023/22) funding programs.

## REFERENCES

- [1] Santos EM dos, Aparecida Fernandes C, Castellano G. Performance of stroke patients using a brain-computer interface during motor imagery: A systematic review. *Research on Biomedical Engineering*. 2023;39:451–465.
- [2] Prasad G, Herman P, Coyle D, McDonough S, Crosbie J. Applying a Brain-Computer Interface to Support Motor Imagery Practice in People with Stroke for Upper Limb Recovery: A Feasibility Study. *Journal of NeuroEngineering and Rehabilitation*. 2010;1(7):60.
- [3] Pichiorri F *et al.* Brain-computer Interface Boosts Motor Imagery Practice During Stroke Recovery. *Annals of Neurology*. 2015;5(77):851–865.
- [4] Shenoy P, Krauledat M, Blankertz B, Rao R, Müller K. Towards Adaptive Classification for BCI. *Journal of Neural Engineering*. 2006;3(1):R13–R23.
- [5] Myrden A, Chau T. Effects of User Mental State on EEG-BCI Performance. *Frontiers in Human Neuroscience*. 2015;9.
- [6] Glader EL, Stegmayr B, Asplund K. Poststroke Fatigue: A 2-Year Follow-Up Study of Stroke Patients in Sweden. *Stroke*;33(5):1327–1333.
- [7] Barwick F, Arnett P, Slobounov S. EEG Correlates of Fatigue During Administration of a Neuropsychological Test Battery. *Clinical Neurophysiology*. 2012;123(2):278–284.
- [8] De Gennaro L *et al.* Neurophysiological Correlates of Sleepiness: A Combined TMS and EEG Study. *NeuroImage*. 2007;36(4):1277–1287.
- [9] Li G, Li B, Wang G, Zhang J, Wang J. A New Method for Human Mental Fatigue Detection with Several EEG Channels. *Journal of Medical and Biological Engineering*. 2017;37(2):240–247.
- [10] Yaacob H, Hossain F, Shari S, Khare SK, Ooi CP, Acharya UR. Application of artificial intelligence techniques for brain-computer interface in mental fatigue detection: A systematic review (2011–2022). *IEEE Access*. 2023;11:74736–74758.
- [11] Eoh HJ, Chung MK, Kim SH. Electroencephalographic Study of Drowsiness in Simulated Driving with Sleep Deprivation. *International Journal of Industrial Ergonomics*. 2005;35(4):307–320.
- [12] Stern JM. *Atlas of eeg patterns*. Lippincott Williams & Wilkins (2005).
- [13] Jap BT, Lal S, Fischer P, Bekiaris E. Using EEG Spectral Components to Assess Algorithms for Detecting Fatigue. *Expert Systems with Applications*. 2009;36(2):2352–2359.
- [14] Trejo L, Kubitz K, Rosipal R, Kochavi R, Montgomery L. EEG-Based Estimation and Classification of Mental Fatigue. *Psychology*. 2015;6:572–589.
- [15] Tran Y, Craig A, Craig R, Chai R, Nguyen. The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. *Psychophysiology*. 2020;57:e13554.
- [16] Foong R *et al.* Assessment of the Efficacy of EEG-Based MI-BCI With Visual Feedback and EEG Correlates of Mental Fatigue for Upper-Limb Stroke Rehabilitation. *IEEE Transactions on Biomedical Engineering*. 2020;67(3):786–795.
- [17] Talukdar U, Hazarika SM, Gan JQ. Motor imagery and mental fatigue: Inter-relationship and EEG based estimation. *Journal of Computational Neuroscience*. 2019;46(1):55–76.
- [18] Rosipal R, Korečko Š, Rošťáková Z, Porubcová N, Vanko M, Sobota B. Towards an Ecologically Valid Symbiosis of BCI and Head-mounted VR Displays. In: *Proc. of the 16th International Scientific Conference on Informatics*. Poprad, Slovakia, 2022, 251–256.
- [19] Wen H, Liu Z. Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal. *Brain Topography*. 2016;29(1):13–26.
- [20] Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*. 2014;44:58–75.

# RECORDING THE SSSEP WITH THE CEEGRID

Jimmy Petit<sup>1</sup>, Matthias Eidel<sup>2</sup>, José Rouillard<sup>1</sup>, and Andrea Kübler<sup>2</sup>

<sup>1</sup>Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

<sup>2</sup>Institute of Psychology, Department of Psychology I, University of Würzburg, Würzburg, Germany

E-mail: Matthias.Eidel@uni-wuerzburg.de

**ABSTRACT:** The feasibility of EEG systems in real-world scenarios, particularly as assistive devices for people with impairments, remains limited by practical issues of conventional cap EEG. However, the emergence of the *cEEGrid*, an unobtrusive around-the-ear EEG system, might offer a solution. While the *cEEGrid* has demonstrated success in measuring event-related potentials, essential for brain-computer interfaces (BCIs) in a variety of settings, its ability to measure steady-state somatosensory evoked potentials (SSSEPs) remains unexplored. Here, we recorded SSSEPs from seven stimulation frequencies in six participants. To allow for a direct comparison, the signal was recorded from a conventional scalp EEG (Brain Products Acticap) and two *cEEGrids* under the same conditions. Results indicate significant SSSEP elicitation with the Acticap, whereas this was only found for one participant with the *cEEGrid*. Amplitudes measured with *cEEGrids* are generally smaller, however, their relative discreet design make them an interesting alternative. Further exploration is necessary to characterise the capabilities of the *cEEGrid* in a potential SSSEP-based BCI application.

## INTRODUCTION

Conventional scalp EEG is often considered impractical in ecological conditions, *i.e.*, daily use in patients' home-care settings. The *cEEGrid* (Fig. 1), developed since 2015, concealed and unobtrusive around-the-ear EEG system, and may thus be a promising alternative to cap EEG systems [1]. The *cEEGrids* have successfully measured event-related potentials, such as the N100, P100 and P300 [1–3]. The P300 in particular is an important input signal for many Brain-Computer Interfaces (BCIs), and has already been recorded by the *cEEGrid* in the visual [4], auditory [2], and even tactile modality [5]. However, its capacity to measure steady-state somatosensory-evoked potentials (SSSEP) has not yet been explored.

The amplitude and the individual frequency of an SSSEP is highly variable between participants [6]. A recent literature review of SSSEP-based BCI observed that a screening procedure is often performed to identify the optimal frequency of stimulation (FOS) for each participant [7].

Here, we report preliminary data from the first six participants of our study. The SSSEP was measured from seven different FOS using two different EEG systems: a con-

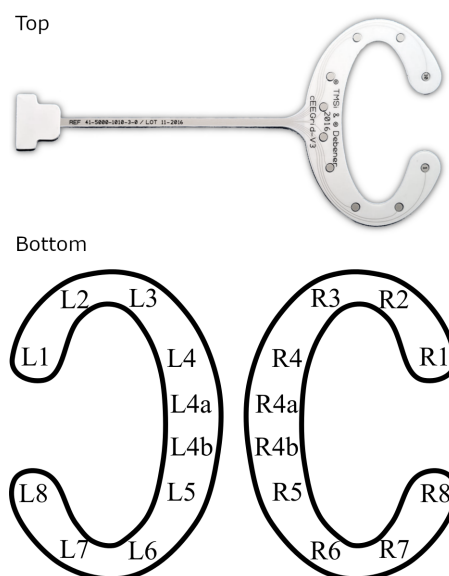


Figure 1: **Top:** A single *cEEGrid*. **Bottom:** Electrode positions of the left and right ear *cEEGrids*.

ventional cap EEG and two *cEEGrids*. We hypothesised that the *cEEGrid* systems can measure SSSEP, but with a lower signal-to-noise ratio (SNR) than the cap EEG.

## METHODS AND MATERIALS

**Methods:** Six healthy participants (5 female, 1 male,  $26.5 \pm 2.3$  years) performed one SSSEP screening procedure per EEG system, starting with either cap or *cEEGrid* in a balanced design. Participants had a 15-minute break in-between. Participants were seated in front a computer screen and equipped with one tactile actuator taped to each wrist (see Fig. 2). Finally, to prevent auditory-evoked potentials from appearing in the EEG data due to the sound of the vibration, the participant wore disposable earplugs during the recordings.

In each trial of the screening procedure, the participant received a train of seven simulations with a duration of 2 seconds, spaced by 0.5 seconds allowing the sensory system to return to an idle state. This train of simulations was preceded by a reference period of 4 seconds. Per stimulation, one of the following FOS were applied: 14, 17, 20, 23, 26, 29, and 32 Hz, on one of the wrists. The stimulation sequence, *i.e.* the combination of FOS and

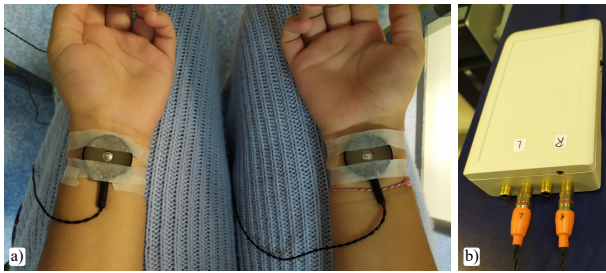


Figure 2: (a) C-2 tactors were taped on the left and right wrists. (b) Stimulation box: a micro-controller establishes the connection to the computer and powers the tactors.

wrist, followed a pseudo-random order, without allowing the same combination of FOS x wrist in direct succession. The protocol is summarised in Figure 3.

During a block in the screening, the participant received 20 trials spaced by a random-length break of 6 to 8 seconds. After that, four blocks with one EEG system (32-channel cap EEG or two cEEGrids, one per ear), plus four additional blocks using the other EEG system were performed. The completion of the full protocol resulted in a total of 40 stimulation epochs of 2 seconds for each FOS, wrist, EEG system, and participant. Data were recorded using a sampling of 500 Hz. AFz was the ground electrode in the cap EEG while R4b was the ground electrode on a cEEGrid pair.

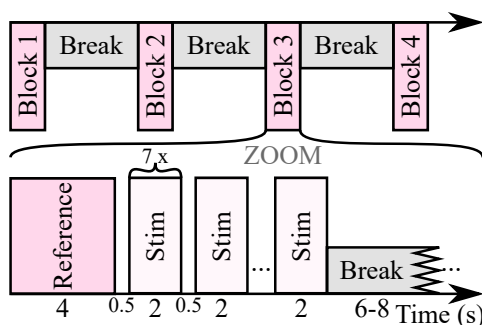


Figure 3: Timeline of the screening session with one EEG system. Each block comprised 20 trials and blocks were separated by 3-minute breaks. This protocol was repeated with the second EEG system 15 minutes after the end of the first screening.

#### Data Analysis:

Before further processing, we applied an artefact rejection algorithm to the raw EEG. If one or more instances of an electrode peak-to-peak amplitude higher than  $100 \mu V$  for at least 5 ms was detected in an epoch, that epoch was excluded from further analysis. We then applied two causal filters: a notch filter at 50 Hz to remove power line noise and a bandpass filter from 5 to 35 Hz.

To analyse the data from the cap EEG, the bipolar channels FC3-CP3 and FC4-CP4 were derived, which cover the primary somatosensory cortex. These channels are known to record SSSEPs originating from the right and left wrist stimulation, respectively [8]. For cEEGrid data, we used linear combinations of cEEGrid channels to approximate cap EEG positions over the same cortical ar-

reas [9], namely  $\widehat{C3}$  and  $\widehat{C4}$ . This is also in line with previous studies, which showed that the cEEGrid measured the highest event-related potential on vertical bipolar channels [1, 5].  $\widehat{C3}$  and  $\widehat{C4}$  are obtained using the following formula:

$$\begin{cases} \widehat{C3} = (L2 + L3)/2 - (L6 + L7)/2 \\ \widehat{C4} = (R2 + R3)/2 - (R6 + R7)/2 \end{cases},$$

using the electrode positions displayed in Figure 1.

We evaluated the SSSEP by two components of the frequency spectrum: the amplitude at the frequency of interest, i.e. the FOS, in relation to the mean amplitude of its neighbouring frequencies in the spectrum. The ratio of these components is a commonly used definition of the SNR in steady-state visually-evoked potentials [10]. The spectrum estimation is performed using the discrete-time Fourier transform algorithm.

To assess the statistical difference between the two components, i.e. to test whether SSSEP amplitude is larger than at neighbouring frequencies, we used the non-parametric<sup>1</sup> one-tailed Wilcoxon signed-rank test. The significance level was corrected using the Benjamini-Hochberg procedure to control the false discovery rate (FDR).

#### Materials:

C-2 Tactors (Engineering Acoustic Inc., Casselberry, USA) were used to administer mechanical vibrations to the participant's wrists, as depicted in Figure 2. The two tactors were powered by a single micro-controller, connected to a host computer. The micro-controller box design was inspired by the work of Pokorny *et al.* [8]. Further information on the initial implementation of our stimulation system can be found in [11]. The stimulation signal comprised a high-frequency sine-wave, around 275Hz, closely aligned with the resonance frequency of the C-2 Tactor. This signal was amplitude-modulated by a square signal at a lower frequency, our FOS. EEG data was recorded using a Brain Products EEG cap (ActiCap) with 32 active electrodes positioned according to the 10-20 international system [9]. Positions AFz and CPz were used as Ground and reference, respectively. The signal was sampled at 500 Hz using a *Brain Products BrainAmp*. Data acquisition and real-time processing were conducted through OpenViBE [12]. Offline signal processing and statistical analysis were carried out using the MNE and SciPy libraries in Python.

## RESULTS

Spectrum analysis revealed a significant SSSEP elicitation for at least one FOS for all participants with the conventional cap system. Participant #1 showed significant SSSEPs at 20, 26, and 32 Hz on the right wrist, while 20 and 32 Hz yielded significant SSSEP amplitudes on the left wrist. Participant #2 displayed a significant SSSEP

<sup>1</sup>Using Shapiro-Wilk tests for normality and an alpha risk level at 0.05, the normality hypothesis could not be maintained on 39.6% of our conventional EEG samples and 55.3% for the cEEGrid samples.

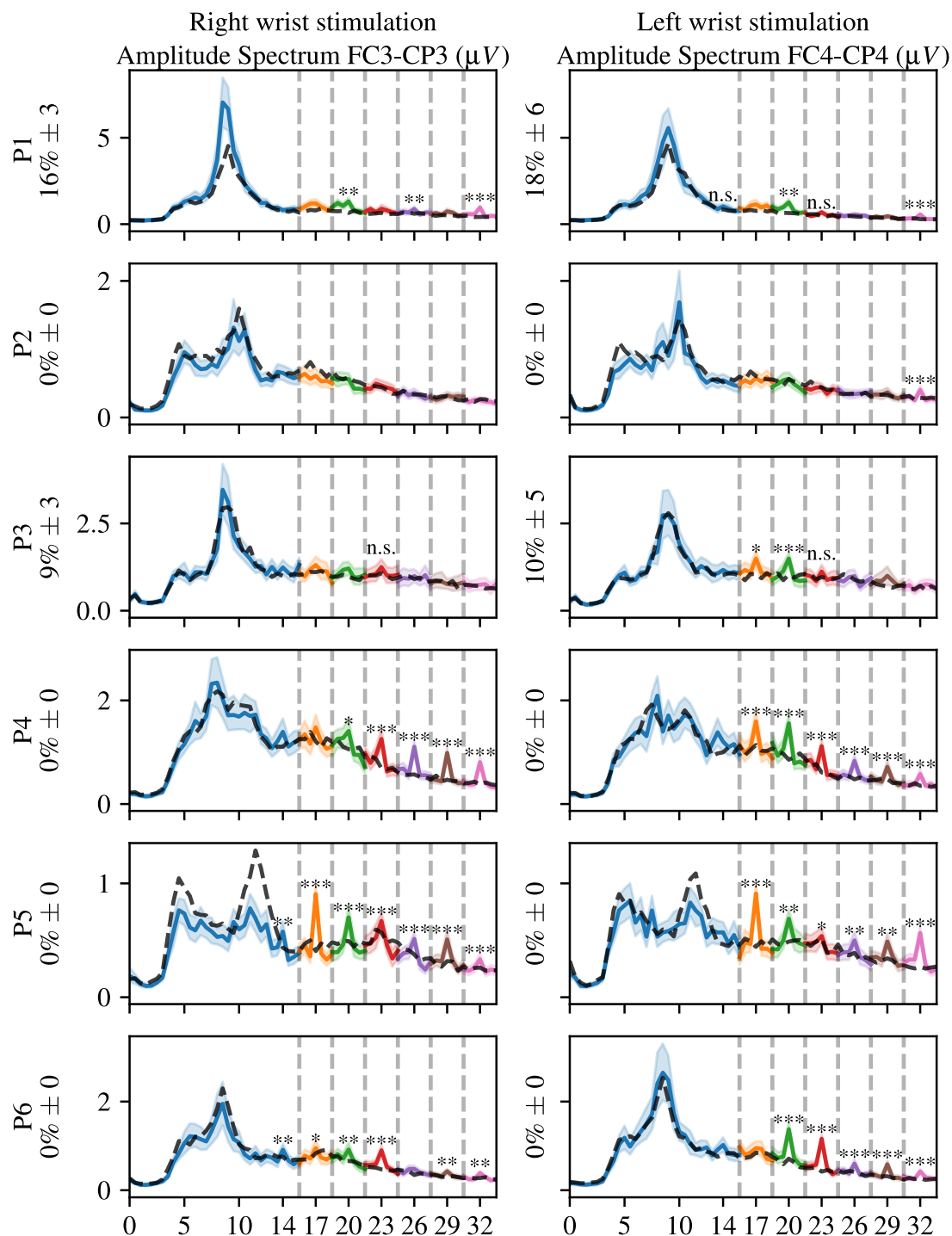


Figure 4: Cap EEG data from channels FC3-CP3 and FC4-CP4. Frequency spectra were concatenated for FOS 14 to 32 Hz (color coded), with lighter shades representing 95% CI. Dashed black lines show the mean spectrum from reference period. Significant results (\*, \*\*, \*\*\*) marked for  $p \leq 0.05, 0.01, 0.001$ , FDR controlled by Benjamini-Hochberg procedure. Null hypotheses rejected at 0.05 alpha level, but not surviving FDR procedure are denoted "n.s.". Numbers next to the participant code show the mean percentage of rejected trials after artefact detection using peak-to-peak analysis.

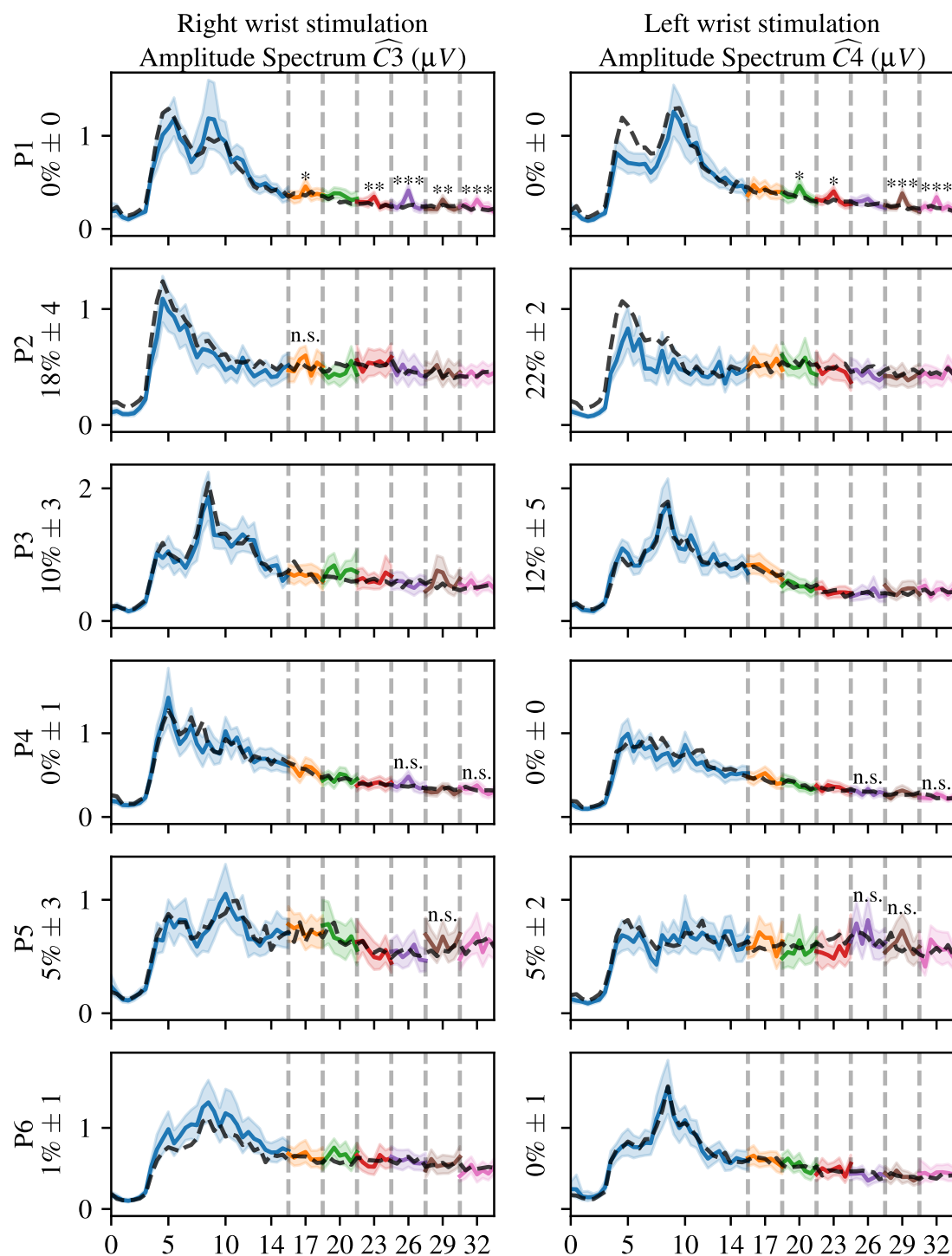


Figure 5: cEEGrid data from channels  $\widehat{C3}$  and  $\widehat{C4}$ . Frequency spectra were concatenated for FOS 14 to 32 Hz (color coded), with lighter shades representing 95% CI. Dashed black lines show the mean spectrum from reference period. Significant results (\*, \*\*, \*\*\*) marked for  $p \leq 0.05, 0.01, 0.001$ , FDR controlled by Benjamini-Hochberg procedure. Null hypotheses rejected at 0.05 alpha level, but not surviving FDR procedure are denoted "n.s.". Numbers next to the participant code show the mean percentage of rejected trials after artefact detection using peak-to-peak analysis.

only at 32 Hz on the right wrist. Participant #3 exhibited only two significant SSSEPs at 17 and 20 Hz on the left wrist. Participant #4 showed significant SSSEPs at every FOS except 14 Hz on both wrists and 17 Hz on the right wrist. Participant #5 displays significant SSSEPs at every FOS, except 14 Hz on both arms. Finally, participant 6 showed significant SSSEPs at almost all FOS, except at 26 Hz on the right wrist, as well as at 14 and 17 Hz on the left wrist. Every spectrum calculated from the cap EEG data are displayed in Figure 4.

Frequency spectra from the cEEGrid recordings are displayed in Figure 5. We found significant SSSEPs for participant #1 at 17, 23, 26, 29, and 32 Hz on the right wrist, while at 20, 23, 29, and 32 Hz on the left wrist exhibited significant SSSEPs. The remaining five participants showed no significant SSSEPs at any FOS, however, some SSSEPs were significant before FDR correction, those SSSEPs are highlighted as "n.s.". For example, a descriptive but finally insignificant SSSEP was found at FOS at 17 Hz (right wrist) for participant #2.

## DISCUSSION AND CONCLUSION

The present study investigated whether the cEEGrid system can be used to record SSSEP. In this preliminary study, we performed two sequential EEG recordings using two different EEG systems: a conventional ActiCap cap system and two cEEGrids around the ear. Our results show that for each participant, a significant SSSEP was recorded with the cap EEG, while only one participant exhibited significant SSSEP amplitude using the cEEGrid. Three other participants showed descriptive increases in activity at the FOS, but this observation was not significant after the false discovery rate correction procedure. Collectively these results indicate that the cEEGrid may help to bridge the translational gap in EEG-based BCI studies, assuming that as a smaller and less obtrusive system, it could be accepted more readily by potential end-users.

Overall, the amplitudes of the SSSEPs found on the cEEGrid were smaller as compared to the cap-recorded SSSEPs, which is well in line with previous observations [2, 4]. This could be explained by the fact that, unlike the cEEGrid, the cap EEG recorded directly over the region of interest in somatosensory-evoked potentials, *i.e.* primary somatosensory cortex. Remarkably however, the cEEGrid still recorded significant SSSEPs under some conditions, using vertical linear combination of electrodes to approximate positions C3 and C4. In addition, the amount of artefacts leading to rejection of a trial using peak-to-peak amplitude analysis was similar between EEG systems. This demonstrates that recording this signal is generally possible with the cEEGrid, without the practical disadvantages of a gel based EEG cap (*e.g.*, gel in the hair, visual appeal).

There were certain limitations to this preliminary study. The EEG systems were applied in a pseudo-random sequential order, unlike in [5, 13]. Thus, without simulta-

neous recordings, we are limited in our ability to examine the correlation between the two systems. This approach was chosen to prevent interference between the active cap electrodes and the unshielded cEEGrid. A limitation arising from the cEEGrid is that the cEEGrids were disinfected and reused (using new sticker) while electrode quality was remained good, as assessed with bucket tests. Generally, cEEGrid application to individuals with facial hair was difficult and required additional tape to ensure a satisfactory impedance. Future versions of the cEEGrid could benefit from solving this problem.

The present study adds to the growing body of literature comparing the cEEGrid with conventional EEG systems, and provides first evidence that the cEEGrid may record SSSEPs. Future studies should address this potentially limiting aspect of the cEEGrid hardware. Finally, for the full study, we will increase the number of included participants to improve statistical power. As for now, our preliminary results are moderately encouraging for SSSEP recording endeavours.

## ACKNOWLEDGEMENT

We thank our dear colleague Pr. François Cabestaing (BCI team, CRISTAL lab., Univ. Lille) for his insightful advices on the experiment and for building the stimulation box.

## REFERENCES

- [1] Bleichner MG, Mirkovic B, Debener S. Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison. *Journal of Neural Engineering*. 2016;13(6):066004.
- [2] Debener S, Emkes R, De Vos M, Bleichner M. Unobtrusive ambulatory eeg using a smartphone and flexible printed electrodes around the ear. *Scientific reports*. 2015;5(1):16743.
- [3] Denk F, Grzybowski M, Ernst SM, Kollmeier B, Debener S, Bleichner MG. Event-related potentials measured from in and around the ear electrodes integrated in a live hearing device for monitoring sound perception. *Trends in hearing*. 2018;22:2331216518788219.
- [4] Pacharra M, Debener S, Wascher E. Concealed Around-the-Ear EEG Captures Cognitive Processing in a Visual Simon Task. *Frontiers in Human Neuroscience*. 2017;11:290.
- [5] Eidel M, Pfeiffer M, Ziebell P, Kübler A. Recording the tactile p300 with the ceegrid – good, but not yet perfect. In: 10th International Brain-Computer Interface Meeting. Brussels, Belgium, 2023, 62.
- [6] Müller-Putz G, Neuper C, Pfurtscheller G. "Resonance-like" Frequencies of Sensorimotor Areas Evoked by Repetitive Tactile Stimulation - Resonanzeffekte in sensomotorischen Arealen, evoziert durch rhythmische taktile Stimulation. *Biomedizinische Technik - BIOMED TECH*. 2001;46:186–190.

- [7] Petit J, Rouillard J, Cabestaing F. EEG-based brain–computer interfaces exploiting steady-state somatosensory-evoked potentials: A literature review. *Journal of Neural Engineering*. 2021;18(5):051003.
- [8] Pokorny C, Breitwieser C, Müller-Putz GR. A Tactile Stimulation Device for EEG Measurements in Clinical Use. *IEEE Transactions on Biomedical Circuits and Systems*. 2014;8(3):305–312.
- [9] Report of the committee on methods of clinical examination in electroencephalography: 1957. *Electroencephalography and Clinical Neurophysiology*. 1958;10(2):370–375.
- [10] Srinivasan R, Russell DP, Edelman GM, Tononi G. Increased Synchronization of Neuromagnetic Responses during Conscious Perception. *The Journal of Neuroscience*. 1999;19(13):5435–5448.
- [11] Rouillard J, Cabestaing F, Bekaert MH, Vannobel JM. Toward a SSSEP-Based BCI Using the Sensory Gating Phenomenon. *Journal Of Clinical Neurology, Neurosurgery And Spine*. 2021.
- [12] Renard Y *et al*. OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments. *Presence: Teleoperators and Virtual Environments*. 2010;19(1):35–53.
- [13] Souto C *et al*. Flex-printed ear-eeeg sensors for adequate sleep staging at home. *Frontiers in Digital Health*. 2021;3:688122.

# INVESTIGATING COORDINATES REPRESENTATION DURING REACHING VIA LOW-FREQUENCY EEG: A PRELIMINARY STUDY

Nitikorn Srisrisawang<sup>1</sup>, Gernot R. Müller-Putz<sup>1,2</sup>

<sup>1</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>2</sup>BioTechMed Graz, Graz, Austria

E-mail: gernot.mueller@tugraz.at

**ABSTRACT:** Understanding how the brain plans reaching movements is crucial in designing a brain-computer interface (BCI) system for motor control. It is still unclear which referencing frame the brain uses to plan the movement. In this study, we investigated the global representation of a referencing frame during reaching planning via a low-frequency electroencephalogram (EEG). Participants were asked to perform directional reaching inward (from the outer target towards the center point) and outward (from the center point towards the outer target) while maintaining gaze on a target such that the reaching in inward and outward conditions should be represented similarly in eye-centered coordinates but differently in shoulder-centered coordinates. We could classify the direction with a peak accuracy of 40.59% but not the inward and outward conditions. The preliminary results confirmed that low-frequency EEG may be globally represented in the eye-coordinates. The classification results suggested that the difference between inward-outward conditions was negligible in low-frequency EEG and could be combined in further analysis.

## INTRODUCTION

When reaching for objects (e.g., a glass of water), our brain must rely on sensory information from several sources to execute the reaching. First, the brain must locate the target and the hand via visual and kinematic information. The displacement and direction between the target and the hand can be calculated, and the reaching can eventually be executed.

The evidence from several studies supports the representation of the target in eye-centered coordinates (the point of origin is at the center of the gaze) in the posterior parietal cortex (PPC) in non-human primates [1], [2], [3], and in humans [4], [5]. In contrast, the brain transforms reaching planning into muscle activations in the shoulder-centered coordinates (the point of origin is at the shoulder) in the sensorimotor area (SMA) [6]. The question remains in which coordinates the brain computes the displacement vector necessary for reaching. Three theories arise: the brain could compute the displacement vector in the shoulder-centered, eye-centered coordinates, or other intermediate coordinates [5], [7].

Several electroencephalographic (EEG) studies provide strong evidence supporting that the information

regarding hand kinematics is represented primarily in a low-frequency EEG in discrete reaching [8], [9], [10] and continuous tracking [11], [12], [13], [14], [15], [16]. Recent studies have also investigated other important aspects of hand kinematics decoding, which could improve the usability of the system: learning effect and adaptation over sessions [17], [18], decoding performance [19], and continuous error processing [20]. As proven useful in the majority of the study, we would like to focus on low-frequency EEG.

Let us consider inward and outward reachings of the same direction where the eyes are always fixated on the target prior to the initiation of the movement. The internally estimated displacement from hand to the target should always point towards the origin of the eye-centered coordinates and the reaching is done towards the target, regardless of the direction of reaching. On the other hand, the displacement should be represented differently in the shoulder-centered coordinates as the inward and outward displacement are represented with different sets of kinematics. We hypothesize that the low-frequency EEG does not carry enough information to distinguish the inward and outward conditions, which also implies weakly that the reach planning is represented in eye-centered coordinates in low-frequency EEG. If this is the case, then a classifier should distinguish the inward and outward conditions with accuracy no better than chance. To answer this, we performed a preliminary study where we collected EEG from five healthy participants performing directional with inward and outward reaching.

## MATERIALS AND METHODS

The experiment was designed based on discrete center-out reaching in 4 directions (up, down, left, and right). Participants performed outward reaching in one direction (e.g., left direction from center) and then inward reaching in the opposite direction (e.g., right direction to center) while fixating their eyes on the target (see Fig. 1c and d). When contrasting the inward and outward reaching from the same direction (e.g., right inward and right outward), the target will always be at the center of the eye-centered coordinates, while the hand position will be in different locations in the shoulder-centered coordinates.

*Participants:* In this preliminary study, data from five healthy participants (1 female) were recorded.

Three were right-handed, and two were ambidextrous, who normally used their right hand to control a computer mouse. The age range was between  $30 \pm 3.16$  years old (mean  $\pm$  std).

**Biosignal recording:** We measured 60 EEG channels and 4 electrooculogram (EOG) channels with a sampling frequency of 500 Hz using BrainAmp amplifiers (Brain Products GmbH, Gilching, Germany). The ground electrode was placed at Fpz, while the reference electrode was placed at the right mastoid. The EOG electrodes were positioned on the outer canthi of both eyes, above and under the left eye.

**Hand movement recording:** A custom-made motion capture system was employed to track hand position via a marker attached to the index finger, and a camera recorded at 30 Hz. The hand position was mapped to the cursor position so that the space between each grid marker (see Fig. 1) was roughly equal to 5cm on the physical plane. Additionally, the hand position was utilized in real time to estimate the direction of the movement. The detected direction of hand movement was used to provide feedback during the trial.

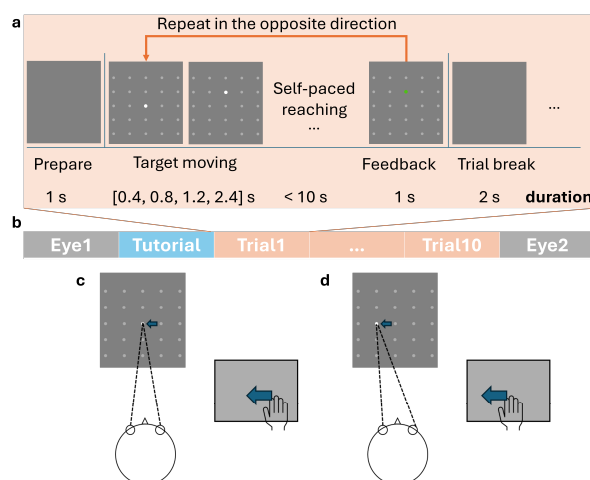


Figure 1: Overview of the experiment (a) timing of a trial, (b) structure of the experiment, (c) and (d) example of left directional reaching from the inward (c) and outward conditions (d) the grey block indicates a motion capture box used in the experiment.

**Experimental paradigm:** The overview of the experiment is visualized in Fig. 1. The experiment was divided into 13 blocks: 2 Eye blocks, 1 Tutorial block, and 10 Trial blocks. For the eye blocks, participants were asked to blink, move their eyes, or rest according to the visual cue. The data was used to correct the eye artifacts in the trial blocks, according to [21]. Another eye block was repeated at the end of the experiment.

During the tutorial, participants practiced reaching in different conditions. The participants were instructed to fixate their eyes on the target before initiating the reaching. The hand position was visualized only during the tutorial so that the participants got used to moving in 2 levels of distance. During the trial blocks, the task was to perform reaching conditions indicated by the target's movement. There were 4 directions (up, down, left, and

right), 2 levels of speed (slow and quick), and 2 levels of distance (near and far), which summed up to 16 different reaching conditions. At the beginning of the trial, a blank screen was visualized for 1 s, indicating a preparation phase. Then, a 5-by-5 grid and a white circle ("target") at the center position appeared. The distance between each dot on the grid was calibrated to match 5 cm in real space, but the hand position was not shown to the participants to force them to always look at the target. The grid provided a guide on possible positions the target could move to (see Fig. 1a). The conditions of the reaching (direction, speed, distance) were randomized for each trial. The direction determined which direction the target moved toward; the speed determined how fast the target moved for 1 grid unit, either within 0.4 s for quick or 1.2 s for slow condition; the distance determined how far the target moved, either 1 or 2 grid units. The time required to move in the far condition was double that required in the near condition (e.g., 0.4 s in the quick near condition but 0.8 s in the quick far condition). As soon as the target stopped moving, the participants were instructed to wait for at least 1 s before initiating the same movement. The color of the target gave feedback: green for the correct direction of the hand movement within 10 s, or red for incorrect (wrong direction, no movement detected within 10 s). The feedback was given for 1 s before the target turned to white, prompting the preparation phase for the participants. The target then moved at the same distance and speed but in the opposite direction, towards the center position. The participants then waited for at least 1 s before initiating movement toward the center position. The feedback was given again for 1 s, and then the screen turned blank, indicating the trial break period. There were 480 trials (48 trials per block) but 960 movements because each trial comprised 2 movements (outward and inward). For simplicity, we treated outward and inward movements as separate trials, so the total number of trials is 960 trials. There were 480 trials for inward and outward conditions and 240 for direction conditions. On average,  $157.2 \pm 88.35$  (mean  $\pm$  SD) trials were rejected due to artifacts, incorrectly performed direction, distance, and speed.

**Processing pipeline:** The measured EEG and EOG signals were processed via a custom script based on EEGLAB [22] on MATLAB version R2019b. The EEG signals were visually inspected, and the bad channels were identified and interpolated. The powerline noise at 50 Hz was removed via 2<sup>nd</sup> order Butterworth notch filter. The signals were downsampled to 200 Hz and then bandpass filtered via the 2<sup>nd</sup> order Butterworth filter between 0.3 – 70 Hz. Eye artifacts were corrected via sparse generalized eye artifact subspace subtraction (SGEYESUB) [21], trained on the eye blocks data. Independent component analysis (ICA) was employed via the FastICA algorithm [23] to identify and remove the artifact component. The IClab plugin was utilized to estimate the probability of each IC component being the artifact. Any IC components that had the probability

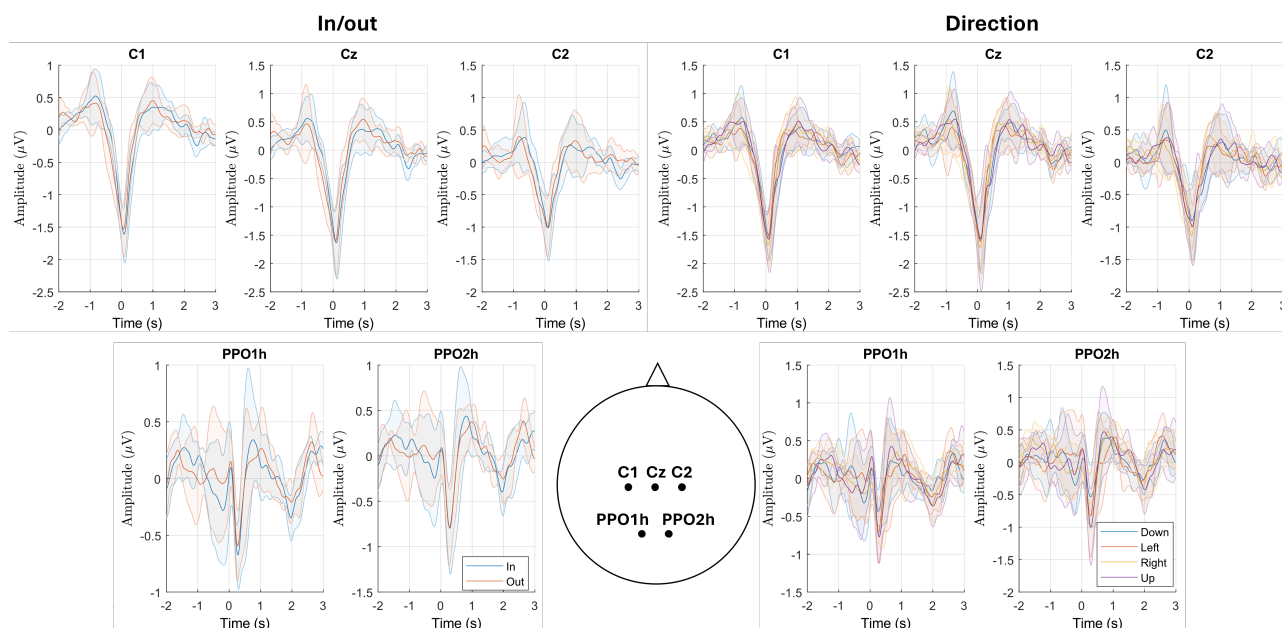


Figure 2: Average MRCP from C1, Cz, C2, PPO1h, and PPO2h in inward/outward and direction conditions. The shaded bands indicate standard deviation (SD) across participants.

of being any non-brain activity higher than 90% were removed and the cleaned signals were recomputed again. The EOG signals were included in the computation of the ICA to aid the identification of the eye artifact. The signals were epoched between -2 to 5 s around when the target stopped, and bad trials were identified based on the amplitude (rejected if amplitude was higher than  $\pm 150 \mu\text{V}$ ) and the statistics of the signals (joint probability, kurtosis). After trial rejection, signals were then low pass filtered with 2<sup>nd</sup> order Butterworth filter at 3 Hz and then epoched around the detected movement onsets between -3 to 3 s. The signals were then re-referenced to a common average reference (CAR).

**Movement onset detection:** The position of the hand was smoothened with a 1st-order Savitzky-Golay filter. The speed was computed by taking the derivative of the vector norm of the smoothened position in 2D. The onsets were detected when the speed exceeded the threshold at 30 pixels per second. The signals were epoched around these movement onsets, and the maximum movement speed and total distance were computed per movement. The median of the maximum speed across all movements was used as a threshold to identify bad movements (too fast or too slow for the speed conditions).

Similarly, the median of the total distance was used to classify bad movements (too far or too near for the distance conditions). Short movement (lasting less than 0.2 s), movement with "incorrect" feedback, and movement outside the movement period were excluded. Finally, the movements that were performed too soon (less than 0.5 s after the target stopped) were excluded as well. On average,  $19.60 \pm 14.04$  trials were rejected due to incorrect behavior (e.g., no movement detected, wrong direction), while  $61.80 \pm 37.10$  and  $55.40 \pm 59.25$  trials were rejected due to incorrect distance and speed, respectively.

**Movement-related cortical potential (MRCP) analysis:** The signals were averaged over the same conditions in directions and inward/outward time-locked to the movement onset. The distance and speed were excluded from further analysis. The MRCPs were averaged over participants, but no statistical tests were performed due to the low number of participants.

**Point-wise Classification:** A shrinkage linear discriminant analysis (sLDA) [24] was employed for classification. Only the EEG signals were considered in this case. AF row channels were excluded due to the residual artifacts that could not be corrected from the eye artifact correction model. The signals were downsampled to 10 Hz to reduce the computational time. The input features for the training were the amplitude of EEG signals within a sliding window with a size between 1 sample (spontaneous) and 10 samples (1-s windows). Due to the maximum window size, the first second of the trial was omitted from the classification. We tested the classification on directions and inward/outward conditions. An sLDA model was trained per time point, so there were 51 sLDA models per condition. The training of each model was done with stratified 10-fold cross-validation. To estimate the effect of the classification, the chance level was simulated with a label shuffling approach [25] according to the number of classes and number of trials per class (TPC). In the direction case, the chance level was determined with 4 classes with 240 TPCs, and in the inward/outward, with 2 classes with 480 TPCs. No statistical tests were performed due to the low number of participants.

## RESULTS

**MRCP analysis:** The shape of MRCPs of different directions and inward/outward conditions is illustrated in Fig. 2. In both the direction and inward/outward

conditions, the amplitude of the MRCPs was strongest at Cz and C1 channels. The minimum peaks in PPO1h and PPO2h lagged behind the ones in C1, Cz, and C2 by 200 ms. The shape of MRCPs differed in C1, Cz, and C2 but not in PPO1h and PPO2h. Only the amplitude of the minimum peaks in the PPO1h and PPO2h differed slightly. Similar effects could also be seen in the direction conditions. Additionally, the shape of MRCPs was similar in the horizontal (left and right) and vertical directions (up and down).

*Point-wise classification:* Fig. 3 visualizes the classification accuracy of the direction and the inward/outward condition. The estimated chance level was determined to be 27.39% and 52.70% for the direction and inward/outward conditions, respectively. In the inward/outward conditions, the average accuracy hovered around the chance level. The peak accuracy was around 0.3 s at 55.53% for the spontaneous classification and at 0 s at 58.16% for 1 s window classification. On the other hand, the accuracy curve of the direction stayed around the chance level before rising well beyond it and then peaks at 0.6 s at 34.24% for the spontaneous classification and at 0.9 s at 40.59% for the 1 s window classification after movement onset and then dips below the chance level again after 1s. The accuracy generally improved with increasing window size.

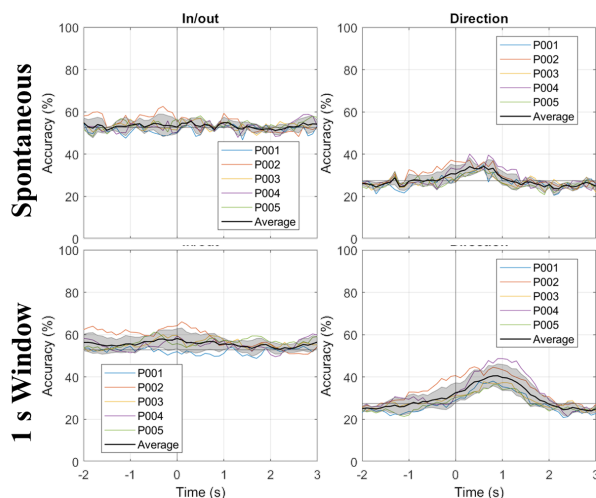


Figure 3: Accuracy curve from spontaneous and 1-s window point-wise classification. Each line indicates the accuracy of each participant. The black line and the shaded area indicate the average accuracy over 5 participants and  $\pm$  standard deviation (SD)

## DISCUSSION

The results suggested that the low-frequency EEG does not carry enough information to distinguish between inward and outward conditions, whereas the amount of information seemed to suffice to distinguish the direction of reaching, as seen in Fig. 3. This seemed to support the hypothesis that the low-frequency EEG carries information that is likely to be represented in the eye-centered coordinates.

Interestingly, we observed in the inward and

outward conditions the similar shape of the MRCP in PPO1h and PPO2h, whereas the shape differed slightly in the C1, Cz, and C2. As discussed earlier, the main difference between inward and outward conditions was in the different representations of the planned displacement vector in the shoulder-centered but not in the eye-centered coordinates. We speculated that this could be related to the underlying referencing frame in shoulder-centered coordinates in SMA [6] (represented as C1, Cz, and C2) and in eye-centered coordinates in PPC [1] (represented as PPO1h and PPO2h). However, we must be careful in drawing any strong conclusion on this topic due to the low number of participants and the lack of evidence linking the non-invasive and EEG measurements.

On the other hand, the directional information of the reaching seemed to be well represented in the EEG signal. The different shapes of MRCP in the horizontal (up and down) and vertical reaching (left and right) could be explained by the different muscle/joint activations when reaching in the left and right in comparison to the up and down direction. Alternatively, this could be due to a mismatch in the actual plane of the movement and the plane of the screen in up and down directions (as the left and right were correspondingly in the same plane) [16], [19]. Nevertheless, the directional information of the reaching could be distinguished with above-chance accuracy.

For further analysis in a subsequent study, the data from the inward and outward can safely be combined. The model showed promising peak accuracy for the direction at 40.59%, which was lower than the accuracy reported in similar directional decoding studies [8], [9], [10]. The classification in this study only represented how informative each time point was in discriminating between conditions but did not reflect the actual decoding accuracy of the detection of the directional reaching. Another major difference was that the participants were specifically asked to fixate their eyes on the target so that the eye movement was completely separated from the movement. This might reduce the amount of information that the decoder could utilize. Further analysis with more participants would be needed to draw a stronger conclusion.

## CONCLUSION

We provided a preliminary result that low-frequency EEG may be globally represented in the eye-coordinates. We speculated that the shape of MRCP in PPC and SMA may represent the underlying referencing frames as reported in earlier studies. The classification results also confirmed that the directional information was encoded in the movement planning, which was sufficient to differentiate between directions but suggested that the inward and outward movement were not differentiable and could be combined for further analysis.

## ACKNOWLEDGEMENTS

We thank Markus Crell for the hand motion-capturing system. We would like to thank the Graz BCI

Team, especially Hannah Pulferer and Markus Crell, for fruitful discussions. NS received funding from the Royal Thai Government.

## REFERENCES

- [1] A. P. Batista, C. A. Buneo, L. H. Snyder, and R. A. Andersen, "Reach Plans in Eye-Centered Coordinates," *Science*, vol. 285, no. 5425, pp. 257–260, Jul. 1999, doi: 10.1126/science.285.5425.257.
- [2] C. A. Buneo, M. R. Jarvis, A. P. Batista, and R. A. Andersen, "Direct visuomotor transformations for reaching," *Nature*, vol. 416, no. 6881, Art. no. 6881, Apr. 2002, doi: 10.1038/416632a.
- [3] A. Battaglia-Mayer *et al.*, "Early Coding of Reaching in the Parietooccipital Cortex," *J. Neurophysiol.*, vol. 83, no. 4, pp. 2374–2391, Apr. 2000, doi: 10.1152/jn.2000.83.4.2374.
- [4] S. M. Beurze, S. Van Pelt, and W. P. Medendorp, "Behavioral Reference Frames for Planning Human Reaching Movements," *J. Neurophysiol.*, vol. 96, no. 1, pp. 352–362, Jul. 2006, doi: 10.1152/jn.01362.2005.
- [5] J. D. Crawford, W. P. Medendorp, and J. J. Marotta, "Spatial Transformations for Eye–Hand Coordination," *J. Neurophysiol.*, vol. 92, no. 1, pp. 10–19, Jul. 2004, doi: 10.1152/jn.00117.2004.
- [6] S. Kakei, D. S. Hoffman, and P. L. Strick, "Sensorimotor transformations in cortical motor areas," *Neurosci. Res.*, vol. 46, no. 1, pp. 1–10, May 2003, doi: 10.1016/S0168-0102(03)00031-2.
- [7] R. Shadmehr and S. P. Wise, *The Computational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning*. in Computational Neuroscience Series. MIT Press, 2004.
- [8] S. Waldert *et al.*, "Hand Movement Direction Decoded from MEG and EEG," *J. Neurosci.*, vol. 28, no. 4, pp. 1000–1008, Jan. 2008, doi: 10.1523/JNEUROSCI.5171-07.2008.
- [9] R. J. Kobler, E. Kolesnichenko, A. I. Sburlea, and G. R. Müller-Putz, "Distinct cortical networks for hand movement initiation and directional processing: An EEG study," *NeuroImage*, vol. 220, p. 117076, Oct. 2020, doi: 10.1016/j.neuroimage.2020.117076.
- [10] V. Shenoy Handiru, A. P. Vinod, and C. Guan, "EEG source space analysis of the supervised factor analytic approach for the classification of multi-directional arm movement," *J. Neural Eng.*, vol. 14, no. 4, p. 046008, Aug. 2017, doi: 10.1088/1741-2552/aa6baf.
- [11] T. J. Bradberry, R. J. Gentili, and J. L. Contreras-Vidal, "Reconstructing Three-Dimensional Hand Movements from Non-invasive Electroencephalographic Signals," *J. Neurosci.*, vol. 30, no. 9, pp. 3432–3437, Mar. 2010, doi: 10.1523/JNEUROSCI.6107-09.2010.
- [12] A. Korik, R. Sosnik, N. Siddique, and D. Coyle, "Imagined 3D Hand Movement Trajectory Decoding from Sensorimotor EEG Rhythms," presented at the 2016 IEEE International Conference on Systems, Man and Cybernetics, 2016, p. 7.
- [13] B. J. Edelman *et al.*, "Non-invasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, p. eaaw6844, Jun. 2019, doi: 10.1126/scirobotics.aaw6844.
- [14] V. Mondini, R. J. Kobler, A. I. Sburlea, and G. R. Müller-Putz, "Continuous low-frequency EEG decoding of arm movement for closed-loop, natural control of a robotic arm," *J. Neural Eng.*, Jul. 2020, doi: 10.1088/1741-2552/aba6f7.
- [15] R. J. Kobler, A. I. Sburlea, V. Mondini, M. Hirata, and G. R. Müller-Putz, "Distance- and speed-informed kinematics decoding improves M/EEG based upper-limb movement decoder accuracy," *J. Neural Eng.*, Aug. 2020, doi: 10.1088/1741-2552/abb3b3.
- [16] V. Martinez-Cagigal, R. J. Kobler, V. Mondini, R. Hornero, and G. R. Müller-Putz, "Non-Linear Online Low-Frequency EEG Decoding of Arm Movements During a Pursuit Tracking Task," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2020, p. 5. doi: 10.1109/EMBC44109.2020.9175723.
- [17] H. S. Pulferer, B. Ásgeirsdóttir, V. Mondini, A. I. Sburlea, and G. R. Müller-Putz, "Continuous 2D trajectory decoding from attempted movement: across-session performance in able-bodied and feasibility in a spinal cord injured participant," *J. Neural Eng.*, vol. 19, no. 3, p. 036005, Jun. 2022, doi: 10.1088/1741-2552/ac689f.
- [18] N. Srisrisawang and G. R. Müller-Putz, "Transfer Learning in Trajectory Decoding: Sensor or Source Space?," *Sensors*, vol. 23, no. 7, p. 3593, Mar. 2023, doi: 10.3390/s23073593.
- [19] N. Srisrisawang and G. R. Müller-Putz, "Applying Dimensionality Reduction Techniques in Source-Space Electroencephalography via Template and Magnetic Resonance Imaging-Derived Head Models to Continuously Decode Hand Trajectories," *Front. Hum. Neurosci.*, vol. 16, p. 830221, Mar. 2022, doi: 10.3389/fnhum.2022.830221.
- [20] H. S. Pulferer, K. Kostoglou, and G. R. Müller-Putz, "Getting off track: Cortical feedback processing network modulated by continuous error signal during target-feedback mismatch," *NeuroImage*, vol. 274, p. 120144, Jul. 2023, doi: 10.1016/j.neuroimage.2023.120144.
- [21] R. J. Kobler, A. I. Sburlea, C. Lopes-Dias, A. Schwarz, M. Hirata, and G. R. Müller-Putz, "Corneo-retinal-dipole and eyelid-related eye artifacts can be corrected offline and online in electroencephalographic and magnetoencephalographic signals," *NeuroImage*, vol. 218, p. 117000, Sep. 2020, doi: 10.1016/j.neuroimage.2020.117000.
- [22] A. Delorme and S. Makeig, "EEGLAB: an open

- source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004, doi: 10.1016/j.jneumeth.2003.10.009.
- [23] A. Hyvärinen and E. Oja, “Independent component analysis: algorithms and applications,” *Neural Netw.*, vol. 13, no. 4, pp. 411–430, Jun. 2000, doi: 10.1016/S0893-6080(00)00026-5.
- [24] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, “Single-trial analysis and classification of ERP components — A tutorial,” *NeuroImage*, vol. 56, no. 2, pp. 814–825, May 2011, doi: 10.1016/j.neuroimage.2010.06.048.
- [25] G. R. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller, “Better than random? A closer look on BCI results,” *Int. J. Bioelectromagn.*, vol. 10, no. 1, pp. 52–55, 2008.

# IDENTIFYING NEW FEATURES FOR BCI CONTROL: SPECTRAL CHANGES IN THE MOTOR THALAMUS REVEAL HAND REPRESENTATION DURING OVERT AND IMAGINED MOVEMENT

Matthew R. Baker<sup>1</sup>, Bryan T. Klassen<sup>2\*</sup>, Michael A. Jensen<sup>1</sup>, Gabriela Ojeda Valencia<sup>1,3</sup>, Samantha A. Banks<sup>2</sup>, Kai J. Miller<sup>1,3\*</sup>

<sup>1</sup>Department of Neurosurgery, Mayo Clinic, Rochester, MN, USA

<sup>2</sup>Department of Neurology, Mayo Clinic, Rochester, MN, USA

<sup>3</sup>Department of Biomedical Engineering, Mayo Clinic, Rochester, MN, USA

\*equal contribution

E-mail: baker.matthew@mayo.edu; miller.kai@mayo.edu

**ABSTRACT:** This study explores the potential of the ventral intermediate nucleus (VIM) of the thalamus as a subcortical signal source for brain-computer interfaces (BCIs). We analyzed spectral changes in the VIM for overt and imagined hand movements during deep brain stimulation (DBS) lead implantation surgery. During task periods, we found suppression of power in the stereotypical beta range (13-30 Hz). Only in one recording site did we find a significant increase in broadband power (65-115 Hz) with overt hand movement, but not for imagined movement. We provide evidence that motor representation in the VIM could act as a subcortical control signal for future BCI applications.

## INTRODUCTION

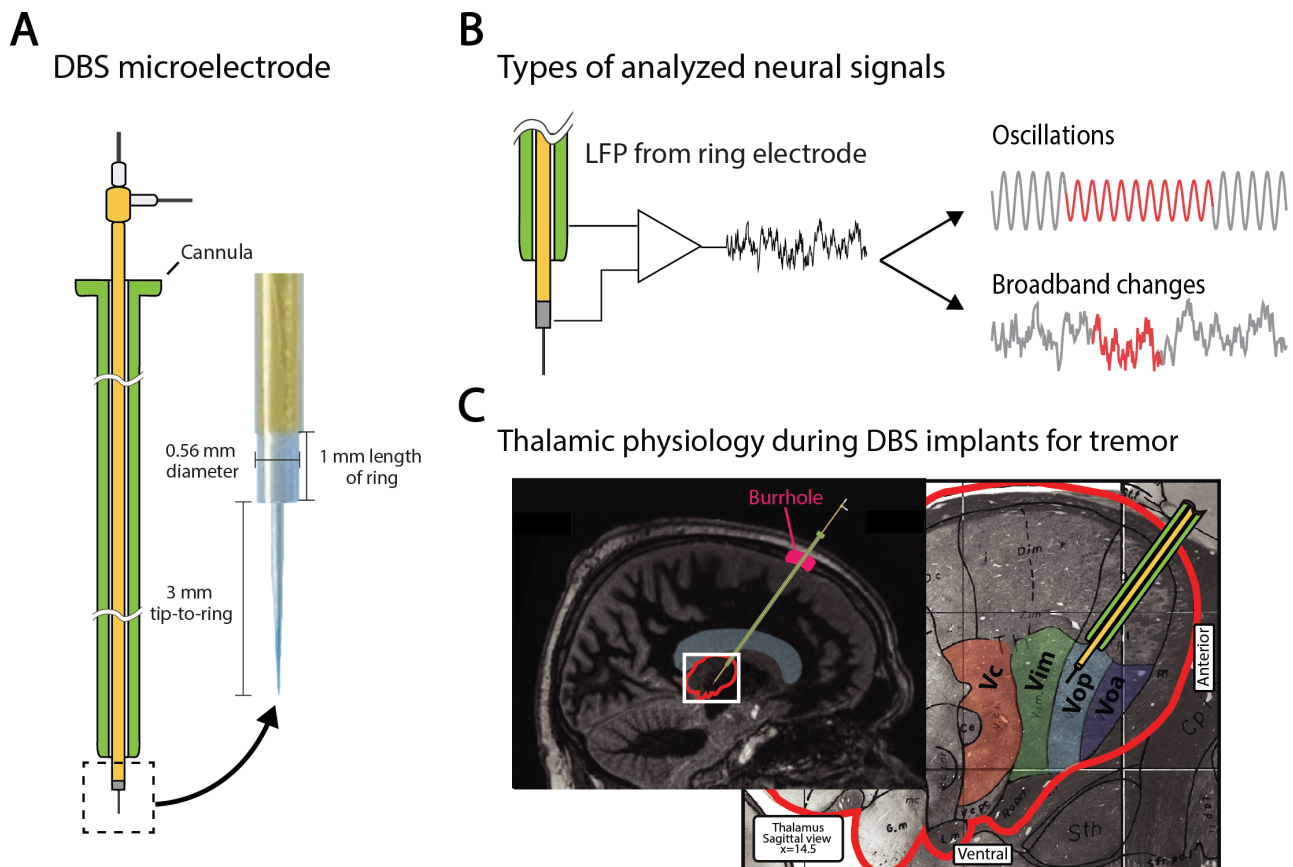
In recent years, brain-computer interfaces (BCIs) have seen remarkable advances, particularly in motor control and rehabilitation tasks. A critical component of BCIs is the ability to sense and accurately decode neural signals into actionable commands, restoring the dysfunction between intent and action seen in many disorders of the motor system. Current BCI applications typically involve electroencephalography (EEG), magnetoencephalography (MEG), electrocorticography (ECoG), or intracortical microelectrode array recordings of the motor or sensory cortices [1–4]. However, identification and development of BCI's using subcortical neural signals may provide several advantages for future applications: (1) subcortical targets may provide unique information regarding motor planning and coordination (2) can be integrated into existing neuromodulation technologies like deep brain stimulation (DBS), and (3) may be necessary in patients with focal damage to cortical regions [5]. The ventral intermediate nucleus (VIM) of the thalamus plays a pivotal role in the modulation and relay of motor signals between the cerebellum and the motor cortex and is a major target of DBS therapy for movement disorders [6]. However, its potential as a BCI signal is unclear. In this case study, we measured spectral changes in the

VIM during overt and imagined hand movements during DBS intraoperative microelectrode macro recordings. During task periods we found significant suppression of power in the stereotypical beta range (13-30 Hz) compared to rest periods. In one recording site, we found a significant increase in broadband power (65-115 Hz) with overt hand movement, but did not observe this as a general phenomenon or in imagined movement conditions. Overall, our findings suggest that the VIM could act as a control signal for future BCI applications.

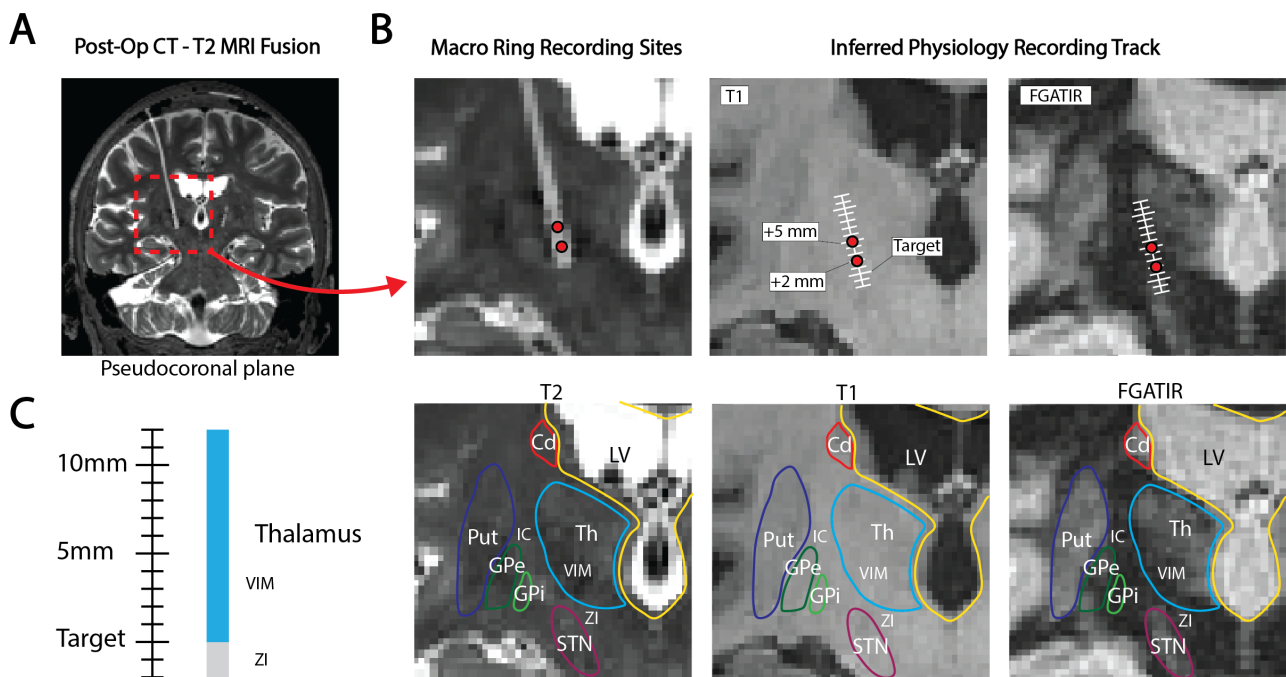
## MATERIALS AND METHODS

*Patient and surgical implantation:* A 49-year-old male patient with multiple sclerosis (MS) presented for bilateral DBS electrode implantation of the ventral intermediate nucleus of the thalamus (VIM). The patient was diagnosed with MS at age 21 after optic neuritis, followed by multiple relapses. He previously received disease modifying therapy including interferon beta-1b and mitoxantrone. MRI revealed multiple lesions throughout his cervical spine, and supratentorial periventricular and juxtacortical white matter, no lesions were seen in the posterior fossa. At age 27 he developed gradual onset of right greater than left sided tremor, with kinetic greater than postural component. At the time of his present evaluation, his only disability was related to his tremor. The patient consented to participate in a research protocol during the awake surgery for implantation of these leads. Mayo Clinic's internal review board approved the study and the consent process (IRB no. 19-009878). Stereotactic targeting and alignment to the left VIM was performed with the Leksell G frame (Elekta, Stockholm, Sweden) and Stealth system (Medtronic, Minneapolis, MN). A cannula was stereotactically passed to the VIM (Fig. 1-2). From the tip of the cannula, a microelectrode (0.5–1 M $\Omega$  platinum–iridium; FHC, Bowdoin, ME) was advanced 15 mm to a target in the VIM (Fig. 1).

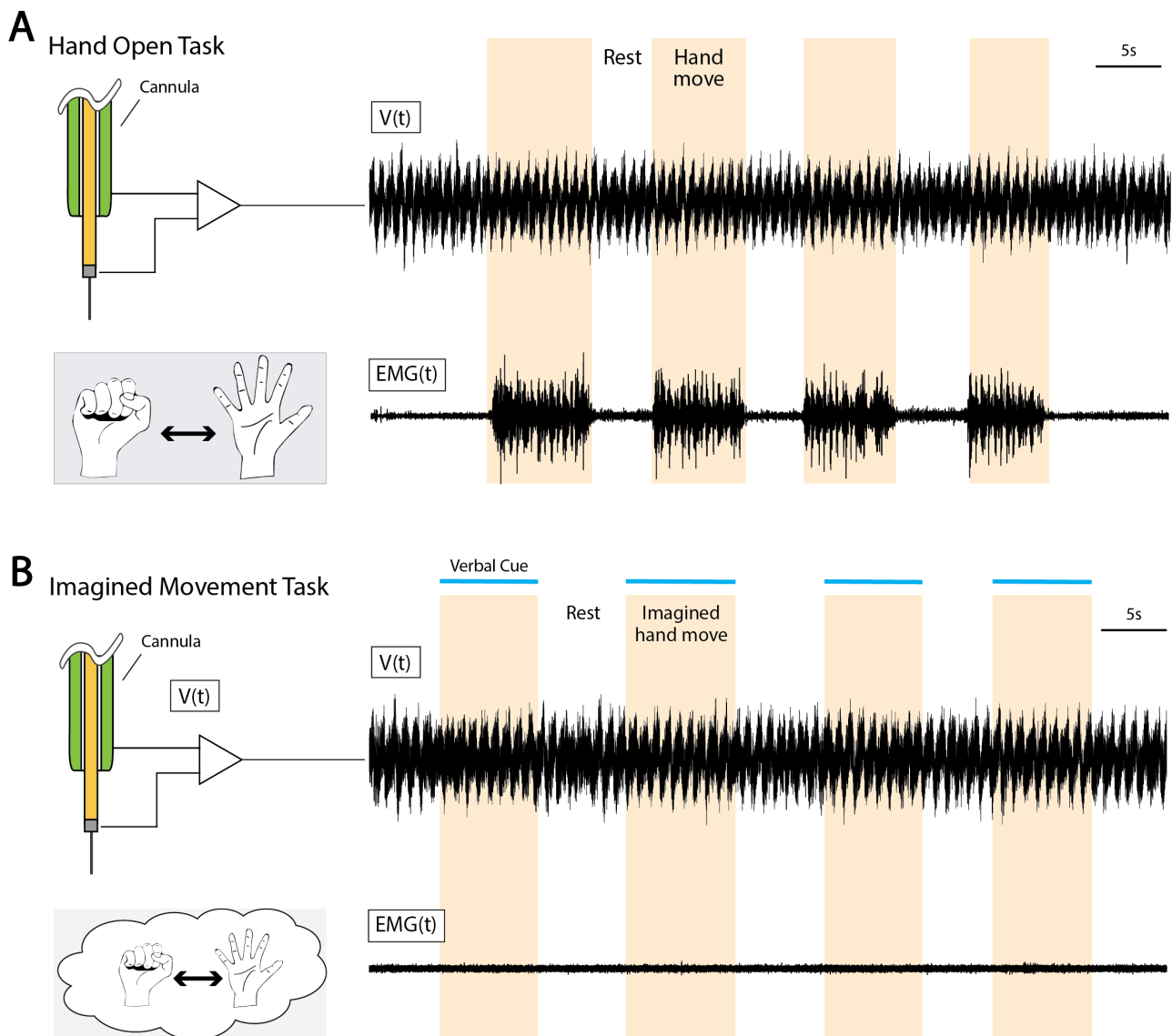
*Motor Task:* Data were collected during a motor task involving opening and closing of the hand. The patient



**Figure 1: Deep brain stimulation (DBS) microelectrode recordings** (A) Microelectrode recording schematic, micro tip, and macro ring dimensions. (B) Micro and macro neural signals are referenced to the cannula and can inform about specific signal modalities. Local field potentials (LFP) analyzed from the macro ring can uncover neural oscillations and broadband changes. (C) Microelectrode recordings are used during thalamic DBS implants to map thalamic neurophysiology. Serial microelectrode recordings may help inform about thalamic anatomical subregions and DBS lead placement. Thalamic anatomy background included from [7].



**Figure 2: DBS lead localization and anatomical segmentation** (A) Post-op CT - T2 MRI fusion showing final lead position on brain anatomy. Pseudocoronal image is resliced in-plane with DBS lead while maintaining midline symmetry. (B) Identification of thalamic microelectrode macro ring recording sites through lead localization. (C) Inferred anatomical segmentation along physiologic recording track.



**Figure 3: Physiologic recordings and behavioral tasks** (A) Microelectrode ring recordings are referenced to the cannula and are recorded over rest and movement periods for a hand open task. Movement periods are segmented by EMG signal. (B) For the imagined movement task, the patient was verbally cued to imagine hand opening.

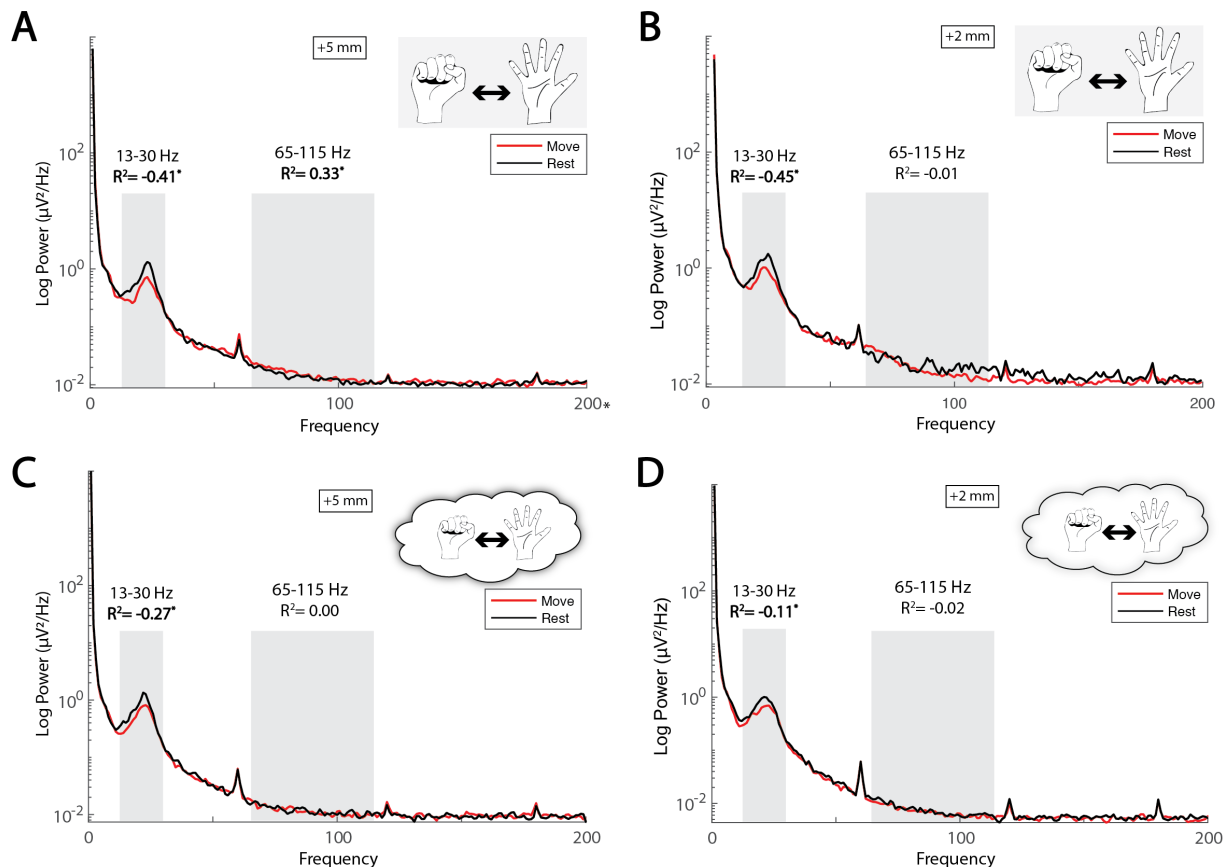
was verbally cued to perform a simple self-paced movement (or imagined movement) with interleaved rest periods. After twenty trials of movement were performed, the task was repeated for kinesthetically imagined movement (Fig. 3). These tasks were chosen based on prior work, which has produced clear results in recordings from brain surface and depth electrodes [8, 9].

**Electrophysiological recordings:** Microelectrode recordings were referenced to the cannula. Voltage time series were recorded with an Alpha Omega system (Alpha Omega, Israel). Recordings were sampled at 44 KHz. EMG was measured from the forearm extensors/flexors and synchronized with microelectrode ring recordings.

**Lead localization:** As illustrated in Fig. 2, microelectrode recording position was determined by co-registration of the postsurgical CT, which includes elec-

trode artifact, and the pre surgical MRI using a normalized mutual information approach. This fused image was then resliced in plane with the DBS lead position using custom MATLAB code [10]. T1, T2, and FGATIR MRI series were overlaid to reveal thalamic and surrounding anatomical borders.

**Signal processing and analysis:** Within each movement or movement imagery trial, averaged power spectral densities (PSDs) were calculated from 1 Hz to 300 Hz every 1 Hz using Welch's averaged periodogram method with 1 second Hann windows and 0.5 second overlap to attenuate edge effects. These trials were defined based on 1) EMG during the movement task, and 2) timing of verbal cuing during the kinesthetic imagery task. Average power from 65-115Hz was used to localize broadband activity as done previously in cortical studies [9]. This range was chosen as it is well above the known range of



**Figure 4: Power spectral densities for behavior tasks** (A) The open hand task at +5mm from target showed significant suppression in the stereotypical beta range (13-30 Hz) and a significant increase in the broadband frequency (65-115 Hz) with hand movement. (B) At +2mm from target, hand movement was associated with a beta range suppression, but no significant change in broadband. (C) Imagined hand movement at +5mm and +2mm (D) from target were associated with significant beta range suppression, but no changes in broadband frequencies. \* and bold text indicates  $p < 0.05$  for paired-samples t-test of rest and move conditions.

most oscillations and avoids ambient line noise at 60 Hz and 120 Hz.

For each task and recording site, we calculated separate signed  $r^2$  cross-correlation values ( $r^2$ ) of the mean spectra from 13-30 Hz and 65-115 Hz. the sign indicates whether power is increasing or decreasing with each task. To calculate a p value for each site and task, we performed a paired-sample t-test comparing 13-30 Hz and 65-115 Hz power for task trials and the rest trials that immediately follow.

## RESULTS

Field potentials were measured at two VIM locations, +2 and +5mm from the DBS lead implant depth (Fig. 2) during the overt and imagined hand movement tasks. In all recording sites a clear oscillation peak emerged in the stereotypical beta range between 13-30 Hz (Fig. 4). During overt hand movement, there was a significant decrease in beta power at both +2 mm ( $r^2 = -0.45$ ,  $p = 1.2 \times 10^{-6}$ ) and + 5 mm ( $r^2 = -0.41$ ,  $p = 1.4 \times 10^{-6}$ ) recording sites. At the +5mm recording site there was a significant increase in broadband power (65-115 Hz;  $r^2 = 0.33$ ,  $p$

$= 2.3 \times 10^{-5}$ ), but no significant increase at the +2mm site ( $r^2 = -0.01$ ,  $p = 0.62$ ). For the imagined hand movement task, we also found significant decreases in beta power at both +2 mm ( $r^2 = -0.11$ ,  $p = 0.04$ ) and + 5 mm ( $r^2 = -0.27$ ,  $p = 4.26 \times 10^{-4}$ ) recording sites. We did not observe any significant increases in broadband power at either +2mm ( $r^2 = -0.02$ ,  $p = 0.42$ ), nor +5mm ( $r^2 = 0.00$ ,  $p = 0.86$ ) depths during imagined movements.

## DISCUSSION

Beta power suppression, characterized by a decrease in the beta frequency band (13-30 Hz) is a well-documented feature in cortical and subcortical motor circuits related to motor planning, execution, and suppression [11]. This beta band has been utilized for a variety of BCI applications, however primarily from motor cortical regions [12]. Our findings suggest that the beta suppression observed with movement in the VIM may also be a suitable control source for future applications.

In primary motor cortex, studies have also found a broadband spectral increase above 50 Hz that we measure between 65 Hz and 115 Hz, which is generally correlated

to neural population firing. These broadband increases are typically more focal than the beta suppression phenomenon, and have similarly been utilized as BCI control signals. While we observed an increase in broadband at the +5mm recording site for overt movement, this was not a general phenomenon in both depths or across tasks - contrary to other motor cortex BCI studies [13]. Similar to other motor structures, the motor thalamus has somatotopic organization. Specifically, the VIM is organized continuously rostrocaudally [14]. It is unclear whether an increase in broadband does not occur in the VIM with imagined movements, or whether our lack of findings could be due to spatial sampling or amplifier noise floor limitations in the intraoperative DBS setting. Future studies will examine the directional somatotopic organization for different types of overt and imagined movements to further investigate this feature.

## CONCLUSION

Overall, we found beta band suppression in the VIM with overt and imagined movements which could act as a subcortical control signal for future BCI applications.

## ACKNOWLEDGEMENTS

We are grateful to the patient who volunteered their time to participate in this research. This work was supported by the NIH U01-NS128612 (KJM), Brain and Behavior Research Foundation (KJM), and the Foundation for OCD Research (KJM). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Our funders played no role in data collection and analysis, study design, decision to publish, or manuscript preparation.

## REFERENCES

- [1] Wolpaw JR, McFarland DJ. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the national academy of sciences*. 2004;101(51):17849–17854.
- [2] Collinger JL *et al*. High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*. 2013;381(9866):557–564.
- [3] Mellinger J *et al*. An meg-based brain-computer interface (bci). *Neuroimage*. 2007;36(3):581–593.
- [4] Snyder LH, Batista A, Andersen RA. Coding of intention in the posterior parietal cortex. *Nature*. 1997;386(6621):167–170.
- [5] Zhao ZP *et al*. Modulating brain activity with invasive brain-computer interface: A narrative review. *Brain Sciences*. 2023;13(1):134.
- [6] Klein J *et al*. The tremor network targeted by successful vim deep brain stimulation in humans. *Neurology*. 2012;78(11):787–795.
- [7] Schaltenbrand G, Wahren W. Atlas for stereotaxy of the human brain. Thieme: Stuttgart (1977).
- [8] Miller KJ *et al*. Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*. 2007;27(9):2424–2432.
- [9] Jensen MA *et al*. A motor association area in the depths of the central sulcus. *Nature neuroscience*. 2023;26(7):1165–1169.
- [10] Richner TJ, Klassen BT, Miller KJ. An in-plane, mirror-symmetric visualization tool for deep brain stimulation electrodes. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020, 1112–1115.
- [11] Pfurtscheller G, Zalaudek K, Neuper C. Event-related beta synchronization after wrist, finger and thumb movement. *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*. 1998;109(2):154–160.
- [12] Miller KJ, Ojemann JG. A simple, spectral-change based, electrocorticographic brain-computer interface. In: *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. Springer, 2010, 241–258.
- [13] Miller KJ, Schalk G, Fetz EE, Den Nijs M, Ojemann JG, Rao RP. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*. 2010;107(9):4430–4435.
- [14] Nambu A. Somatotopic organization of the primate basal ganglia. *Frontiers in neuroanatomy*. 2011;5:26.

# FEASIBILITY OF STEREO EEG BASED BRAIN COMPUTER INTERFACING IN AN ADULT AND PEDIATRIC COHORT

Michael A Jensen<sup>1</sup>, Gerwin Schalk<sup>1</sup>, Nuri Ince<sup>2</sup>, Dora Hermes<sup>2</sup>, Peter Brunner<sup>3</sup>, Kai J Miller<sup>1,2</sup>

<sup>1</sup>Department of Neurosurgery, Mayo Clinic, MN, USA

<sup>2</sup>Department of Biomedical Engineering, Mayo Clinic, MN, USA

<sup>3</sup>Department of Neurosurgery, Washington University School of Medicine, St Louis, MO, USA

E-mail: jensen.michael1@mayo.edu

## ABSTRACT:

**Introduction:** Stereoelectroencephalography (sEEG) is a mesoscale intracranial monitoring method which records from the brain volumetrically with depth electrodes. Implementation of sEEG in BCI has not been well-described across a diverse patient cohort.

**Methods:** Across eighteen subjects, channels with high frequency broadband (HFB, 65-115Hz) power increases during hand, tongue, or foot movements during a motor screening task were provided real-time feedback based on these HFB power changes to control a cursor on a screen.

**Results:** Seventeen subjects established successful control of the overt motor BCI, but only nine were able to control imagery BCI with  $\geq 80\%$  accuracy. In successful imagery BCI, HFB power in the two target conditions separated into distinct subpopulations, which appear to engage unique subnetworks of the motor cortex compared to cued movement or imagery alone.

**Conclusion:** sEEG-based motor BCI utilizing overt movement and kinesthetic imagery is robust across patient ages and cortical regions with substantial differences in learning proficiency between real or imagined movement.

## INTRODUCTION

Brain-computer interfacing (BCI) requires a signal that is strongly correlated to a behavioral state such as movement or speech. Many types of electrical signals can be used for real-time BCI, including scalp electroencephalography (EEG)[1], magnetoencephalography (MEG)[2], electrocorticography (ECoG)[3, 4], and single neuron recordings[5, 6]. Stereoelectroencephalography (sEEG) is a mesoscale measurement that records from the brain volumetrically using depth electrodes[7]. Like ECoG, it represents an intracranial population measure of the summation of local field potentials generated from the n-poles of 100,000s of neurons surrounding the recording electrode. Compared to ECoG, sEEG is not limited to the surface of the cortex. Thus, sEEG allows for sampling from distance cortical and subcortical regions that were not previously possible with ECoG.

Currently, sEEG is utilized in the treatment of drug-resistant epilepsy. Once implanted with sEEG depth

electrodes, patients remain in the hospital for characterization of their seizures. This often takes days to weeks, allowing patients to participate in experiments including brain computer interfaces, if they wish to. Historically, researchers have used spectral changes on the cortical surface to provide feedback [3, 4], allowing individuals to control a cursor on a computer screen in a matter of minutes. Our work describes the extension of this work to sEEG, including its design, implementation, and feasibility.

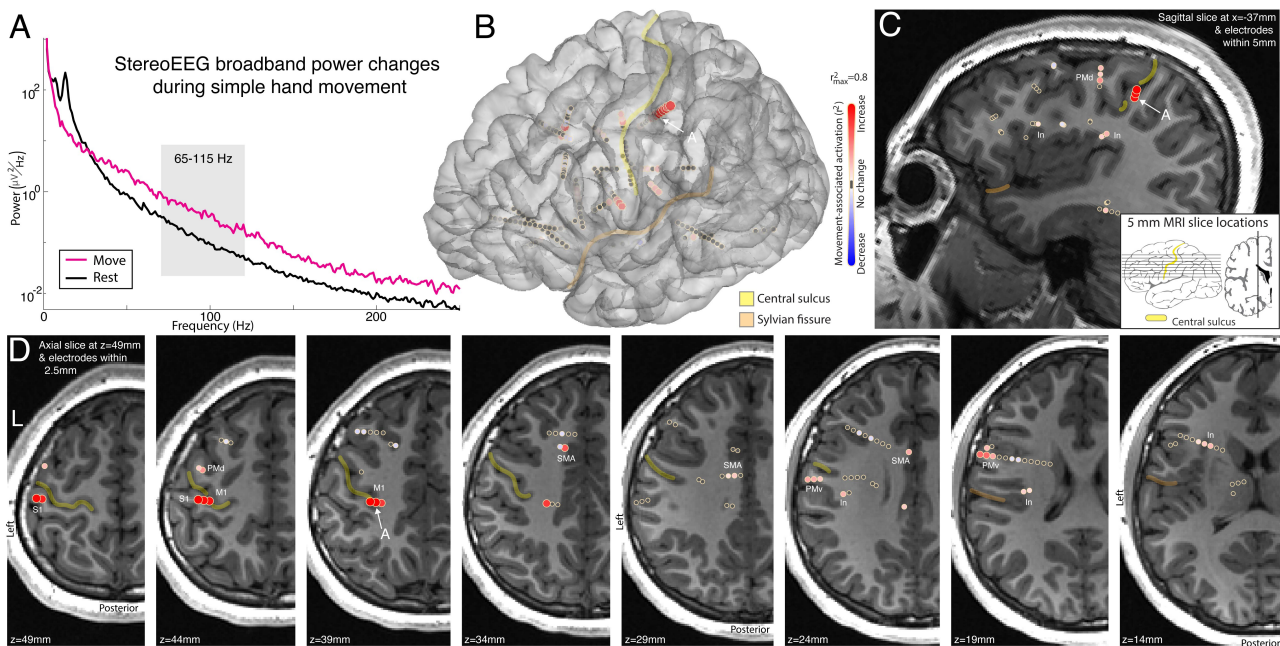
## MATERIALS AND METHODS

*Ethics statement:* The study was approved by the Institutional Review Board of the Mayo Clinic (IRB 15-006530) and conducted according to the guidelines of the Declaration of Helsinki. Each patient or their parental guardian provided informed consent as approved by the IRB.

*Subjects:* Eighteen patients (8 females, 6-37 years of age, Table 1) with drug resistant epilepsy participated in this study, after implantation with 10-17 sEEG electrode leads. Electrode planning was performed by the clinical epilepsy team using brain imaging, typical semiology, and scalp EEG. Electrode locations were not modified to accommodate research; no extra electrodes were added. All experiments were performed in the epilepsy monitoring unit (EMU) or Pediatric Intensive Care Unit (PICU) at the Mayo Clinic in Rochester, MN.

*Lead Placement, Electrode Localization, Re-referencing:* Platinum depth electrode contacts (DIXI Medical) were 0.8mm in diameter with 10-18 2mm length circumferential contacts separated by 1.5mm (Fig 1). Surgical targeting and implantation were performed in the standard clinical fashion. Anatomic locations of electrodes were determined using the steps and tools described previously[8, 9]. All data were bipolar re-referenced such that channels reflect mixed activity at two adjacent electrode contact sites (Figs 1-4). These dipolar channels were plotted using SEEGVIEW, which slices brain renderings, and projects channels to the center of the closest slice [9] in order to present analyses in a more clinically familiar manner.

*Motor Screening Task:* Our motor task involved 3 sec-



**Figure 1. sEEG recordings during movement - Subject 1.** (A) Power spectrum from a sEEG electrode at the sulcal base of primary motor cortex during hand movement (pink) and rest (black), from the recording site noted in panels B-D. (B) Broadband power (approximated by 65-115Hz band) increases during movement compared to rest. (C) Sagittal slice showing electrodes within 5mm of this slice allows viewing broadband power increases on the surface and at depth. (D) As in (C), but for axial slices and electrodes within 2.5mm. Activation maps for movement are shown in the central colorbar (signed  $r^2$ , scaled to 1 maximum, with red/blue reflecting power increase/ decrease with movement). Yellow and peach in B-D indicate the central & sylvian fissures. Note the simultaneous measurement of M1, PMd, PMv, Insula (In), SMA, and S1 (primary sensory), which all show movement-associated broadband power increases.

onds of 1) opening and closing of the hand, 2) side-to-side movement of the tongue with mouth closed, and 3) alternating dorsi- and plantar flexion of the foot with 3 second rest periods interleaved as described previously [8]. The BCI2000 software was used for stimulus presentation and synchronization of (EMG) and sEEG signals [10].

**Offline Signal Processing and Analysis:** All analyses were performed in MATLAB. EMG signal was recorded in parallel to determine the precise timing of movement onset and offset in response to a visual cue. Within each movement trial, averaged power spectral densities (PSDs) were calculated from 1 to 300 Hz every 1 Hz using Welch's averaged periodogram method with 1 second Hann windows to attenuate edge effects and 0.5 second overlap[11]. The averaged PSD for each movement or rest trial was normalized to the global mean across all trials. The PSDs were normalized in this way since brain signals of this type generally follow a  $1/f$  power law and shape[12], so that lower frequency features dominate in the absence of normalization. From each of these normalized single trial PSDs, averaged power in a broadband high frequency band (65-115 Hz) was calculated for subsequent analysis, as previously described [8]. This band was chosen as it captures broadband activity above most oscillations and avoids ambient line noise at 60 Hz and 120 Hz.

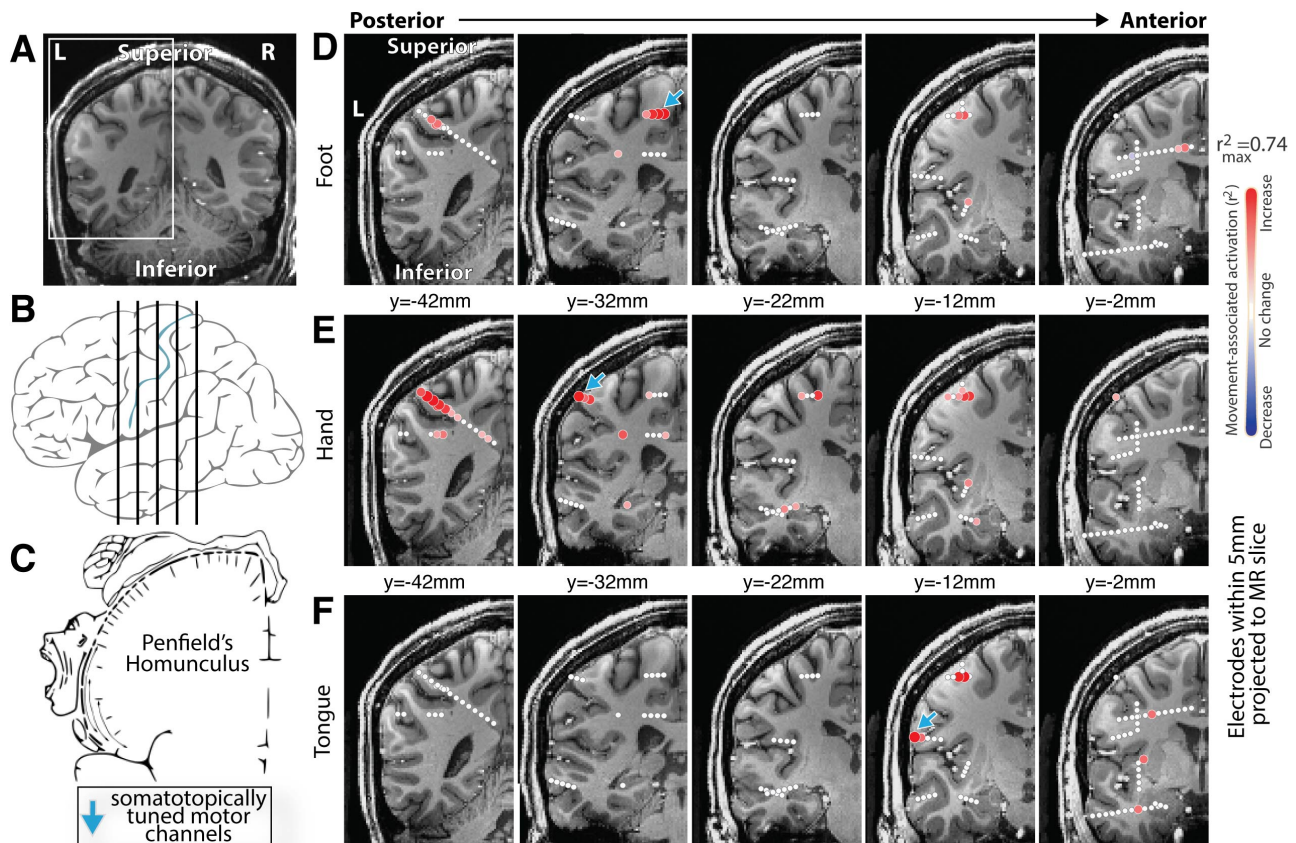
For each bipolar re-referenced channel, signed  $r^2$  cross-correlation values ( $r^2$ ) of the mean spectra from 65-115 Hz were calculated for each movement modality. The  $r^2$  value of each channel was determined by comparing mean power spectra between rest and movement trials

separately. The sign of each  $r^2$  indicates whether power is increasing or decreasing with movement, as illustrated by red and blue circles, respectively, in each figure.

**BCI Task:** We implemented our BCI using the BCI2000[10] software, which applies a spectral estima-

**Table 1. Subject Information,** DNP = did not participate, "/" indicates that modalities pushed cursor in opposing directions.

ID	Age/Sex	Hand	No. Elecs Leads/Side	BCI Modality	Accuracy Overt	Accuracy Imagery
1	10/F	R	198/14/L	DNP	DNP	DNP
2	16/F	L	217/14/B	T/ RF	100	93
3	17/F	R	168/13/R	LH	95	97.5
4	18/M	L	231/14/R	LH	100	89.5
5a	15/F	R	159/10/L	RH	100	50
5b	15/F	R	159/10/L	RH/T	95	70.5
6	13/M	R	196/13/R	H/F	100	73
7a	15/M	L	185/12/L	RH	100	47.5
7b	15/M	L	185/12/L	RH	83.5	50
8a	36/F	R	199/14/B	T/F	100	45
8b	36/F	R	199/14/B	H/F	96	86
9	8/F	R	230/17/B	RH	90	DNP
10	19/M	R	211/15/B	RH	100	100
11	6/M	R	193/13/R	LH	100	DNP
12	37/M	R	237/15/B	T	100	45
13	15/M	R	215/16/B	LH	100	100
14	35/M	R	252/15/B	RH	100	100
15	17/F	R	232/15/B	RH	90.5	DNP
16	36/M	B	195/15/R	LH	100	84
17	12/F	R	232/16/B	LH	100	66
18	16/M	R	254/16/R	LH	96	100



**Figure 2. The homunculus in 3 dimensions - Subject 2.** SEEG allows us to measure the volumetric structure of the homunculus electrophysiologically, shown here for the first time. (A&B) Locations of coronal insets in D-F. (C) Reproduction of Penfield's classic motor homunculus (Wikipedia.org). (D) Comparison of blocks of foot movement vs rest from an SEEG array, plotting movement associated broadband (65-115Hz) change. (E&F) As in D, for Hand and Tongue movement. Note that the classic 2-dimensional homunculus extends into the brain depths, reflecting the volumetric nature of motor representation.

tor to incoming signals using an autoregressive model of the input, operating like a Fast Fourier Transform with a limited number of coefficients. A linear classifier was applied to the feature space of 70-110 Hz power in the channel(s) chosen for BCI to differentiate between movement (or imagined movement) and rest periods allowing for cursor control. During the initial experimental run, BCI2000 adapted this classifier based on the mean and variance of a data buffer (previous 30 sec of incoming data). The threshold was then set to the mean of the data buffer, and the velocity is set to the inverse square root of the variance of the data buffer. These parameters were then fixed for the remainder of the experiment to allow for online learning by each subject.

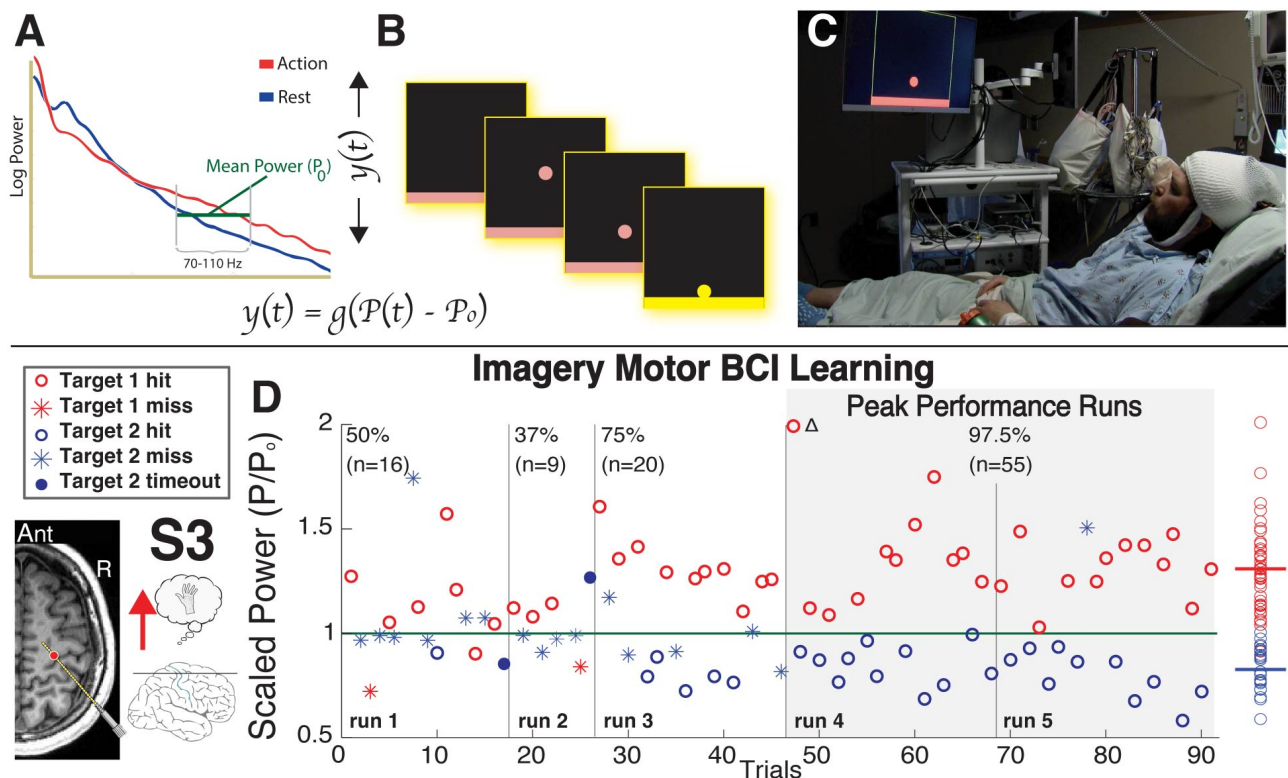
In overt BCI, patients controlled the cursor by moving in order to modulate cortical activity in the pre-selected channels, and in imagery BCI, the cursor was controlled using kinesthetic imagery alone (confirmed by EMG. Both overt and imagery BCIs in this study provided feedback to channels that demonstrated the highest somatotopic tuning based on  $r^2$  values [8] during the motor screening task (Fig. 2).

Prior to BCI, subjects were instructed to associate a target (up vs. down, left vs. right) on the screen with rest or movement (e.g. hand open/close). The target appeared at the top or bottom of the screen 1.5 seconds prior to a red cursor, at which point subjects proceeded to

move/imagined moving or remained still once the cursor appeared (Fig. 3). Subjects were allowed 5 seconds to move the cursor to the target. If the trial was not completed (the cursor hits neither the target nor the opposite edge of the screen), this trial was not considered in the accuracy calculation and a new trial began. Each run was 2 min and allowed subjects to complete as many trials as possible. The first run was for calibration such that the computer could adapt to the power changes in the control channel(s) as subjects alternated between movements/imagined movements or rest.

## RESULTS

**Movement:** After participating in our motor screening task, changes in the power spectral density (PSD) within each sEEG channel were compared between movement and rest periods, and as in previous studies [3, 4, 8]. Movement resulted in suppression of oscillatory activity and an increase in high frequency broadband power (Fig. 1). As broadband power is correlated to local neuronal activity, it served to localize functional representation of movement across the sEEG montage (Fig 1). Germane to our goal of implementing a BCI, this enabled the identification of the somatotopically tuned cortical regions which could generate the control signal in a closed-loop feedback task (Fig. 2).



**Figure 3. Schematic of online BCI feedback** (A) Power from 70-110 Hz in the channel chosen in A determines the direction and velocity of the cursor on screen. (B) Targets are displayed prior to cursors to cue movement or rest and subjects attempt to direct cursors toward the rectangular target. (C) Subjects perform the BCI within their bed viewing a monitor 80-100 cm from their head. (D) Subject 3 learned imagery BCI using a channel in the precentral gyrus across five consecutive runs. The subpopulations of power during trials of opposing targets gradually separated across the learning process until an accuracy of 97.5% was obtained (average accuracy across last two runs).

**Imagery:** Subjects repeated the movement task, but were instructed to kinesthetically imagine performing the cued movement [13, 14]. As demonstrated in ECoG [3, 4], kinesthetic imagery produced an increase in broadband power within motor regions just as during movement (Fig 5).

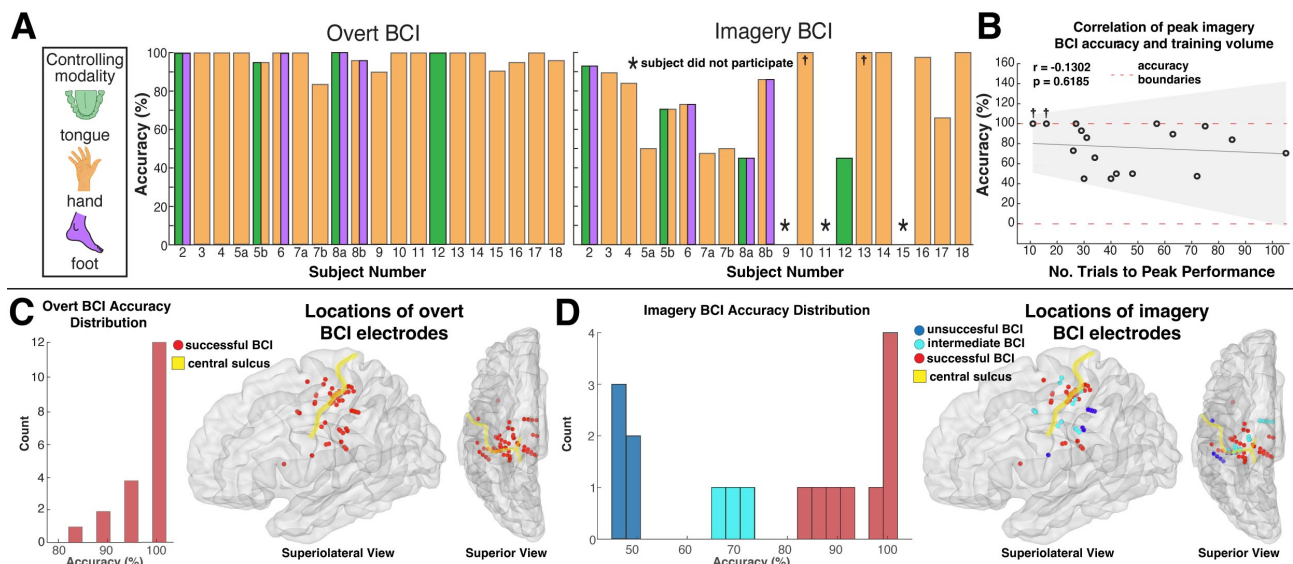
**BCI closed-loop feedback:** Successful BCI control was defined as runs in which the cursor was moved to the correct target in  $\geq 80\%$  of trials for a minimum of 20 trials. The control channels were chosen based on the changes in HFB power associated with movement during the motor screening tasks, and it was modulation of HFB power which controlled the speed and 1-dimensional movement of a cursor on a computer monitor a few feet from the patient's head (Fig. 3). All seventeen subjects established successful control of the overt BCI within minutes. Of these seventeen subjects, fourteen attempted to perform imagery BCI with three subjects attempting two separate BCIs for a total of seventeen. Among these subjects, nine were able to attain successful BCI control, and three controlled the cursor with above chance accuracy (Tab. 1, Fig. 4a). As represented by subject 3, learning imagery BCI occurs across several runs and results in the gradual separation of the average 70-110 Hz power within control channels between trials with opposing targets (Fig. 3). While some patients required many trials to learn the imagery BCI (Fig. 3), there was no relationship between the number of training trials and accuracy (Fig. 4b). The location of the control channel varied across patients (Tab.

1), but the majority of control channels were within the precentral gyrus (PCG). Although control channels outside of the PCG may be assumed to lead to lower accuracies, control channel location did not have a strict relationship to cortical location (Fig. 4c,d).

**Differential cortical engagement across tasks:** Although successful BCI control necessitates broadband power modulation within the pre-selected sEEG channels controlling the BCI, activity patterns within the rest of the motor network are unconstrained. Across several subjects, we see selective engagement and differential activation based on the task being performed. For example, in Subject 3, we see maximal activity in the dorsal pre-motor area during the kinesthetic imagery screening task, and parietal engagement only when feedback is provided (Fig 5). This not only demonstrates that due to the volumetric configuration of sEEG, cortical activity across movement tasks and BCIs can be assessed on the network level, but that cortical subnetworks can be differentially engaged across tasks.

## DISCUSSION

Similar to ECoG studies utilizing high-frequency power to control a motor BCI[3, 4, 15], we demonstrate that both overt and imagery motor BCI can be implemented using sEEG. Even more, sEEG BCI is robust, enabling successful control of overt and imagery BCI in patients as young as six and thirteen years old, respectively. In



**Figure 4. Overt and Imagery BCI accuracy across subjects.** **A** Both overt (left) and imagery BCI accuracies are displayed for each subject. In some cases, subjects performed multiple BCIs that differed in either the pre-selected sEEG channels, controlling modalities, or both. Each BCI within these subjects were assigned a unique bar (e.g. subject 5a vs. 5b). † indicates subjects who attained 100% accuracy in less than 20 trials. **B** The relationship between number of training trials and peak accuracy during imagery BCI was not significant ( $p = 0.6185$ ). **C** Distribution of accuracies during overt BCI (left) with locations of BCI controlling electrodes transformed to the left hemisphere of the MNI152 brain. **D** As in B, but during imagery BCI. Note that electrodes used for BCI control in subject 7 are not shown as their anatomy did not allow for accurate transformation into MNI space.

the majority of cases in which imagery BCI accuracy was not  $\geq 80\%$ , this was due to either a lack of interest from the patient, insufficient time due the rapidly progressing clinical schedule, or an inability to learn the BCI in the allotted time.

Across the learning process, power distributions specific to time periods when each target was shown separated into two clear sub-distributions (Fig 3), with the active targets being more easily hit than inactive targets early on in learning. Presumably this may be due to the more concrete nature of kinesthetic imagery compared to rest which allows subjects to anchor to a tangible process. This is supported by results in subject 5 where the combination of imagined tongue and hand movement increased accuracy. All but two of the sEEG channels controlling the BCI were located in the pre-central gyrus (PCG). This said, selection of control channels within the PCG is not necessary nor sufficient for successful BCI control as 5 out of 8 subjects who failed to establish successful imagery BCI control used control channels in the PCG. In addition, successful imagery BCI control using channels outside of the PCG was performed in several subjects. For example, control channels for subject 4 were in the parietal operculum (Tab. 1), and control channels for subject 6 were in both primary and cingulate areas. Additionally, 14 of 17 BCI modalities involved hand movement, and of the 3 BCIs that did not utilize hand movement, 2 were unsuccessful. The disproportionate representation of the hand in our BCI is due to the sEEG trajectories chosen by the clinicians, but future work should continue to explore motor BCI modalities outside of the upper limb to allow for more powerful studies into the unique characteristics of each modalities.

Although training time was limited, there was no correlation between the number of training trials and the peak

accuracy achieved by each subject (Fig 4b). This indicates that there may be a qualitative difference in the ability to learn imagery BCI across subjects independent of training volume. Certainly, the causal mechanism underlying this difference should be explored further in future work.

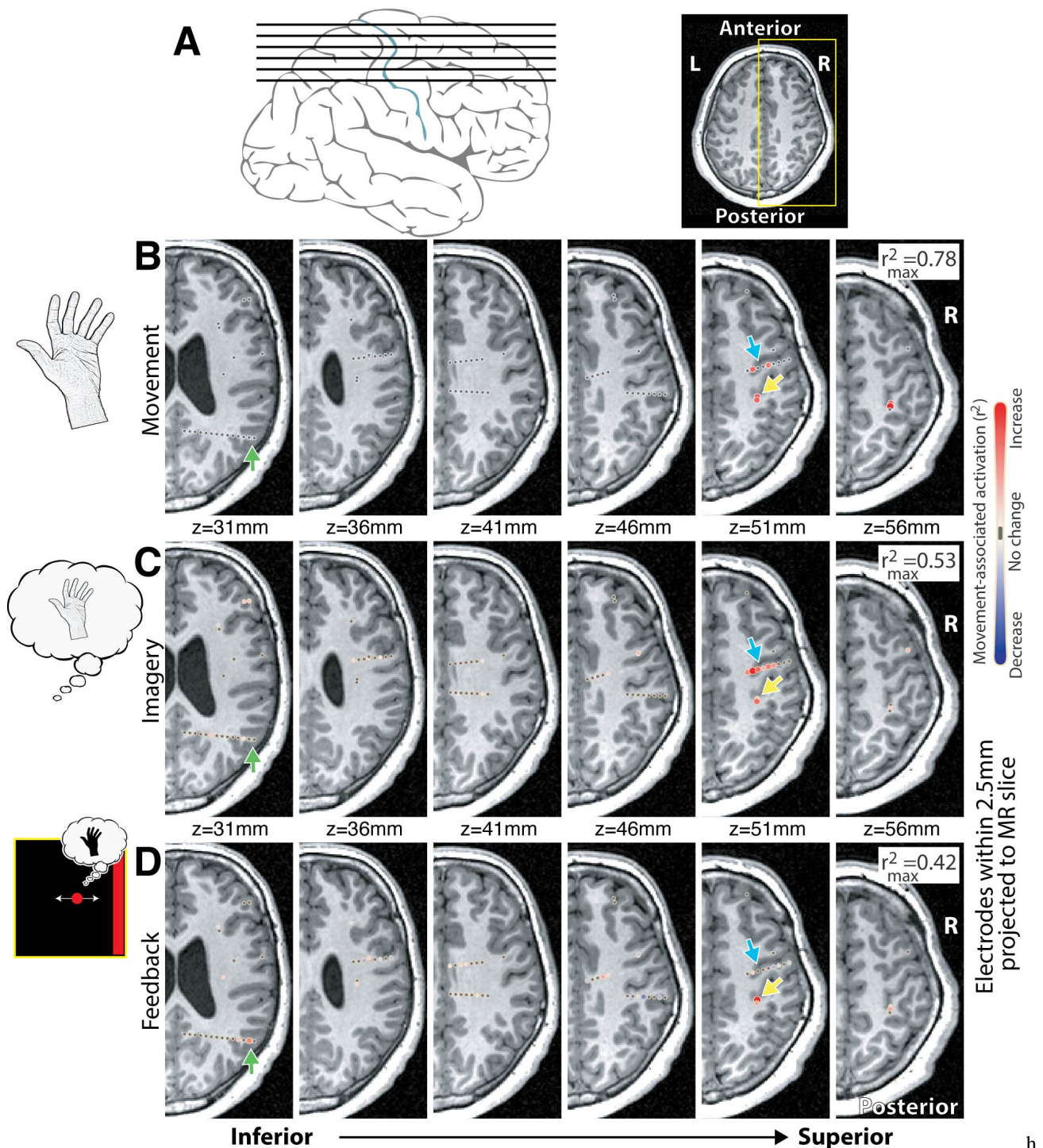
Movement, kinesthetic imagery, and imagery BCI appear to differentially engage the motor network (Fig 4). While detailed exploration of this concept is outside the scope of this work, the examination of the unique roles of non-primary subnetworks of the motor network in abstract learning is a critical advantage of sEEG-based BCI.

## CONCLUSION

One-dimensional motor, sEEG-based BCI utilizing overt movement and kinesthetic imagery is robust across patient ages and cortical regions. Subjects differ in their ability to learn imagery BCI, and further work should explore the mechanism behind this difference.

## ACKNOWLEDGEMENTS

We extend our deepest gratitude to the patients who volunteered their time in the EMU to participate in our research, to Bambi Wessel, Cindy Nelson, and the staff at St. Mary's hospital. This work was supported by the NIH U01-NS128612 (KJM, PB), Brain and Behavior Research Foundation (KJM), and the Foundation for OCD Research (KJM). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Our funders played no role in data collection and analysis, study design, decision to publish, or manuscript preparation.



**Figure 5. Hand movement, movement imagery, and one-dimensional BCI cursor control using sEEG - Subject 4.** (A) Axial insets in (B-D) are as shown here. (B)  $r^2$  maps of hand movement vs rest, broadband 65-115Hz power, as in Fig 1D. (C) Hand movement kinesthetic imagery in the same patient. (D) Map of left hand imagery-based cursor control, comparing left-to-right target presentation times (cursor velocity linked linearly to 65-115Hz power from M1 site indicated by yellow arrow). Note 1) the selective augmentation in recruitment of the PMd (blue arrow) during movement imagery, and 2) the PRR (green arrow) activity selectively during BCI, but not during movement or imagery.

## REFERENCES

- [1] Wolpaw JR, McFarland DJ. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the national academy of sciences*. 2004;101(51):17849–17854.
- [2] Mellinger J *et al*. An meg-based brain-computer interface (bci). *Neuroimage*. 2007;36(3):581–593.

- [3] Miller KJ, Schalk G, Fetz EE, Den Nijs M, Ojemann JG, Rao RP. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*. 2010;107(9):4430–4435.
- [4] Miller KJ, Ojemann JG. A simple, spectral-change based, electrocorticographic brain-computer interface.

In: Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction. Springer, 2010, 241–258.

[5] Snyder LH, Batista A, Andersen RA. Coding of intention in the posterior parietal cortex. *Nature*. 1997;386(6621):167–170.

[6] Collinger JL *et al.* High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*. 2013;381(9866):557–564.

[7] Bancaud J *et al.* Functional stereotaxic exploration (seeg) of epilepsy. *Electroencephalography and clinical neurophysiology*. 1970;28(1):85–86.

[8] Jensen MA *et al.* A motor association area in the depths of the central sulcus. *Nature neuroscience*. 2023;1–5.

[9] Huang H, Valencia GO, Hermes D, Miller KJ. A canonical visualization tool for seeg electrodes. *Annu Int Conf IEEE Eng Med Biol Soc*. 2021;2021:6175–6178.

[10] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. Bci2000: A general-purpose brain-computer interface (bci) system. *IEEE Trans Biomed Eng*. 2004;51(6):1034–43.

[11] Nuttall A. Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1981;29(1):84–91.

[12] Miller KJ, Sorensen LB, Ojemann JG, Nijs M den. Power-law scaling in the brain surface electric potential. *PLoS Comput Biol*. 2009;5(12):e1000609.

[13] Neuper C, Scherer R, Reiner M, Pfurtscheller G. Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial eeg. *Cognitive brain research*. 2005;25(3):668–677.

[14] Guillot A, Collet C, Nguyen VA, Malouin F, Richards C, Doyon J. Brain activity during visual versus kinesthetic imagery: An fmri study. *Human brain mapping*. 2009;30(7):2157–2172.

[15] Schalk G *et al.* Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *Journal of neural engineering*. 2007;4(3):264.

# DETECTION OF MOTION TERMINATION FROM EEG DURING THE EXECUTION OF CONTINUOUS HAND MOVEMENT

Markus Crell<sup>1</sup>, Gernot R. Müller-Putz<sup>1,2</sup>

<sup>1</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>2</sup>BioTechMed Graz, Graz, Austria

E-mail: gernot.meuller@tugraz.at

**ABSTRACT:** Recent advances in the decoding of hand kinematics from neural data and the usage for the control of cursors also prompt the need to detect the begin and end of continuous movements. This study investigates the asynchronous detection of the termination of a continuous hand movement in a handwriting task using electroencephalography data and the power of frequencies in the  $\mu$  and  $\beta$  band. Results obtained with a shrinkage linear discriminant analysis classifier yield a correct determination of the offset in 53.5% (chance level:  $\approx 18\%$ ) of the trials. We show the general feasibility of the proposed method in the detection of the termination of a continuous hand movement and visualize the benefit of the information of the moment of movement termination in a simulated application.

## INTRODUCTION

Communication capabilities for patients using brain-computer-interfaces (BCIs) have recently been greatly enhanced through the decoding of continuous motion from neural data during imagined hand movements [1]. For people in the late stages of amyotrophic lateral sclerosis (ALS) or with any form of locked-in syndrome, such methods can provide highly desirable ways for the interaction with the outside world [2]. Although most improvements have been driven by advances in implantable devices and the resulting improvement of signal quality [1], non-invasive BCIs employing electroencephalography (EEG) have lately shown promising advances in the field of continuous movement decoding [3] and could potentially be adopted in a comparable way in the future. An important aspect of the decoding of continuous movement and the reconstruction of hand trajectories for the usage in cursor control is the discontinuation of the movement. For imperfect control, as is to be expected at the current state of continuous movement decoding, trajectories have to be constantly corrected by the user [4]. While this can be acceptable during the execution of controlling the cursor where users are engaged with the task, unintended movement should be avoided during periods in which patients are not engaged with the cursor control and where an incessantly moving cursor could be irritat-

ing. One way of overcoming this problem is the detection of voluntary movement initiation and termination and the utilization for starting and stopping of the cursor control. The execution of movement is accompanied by different neural phenomena of which mostly the movement-related cortical potential (MRCP) and event-related de-/synchronization (ERD/ERS) have received major attention. While MRCPs [5] are detectable directly in the time-series EEG and are time- and phase-locked to the start of the movement [6], oscillatory components (ERD) in  $\mu$  and  $\beta$  frequencies decrease during planning and execution of a movement until the motion is terminated [7, 8]. At the point of termination, the desynchronization is followed by a period of increased  $\beta$  synchronization before returning to the baseline, also known as post-movement  $\beta$  synchronization (in short  $\beta$  rebound). ERD and ERS are phenomena which are typically observed in  $\mu$  and  $\beta$  frequencies of the EEG and have been extensively studied for different movement tasks. While a large body of research has focused on the detection of movement onsets [6, 9, 10], few studies have explored the detection of movement termination. Only limited work examined the usage of ERD/ERS patterns for the termination of movement imagination of single, short foot dorsiflexion [11, 12]. They showed that the classification of the  $\beta$ -ERS patterns proved more reliable and resulted in a higher detection accuracy for the imagined movement. Hortal et al. reported similar classification accuracies for the detection of start and stop of the gait cycle from  $\mu$  and  $\beta$  frequency power [13]. In a related approach, Bai et al. employed  $\beta$ -ERS to classify repetitive imagined and executed wrist movement [14]. They reported overall high discriminability between execution and termination and better performance in the execution case than for imagination. Noticeably, the mentioned studies used either short single or repetitive movements. More recently, Orset et al. investigated the detection of termination of sustained movement imagination of the hand, achieving an accuracy of 76.2% [15]. While these studies form a basis for the detection of discontinuation of movements as needed for the stop of cursor control from EEG, they do not integrate the self-initiation of the termination of a continuous movement as is inherent to the control of a

computer cursor.

We aimed to detect the movement termination (herein termed *offset* in contrast to the *onset* of a movement) of a continuous and autonomously executed hand motion. We designed a paradigm in which participants were tasked with the writing of letter trajectories, which constitute a continuous hand and finger movement. Although the movement onset was defined based on a cue, the stop of the motion was dependent on the letter itself as well as on the speed and size of the writing trajectory, which varied among participants and over the duration of the paradigm.

## MATERIALS AND METHODS

**Experimental Paradigm:** The study was conducted among 22 healthy, right-handed participants with a mean age of  $27.5 \pm 3.92$  years (mean  $\pm$  standard deviation). Each subject participated in a paradigm with an approximate duration of 2.5 h. Participants were fitted with an EEG cap equipped with 60 electrodes which were placed on the head according to a standard 10-10 EEG montage. Four electrodes were positioned on the outer canthi of the eyes as well as above and below the left eye to measure electrooculogram (EOG) signals induced by eye movements. The experiment consisted of (1) an instruction phase during which participants were informed about the measurement and the paradigm, (2) a measurement of specified eye movements for the elimination of eye artifacts in EEG data and (3) the execution of a session of the paradigm. Steps 2 and 3 were repeated twice and separated by a break of approximately 10-20 min. The specified eye movements were part of the SGEYESUB algorithm and are described further in [16]. One session of the paradigm contained seven (first session) and eight (second session) runs separated by breaks of 60 s. During each run 40 trials were performed in each of which one out of 10 letters (*a, d, e, f, j, n, o, s, t, v*) was written with the index finger of the right hand. Finger movements were recorded with a custom motion capture algorithm tracking a visual marker applied to the right index finger of the participant. During trials, participants observed a screen and waited for a letter to fade onto the screen (duration: 2 s), stay on the screen at full opacity (0.5 s) and fade out again (2 s). As soon as the letter was invisible, participants started to write the displayed letter with their finger and stopped at the last point of the letter without any further movement. The maximum duration of writing a letter was limited to 4 s, after which the next letter would be faded onto the screen. Every letter was written 60 times, resulting in a total of 600 executed movements. Due to technical problems and early termination, two participants had to be excluded from the study.

**Data Acquisition and Preprocessing:** EEG was acquired from two biosignal amplifiers (BrainAmp, Brain Products GmbH, Germany) at a sampling rate of 500 Hz. The signals were bandpass-filtered between 0.3-70 Hz and Notch-filtered at 50 Hz to eliminate powerline noise. Eye artifacts were removed using the SGEYESUB algo-

rithm [16] and signals from the EOG channels. Residual eye and muscle artifacts were removed by applying independent component analysis. The EEG was finally re-referenced using a Laplacian reference and 24 channels on the outer periphery of the montage including the EOG channels were removed from the dataset. We further extracted the start and stop of the motion from the movement data using the falling and rising edge of the trajectory speed. The kinematic data was recorded at 30 Hz and smoothed with a Savitzky-Golay filter (first order polynomials and window length of 200 ms).

**Neural Correlates of Movement Termination:** An analysis of the neural correlates of the offset of continuous movements was conducted using ERD/ERS patterns. 20 frequency bands between 0.3 and 40 Hz were extracted using Morlet wavelets [17]. The data was split into trials of length  $[-2.5; 4.0]$ s relative to (a) the start of the movement or (b) the stop of the movement. We then extracted the relative change in power of the time-frequency data  $A$  to the reference period  $R$ , in this case the whole trial [7].

$$ERD/ERS\% = \frac{A - R}{R} * 100\% \quad (1)$$

Finally, the trials were averaged aligned according to (a) and (b) for each channel. We additionally examined the delay between movement onset and offset from the kinematics to compare with the spread of ERD/ERS patterns when aligned to the onset of the movement.

**Movement Offset Detection:** Detection of the movement offset was implemented by classifying windows of neural data into *movement* and *no movement* and determining the falling edge. We extracted time-frequency data around six center frequencies (8, 12, 16, 20, 24 and 28 Hz) using Morlet wavelets to cover the range of  $\mu$  and  $\beta$  frequency bands. We obtained the slow changes in power, equivalent to the envelope of the signal, of specific frequency bands by lowpass filtering the band-filtered data to 12 Hz and subsequently downsampled the features to 30 Hz to reduce the amount of data. The relative change in frequency power was calculated according to Eq. 1 with the reference period set to  $[-2.5; 0]$ s relative to the movement onset. The trials were then cut to include the time period  $[0.2, 4.0]$ s relative to the start of the movement. Trials in which the offset of the movement happened more than 4 s after the cue onset were discarded. The data was then labeled as *movement* until the movement offset and as *no movement* between the movement offset and the end of the trial. Windows of length  $w_L$  between  $[t_0 - w_L; t_0]$  were created to classify the label at  $t_0 - lag$  with a lag of  $\frac{w_L}{2}$  to incorporate non-causal information for the classification. (see Fig. 1a).  $t_0$  was shifted with a stride of one sample to generate the labeled data. A 5-fold cross-validation procedure was employed and a shrinkage linear discriminant analysis was used for the classification. For each participant, a separate model was trained and evaluated. After training the models, windows were classified sequentially per trial to retain the order and allow for a reconstruction of the labeling per trial from the classified data. We then defined

a detection strategy for the definition of movement offsets from the classified data by generating the majority vote from five consecutive classifications and choosing the first falling edge as movement offset after which the subsequent ten samples (i.e.  $\approx 330$ ms) were successively classified as *no movement*. For each trial, a movement offset was defined as correctly identified when the detected offset was within a range of  $\pm 330$  ms of the actual offset. Since the window length of each trial was 3.8 s, the chance level for a random classifier was approximately 18% with a level of statistical significance of 21.1% calculated according to [18]. We tested window sizes of  $w_L \in [200, 400, 500, 600, 800]$ ms and a lag of  $\frac{w_L}{2}$ .

**Simulated Application - Handwritten Letters:** To visualize the benefit of the information of the moment of termination of the continuous movement, we employed a simulated application in which handwritten letters were classified from reconstructed trajectories. We simulated imperfect reconstruction of trajectories from neural data by adding noise to the measured  $x$  and  $y$  positions obtained during the handwriting task. The noise, derived from lowpass-filtered white Gaussian noise, ranged from zero to a maximum amplitude of 1.5 times the average letter size. We then constructed windows of 4 s starting from the movement onset and constructed images of the handwritten trajectory. This procedure was executed for different conditions:

- **NN:** No additional noise was added to the recorded trajectories and the full four second window was used for the image construction.
- **AN:** Additional noise as described above was applied to the trajectories and the full four second window of noisy trajectory data was used for the image construction.
- **AN-C:** Additional noise as described above was applied to the trajectories, which were cut after the movement offset, and only the shortened window was used for the image construction.
- **AN-P:** Additional noise as described above was applied to the trajectories, which were cut after the movement offset according to the predictions from the classifier, and only the shortened window was used for the image construction.

These conditions are visualized in Fig. 1b. A convolutional neural network (CNN) for the detection of handwritten letters from images was employed to classify the constructed letters. CNNs are often utilized in the recognition of handwritten characters since they have proven to yield high accuracies [19]. We repeated a 5-fold cross validation procedure twice to generate reliable classification results.

## RESULTS

**Neural Correlates of Movement Termination:** An analysis of the lag between movement onset and offset per

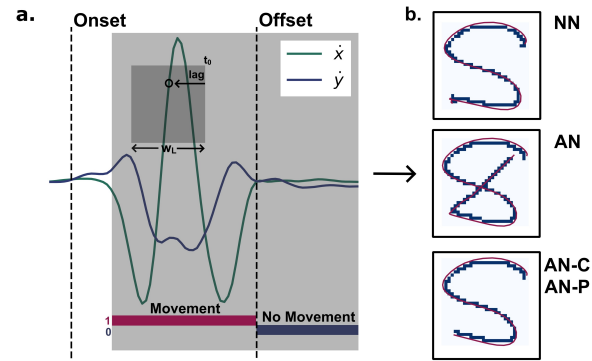


Figure 1: (a) Schematic of the movement kinematics used for the detection of onset and offset and the labeling of the data for the classification. Labels are chosen 200 ms after the movement onset until 4 s after the movement onset and are classified from frequency data of length  $w_L$  using a lag to incorporate non-causal information. (b) The illustration shows letters reconstructed from kinematics and transformed into an image for classification using a CNN. Letters are shown for the conditions no-noise (NN), additional-noise (AN) and additional-noise-cut (AN-C/AN-P).

trial showed that the average movement was terminated after 1.9 s with a standard deviation of 0.6 s. We also investigated the difference between movement onset and offset for individual letters with the distribution shown in Fig. 2a. The largest difference in the average duration of movements between letters occurred for letters  $d$  ( $2.44 \pm 0.62$ s) and  $v$  ( $1.33 \pm 0.43$ s). We also found an influence of the writing duration on the distribution since the coefficient of variation increased with longer average duration of the letters (Pearson's  $r$ : 0.73,  $p$ -value: 0.017). We then calculated the ERD/ERS maps for the movement onset (Fig. 2b upper images) and movement offset (Fig. 2b lower images) aligned trials. Desynchronization (relative decrease of power in the frequencies) is shown in red while synchronization (relative increase of power in the frequencies) is shown in blue. Movement onset centered ERD/ERS maps are also overlayed with an aligned distribution map of the movement offset to indicate the moment of termination of the writing motion. An ERD in  $\mu$  and  $\beta$  frequencies during the movement and a post-movement  $\beta$  synchronization can be observed. Furthermore, a  $\mu$ -ERS can be observed, however, delayed to the  $\beta$ -ERS. Those patterns are more pronounced at channel C1, contralateral to the movement of the right finger.

**Movement Offset Detection:** The maximum accuracy for the detection of movement offsets was achieved for a window length of  $w_L = 500$  ms with an average accuracy of  $53.3 \pm 11.9\%$  (mean  $\pm$  standard deviation) over all participants. The subject with the best performance reached an accuracy of 77.5%. Average classification accuracies achieved with different window lengths yielded lower but comparable results ( $49.9 \pm 13.0\%$ ,  $53.2 \pm 12.4\%$ ,  $52.7 \pm 12.0\%$ ,  $51.4 \pm 11.0\%$  for window sizes 200, 400, 600 and 800 ms, respectively). For all participants, the classification accuracy was above chance. The influence of the

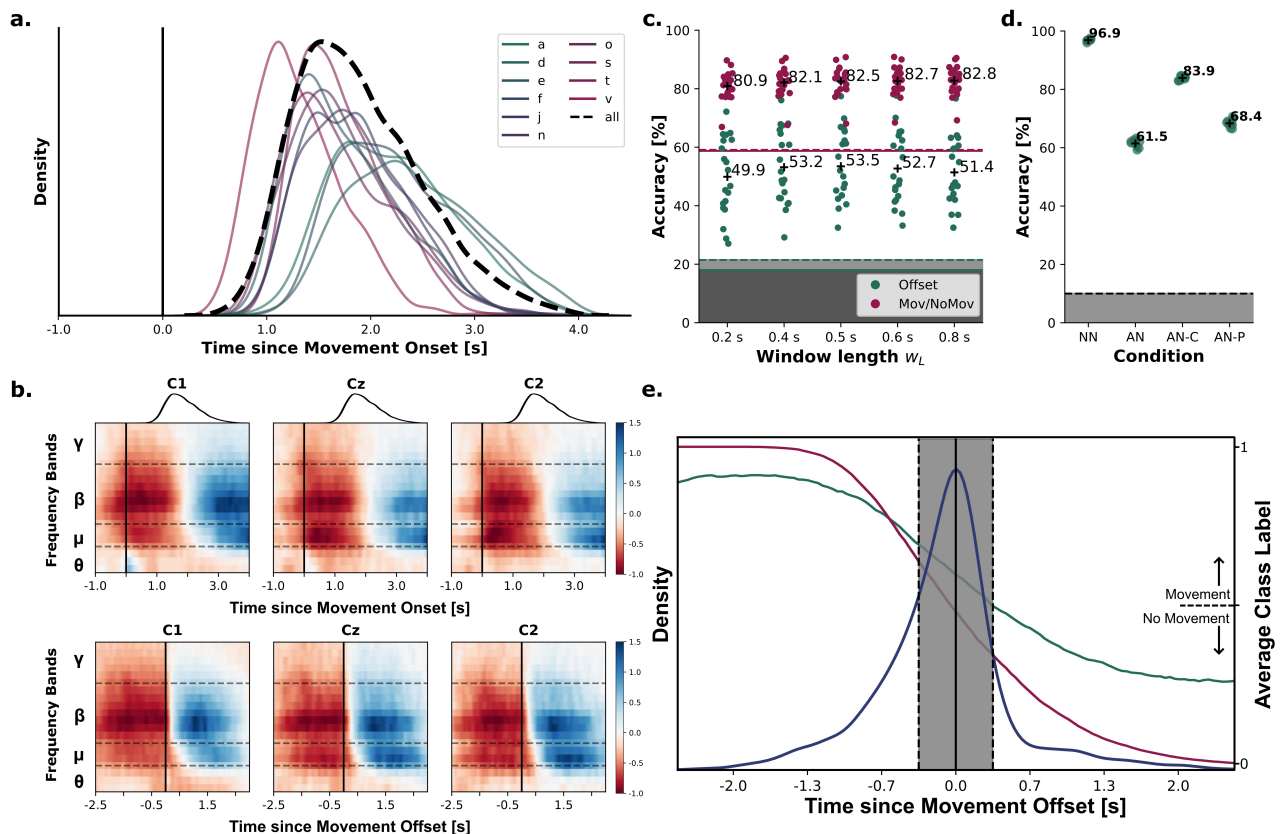


Figure 2: (a) Density plot of the distribution of the movement offset of the handwritten letters relative to the start of the movement. The distributions for the ten individual letters are shown together with the overall distribution. (b) ERD/ERS maps of electrodes C1, Cz and C2 for frequency bands aligned to the movement onset (upper plots) and movement offset (lower plots). Black straight lines indicate the movement onset/offset while dashed lines indicate the frequency band borders. (c) Classification accuracies for the detection of the movement onset (green) and classification between *movement* and *no movement* (red) for the different window lengths. Results for the individual subjects are plotted as dots, average results are marked by a cross. Theoretical and statistical chance levels for both classifications are indicated by lines in the corresponding color. (d) Results for the simulated application of classifying handwritten letters under different conditions of noise and length of the kinematic trajectories. (e) The density plot for the detected movement offsets displaying the difference between actual and detected movement offset per trial is shown in blue. The image also depicts the average predicted class of *movement* or *no movement* in green and the actual average class in red.

window length  $w_L$  was not significant ( $p > 0.05$ ) between window lengths 200-800 ms, 400-500 ms and 400-600 ms as assessed with a paired t-test. The density distribution for the temporal difference between actual and detected movement offset is shown in Fig. 2e in blue. The grey area displays the time range around the actual offset in which detected offsets were considered to be correct. This area contains 53.5% of the detected offsets, the area to the left contains 24.8% and the area to the right 10.9%. In 10.8% of the trials, no movement offset was detected. When increasing the threshold for considering a predicted offset to be correct to  $\pm 500$  ms, the accuracy could be increased to 64.3%. We also analyzed the binary classification accuracy for the classification of *movement* vs. *no movement*, which acted as a basis for the detection of the movement offset. The highest accuracy was achieved for a window length of  $w_L = 800$  ms with an average accuracy of  $82.8 \pm 4.8\%$  over all participants. The best performance of a single participant was achieved for

the same subject as in the offset detection at an accuracy of 90.5%. The average classification accuracies for different window lengths were  $80.9 \pm 4.8\%$ ,  $82.1 \pm 4.9\%$ ,  $82.5 \pm 4.9\%$ ,  $82.7 \pm 4.9\%$  for window sizes 200, 400, 500 and 600 ms, respectively. The average class label over the trials is shown in Fig. 2e in green (actual class label) and red (predicted class label). The probability for a prediction of movement at the beginning of a trial was 0.92 (actual probability: 1.0) and 0.25 (actual probability: 0.0) at the end of a trial. Results for all subjects are shown in Fig. 2c in green for the offset detection and in red for the classification between *movement* and *no movement*.

*Simulated Application - Handwritten Letters:* Handwritten letters could be correctly classified from the recorded kinematics with an average accuracy of 96.9% over all subjects in the NN condition in which no additive noise was applied and the trajectories were not cut at the movement offset. When additive noise was applied to the trajectories (AN condition) the classification accuracy

dropped to 61.5%. Cutting the noisy trajectories after the actual movement offset (**AN-C** condition) increased the accuracy to 83.9%. The usage of the detected motion offset for cutting the noisy trajectories (**AN-P** condition) enabled an increase of the classification accuracy by 6.9% compared to the **AN** condition and achieved a total accuracy of 68.4%. The results of the 2-times repeated 5-fold cross-validation are shown in Fig. 2d.

## DISCUSSION

Continuous decoding of hand movement, especially in EEG, stands to benefit from the detection of movement offsets. So far, the limited research in this area showed promising results [11–15]. We attempted the detection of the movement offset from time-frequency data after continuous movement and obtained a moderate accuracy of 53.5% against a chance level of 21.1%. We also showed the benefit of including this information into an exemplary application of classification of handwritten letters.

*Neural Correlates of Movement Termination:* The constructed ERD/ERS maps show the connection of the desynchronization in  $\mu$  and  $\beta$  frequencies with the continuous movement as well as the synchronization in these frequencies as soon as the movement terminates. The ERD/ERS maps exhibit broader distribution and reduced intensity of synchronization when trials are aligned to movement onset compared to alignment with movement offset. This is caused by the time-locking of the synchronization to the termination of the movement and the dispersion of the movement offset as shown in Fig. 2a. These findings are also in accordance with literature [7, 8, 20, 21]. The dependency of frequency power on the movement on- and offset generally shows that a classification of *movement* vs. *no movement* based on the time-frequency data is possible. Since the effects are mostly occurring in the  $\mu$  and  $\beta$  frequencies, the usage of these frequencies is appropriate.

*Movement Offset Detection:* The maximum average accuracy of 53.5% is comparable to those achieved in other movement offset detection BCIs [11, 13]. The classification accuracy of around 82% for the binary classification is also in the range of performances of similar BCIs [22]. Interestingly, the accuracies for different subjects were more dispersed in the detection of the movement offset than for the underlying binary classification. Also, the influence of the window length proved to be different for the detection of the movement offset and the binary classification with an optimal window length of 500 ms for the offset detection and an increasing accuracy of the classification with longer window lengths. It is possible that these differences occurred due to the decision strategy with which movement offsets were defined being suboptimal for some participants. A closer analysis of the distribution of *movement* vs. *no movement* classifications might yield insight into this and allow for the formulation of a better strategy. The distribution of differences between detected and actual movement off-

sets as given in Fig. 2e shows a tendency of predicting the movement offset earlier than the actual offset. This could similarly be influenced by the detection strategy, which used the first feasible predicted offset as the detected offset and discarded every other following feasible offset. Since the *movement* vs. *no movement* classification is imperfect, this strategy naturally leads to a higher rate of false positives in the beginning than at the end of a trial. Due to the non-causal window length and the detection strategy requiring ten continuous predictions being classified as *no movement*, the causal latency between a detected and actual movement offset amounts to 580 ms. Including the range of  $\pm 330$  ms in which the detected offset was considered correct, the minimal and maximal latency between stating a correctly detected and actual offset is between 250 and 910 ms. This calculation also shows that shorter window lengths should generally be preferred to longer ones in order to minimize the latency. Since the drop in detection performance proved to be limited for shorter window lengths in this study, other studies that employ a similar paradigm online should consider choosing a short window length for the classification.

*Simulated Application - Handwritten Letters:* The increase in classification accuracy of the **AN-C** and **AN-P** conditions compared to the **AN** condition in the simulated handwritten letter classification task shows the positive influence of the inclusion of motion termination information. While the actual movement offset information generated an increase of 22.4%, the offsets detected from neural data still allowed for an increase of 6.9%. Although this increase is modest, it demonstrates that classifiers also benefit from the inclusion of the moment of movement termination for imperfect detection performances. While the slight increase in accuracy compared to the **AN-C** condition might be improved by increasing the accuracy of the offset detection, other methods could also be effective: since the movement offset tended to be predicted earlier than the actual offset, we tested a modification in which all offsets predicted before 1.5 s after the movement onset were set to 1.5 s. With this method, we were able to increase the classification accuracy in the **AN-P** condition to 72.6% (i.e., an increase of 11.1% compared to **AN**). Other, more sophisticated approaches could increase the accuracy even further. It needs to be noticed that the simulated additive noise might not be representative of the distortion of movements in real-world decoding of hand trajectories from neural data. The benefit of the movement offset inclusion under the influence of different noise must be evaluated depending on the actual task at hand. However, in cases where the distortion leads to a residual, erroneous motion after the actual offset, the additional detection of the movement termination and incorporation of this information can be of great use.

## CONCLUSION

This study showed the general feasibility and benefit of the detection of self-initiated movement offsets during

continuous hand motion. While the accuracy of the detection was modest, it showed that detection is generally possible with a low latency. We also proved that a movement offset detection can be useful to increase the accuracy in specific movement-related tasks even for limited performance of the detection model. Since the benefit of including the movement termination detection grows with its accuracy, we aim to enhance the performance in the future. Recently, a new method for the detection of *movement* vs. *no movement* classes has been proposed using a pole tracking algorithm [23]. Although this method has not yet been applied to EEG, it could pose a way of increasing the classification and detection performances. While the current study focused on the detection of self-initiated motion termination, the start of the movements was based on an external cue, which prevented the application of a classifier to identify both movement onset and offset. Future work should focus on self-paced continuous movement tasks to investigate the detection of both start and end of continuous movements.

## ACKNOWLEDGEMENTS

This project is funded by the European Union's HORIZON-EIC-2021-PATHFINDER CHALLENGES program under grant agreement No 101070939 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00198.

## REFERENCES

- [1] Willett FR, Avansino DT, Hochberg LR, Henderson JM, Shenoy KV. High-performance brain-to-text communication via handwriting. *Nature*. 2021;593(7858):249–254.
- [2] Vansteensel MJ *et al.* Towards clinical application of implantable brain–computer interfaces for people with late-stage ALS: medical and ethical considerations. *Journal of Neurology*. 2023;270(3):1323–1336.
- [3] Mondini V, Kobler RJ, Sburlea AI, Müller-Putz GR. Continuous low-frequency EEG decoding of arm movement for closed-loop, natural control of a robotic arm. *Journal of Neural Engineering*. 2020;17(4):046031.
- [4] Pulferer HS, Kostoglou K, Müller-Putz GR. Getting off track: Cortical feedback processing network modulated by continuous error signal during target-feedback mismatch. *NeuroImage*. 2023;274:120144.
- [5] Shibasaki H, Hallett M. What is the Bereitschaftspotential? *Clinical Neurophysiology*. 2006;117(11):2341–2356.
- [6] Pereira J, Kobler R, Ofner P, Schwarz A, Müller-Putz GR. Online detection of movement during natural and self-initiated reach-and-grasp actions from EEG signals. *Journal of Neural Engineering*. 2021;18(4):046095.
- [7] Pfurtscheller G, Silva FLd. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*. 1999;110(11):1842–1857.
- [8] Pfurtscheller G, Brunner C, Schlögl A, Silva FLd. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*. 2006;31(1):153–159.
- [9] Liu D *et al.* EEG-Based Lower-Limb Movement Onset Decoding: Continuous Classification and Asynchronous Detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2018;26(8):1626–1635.
- [10] Müller-Putz GR, Scherer R, Pfurtscheller G, Rupp R. EEG-based neuroprosthesis control: A step towards clinical practice. *Neuroscience Letters*. 2005;382(1-2):169–174.
- [11] Pfurtscheller G, Solis-Escalante T. Could the beta rebound in the EEG be suitable to realize a “brain switch”? *Clinical Neurophysiology*. 2009;120(1):24–29.
- [12] Müller-Putz GR, Kaiser V, Solis-Escalante T, Pfurtscheller G. Fast set-up asynchronous brain-switch based on detection of foot motor imagery in 1-channel EEG. *Medical & Biological Engineering & Computing*. 2010;48(3):229–233.
- [13] Hortal E, Úbeda A, Iáñez E, Azorín JM, Fernández E. EEG-Based Detection of Starting and Stopping During Gait Cycle. *International Journal of Neural Systems*. 2016;26(07):1650029.
- [14] Bai O, Lin P, Vorbach S, Floeter MK, Hattori N, Hallett M. A high performance sensorimotor beta rhythm-based brain–computer interface associated with human natural motor behavior. *Journal of Neural Engineering*. 2008;5(1):24.
- [15] Orset B, Lee K, Chavarriaga R, Millán JdR. User Adaptation to Closed-Loop Decoding of Motor Imagery Termination. *IEEE Transactions on Biomedical Engineering*. 2019;68(1):3–10.
- [16] Kobler RJ, Sburlea AI, Lopes-Dias C, Schwarz A, Hirata M, Müller-Putz GR. Corneo-retinal-dipole and eyelid-related eye artifacts can be corrected offline and online in electroencephalographic and magnetoencephalographic signals. *NeuroImage*. 2020;218:117000.
- [17] Morlet Wavelets and Wavelet Convolution. In: *Analyzing Neural Time Series Data: Theory and Practice*. The MIT Press, Jan. 2014.
- [18] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random? a closer look on bci results. *International Journal of Bioelectromagnetism*. 2008;10(1):52–55.
- [19] Baldominos A, Saez Y, Isasi P. A Survey of Handwritten Character Recognition with MNIST and EMNIST. *Applied Sciences*. 2019;9(15):3169.
- [20] Müller G, Neuper C, Rupp R, Keinrath C, Gerner H, Pfurtscheller G. Event-related beta EEG changes during wrist movements induced by functional electrical stimulation of forearm muscles in man. *Neuroscience Letters*. 2003;340(2):143–147.
- [21] Seeber M, Scherer R, Müller-Putz GR. EEG Oscillations Are Modulated in Different Behavior-Related Networks during Rhythmic Finger Movements. *The Journal of Neuroscience*. 2016;36(46):11671–11681.
- [22] Rashid M *et al.* Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. *Frontiers in Neurorobotics*. 2020;14:25.
- [23] Müller-Putz G, Crell M, Egger J, Suwandjieff P, Kostoglou K. Towards Implantable Brain-Computer Interface for Communication in Locked-In Syndrome patients. *Current Directions in Biomedical Engineering*. 2023;9(2):1–4.

# AN EMG-BASED BRAIN-COMPUTER INTERFACE FOR COMMUNICATION-IMPAIRED PATIENTS: A CASE STUDY

P. Raggam<sup>1,2</sup>, M. Eder<sup>1</sup>, A.-T. Popa<sup>1</sup>, P. Fugger<sup>1</sup>, M. Grosse-Wentrup<sup>1,3,4</sup>

<sup>1</sup>Research Group Neuroinformatics, Faculty of Computer Science, University of Vienna, Vienna, Austria

<sup>2</sup>Doctoral School Computer Science, Faculty of Computer Science, University of Vienna, Vienna, Austria

<sup>3</sup>Research Network Data Science, University of Vienna, Vienna, Austria

<sup>4</sup>Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria

E-mail: philipp.raggam@univie.ac.at

**ABSTRACT:** Electromyography (EMG)-based brain-computer interface (BCI) systems primarily rely on electrical signals generated by muscle activity instead of the typically used brain activity measured via electroencephalography (EEG). Such EMG-BCIs are promising systems that enhance communication and control. This study introduces a simple EMG-BCI communication system developed as a football game for a communication-impaired participant. The football in the game can be moved to a left-side or a right-side goal, representing answers to two-state queries, i.e., yes-or-no-questions. By using restricted game controls, correctly following verbal instructions, and showing movement-related brain activity preceding muscle contractions, our participant can deliberately control the directions of the ball movements and, thus, successfully use our game for communication.

## INTRODUCTION

BCI technology has witnessed significant advancements by integrating a diversity of neurophysiological signals besides the traditionally used EEG signals [1, 2]. EMG-based BCI systems have emerged as a promising approach among these neurophysiological signals. Making use of the electrical activity generated by skeleton muscle contraction, the integration of EMG enhances the scope and precision of BCI applications, unlocking new possibilities for communication and control [3, 4]. Initially used for prosthetic control and rehabilitation, EMG-BCI systems have expanded their scope to include assistive technology, gaming, and communication [5, 6]. Zhang et al. introduced an EMG-based wearable multifunctional eye-control glass to control home appliances and communicate by voluntary blinks [7], Chai et al. and Rashid et al. combined steady-state visually evoked potentials (SSVEPs) and EMG to control communication interfaces [8, 9].

This study introduces an EMG-BCI communication system designed as a simple football game developed for a communication-impaired participant. The idea behind

developing this communication system was twofold: 1. test whether our participant was intellectually and physically capable of communicating with others, and 2. if so, provide a very simple yet engaging game as a communication basis. The following sections introduce our participant, the game design and controls, the recording modalities, and the implemented signal processing procedures. Furthermore, we demonstrate with our results that our participant understood verbal instructions and intentionally controlled arm muscle activity to move the ball to the left or to the right.

## MATERIALS AND METHODS

*Participant:* The study was designed for one participant (seven years old, male) who suffered through an accident from a severe hypoxic-ischemic encephalopathy (especially in the basal ganglia), dysphagia, dysarthrophonia, and a severe bilateral spastic and dystonic cerebral movement disorder. Based on our interactions, we learned that our participant communicates by looking and smiling at someone to show joy or contentment or by looking displeased if otherwise. During the whole study, our participant's parents were present, and the comfort and safety of our participant were our highest priorities. The study was approved by the University of Vienna's ethics committee.

*Game design:* We designed our communication system to resemble a football game since it was one of our participant's biggest interests before the accident. With that, we wanted to ensure that the game was engaging enough to be played over a longer period of time. The game was designed in Python<sup>1</sup> using the PsychoPy<sup>2</sup> library. Fig. 1 shows the interface of the football game. The game's aim is to move the football to the left-side or the right-side goal and can be played in two modes: practice or playing. During practice, the distances to the goals are shortened to learn how the game is controlled

<sup>1</sup><https://www.python.org/>

<sup>2</sup><https://www.psychopy.org/>

more easily. At the beginning of each round, the ball is placed at the center of the field.

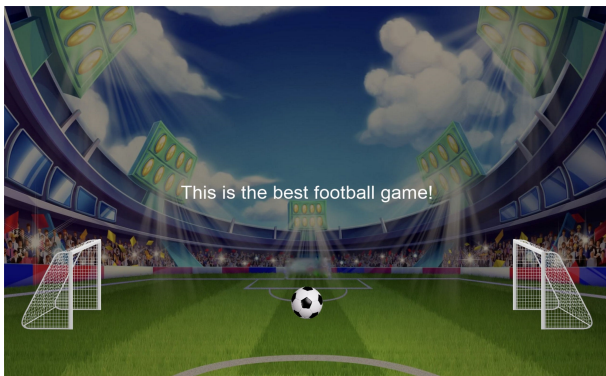


Figure 1: The interface of the football game.

**Game controls:** The football's movement is controlled by the EMG. To move the ball to the left or the right, the left-arm or the right-arm muscles must be activated, causing an increase in the EMG signal amplitude. The ball can move in single steps or continuously, depending on the duration of the muscle contraction. Two restrictions were introduced to avoid random, unintended movements: a signal threshold window and one-sided contraction. The signal threshold window ensures a controlled movement of the ball by only moving the ball if the signal amplitude is within a lower and an upper limit. Furthermore, the ball only moves if the arm muscles are contracted only on one side and stops if the muscles are contracted at both arms simultaneously. As feedback for the user, the ball turns red if a restriction is applied, i.e., if the signal amplitude is above the upper limit of the threshold window or if both arms are contracted simultaneously.

**Recording sessions:** Fig. 2 shows the recording setup. Our participant was sitting in a wheelchair, looking at a monitor to play the football game. The game was played over four sessions. The first two sessions were used to accustom our participant to the game and its controls. In the beginning, a squeeze bulb was used to move the ball. After establishing that the principles of the game were understood, the game controls were switched to the EMG since it did not require the coordinated muscle activation necessary to squeeze a bulb and, hence, was easier to use. The third session was split into practice runs and a playing run. During the playing run, our participant was instructed verbally to move the ball to the left or the right goal. In session four, we recorded both the EMG and the EEG. The session was divided into a resting-state run and two playing runs. Playing run one (run P1) was further split into six trials, where our participant was asked again to move the ball to the left or the right goal (three trials for each side).

**Recording modalities:** EMG and EEG signals were recorded with the Bittium NeurOne™ Tesla EEG system<sup>3</sup>, with a sampling frequency of 1 kHz. The EMG was

<sup>3</sup><https://www.bittium.com/medical/bittium-neurone>

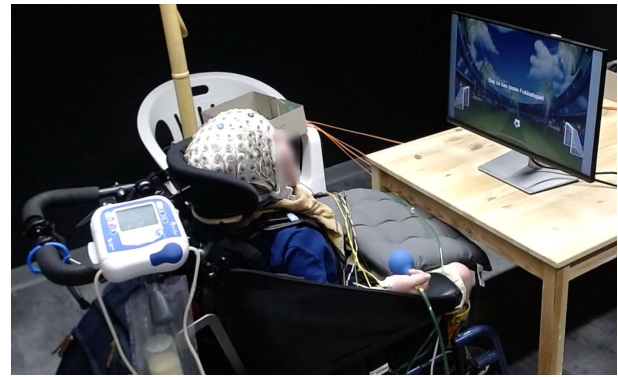


Figure 2: Setup for our participant playing the football game.

recorded with bipolar electrode channels at the following arm muscles (for each side): flexor digitorum profundus (FDP), extensor digitorum (ED), and abductor pollicis longus (APL). For the EEG, passive, gel-based electrodes were used at the following channels: F1, Fz, F2, FC3, FC4, C3, C1, Cz, C2, C4, CP3, CP4, and Pz. In addition to the physiological signals, event markers were recorded to put time stamps on certain events or phases of the game, e.g., when a new trial started or a goal was scored.

**Online signal processing:** To access the recorded signals in (near-)real-time, the lab streaming layer (LSL)<sup>4</sup> and its Python interface pylsl<sup>5</sup> were used. With the pylsl library, the EMG signals could be streamed into the PsychoPy game framework for further processing. The EMG signals were processed in three steps: 1. applying a 4<sup>th</sup>-order Butterworth bandpass filter between 20 and 40 Hz, 2. calculating the envelope via Hilbert transform, and 3. smoothing the signal with a Savitzky-Golay filter [10]. Since our participant suffered from a spastic and dystonic movement disorder, we decided to define a personalized EMG signal band. The EMG bandwidth was chosen by maximizing the cross-correlation coefficients between the squeeze bulb signal and the EMG signals. After processing the EMG signals and checking the movement restrictions, the position or color of the ball on the screen was updated, giving feedback to the user on whether the movement attempt was successful.

**Offline data analysis:** The offline data analysis was also implemented in Python. Similar to the EMG signal processing, we also personalized the EEG frequency bands. After inspecting the power spectral density (PSD) function of the resting-state EEG, the following frequency bands were chosen for further investigation: 5–7 Hz for the mu band and 15–25 Hz for the beta band. The EMG was again filtered between 20 and 40 Hz. The band power was calculated for each frequency band (mu, beta, and EMG) by squaring the amplitude values. All applied filters were 4<sup>th</sup>-order Butterworth filters.

Due to our participant's involuntary repeated head movements during the recording session, the EEG cap was pressed and shifted against the headrest, which led to a

<sup>4</sup><https://github.com/sccn/labstreaminglayer>

<sup>5</sup><https://github.com/chkothepy/pylsl>

low signal-to-noise ratio (SNR) and gel bridges between channels. Artifact correction with independent component analysis (ICA) proved to be ineffective. However, a simple bipolar derivation, i.e., subtracting channels from one another, led to clean EEG signals of a few channel pairs.

After cleaning the EEG, we investigated the mu and beta rhythms. First, the cross-correlation functions (CCFs) of mu and beta band power vs. right-hand and left-hand EMG were calculated. After inspecting the results, we decided to continue with the mu band only since the CCFs of the beta band were inconclusive. Next, the continuous signals of run P1 were split into left-goal and right-goal trials for calculating the cross-correlation functions of mu band power vs. right-hand and left-hand EMG. A permutation test with cyclical shifts and  $n = 1000$  permutations was applied to generate p-values for the CCFs, i.e., finding significance in our results. The p-values were corrected using the false discovery rate (FDR) correction with the Benjamini-Hochberg procedure [11].

## RESULTS

All results in this section were generated from EMG and EEG signals recorded in session four's resting-state run and run P1 since this was the only session with EEG recordings, and only run P1 included verbal instructions.

**Resting-state EEG:** The resting-state EEG signals were used to find personalized frequency bands for our participant's mu and beta rhythms. Fig. 3 shows the power spectral density (PSD) function of the resting-state EEG at channel pair Cz-C4. We can clearly see the alpha/mu peak between 5 and 7 Hz and the beta bump between 15 and 25 Hz. The alpha/mu rhythm is slower than an average adult's (8–13 Hz [12]). However, this is not a pathological indicator since the alpha rhythm increases with age during childhood and adolescence [13].

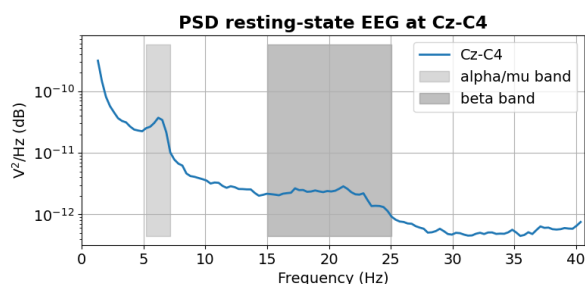


Figure 3: PSD of resting-state EEG at channel pair Cz-C4.

**Run-level analysis:** We first looked at run P1 as a whole. In Fig. 4, we see the power of the right-hand (blue lines) and left-hand (orange lines) EMG. Individual EMG channels (FDP, ED, and APL) were averaged on each side. The left-hand EMG power is much lower than the right-hand EMG power, possibly due to a Botox treatment on our participant's left arm before the record-

ing session.

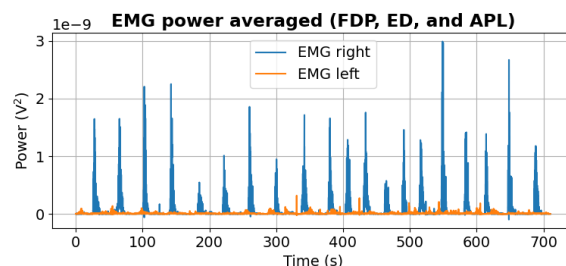


Figure 4: Right-hand (blue) and left-hand (orange) EMG power of run P1. Individual EMG channels (FDP, ED, and APL) were averaged on each side.

Fig. 5 shows the CCFs of EEG mu band power (subfigure A) and beta band power (subfigure B) at channel pair Cz-C4 vs. right-hand (blue lines) and left-hand (orange lines) EMG power. In the mu band, we can observe a negative correlation between EEG power and both left-hand and right-hand EMG at time lag = 0. A negative correlation means that the mu rhythms desynchronize (decrease in EEG mu power) when the arm muscles are activated (increase in EMG power), which displays typical, non-pathological event-related desynchronization (ERD) [14]. Also, having mu rhythm ERD on the right hemisphere (Cz-C4) for both left-hand and right-hand EMG indicates bilateral cortical activation for one-sided movements. Even though bilateral mu rhythm ERD is uncommon, it can occur during one-sided hand movements, especially in the context of motor planning and execution [15].

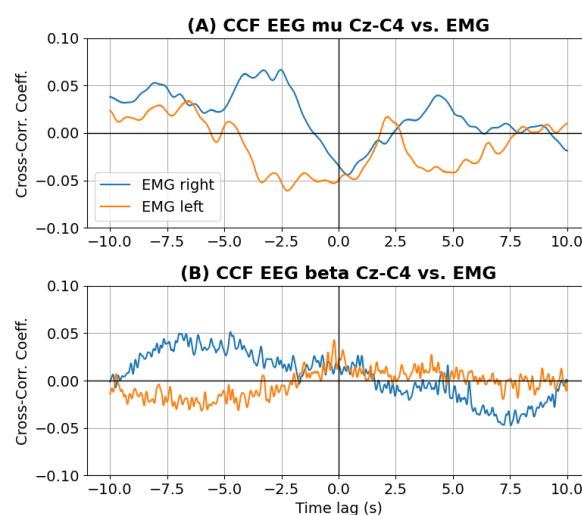


Figure 5: CCFs of EEG mu band power (A) and beta band power (B) at channel pair Cz-C4 vs. right-hand (blue) and left-hand (orange) EMG power.

When we look at the time course of the CCFs, however, we can observe very unusual behavior: For the right hand, the desynchronization process starts roughly 2.5

seconds before the muscles are activated, whereas for the left hand, it starts about five seconds before muscle activation. The mu rhythm desynchronization is very slow compared to healthy people [15]. This means that our participant can react quickly to instructions, but it takes very long to activate the motor system to cause a muscle contraction. The initially stated damage in the basal ganglia could be a possible reason for that. The Botox treatment on the left arm again may have caused the difference between right-hand and left-hand ERD time.

The CCFs of EEG beta band power with EMG power didn't show any conclusive results, and hence, only the mu band was used for further investigations.

*Trial-level analysis (left-goal/right-goal split):* Next, EMG and EEG signals were split into trials. Our participant successfully moved the football to the correct goal in all six trials of run P1 (three left-goal trials and three right-goal trials). Therefore, we used the left-goal and right-goal trials for ball movements to the left and the right, respectively.

In Fig. 6, we can see the CCFs of EEG mu band power at channel pair Cz-C4 vs. right-hand (blue lines) and left-hand (orange lines) EMG power at left-goal trials (subfigure A) and right-goal trials (subfigure B). The cross-correlation coefficients are rather small, but the CCFs are highly significant ( $p < 0.05$ ) around time lag = 0, as shown in Fig. 7 by the corresponding FDR-corrected p-values.

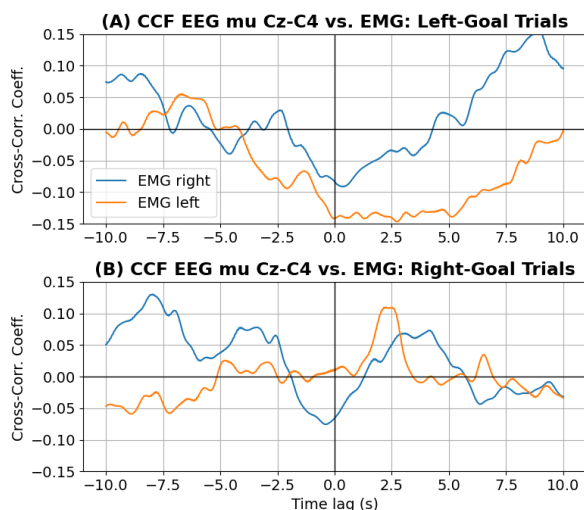


Figure 6: CCFs of EEG mu power at channel pair Cz-C4 vs. right-hand (blue) and left-hand (orange) EMG power at left-goal trials (A) and right-goal trials (B).

Looking at the right-goal trials, we can observe a negative correlation between EEG mu band power and right-hand EMG power but no correlation between mu power and left-hand EMG. These results suggest that there is only a clear mu rhythm ERD for right arm muscle activity, meaning only the right hand was intentionally used for moving the ball to the right goal, which is expected behavior.

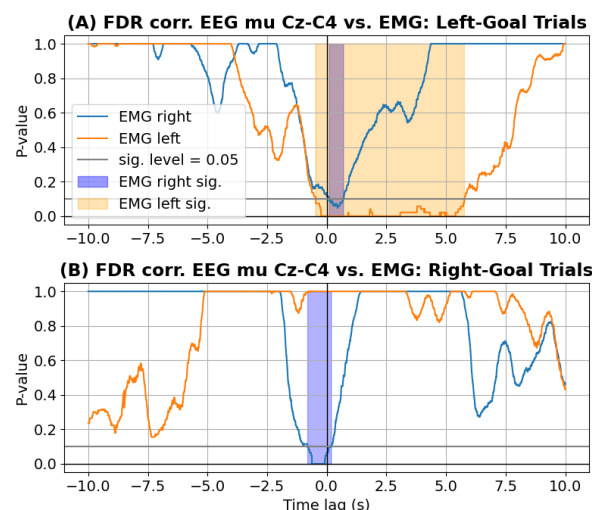


Figure 7: FDR-corrected p-values of CCFs of EEG mu band power (A) and beta band power (B) at channel pair Cz-C4 vs. right-hand (blue) and left-hand (orange) EMG power. The p-values were calculated by a permutation test with a cyclical shift and  $n = 1000$  permutations, including an FDR correction with the Benjamini-Hochberg procedure.

For the left-goal trials, however, we see something very interesting: There is a negative correlation between EEG mu band power and both left-hand and right-hand EMG. This indicates clear mu rhythm ERD for left and right arm muscle activity, meaning both hands were intentionally used for moving the ball to the left goal. This could mean that our participant used the right-hand activity to trigger a left-hand activity, which we could actually observe during the recording sessions. A reason for that could be again the Botox treatment on the left arm, which is also manifested in the slow and long-lasting ( $\sim 5$  s) mu rhythm ERD, compared to the faster ( $\sim 2.5$  s) desynchronization for right-hand muscle activity.

## DISCUSSION

In the first two recording sessions, we tested whether our participant could use the game controls. Both the squeeze bulb and the EMG were successfully used to move the ball to the left or right. We decided to continue controlling the game with the EMG because it does not require coordinated muscle activations necessary to squeeze a bulb and, hence, was easier to use. In the third session, our participant could follow our verbal instructions to go to the left or the right goal. In run P1 of the fourth session, our participant again successfully followed verbal instructions, this time split into six trials, with three left-goal and three right-goal trials in random order.

Overall, we found three indicators that demonstrated that the ball movements did not occur randomly but were the results of deliberate control of our participant:

1. *Game control restrictions:* Every ball movement resulted from precise muscle activity since a lower and an upper threshold defined an EMG amplitude/power window. Furthermore, only one-sided EMG activity led to a

ball movement.

2. *Successful task completion:* In session four, run P1, all six trials were completed successfully. That means our participant could understand the verbal instructions and generate the appropriate response to suit the game controls.

3. *Mu rhythm ERD preceding muscle contraction:* By simultaneously recording EMG and EEG signals in the fourth session, we could further demonstrate that the muscle activity to move the ball did not occur through random or spastic contractions but followed movement-related brain activity. Fig. 6 and Fig. 7 show a clear decrease in mu band power that precedes muscle contraction for ball movements on both sides. The slow mu rhythm desynchronization could be due to the damaged basal ganglia, which could cause a delay in activating the motor system. The left arm was also treated with Botox, which would explain the even slower and long-lasting left-hand mu rhythm ERD. It is also noteworthy that our participant developed a strategy to overcome the increased difficulty of activating the left arm muscles by involving the right arm, which eventually triggered the left-hand muscle contraction.

## CONCLUSION

This study demonstrated that our participant could deliberately control a football game and follow verbal instructions despite the severe impairments. Furthermore, the combination of EEG and EMG revealed normal reaction times to instructions but a slow motor system activation. This provided important information about our participant's mental abilities for the family.

Currently, the game can be used for simple two-state queries, e.g., answering yes-or-no-questions by moving the ball to the left or the right. Future game adaptations could facilitate the controls or increase the number of goals, i.e., the number of answers to select. Combining EMG and EEG signal features could further improve our communication system's precision and robustness.

Finally, we also want to emphasize that even simple systems can be very effective. Straightforwardness and convenience are key features for people with mobility and/or communication impairments.

## REFERENCES

- [1] Muller-Putz G *et al.* Towards noninvasive hybrid brain-computer interfaces: Framework, practice, clinical application, and beyond. *Proceedings of the IEEE*. 2015;103(6):926–943.
- [2] Käthner I, Kübler A, Halder S. Comparison of eye tracking, electrooculography and an auditory brain-computer interface for binary communication: A case study with a participant in the locked-in state. *Journal of NeuroEngineering and Rehabilitation*. 2015;12(1):76.
- [3] Balasubramanian S, Garcia-Cossio E, Birbaumer N, Burdet E, Ramos-Murguialday A. Is EMG a viable alternative to BCI for detecting movement intention in severe stroke? *IEEE Transactions on Biomedical Engineering*. 2018;65(12):2790–2797.
- [4] Li K, Zhang J, Wang L, Zhang M, Li J, Bao S. A review of the key technologies for sEMG-based human-robot interaction systems. *Biomedical Signal Processing and Control*. 2020;62:102074.
- [5] Rouillard J *et al.* Hybrid BCI coupling EEG and EMG for severe motor disabilities. *Procedia Manufacturing*. 2015;3.
- [6] Althekair A, Odeh M, AlBayaa M, Sharawi M, Doush IA. Mobile gaming emg-based brain computer interface. In: *International Conference on Computer-Human Interaction Research and Applications*. 2023, 40–52.
- [7] Zhang S *et al.* An EMG-based wearable multifunctional eye-control glass to control home appliances and communicate by voluntary blinks. *Biomedical Signal Processing and Control*. 2023;86:105175.
- [8] Chai X *et al.* A hybrid BCI-controlled smart home system combining SSVEP and EMG for individuals with paralysis. *Biomedical Signal Processing and Control*. 2020;56:101687.
- [9] Rashid M *et al.* A hybrid environment control system combining EMG and SSVEP signal based on brain-computer interface technology. *SN Applied Sciences*. 2021;3(9):782.
- [10] Savitzky A, Golay MJ. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*. 1964;36(8):1627–1639.
- [11] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*. 1995;57(1):289–300.
- [12] Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*. 1999;29(2):169–195.
- [13] Cellier D, Riddle J, Petersen I, Hwang K. The development of theta and alpha neural oscillations from ages 3 to 24 years. *Developmental Cognitive Neuroscience*. 2021;50:100969.
- [14] Pfurtscheller G, Lopes Da Silva F. Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*. 1999;110(11):1842–1857.
- [15] Pfurtscheller G, Neuper C. Event-related synchronization of mu rhythm in the EEG over the cortical hand area in man. *Neuroscience Letters*. 1994;174(1):93–96.

# FINDING THE OPTIMAL SIX: DECODING FROM A LARGE SET OF HAND GESTURES WITH 7T FMRI FOR IMPROVED BCI CONTROL

M. Kromm<sup>1</sup>, S. Schellander<sup>1,2</sup>, M.P. Branco<sup>1</sup>, M.A.H.L.L. Raemaekers<sup>1</sup>, N.F. Ramsey<sup>1</sup>

<sup>1</sup>UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, University Medical Center, Utrecht, The Netherlands.

<sup>2</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria.

E-mail: m.kromm-2@umcutrecht.nl

**ABSTRACT:** Decoding movements from the human sensorimotor cortex has been of great interest for brain-computer interfaces (BCIs). To establish the possibility of increasing the degrees of freedom of a sensorimotor-driven BCI, we investigated the decodability of 20 hand gestures using 7-Tesla fMRI and narrowed it down to a set of six best distinguishable gestures. Six able-bodied volunteers performed gestures from the American Sign Language alphabet and single-digit movements. Results indicated significant classification accuracies across all 20 gestures (*mean* = 46%, *range* = 39.5% – 51.5%, *chancelevel* = 5%), with some differences in decodability across gestures. Subsequently, optimal sets of six gestures were identified by establishing classification performance for all possible permutations, and applying the identified set in a leave-one-subject-out cross-validation scheme. The results showed a near-optimal classification in five out of six subjects. Our findings contribute to the understanding of the generalizability of gesture decoding performance and offer insights for refining BCI control strategies to enhance communication for individuals with motor impairments.

## INTRODUCTION

Fully implantable brain-computer interfaces (BCIs) intend to establish a communication pathway between signals directly measured from the surface of the brain and a computer [1]. This can be of great use for individuals with locked-in syndrome, a condition that can result from Amyotrophic Lateral Sclerosis (ALS) or brainstem stroke [2]. A common target for BCI-readout is the sensorimotor cortex due to its well-established topographic representations and the fact that it shows activity during attempted movement even years after paralysis, despite the absence of the actual movement [3, 4]. For BCI control, different attempted movements need to be classified and coupled to intended commands. However, if the cortical activity related to the selected movements is not distinct enough, the BCI may misclassify the intended action – resulting in outcomes that are not desired by the user. Thus, it is crucial to ensure that the command-coupled movements

are well-decodable and not easily confused by the BCI. While fully implanted electrocorticography (ECoG)-BCIs used at home have shown considerable success [5], their degrees of freedom have been limited so far (i.e., opening and closing the hand to produce a ‘brain-click’). Expanding the range of BCI control signals could significantly speed up communication, thereby improving their usability. A substantial increase in the degree of control in a home-use ECoG-BCI could be provided by the ability to decode six different movements, each corresponding to a specific command (i.e., “up”, “down”, “left”, “right”, “select”, and “escape”). These six commands would be produced through six different attempted movements with the hand. Particular sets of attempted movements may be more or less suitable for this purpose based on the similarity of the elicited cortical activity patterns. To optimize the performance of such BCI, we require a set of hand movements that is maximally distinct based on brain activity patterns in the sensorimotor cortex. As there is potentially a huge number of possible movements, the options need to be narrowed down at the outset.

Functional magnetic resonance imaging (fMRI) allows us to measure the representations of different hand movements in the sensorimotor cortex non-invasively and with high spatial resolution. This method provides the opportunity to explore activity in the sensorimotor cortex for various movements and across multiple individuals. Using fMRI, it is feasible to distinguish individual finger movements [6, 7], but also hand gestures consisting of the flexion and extension of multiple fingers [8, 9], even for attempted movements without an actual motor output [10]. Furthermore, fMRI results can be extrapolated to an implanted BCI as previous work has shown that fMRI activity patterns map consistently to the gamma band of ECoG recordings [11–13], for review, see [14].

Here, we investigate the decodability of 20 unimanual hand gestures in six healthy individuals, using 7-Tesla fMRI. From these 20 gestures, we identify the set of six that results in the most accurate classification. Furthermore, we explore the consistency of classification performance across individuals and look at the potential compromise of choosing hand movements based on group averages as opposed to individual results. These in-

sights can serve as a starting point to predict which hand movements can be well-decoded from the sensorimotor cortex.

## MATERIALS AND METHODS

**Participants:** Six healthy, able-bodied volunteers (age: *mean* = 23 years, *SD* = 1.8; 4 females; all right-handed) performed a hand gesture task during the acquisition of functional scans in a 7-Tesla MRI scanner. All participants gave written informed consent to participate, which was approved by the Medical Research Ethics Committee according to the Declaration of Helsinki (2013).

**Data acquisition:** MRI data were recorded using a Philips Achieva 7-T MRI system with a 32-channel head coil. Functional data were recorded using an EPI sequence (TR/TE = 1400/29 ms, FA = 60°, multiband factor 2, voxel size = 1.5 x 1.5 x 1.5 mm<sup>3</sup>, in-plane resolution = 200 x 200 mm<sup>2</sup>, 40 slices). A high-resolution anatomical T1-weighted MP2RAGE [15] was acquired for anatomical reference.

**Experimental task:** Participants performed a gesture task with their right hand. The movements were a sub-selection of 15 gestures from the American Sign Language alphabet based on ease of execution. In addition, we included individual flexion of each finger, resulting in a total of 20 gestures (Fig. 1). Participants practiced the hand gestures at home during the week prior to scanning to ensure familiarity with the movements.

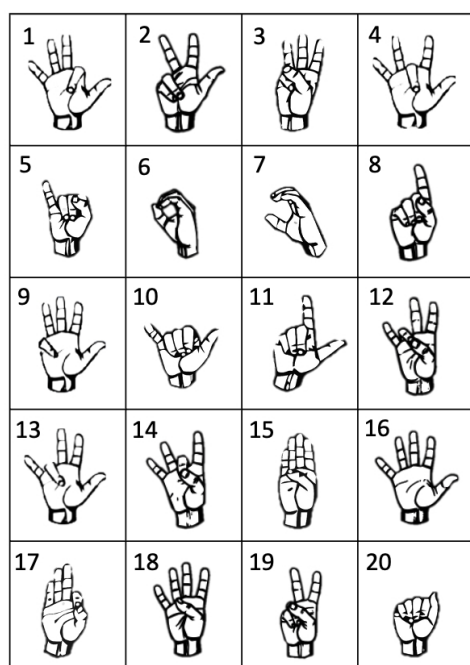


Figure 1: The task contained 20 right-hand gestures, including single-finger flexions (gestures “1”, “4”, “9”, “13”, “18”) and gestures from the American Sign Language alphabet.

During the scan, the stimuli were projected onto a screen that was visible to the participants through a

mirror and prism glasses. Each trial consisted of three different images, illustrating either the preparation phase, execution phase, or resting phase of the gestures to be performed (Fig. 2). After the preparation phase (2 s), the participant was instructed to execute the presented gesture and hold it for 6 s (execution phase) before returning the hand to the baseline position (hand relaxed, fingers slightly bent, palm face up; resting phase). The resting phase lasted for 8.8 s, to prevent blood-oxygen-level-dependent (BOLD) responses to bias the activity estimates of the subsequent trial. The stimuli were presented in a pseudo-random order.

Each gesture was performed once per run. Participants completed 10 runs in total, split across two scanning sessions on separate days (5 runs/session). This yielded a total of 10 repetitions per gesture for each participant. MRI-compatible data gloves (5 DT Inc, Irvine, USA) were worn during the task on both hands to record kinematic data. The data glove measurements were visually inspected for correct bending of the fingers and the absence of additional movements.



Figure 2: Trial schematic. An image of a gesture inside a red rectangle signaled the onset of a preparation phase, which was included to minimize error in the execution of the movement. The change of the rectangle's color to green indicated to the participant to make the displayed gesture and hold it for the duration that the gesture was presented. After returning the hand position to baseline, there was an 8.8 s pause until the onset of the next stimulus.

**Data preprocessing:** Functional scans from the Gesture Task were preprocessed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>) and custom MATLAB (<https://www.mathworks.com>) scripts. Scans from both sessions were aligned with each other and coregistered with the T1-weighted image. A General Linear Model was created including factors for each gesture. T-maps were computed for each gesture type while using a leave-one-run-out procedure, resulting in a total of 200 t-maps (20 gestures x 10 run-combinations).

**Region of Interest:** The left precentral and postcentral gyrus were defined as regions of interest through the Freesurfer surface reconstruction pipeline (<https://surfer.nmr.mgh.harvard.edu>), based on the Desikan-Killiany atlas.

**Gesture classification:** To assess the discriminability of hand gestures in the contralateral sensorimotor cortex, we used a support vector machine (SVM). The 500 voxels with the highest absolute t-values across gestures were selected as features. The BOLD signal in these voxels was detrended and transformed into z-scores for each run separately. For each trial, the peak signal in the 5th,

6th, and 7th volume after trial onset was extracted, which corresponds to the amplitude of the peak of the BOLD signal.

The SVM was run with a linear kernel and constraint parameter  $C = 1$ . A leave-one-run-out cross-validation scheme was used, meaning that with each iteration, one run was left out for training the model, and the left-out run was subsequently used to test the model. For each training/test set, the classification accuracy was calculated as the proportion of correctly classified gestures. These classification scores were then averaged across iterations, resulting in a single classification score per participant.

Classification performance was further evaluated using confusion matrices, which contain the details on correct and incorrect classifications. Confusion matrices were computed per subject and subsequently averaged (Fig. 3).

**Choosing an optimal set of six gestures:** For maximizing the performance of a BCI, we aim to select the six best distinguishable gestures and estimate if the performance of this set is generalizable across subjects. For this, we created SVMs for all possible combinations of a set of six out of the 20 gestures (in total 38760). This resulted in 38760 classification accuracies for each participant, containing the classification accuracy per possible gesture set. The optimal sets were selected based on the mean accuracy across participants. These sets were then evaluated using a leave-one-subject-out cross-validation, by testing their performance relative to that of all other combinations.

## RESULTS

**Classification performance of 20 gestures:** The classification accuracy for all gestures across all participants ( $mean = 46\%$ ;  $range = 39.5\% - 51.5\%$ ) was significantly above the 5% chance level ( $t(5) = 19.8$ ,  $p < 0.001$ ) (Fig. 3). Visual inspection of the confusion matrix revealed that the decoder often confused gesture "18" with "15" ( $mean = 31.7\%$ ), and gesture "9" with "2" ( $mean = 28.3\%$ ).

**Gesture set selection:** With the aim of finding the set of six gestures that are maximally decodable, we ran the SVM for each of the possible 38760 combinations of six out of 20 gestures. The gesture sets with the, on average, highest classification performance are shown in Fig. 4, in addition to the set with the highest mean ranking.

Next, to estimate the extent to which group-mean performances of optimal gesture sets are generalizable to different participants, we chose a gesture set based on a group-average result (leave-one-subject-out cross-validation) and checked the performance of this set in an individual. A summary of the results can be seen in Tab. 1. The difference in classification accuracy between the chosen gesture set and the set with maximum classification in the left-out subject ranged from 8.33% to 20.00%. The percentile scores of the chosen set in the distribution of all sets ranged between 95.01% to 98.86%

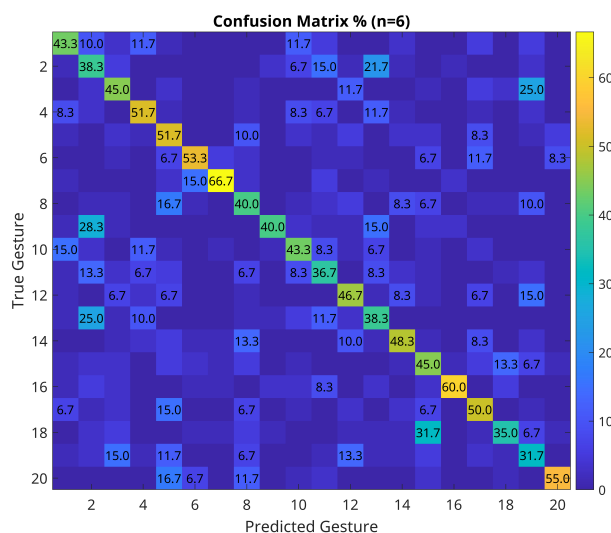


Figure 3: Average group classification in % for all 20 gestures ( $chancelevel = 5\%$ ). The numbers inside the squares correspond to the accuracy values assigned to the classified gesture (for values  $> 5\%$ ).

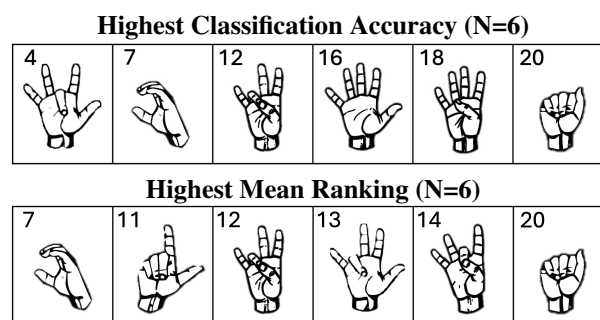


Figure 4: The overall best decodable set of six gestures, based on average classification accuracy (top row;  $mean = 87.22\%$ ,  $SD = 4.91\%$ ) and on average ranking scores (bottom row;  $mean = 76.38\%$ ,  $SD = 6.94\%$ ) of all participants ( $chance level = 16.67\%$ ).

for five subjects, one subject's (sub006) percentile score was 82.15% (Fig. 5). Some gestures were consistently present in the best-performing sets from the group average, with gesture "7", "16", and "20" being selected 100% of the time (Tab. 2). The highest performing subject-specific gesture sets were more varied, however, gesture "7" and "16" were still present in 66.67% of the sets (the highest observed percentage for subject-specific sets), and "20" in 50% of the cases. Gestures "10", "11", and "15" were never present among the highest performing gestures.




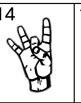
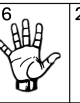

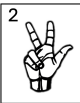


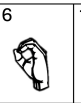
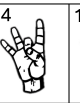

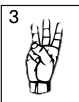


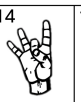




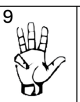
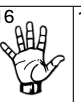


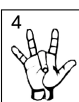

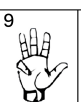
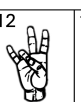
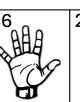




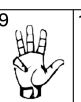
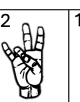



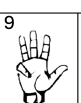
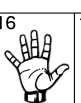
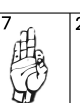
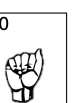

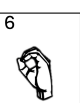

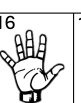
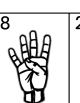
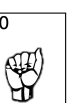
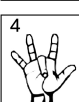


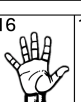
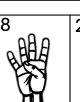
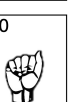

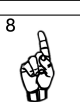

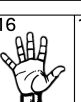

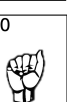
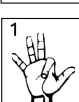



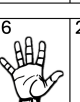
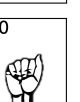
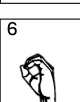

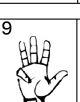
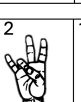

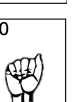
## DISCUSSION

In this study, we examined which sets of gestures are consistently well-decodable across individuals with 7-Tesla fMRI. For this, we first demonstrated the classification performance of 20 gestures and then created subsets of six gestures with the highest classi-

Table 1: Results for leave-one-subject-out gesture set selection. From left to right: Accuracy of the best-performing gesture set for five subjects ("GS Acc. leave-one-out"); accuracy of the selected gesture set for the left-out subject ("GS Acc. sub"); highest accuracy of subject's best-performing gesture set ("Max Acc. sub"); the difference between selected gesture set and subject's best-performing gesture set ("Diff"); subject's average classification performance of all possible gesture sets ("Overall Mean sub").

	GS Acc. leave-one-out (n=5)	GS Acc. sub	Max Acc. sub	Diff ( Max Acc. sub – GS Acc. sub)	Overall Mean sub
sub001	86.67%	88.33%	100.00%	11.67%	76.40%
sub002	86.67%	88.33%	98.33%	10.00%	76.13%
sub003	88.33%	81.67%	90.00%	8.33%	68.37%
sub004	86.67%	88.33%	96.67%	8.33%	73.08%
sub005	87.67%	85.00%	96.67%	11.67%	68.68%
sub006	89.00%	71.67%	91.67%	20.00%	63.35%

Table 2: Comparison of the selected gestures based on group average performance (left column) and the best-performing gestures based on the left-out subject's classification accuracy (right column).

	Highest gesture set leave-one-out (n=5)	Highest gesture set sub
sub001	     	     
sub002	     	     
sub003	     	     
sub004	     	     
sub005	     	     
sub006	     	     

fication performance across participants. Our findings demonstrate the feasibility of decoding a large set of gestures across able-bodied individuals. Twenty gestures could be decoded from sensorimotor activity with, on average, 46% accuracy. Furthermore, we evaluated the generalizability of optimal gesture sets across individuals. By selecting optimal gesture sets based on a group average and testing their performance in individual participants, we demonstrated that those sets still have a well above-average classification accuracy in the respective individual compared to those of all other sets (percentile ranks higher than 95% for five out of six subjects). This suggests that movements identified through group-level analysis are likely to generalize to individuals. One participant showed a relatively low classification accuracy for the group-selected set (percentile rank at 82%). However, it was still higher than the subject's average classification score across all possible sets. Furthermore, we noticed that the general

classification performance in this subject was lower than in the other participants and that some trials contained ambiguous movements. Further investigation is needed to assess if the performance improves upon the exclusion of wrong trials.

Visual inspection of the successful gestures indicated that distinct digit combinations and wrist movements can be best distinguished from each other. In contrast to that, gestures that were especially prone to confusion were "18" (fingers spread, thumb flexed) with "15" (fingers together, thumb flexed); and "9" (pinky flexed) with "2" (pinky and ring finger flexed). Interestingly, this was not the case for gestures "6" (wrist rotation, thumb, and other fingers touch) and "7" (wrist rotation, thumb, and other fingers bend, not touching), which were consistently among the best decodable gestures with a low confusion score with each other. This may be due to the difference in sensory feedback, which can be checked by decoding from the primary motor and somatosensory

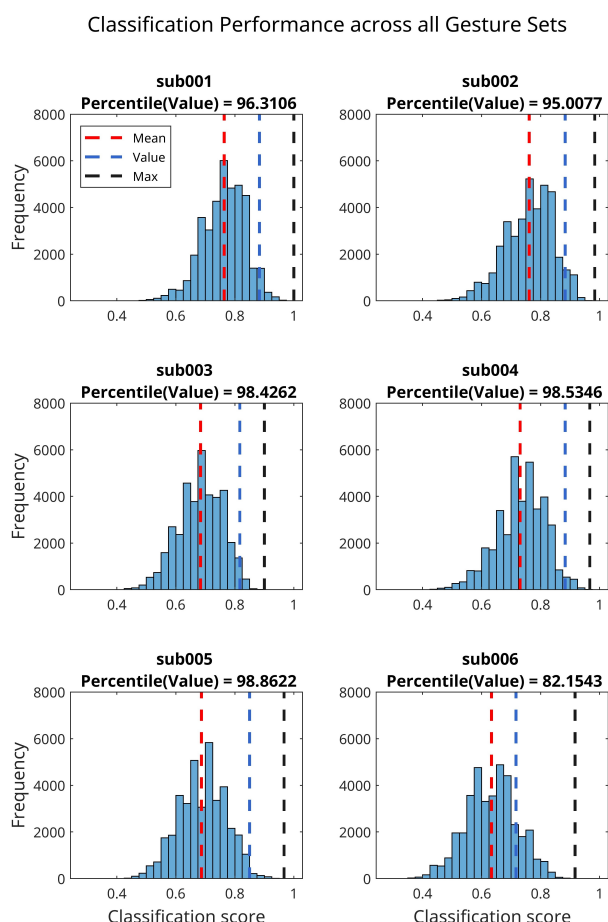


Figure 5: Distribution of the classification accuracies across all possible gesture sets of 6 (38760 combinations) for each participant. The red line (“Mean”) indicates the average classification accuracy across all possible sets. The blue line (“Value”) shows the accuracy of the selected gesture set based on the highest performance in the other five subjects. The black line (“Max”) shows the highest classification score of the respective subject.

cortex separately. In general, gestures that have only subtle movement differences are also more difficult to discriminate.

Partly, this result might not seem very surprising considering the topography of the sensorimotor homunculus containing representations of individual fingers [16, 17]. Thus, a gesture consisting of thumb flexion should be well distinguishable from a pinky flexion gesture. However, previous research has also observed that the activation of complex coordinated finger movements is not a mere linear combination of the activation of individual finger movements [18–21]. The exact nature of movement representation in the sensorimotor cortex and what makes some movements better decodable than others is thus still to be fully elucidated.

**Limitations:** One main limitation of our study is the small sample size. We assume that the predictions can be improved with higher sample sizes, making the compromise between the group and an individual’s optimal result even lower. Additionally, the limitation

of only ten repetitions per gesture may have led to sub-optimal accuracy results, thus, increasing the repetition count could yield higher scores. Furthermore, at the study’s current state, a direct translation to ECoG-BCIs is not possible, as our features were selected from the entire sensorimotor cortex which is not fully accessible by surface recordings. A more restrictive feature selection that overlaps with the recording of an ECoG grid can benefit the translation. We also acknowledge that the acquired data is from able-bodied participants who produced overt motor output. Even though the sensorimotor cortex of paralyzed patients still shows activity [22], potential BCI users might vary more widely in how well they can induce similar activity patterns when attempting certain hand movements. Thus, current results may not apply to every BCI user.

**Future directions:** Some gestures showed high decodability, while others were prone to confusion. However, the exact neuronal mechanisms underlying these variations are still unknown. Future work directed towards understanding which parameters are driving distinct representations in the sensorimotor cortex would not only provide valuable information for optimizing decoding algorithms for BCIs but would also enhance our basic understanding of the nature of movement representations in the sensorimotor cortex.

## CONCLUSION

In this paper, we investigated the decodability and consistency of sets of six gestures extracted from 20 different gestures. Our findings show potential gestures that exhibit robust decodability across individuals. The consistent variations in classification performance across gestures indicate substantial underlying similarity in sensorimotor representation patterns across individuals that makes some gestures more easily decodable than others. The findings highlight the potential for improving BCI control through optimized gesture selection.

## REFERENCES

- [1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*. 2002;113(6):767–791.
- [2] Laureys S *et al*. The locked-in syndrome : What is it like to be conscious but paralyzed and voiceless? *Progress in brain research*. 2005;150:495–511.
- [3] Hochberg LR *et al*. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*. 2012;485(7398):372–375.
- [4] Wang W *et al*. An electrocorticographic brain interface in an individual with tetraplegia. *PloS one*. 2013;8(2):e55344.

- [5] Vansteensel MJ *et al.* Fully implanted brain-computer interface in a locked-in patient with als. The New England journal of medicine. 2016;375(21):2060–2066.
- [6] Beisteiner R *et al.* Finger somatotopy in human motor cortex. NeuroImage. 2001;13(6 Pt 1):1016–1026.
- [7] Dechent P, Frahm J. Functional somatotopy of finger representations in human primary motor cortex. Human brain mapping. 2003;18(4):272–283.
- [8] Bleichner MG, Jansma JM, Sellmeijer J, Raemaekers M, Ramsey NF. Give me a sign: Decoding complex coordinated hand movements using high-field fmri. Brain topography. 2014;27(2):248–257.
- [9] Bruurmijn MLCM, Raemaekers M, Branco MP, Ramsey NF, Vansteensel MJ. Distinct representation of ipsilateral hand movements in sensorimotor areas. The European journal of neuroscience. 2021;54(10):7599–7608.
- [10] Bruurmijn MLCM, Pereboom IPL, Vansteensel MJ, Raemaekers MAH, Ramsey NF. Preservation of hand movement representation in the sensorimotor areas of amputees. Brain : a journal of neurology. 2017;140(12):3166–3178.
- [11] Hermes D, Miller KJ, Vansteensel MJ, Aarnoutse EJ, Leijten FSS, Ramsey NF. Neurophysiologic correlates of fmri in human motor cortex. Human brain mapping. 2012;33(7):1689–1699.
- [12] Leinders S *et al.* Using fmri to localize target regions for implanted brain-computer interfaces in locked-in syndrome. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology. 2023;155:1–15.
- [13] Siero JCW, Hermes D, Hoogduin H, Luijten PR, Petridou N, Ramsey NF. Bold consistently matches electrophysiology in human sensorimotor cortex at increasing movement rates: A combined 7t fmri and ecog study on neurovascular coupling. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2013;33(9):1448–1456.
- [14] Ojemann GA, Ojemann J, Ramsey NF. Relation between functional magnetic resonance imaging (fmri) and single neuron, local field potential (lfp) and electrocorticography (ecog) activity in human cortex. Frontiers in human neuroscience. 2013;7:34.
- [15] Marques JP, Kober T, Krueger G, van der Zwaag W, van de Moortele PF, Gruetter R. Mp2rage, a self bias-field corrected sequence for improved segmentation and t1-mapping at high field. NeuroImage. 2010;49(2):1271–1281.
- [16] Schellekens W, Bakker C, Ramsey NF, Petridou N. Moving in on human motor cortex. characterizing the relationship between body parts with non-rigid population response fields. PLoS computational biology. 2022;18(4):e1009955.
- [17] PENFIELD W, BOLDREY E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain : a journal of neurology. 1937;60(4):389–443.
- [18] Ben Hamed S, Schieber MH, Pouget A. Decoding m1 neurons during multiple finger movements. Journal of neurophysiology. 2007;98(1):327–333.
- [19] Schieber MH. Constraints on somatotopic organization in the primary motor cortex. Journal of neurophysiology. 2001;86(5):2125–2143.
- [20] Schieber MH. Motor cortex and the distributed anatomy of finger movements. Advances in experimental medicine and biology. 2002;508:411–416.
- [21] Shah NP *et al.* Pseudo-linear summation explains neural geometry of multi-finger movements in human premotor cortex. bioRxiv : the preprint server for biology. 2023.
- [22] Shoham S, Halgren E, Maynard EM, Normann RA. Motor-cortical activity in tetraplegics. Nature. 2001;413(6858):793.

## TRANSFERRING BCI MODELS FROM CALIBRATION TO CONTROL: OBSERVING SHIFTS IN EEG FEATURES

Ivo Pascal de Jong, Lüke Luna van den Wittenboer, Matias Valdenegro-Toro, Andreea Ioana Sburlea

Department of Artificial Intelligence, Bernoulli Institute, University of Groningen, The Netherlands

E-mail: ivo.de.jong@rug.nl

**ABSTRACT:** Public Motor Control-based brain-computer interface (BCI) datasets are being used to develop increasingly good classifiers. However, they usually follow discrete paradigms where participants perform Motor Imagery, Attempts or Execution at regularly timed intervals. It is often unclear what changes may happen in the EEG patterns when users attempt to perform a control task with such a BCI. This may lead to generalisation errors. We demonstrate a new paradigm containing a standard calibration session and a novel BCI control session based on EMG. This allows us to observe similarities in sensorimotor rhythms, and observe the additional preparation effects introduced by the control paradigm. In the Movement Related Cortical Potentials we found large differences between the calibration and control sessions. We demonstrate a CSP-based Machine Learning model trained on the calibration data that can make surprisingly good predictions on the BCI-controlled driving data.

### INTRODUCTION

The public availability of various BCI datasets has allowed for more transparent and more reliable progress in the development of Machine Learning models for EEG processing. The most convenient datasets to collect and make Machine Learning models for assume cue-based BCIs. These have cleanly separated instances of the various classes, which increases consistency and makes for a clear classification task.

However, such Machine Learning benchmarks often do not align with the EEG processing that a BCI with high usability needs. BCI competition IV dataset 1 [1] attempts to address this by aiming for Motor Imagery classifiers where the cue is not known in the EEG processing. This dataset has a training section with visual cues for three Motor Imagery class, and a test session with auditory cues for the same classes and a *rest* condition. The candidate models then need to predict for all moments in the test session which of the four states (including the *rest*) the participant is in.

This dataset bridges a gap from the classical trial-based EEG classification to BCI systems that need to make predictions without trial information. However, it also has some limitations that we aim to address with the introduction of a newly collected dataset.

[1] highlights that the participants will have a transient phase between hearing the auditory cue and performing the corresponding task. The timing of this is not precisely known, so the models are not evaluated on these transient phases. As a result, these models will not be optimised to detect the onset of a new state. This can be a problem when BCIs require low latency, which is important for learning to use a BCI [2].

Like most available paradigms, it also does not give a good reflection of the mental state of a user using the BCI to achieve a task. The transition from a BCI model that works well in a controlled paradigm to using the BCI to perform a control task introduces many unknowns. The EEG patterns may change due to planning, eye movements or visual attention and it is generally unclear exactly what does and does not change when shifting from a calibration paradigm to a control task.

*Contributions:* To address these issues we demonstrate a new paradigm with a preliminary data analysis. The paradigm has a visual cued calibration session similar to BCI competition IV dataset 1 [1]. However, the testing session has the participant drive a simulated car. The steering of the car is done through the detection of the flexion of the left and right hand based on electromyography (EMG) signals. By using the EMG to control the car we can observe the EEG of a participant as if they are using a BCI for a control task. The task is then to predict the motor execution state (as measured by the EMG), based on the EEG patterns from the motor cortex.

Motor Execution is chosen over Motor Imagery as an analog for Motor Attempts in paralysed patients because Motor Imagery results in lower BCI accuracy in stroke patients compared to Motor Attempt [3]. Motor Execution might give optimistic results due to affect effects.

The track is designed to have left turns, right turns, and straight sections that can be used as artificially segmented trials. The EMG can then be used to determine the motor execution onset, and allow us to investigate differences in EEG patterns between the calibration session and the driving session.

We show a preliminary analysis of a dataset recorded under this paradigm looking at Sensorimotor Rhythms (SMR) and Movement Related Cortical Potentials (MRCP) for calibration and compare these to the artificially segmented trials during driving. We also demonstrate classification with multiclass CSP [4] in the cali-

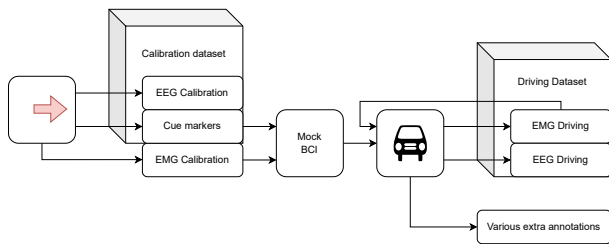


Figure 1: Design of the data acquisition. First participants do a calibration session following the Graz-BCI Motor Imagery paradigm. The EMG for this is used to develop a *mock BCI*, which the users then use to control a simulated car. The EMG, EEG and markers are recorded in both sessions resulting in a driving dataset and a calibration dataset.

bration session, the driving session, as well as a classifier trained on the calibration and applied to the driving session.

We believe that introducing this paradigm will allow for experimentation with motor decoding models that are exceptionally well suited for transferring from the benchmark to the user.

## METHODS

The implementation of the paradigm focuses on building a *mock BCI* that the participants can use to perform a control task. The goal is to collect all the EEG data as if the participants are using a BCI, without the risk of complete loss of control introduced by using a real EEG-based BCI.

The design for the current study consists first of a calibration session following the Graz-BCI Motor Imagery paradigm [5]. The EMG from the calibration session is used to make an EMG-based classifier, which will function as our *mock BCI*. The EEG is recorded for offline analysis. After this, the EMG-based *mock BCI* is used to make online predictions during a simulated driving task. This gives us the EEG patterns of our participants using the *mock BCI*, with EMG as the ground truth of their control intentions.<sup>1</sup> The whole setup of the recording is visualised in Figure 1.

In an offline analysis, we can then look at the EEG during calibration using the cues as the ground truth and the EEG during driving using the EMG as the ground truth. Within the calibration and driving sessions we can look at the MRCP and SMR, and we can develop classifiers on the calibration EEG and apply them to make predictions with the driving EEG.

This study was conducted with 20 healthy participants aged 19-45 years old ( $\mu = 26, \sigma^2 = 40$ ), although the data from one participant was removed from the analysis because a section of the driving was not recorded. None of the participants had prior BCI experience. The EMG, EOG and EEG were recorded using the Biosemi

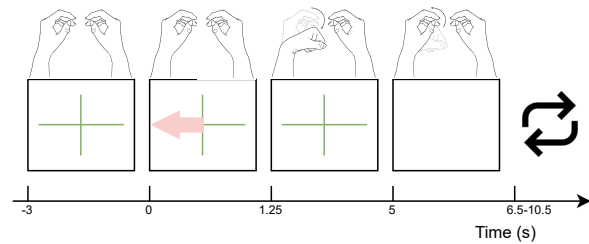


Figure 2: Timing of the calibration paradigm from OpenVibe.

ActiveTwo. 32 EEG channels were collected following the 10-20 system. Two monopolar EMG channels were measured on each forearm to correspond with wrist flexion and wrist extension, resulting in four total EMG channels. Four EOG channels were measured to capture horizontal and vertical eye movements. All 40 channels were recorded at 2048Hz. Participants sat in a chair facing a computer with their arms resting on a desk at a comfortable height. A towel was placed under the forearms to elevate the wrists slightly which allowed for easy wrist flexion.

Whenever participants needed to perform a "left" command they flexed their left hand inward. The "right" command corresponded with flexing the right hand inward. At the end of each trial or turn they returned their hand to a forward resting position.

**Calibration Paradigm:** The Motor Imagery paradigm from OpenVibe[6] version 3.5.0 was used without any modification to the timing. This entails a 30 second preparation time, followed by 2x20 shuffled trials of left and right hand motor execution. Each trial starts with a cross displayed for 3 seconds as a preparation cue. Then an arrow pointing left or right is displayed for 1.25 seconds. When this disappears the participant performs the movement and keeps the wrist flexed while looking at the cross. After 3.75 seconds this cross disappears and the participant moves their hand back to a resting position. A random rest period of 1.5 to 3.5 seconds separates each of these trials. The timing of the paradigm is visualised in Figure 2. We set  $t = 0$ s to the moment that the participant knows the direction, so the movement onset is at  $t = 1.25$ s

We chose to have the movement performed after the arrow disappears, instead of when it appears. This way, the participant is able to prepare the action belonging to the visual cue and initiate the movement when the arrow appears. This should provide a more consistent movement onset, and allows effects of movement planning and inhibition to be included in the calibration data.

In between the left and right hand trials 40 rest samples were extracted, from 0.5s after the end of the previous trial, until 4.5s after. This means it is partly recorded while the participant is looking at a blank screen, and partly when the fixation cross of the next trial is shown. This makes the rest slightly more noisy, but also more similar to rest periods during the driving session. These trials will be used for the online EMG classifier, as well as for the offline analysis of the calibration dataset.

<sup>1</sup>The driving paradigm is available at [https://github.com/lukeeluna/continuous\\_control](https://github.com/lukeeluna/continuous_control), the EMG classifier at [https://github.com/ivopascal/emg\\_classifier](https://github.com/ivopascal/emg_classifier) and the analysis code at [https://github.com/ivopascal/continuous\\_control\\_bci](https://github.com/ivopascal/continuous_control_bci).

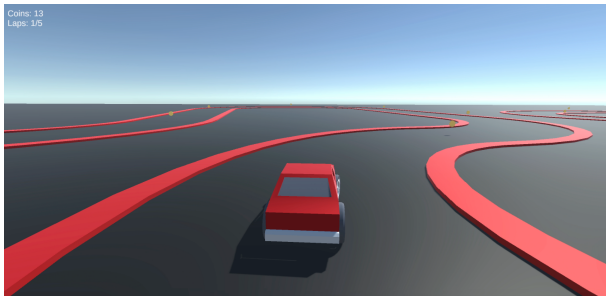


Figure 3: Participant view during the driving session.

**EMG Classifier:** We train a subject-specific online EMG classifier on the EMG collected during the calibration session. This EMG classifier will be used during the driving session to control the simulated car.

The EMG is measured with 4 monopolar electrodes to measure two muscles on each arm. Specifically the flexor carpi radialis and the extensor carpi radialis longus.

First, the EMG of these 4 channels is re-referenced using Common Average Reference (CAR) over the EMG channels. Then the EMG is band-pass filtered between 30–500Hz and notch filtered at 50Hz using causal 4<sup>th</sup> order Butterworth filters. Since the online classification should have low latency, the trials are cropped to the middle 200ms of each movement from  $t = 6.125s$  to  $t = 6.145s$ . We found this to have a minimal impact on classification performance. The mean power of the four EMG channels are used as four features for an LDA classifier. Using 10-fold cross-validation we find that this per-subject EMG classifier has a mean accuracy of 94% with a standard deviation of 5%.

The online implementation of the EMG classification uses Lab Streaming Layer (LSL). LSL allows for real-time streaming of the EMG recordings from BioSemi to the EMG classifier (implemented with MNE-Python [7]), and streaming the classifications to the driving environment. It also saves the multiple streams and provides precise alignment of timestamps from the various streams.

**Driving Paradigm:** As a control task, we chose to control a simulated car in Unity. At the start the car accelerates until it reaches a predefined constant speed. The speed is kept consistent during the driving so the participant needs to initiate and terminate turns at the right time. We believe this is an important factor in developing a BCI for control tasks, as early preparation of a movement may cause patterns in the premotor cortex which may be confused with the actual movement. The ability to distinguish between movement preparation and movement execution is necessary to be able to have a good estimate of the movement onset.

The 3D driving environment is shown in Figure 3. The choice of a 3D design allows for better engagement, but may also affect the EEG with eye movements, visual perception effects, or planning effects. This makes the recorded EEG more ecologically valid for real BCI control.

Each participant drives around the track for 5 laps. After

these 5 laps, they have a break and start again when they feel ready. They perform this task 3 times for a total of 15 laps, resulting in an average of 37.5 minutes ( $\pm 6$  minutes) of driving per participant.

From the driving session, we extract trials of left turns, right turns, and straight sections by identifying periods of at least 3.75 seconds where the EMG classifier made the same prediction. This allows us to investigate the SMR and MRCP for left and right turns, and gives rest trials sufficiently similar to the calibration session. Any trials with peak-to-peak differences exceeding  $100\mu V$  after epoching are rejected. This resulted in an average of  $54 \pm 13$  left turns,  $27 \pm 8$  right turns and  $41 \pm 22$  straight sections.

## RESULTS

**Sensorimotor Rhythms:** We look for the Event Related Desynchronisation (ERD) that is commonly found in Motor Imagery, Motor Attempts [3] and Motor Execution [8]. Specifically, we look for contralateral ERD in the  $\alpha$  band around 8 – 12Hz, which may start before movement onset due to movement planning. We may also find an ERD in the  $\beta$  band around 12.5 – 30Hz.

We apply CAR followed by a non-causal FIR band-pass filter in the range [1, 35]Hz with a lower transition bandwidth of 1Hz, an upper transition bandwidth of 8.75Hz and a filter length of 3.3s. Then we apply artefact removal with FastICA [9], and take the Surface Laplacian [10] around channels C3 and C4. This gives us the relevant frequencies around the parts of the motor cortex responsible for left hand and right hand movement while minimising artefacts. We take epochs from both calibration and driving sessions such that the movement starts at  $t = 1.25$ . Relative to the movement onset we look at the time-frequency effects from  $t = -3$  to  $t = 5$ . This gives us an indication of the activation before movement onset, and for the remainder of the trial.

We then use DPSS multitapers to determine the time-frequency response from [5, 35]Hz, at increments of 1Hz. For this analysis, the epochs are temporarily padded with 0.5s of leading and trailing EEG data to avoid edge effects. Unlike common ERD visualisations we do not subtract the baseline activation. The baseline sections between calibration and driving are very different, which would make it difficult to distinguish changes in the baseline from ERD/ERS effects. Instead, we look at the absolute time-frequency plots.

This pipeline is applied to both left hand and right hand movements, both during calibration and driving. This allows us to see the differences that may affect decoding during calibration and driving. Figure 4 shows the time-frequency response during calibration and driving for both left and right trials averaged over the participants.

In both cases, the movement onset is at  $t = 1.25s$ . We see a contralateral decrease in the  $\alpha$  band for both cases, starting slightly before movement onset. In the calibra-

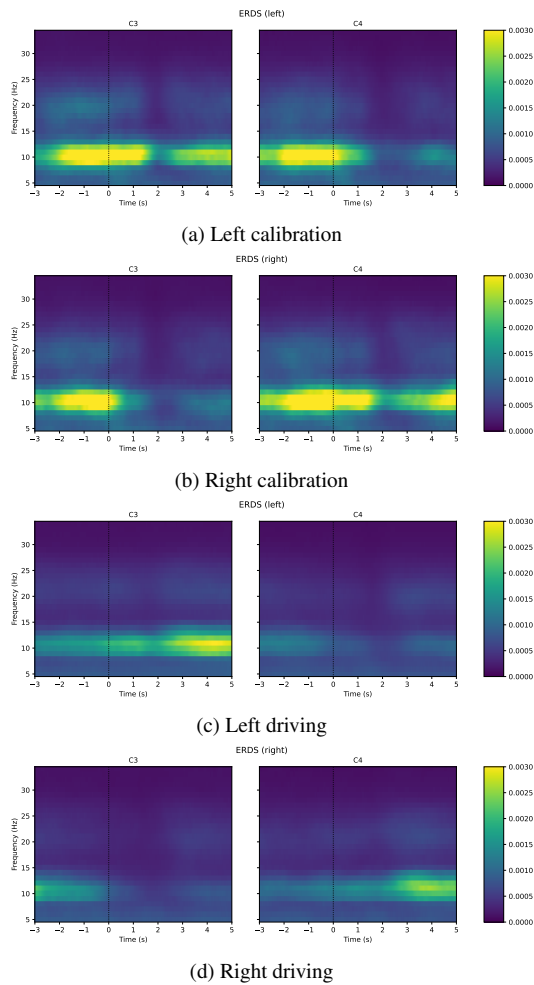


Figure 4: SMR during the left and right trials for calibration and driving. No baseline is used.

tion session, the decrease starts to appear around  $t = 1$ s, roughly 0.25 seconds before movement onset. In the driving session, the  $\alpha$  band decrease starts as early as  $t = 0$ s. Around  $t = 4$ s we see a slight rebound in the  $\alpha$  band for calibration. This rebound also appears in the driving session starting at  $t = 3$ . This early rebound may be because the hand is in a consistent flexed position, so there is no more hand movement.

In all figures we see a short dip in the  $\beta$  band in both channels, aligned with the onset of the movement. The timing of the dip in the  $\beta$  band may allow it to be used to determine the onset of the movement, while the contralateral power decrease in the  $\alpha$  band may be used to determine which hand was moved.

**Movement Related Cortical Potentials (MRCP):** We investigated MRCPs as they are commonly used to study motor preparation effects [11]. Since MRCPs are low-frequency phenomena we apply a causal 8<sup>th</sup> order butterworth band-pass filter in the range  $[0.1, 3.0]$ Hz. After this artefacts are removed with FastICA and the trials are then epoched so that the movement onset is at  $t = 1.25$ s.

Figure 5 shows the MRCPs at C3 and C4 for calibration and driving. During the calibration, we observe a negative peak slightly before the movement onset at  $t = 1$ s,

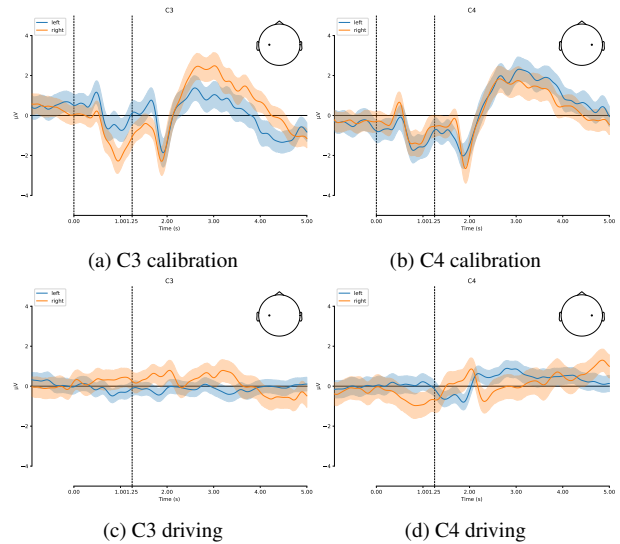
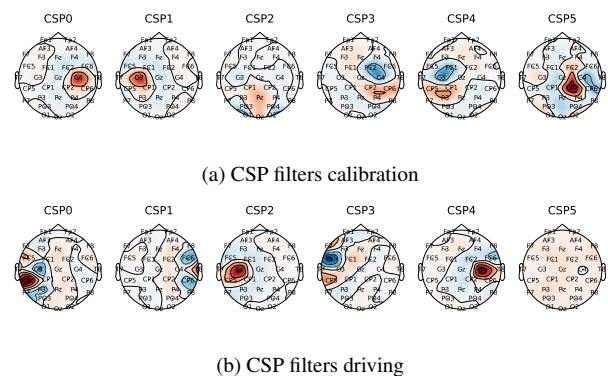


Figure 5: Average MRCP across participants during calibration and driving. The movement starts at  $t = 1.25$ , indicated by the rightmost dashed line. In the calibration plots the dashed line at  $t = 0$  shows when participants are given the directional cue. The blue line indicates the average ERP for left trials, the orange for right trials. The shaded areas indicate the 95% confidence interval of the mean.

which is the contingent negative variation [12]. This effect disappears in the driving session. There are clear patterns that distinguish the classes during calibration, but they do not transfer well to the driving session.

**CSP classifier:** To demonstrate the ability to classify individual samples based on SMR patterns we implemented a simple CSP-based classifier.

First, the EEG data is re-referenced with CAR and the data is filtered with a non-causal FIR band-pass filter in the range  $[1, 35]$ Hz with a lower transition bandwidth of 1Hz, an upper transition bandwidth of 8.75Hz and a filter length of 3.3s. Then, FastICA is used to remove artefacts. The ICs were fitted and rejected or kept based on the calibration data, but used on both the calibration and the driving data. This makes the process suitable for on-line implementations. Epochs are taken from movement onset  $t = 1.25$ s to the end of the trial  $t = 5$ s. Rest samples are also used from between the trials in calibration,



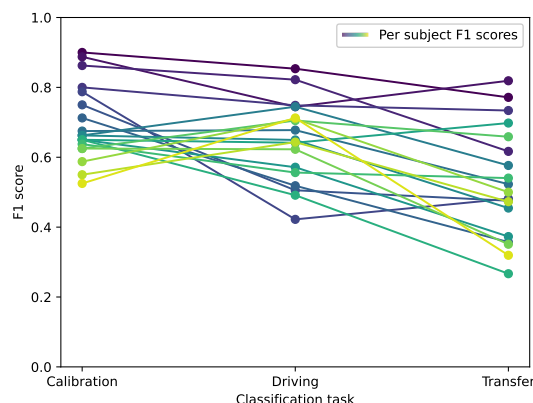


Figure 7: F1 scores of the CSP classifiers in different sessions. We show trained and tested on calibration, trained and tested on driving, and lastly, the transfer scenario is trained on calibration and tested on driving. Each line represents one participant.

or from straight sections in driving.

Six multiclass CSP filters [4] regularised with shrinkage are used to extract the features. The logarithm of the band power for each filter is used by a Logistic Regression classifier to predict the classes. This pipeline is fitted 5 times with 5-fold cross-validation to make the predictions. The model is then trained once more on all the data to show the CSP filters.

This procedure is applied once to the calibration data, and once to the driving data. Lastly, we also train the model on all the calibration data and use it to make predictions on the driving data. Transferring the model in this way means it is never trained on EMG data but is still able to predict the EMG data.

We use the first 10% of driving trials to calibrate a rest threshold, addressing a shift in class imbalance between calibration and driving. The transfer performance is then evaluated on the remaining 90% of driving trials. The threshold that would need to be used in an online scenario should be set manually.

Figure 6 shows the learned CSP filters for the calibration and driving sessions over all participants combined, ordered by Mutual Information. In the calibration CSP0 and CSP1 clearly identify the relevant areas of the motor cortex for left and right hand motor control. CSP3 and CSP4 show bipolar effects around the motor cortex. The CSP filters learned on the driving data are not as clean. However, CSP2 and CSP4 still correspond to the relevant parts of the motor cortex. There are no discernible effects in the prefrontal areas, indicating that the learned patterns are not due to eye movement artefacts.

Figure 7 shows the F1 scores for each participant in each session. The performance is best in the calibration session ( $\mu = 0.69, \sigma^2 = 0.011$ ), followed by the driving session ( $\mu = 0.65, \sigma^2 = 0.013$ ), followed by the transfer scenario ( $\mu = 0.53, \sigma^2 = 0.025$ ). The participants with poor performance on the calibration data are also more likely to have poor performance in transfer, but this is not consistent.

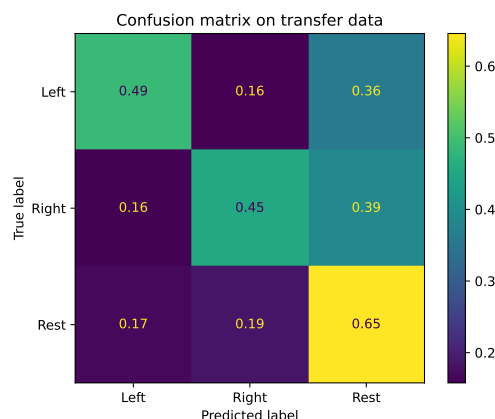


Figure 8: Averaged confusion matrix for model transfer.

Lastly, Figure 8 shows the confusion matrix averaged over all participants in the transfer scenario. A model was trained for each participant based on their calibration data, and then tested on their driving data. The (mis)classifications were then aggregated into this confusion matrix. There are few cases where a left hand is mistaken for a right hand, or the other way around, but many movement samples are still mistaken for rest samples.

## DISCUSSION

The SMR analysis showed that there are patterns to observe in the calibration phase, that remain mostly consistent when people are using a BCI for a control task. We notice that there is a longer movement preparation effect and that the average  $\alpha$  power is lower during the driving task. The dip in the  $\beta$  band corresponds with the movement onset both in the calibration and in the driving session, and is minimally affected by preparation effects. We found that the MRCs are different between calibration and driving. ERPs identified in the calibration data are not identifiable in the driving data. This may be because the two paradigms are too different, or because the 200ms EMG classification interval gives too imprecise movement onsets.

From the MRCs and the SMR we find that a classifier that needs to be transferable from calibration to driving is best suited using band power features. The subsequent CSP-based classifier was able to distinguish between left and right hand trials surprisingly well, but it had quite a few false positives for the rest trials. As expected, accuracy becomes worse in driving due to the added stimulation of the environment, and the accuracy becomes even worse when transferring from calibration to driving. However, both of these factors have a relatively minor impact compared to the individual differences between participants.

The CSP filters showed that the models are picking up phenomena from the motor cortex, indicating that the patterns are originating from movement (intention) and not from eye artefacts or visual attention. The fact that the

ICA was fitted to the calibration, and used in the driving makes this design suitable for use in an online BCI.

## CONCLUSION

The findings from the SMR analysis, MRCP analysis and the CSP classifier show that an SMR-based classifier trained on calibration and applied for a control task is feasible. The SMR patterns are largely similar between the calibration and the driving session, allowing for a surprisingly small decrease in classification accuracy.

However, the current analysis still leaves some hurdles for the implementation of an online EEG-based BCI to control the driving task that can be addressed by future analyses on this dataset.

The primary limitation is that this paper is still making classifications and doing analysis on isolated sections with 3.75s of a consistent movement. To allow continuous control of a BCI we need to have predictions at every timestamp as proposed by [1]. Solutions proposed to this continuous decoding as presented in [13] may be a suitable direction.

The second limitation is that the classification performance of the models that we employed is very low for some participants. For the participants on which the highest performance was achieved an acceptable driving control would be possible. For the participants with low performance ( $F1 < 0.6$ ) it is unlikely that the EEG-based BCI will allow for successful control.

The problem of low-performing participants may be addressed in future experiments by having participants use the EMG and EEG for control in a shared-control BCI [14, 15]. By interleaving EMG predictions with EEG predictions and providing the participants with feedback they can improve the separation of their EEG patterns, without reaching complete loss of control. Such a shared-control BCI should maximise the amount of EEG-based predictions to allow for learning, while still maintaining sufficient control with EMG.

## REFERENCES

- [1] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
- [2] R. Xu, N. Jiang, C. Lin, N. Mrachacz-Kersting, K. Dremstrup, and D. Farina, "Enhanced low-latency detection of motor intention from eeg for closed-loop brain-computer interface applications," *IEEE Transactions on biomedical engineering*, vol. 61, no. 2, pp. 288–296, 2013.
- [3] S. Chen, X. Shu, H. Wang, L. Ding, J. Fu, and J. Jia, "The differences between motor attempt and motor imagery in brain-computer interface accuracy and event-related desynchronization of patients with hemiplegia," *Frontiers in neuro-robotics*, vol. 15, p. 706 630, 2021.
- [4] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, 2008.
- [5] G. Pfurtscheller *et al.*, "Graz-bci: State of the art and clinical applications," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 11, no. 2, pp. 1–4, 2003.
- [6] Y. Renard *et al.*, "Openvibe: An open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments," *Presence*, vol. 19, no. 1, pp. 35–53, 2010.
- [7] A. Gramfort *et al.*, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, no. 267, pp. 1–13, 2013.
- [8] G. Pfurtscheller, "Spatiotemporal erd/ers patterns during voluntary movement and motor imagery," *Supplements to Clinical neurophysiology*, vol. 53, pp. 196–198, 2000.
- [9] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
- [10] J. Kayser and C. E. Tenke, "On the benefits of using surface laplacian (current source density) methodology in electrophysiology," *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, vol. 97, no. 3, p. 171, 2015.
- [11] A. Shakeel, M. S. Navid, M. N. Anwar, S. Mazhar, M. Jochumsen, I. K. Niazi, *et al.*, "A review of techniques for detection of movement intention using movement-related cortical potentials," *Computational and mathematical methods in medicine*, vol. 2015, 2015.
- [12] W. G. Walter, R. Cooper, V. Aldridge, W. McCallum, and A. Winter, "Contingent negative variation: An electric sign of sensori-motor association and expectancy in the human brain," *nature*, vol. 203, no. 4943, pp. 380–384, 1964.
- [13] M. Tangermann *et al.*, "Review of the bci competition iv," *Frontiers in neuroscience*, p. 55, 2012.
- [14] V. Mondini, R. J. Kobler, A. I. Sburlea, and G. R. Müller-Putz, "Continuous low-frequency eeg decoding of arm movement for closed-loop, natural control of a robotic arm," *Journal of Neural Engineering*, vol. 17, no. 4, p. 046 031, 2020.
- [15] A. I. Sburlea, N. Butturini, and G. R. Müller-Putz, "Predicting emg envelopes of grasping movements from eeg recordings using unscented kalman filtering," in *Annual Meeting of the Austrian Society for Biomedical Engineering. Vol. 2021*, Verlag der TU Graz, 2021, pp. 71–74.

# RESTING-STATE BRAIN CRITICALITY AND PERFORMANCE WITH P300-BASED BCIS

T. Settgast<sup>1</sup>, A. Kübler<sup>1</sup>

<sup>1</sup> Institute of Psychology, Section Intervention Psychology, Julius-Maximilians-Universität  
Würzburg

E-mail: tomko.settgast@uni-wuerzburg.de

**ABSTRACT:** Here we present correlations between criticality-related measures calculated from resting-state electroencephalography (EEG) recordings and subsequent performance with a visual P300-based brain-computer interface (BCI) in healthy participants. Results suggest a positive relationship between resting-state brain criticality and subsequent BCI performance using P300-based BCIs.

## INTRODUCTION

The P300-based brain-computer interface (BCI) speller is the most widely used BCI application, and P300-based BCIs are generally described as being convenient, reliable, and easy to use [1] – see also a direct comparison between sensorimotor rhythm-based and event-related potential (ERP)-based BCI applications such as the P300-based BCI speller in Kübler et al. [2]. However, attempts to predict the future performance of a P300-based BCI based on neural brain activity have been rare<sup>1</sup>. Halder et al. [3] and Won et al. [1] showed that in healthy participants<sup>2</sup>, elicited ERPs in a preceding oddball or rapid serial visual presentation (RSVP) task were related to subsequent performance with a P300-based BCI. In the Won et al. [1] sample, the P300 amplitude elicited during the RSVP was positively correlated with subsequent P300 speller performance. The sample of Halder et al. [3] showed that the amplitude of the N2 ERP elicited in the preceding oddball paradigm was related to subsequent performance with a visual P300-based BCI. The latter result was subsequently replicated in patients with amyotrophic lateral sclerosis (ALS) [6]. Nonetheless, these attempts were based on task-related activity and would not allow for performance prediction based on spontaneous brain activity. To establish relationships between spontaneous brain activity and performance

with a P300-based BCI that may be useful for predicting performance in the future, it seems fruitful to investigate correlations between performance with P300-based BCIs and preceding resting-state electroencephalography (EEG) recordings.

While attempts have been made to detect relationships between resting-state brain activity and the subsequent performance achieved with motor imagery BCIs [8, 9, 10, 11], to our knowledge there is only one available study that has attempted to predict BCI performance for the use of a P300-based BCI from resting-state brain activity in healthy participants. Shin et al. [12] found a negative correlation between delta-frequency band power in the resting-state EEG and the subsequently achieved performance with a P300-based BCI. They also reported negative correlations between delta- and alpha-frequency band connectivity at rest and subsequent BCI performance, and a positive correlation between gamma-frequency band connectivity at rest and subsequent BCI performance. Recently, we also investigated the relationship between resting-state brain activity and the subsequently achieved performance with a P300-based BCI [13]. Aiming to establish a relationship between BCI performance and the level of consciousness, we examined the correlations between two theoretically supported measures of consciousness, i.e., the power-law exponent (PLE) and the Lempel-Ziv complexity (LZC), at rest and the subsequently achieved performance with a P300-based BCI. We showed strong and significant correlations between both PLE and LZC at rest and the performance of a locked-in ALS-patient during the subsequent use of a tactile P300-based BCI.

The PLE<sup>3</sup> provides information about the non-periodic, arrhythmic, and scale-free activity of the brain [14, 15, 16] by means of the  $1/f$  aperiodic scaling [17], i.e. the slope of the EEG power spectrum, also called spectral

<sup>1</sup> We have limited our review here to neurophysiological predictors of P300-based BCI performance because of our interest in the relationship between spontaneous brain activity and BCI performance. Research on the influence of psychological factors, e.g., Kleih et al. [4], is not included.

<sup>2</sup> We specifically focus on healthy participants here because many studies investigating the predictability of upcoming performance for the use of a P300-based BCI have examined patients with amyotrophic lateral

sclerosis (ALS) (e.g., [5, 6]). However, we have recently suggested that the brain activity of ALS-patients shows alterations in brain criticality, making the relationship between their spontaneous brain activity and the use of a P300-based BCI a special case [7].

<sup>3</sup> We will use PLE as a catch-all abbreviation throughout this paper, although the cited literature may have referred to this phenomenon as *1/f*, *aperiodic*, or *scale-free activity*, *1/f slope*, *1/f or aperiodic scaling*, *spectral slope*, or *power-law distribution*.

slope [18, 19]. This means that with increasing power in lower frequencies and decreasing power in higher frequencies, the slope of the power spectrum becomes steeper and the PLE value increases. Conversely, with decreasing power in lower frequencies and increasing power in higher frequencies, the slope becomes flatter and the PLE decreases. The LZC applied to neural brain activity provides information about the complexity of a neural signal according to its compressibility [20]. The more easily the brain activity can be compressed, the less complex the corresponding brain activity is. This complexity has been interpreted as reflecting the amount of information content in conscious experience [21], most clearly formulated in the *Entropic Brain Hypothesis* [22]. The PLE plays a prominent role in the *Temporo-Spatial Theory of Consciousness*, where it shows how neural activity of different temporal and spatial scales are nested within a single conscious experience [23]. Both PLE and LZC have recently received empirical support for their ability to discriminate between different states of consciousness [14, 18, 24] or brain states [25, 19]. In addition, recent results have provided evidence for their reactivity to sensory processing [24, 26] as a function of the participants' states of consciousness [24] and the ongoing task demands [26]. Finally, both PLE and LZC are considered to be closely related to the so-called brain criticality, making them *criticality-related measures* that indicate increasing brain criticality with increasing LZC and decreasing PLE, and decreasing brain criticality with decreasing LZC and increasing PLE [18].<sup>4</sup>

Since brain activity at the point of criticality is considered to express remarkable information processing capabilities, with maximal sensitivity to perturbations, an enriched repertoire of system states, and a high capacity to store and transfer information [17], brain criticality may be an interesting concept to be explored in the context of BCI use. Therefore, we have recently discussed the relevance of brain criticality for the use of P300-based BCIs [7]. Based on the available literature, we have argued that an increase in resting-state brain criticality appears to be beneficial for reorganizing brain activity to meet upcoming task demands, such as the use of a P300-based BCI. Central to our argument, Irmischer et al. [27] showed that while a measure of brain criticality during a sustained attention task was negatively related to the performance in that task, the same measure showed a positive relationship with task performance when derived from the preceding resting-state EEG recording. Thus, Irmischer et al. [27] hypothesized that two distinct

processes are at work when it comes to criticality and task performance in attention-demanding tasks. While increased brain criticality at rest is indicative of the brain's ability to adapt to upcoming task demands, the execution of an attention-demanding task appears to favor less critical brain activity. This favorable reduction in brain criticality was also demonstrated by the only study, other than our own [17], that examined a relationship between brain criticality and the use of a P300-based BCI [28]. The authors showed that when using a P300-based BCI, decreasing brain criticality<sup>5</sup> was associated with increased P300 amplitudes. In contrast, Herzog et al. [30] showed that, consistent with Irmischer et al. [27], resting-state brain criticality was positively associated with the P300 amplitude in a subsequent Go/Nogo task. Investigations of the relationship between functional connectivity and the P300 further support the proposed relationship between resting-state brain criticality and performance on an attentionally demanding task. Functional connectivity and brain criticality have been shown to be positively related (see, e.g., [31,32]). Given this, Li et al.'s [33] finding of a positive relationship between increased functional connectivity at rest and the P300 amplitude in a subsequent task, as well as Li et al.'s [34] finding of a positive relationship between decreased functional connectivity during the task and the P300 amplitude, suggest that while the P300 amplitude appears to be positively related with resting-state brain criticality, it appears to be negatively associated with on-task brain criticality.

These results suggest that resting-state brain criticality is likely to be related to subsequent performance with a P300-based BCI, and that increased criticality at rest is related to better performance during the subsequent use of a P300-based BCI. The latter hypothesis was tested in this paper using two selected criticality-related measures, LZC and PLE, and the open-access dataset of Won et al. [35]. In our first study [13], in which we looked for correlations between these variables and BCI performance in a locked-in ALS-patient, we already tried to find similar correlations for healthy participants from the Won et al. [35] dataset. Problematically, the online BCI performance of the participants in this dataset showed a pronounced ceiling effect, which prevented meaningful correlation analyses. To circumvent this ceiling effect in this re-analysis of these data, we here use BCI offline performance as a measure of BCI performance, calculated on the basis of the letter

<sup>4</sup> Note that the *spectral slope*, as calculated by Maschke et al. [18], is positively correlated with brain criticality. However, in contrast to our calculation (see Materials & Methods), they do not use the absolute value of the spectral slope. Accordingly, the PLE, as calculated here, can be expected to be negatively correlated with brain criticality. This difference in calculation also explains the observed strong correlation between the spectral slope and LZC in Maschke et al. [18], which contrasts with our observed anticorrelation between PLE and LZC (see

Results).

<sup>5</sup> Bojorges-Valdez and Yanez-Suarez [28] did not use the word *criticality*, only a measure of brain criticality. This is a more common phenomenon. A recent review highlighted the inconsistent use of *criticality* in the brain criticality literature [29]. Although authors may use concepts or measures that are part of the brain criticality concept, they do not necessarily refer to the concept or use the word.

detection accuracy after only two repetitions of the stimulus sequence (see Materials and Methods for details). We show correlations between our criticality-related measures calculated from resting-state EEG and the subsequent BCI offline performance using a visual P300-based BCI. The positive correlation between LZC at rest and subsequent BCI offline performance, and the negative correlation between PLE at rest and subsequent BCI offline performance, suggest that increased resting-state brain criticality is associated with better subsequent BCI performance when using a P300-based BCI.

## MATERIALS AND METHODS

*Dataset:* The dataset we used is part of the publicly available BCI dataset by Won et al. [35]. The data analyzed here include the BCI performance of 55 participants using a visual P300-based BCI speller, as well as their open-eyes resting-state EEG recordings that were taken prior to the BCI use and that were the closest to the BCI use (see Procedure).

*Participants:* 55 participants took part in the study of Won et al. [35]. 14 of the participants were female and their mean age was 22.91 years ( $\pm$  2.87). None of the participants were excluded for this analysis.

*Procedure:* The data analyzed here, i.e., BCI performance achieved with a visual P300-based speller and a preceding open-eyes resting-state EEG recording, were obtained as part of a larger experimental procedure consisting of 3 blocks of resting-state EEG recordings, each with an open-eyes and closed-eyes condition, a RSVP task, and the use of a visual P300-based BCI speller (for details, see [35]). The entire experiment was presented and recorded using BCI2000 [36]. The EEG data was recorded at a sampling rate of 512 Hz using a Biosemi Active Two system with 32 AG/AgCl active electrodes placed according to the international 10-20 system (Fp1, AF3, F7, F3, Fc1, Fc5, T7, C3, Cp1, Cp5, P7, P3, Pz, Po3, O1, Oz, O2, Po4, P4, P8, Cp6, Cp2, C4, T8, Fc6, Fc2, F4, F8,, AF4, Fp2, Fz, Cz). To investigate possible correlations between preceding resting-state EEG recordings and the performances reached with the visual P300-based BCI, which may have predictive value, we used the resting-state EEG recording that was taken before and closest to the BCI use. The resting-state EEG recordings analyzed here were recorded with opened eyes. The resting-state EEG was recorded for approximately 139 seconds, and the participants were instructed to fixate a cross on the screen in front of them, to remain relaxed, and to minimize movement. The subsequent BCI use consisted of two calibration runs and four test runs. In each of the two calibration runs, the participants were instructed to copy spell a word without visual feedback. In each of the test runs, the participants were instructed to copy spell a word with visual feedback. The visual P300-based BCI speller was based on a 6x6 matrix speller with six columns and six rows and included the letters of the alphabet, digits, and a space. Each of the 12 stimuli – six columns and six rows – of a stimulus sequence was flashed for 125 ms followed

by a 62.5 ms inter-stimulus interval before the next stimulus of the sequence was flashed. For the selection of a single letter, the stimulus sequence was presented 15 times, resulting in a total of 180 stimuli with 30 stimulus flashes comprising the target and 150 non-target stimulus flashes.

*BCI Offline Performance:* BCI offline performance was calculated from the letter detection accuracy provided with the dataset [35]. Letter detection accuracy was calculated as the number of correctly selected letters of the word to be copied in a single test run as a function of the number of repetitions of the stimulus sequence. Provided for each of the 55 participants, consisting of 4 test runs with one word to be copied per run, and 15 repetitions of the stimulus sequence per letter selection, the letter detection accuracy was provided as a matrix with dimensions of 55x4x15. We first calculated a mean letter detection accuracy for each participant as the average of the letter detection accuracies of the four words which had to be copied. As shown in Settgaest et al. [13], the letter detection accuracies after 15 repetitions of the stimulus sequence showed a pronounced ceiling effect, with 25 out of 55 participants having reached 100 percent accuracy and 51 out of 55 participants having reached the 70 percent benchmark for successful BCI performance [37]. To avoid problems related to this ceiling effect, we decided to use the letter detection accuracy whose distribution was closely centered around 50 percent and did not show a violation to normal distribution. We tested the normal distribution for each of the letter detection accuracies across participants as a function of the number of repetitions of the stimulus sequence. We chose the letter detection accuracy after two repetitions of the stimulus sequence as BCI offline performance. This cross-participant letter detection accuracy after two stimulus sequences showed a mean of 0.52 (0.2), showed no significant skewness and kurtosis, and showed no violation of normality according to the Shapiro-Wilk test.

*EEG Pre-Processing (Resting-State):* Resting-state EEG data was pre-processed with EEGLAB [38]. First, the data was band-pass filtered from 1-40 Hz. The 1-Hz low cutoff was applied according to the recommendations of Winkler et al. [39] to ensure good results in the later performed independent component analysis (ICA). The EEGLAB built-in clean\_artifacts function was used to remove flatline, highly correlated, and noisy channels as well as short time bursts and otherwise bad data periods. Deleted channels were then spherically interpolated. The EEG data was re-referenced to the common average because the used EEG recording device does not provide a hardware-based reference [35]. An ICA was performed, and artifact components were automatically flagged and removed using MARA [40].

*Data Analyses:* Both PLE and LZC values were obtained using custom MATLAB scripts. Each of the variables was computed using a sliding-window method with 1-second windows and 50% overlap between windows (for computation details, see [14]) for each of the 32 EEG channels. To account for differences in EEG

recording length due to bad data period rejection in the previous pre-processing step, we took 130 s of EEG data for each participant, starting 5 s after the start of the recording, for further analysis. To calculate the PLE, the power spectral density (for computation details, see [14]) was logarithmically transformed in both frequency and power spectrum domain. The slope of the PSD was then calculated by linear least squares regression. The PLE was then obtained as the absolute value of this slope. For further analysis, we took the average of these values across channels and time windows for each participant. LZC values were computed largely according to the algorithm of Zhang and Roy [41], using the median as the threshold for binarization due to its robustness to outliers [42]. To reflect the number of accruing pattern in the sequence, the LZC was normalized [43]. As for the PLE, the LZC values were averaged across time windows and channels for further analysis. LZC and PLE were tested for normality. None of the variables showed significant kurtosis, skewness, and/or a violation of normality according to the Shapiro-Wilk test. Therefore, the following correlation analysis was performed using parametric Pearson's product-moment correlation (two-tailed). To adjust for multiple comparisons, we report Bonferroni-corrected p-values.

## RESULTS

The results of the correlation analysis are shown in Tab. 1.

Table 1: Correlation Matrix

Variables	LZC	PLE	BCI offline performance
LZC	-		
PLE	-.99** [-.99, -.98]	-	
BCI offline performance	.38* [.13, .59]	-.34* [-.55, -.08]	-

Table 1: Pearson's product-moment correlation (two-tailed) between the selected criticality-related measures, PLE and LZC, and BCI offline performance. \* indicates  $p < .05$ , and \*\* indicates  $p < .001$  after Bonferroni correction for multiple comparison. Confidence interval for the according correlations provided in square brackets.

There was a very high and significant anticorrelation between PLE and LZC ( $r = -.99$ ,  $p < .001$ ). We also observed significant moderate correlations between both PLE and LZC and the BCI offline performance (see Fig. 1). PLE and BCI offline performance showed a negative correlation ( $r = -.34$ ,  $p < .05$ ), whereas LZC and BCI offline performance showed a positive correlation ( $r = .38$ ,  $p < .05$ ). This supports our hypothesis that brain criticality calculated from resting-state brain activity is related to the subsequent BCI performance – as indicated by BCI offline performance – when using a P300-based BCI.

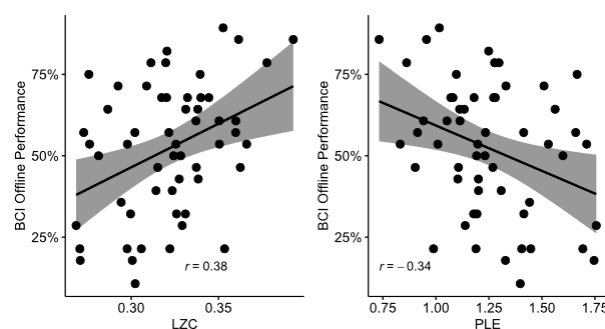


Figure 1: Correlations between the selected criticality-related measures, LZC and PLE, and BCI offline performance (Pearson's product-moment correlation (two-tailed)).

## DISCUSSION

We present here, to our knowledge, the first results indicating a relationship between brain criticality calculated from a preceding resting-state EEG recording and the subsequently achieved performance with a P300-based BCI in healthy participants. The presented correlations between two criticality-related measures, PLE and LZC, at rest and subsequent BCI offline performance suggest that increased resting-state brain criticality appears to be positively related to BCI performance when using a P300-based BCI. This finding is consistent with our hypothesis based on the proposed relationship between resting-state brain criticality and BCI performance with P300-based BCI, which we recently outlined [7]. Increases in EEG-derived measures of brain criticality at rest seem to improve the performance during the subsequent use of a P300-based BCI. This is sound with the idea that brain activity at the point of criticality expresses remarkable information processing capacities [17] and, as shown by Herzog et al. [30], the resting-state brain criticality may even interact with the P300 amplitude itself.

The only other available study that examined the relationship between resting-state brain activity and subsequent performance with a P300-based BCI [12] may support our findings. Shin et al. [12] showed a negative correlation between resting-state delta-frequency band power and subsequent BCI performance with a P300-based BCI. Given that significant increases in low-frequency oscillatory activity in the EEG power spectrum may reflect, or at least be confounded by, a steepened spectral slope [26], i.e., an increased PLE, the results of Shin et al. [12] would suggest that BCI performance with a P300-based BCI decreases with an increasing PLE value, i.e., decreasing criticality. Therefore, their result suggests a positive relationship between resting-state brain criticality and BCI performance with a P300-based BCI, as we have shown here.

The investigation of BCI performance based on brain criticality from preceding brain activity has also been performed with motor imagery (MI) BCI [44]. Samek and colleagues showed that brain criticality during a training session with a MI BCI was positively associated

with the subsequent performance in the test session. Although they did not use resting-state brain activity, this finding further suggests the importance of state-dependent brain criticality for the upcoming task, such as BCI use.

We speculate that one of the classical target groups for P300-based BCIs, i.e., complete locked-in (CLIS) ALS-patients, could benefit from the demonstrated relationship between resting-state brain criticality and subsequent BCI performance with a P300-based BCI. Limiting BCI communication attempts to periods of increased brain criticality might eventually improve their often inadequate performances with non-invasive BCIs (see e.g., Bettencourt et al. [45] for a recent overview). This would be a first attempt to detect the so-called *Windows of Consciousness* postulated by Kübler [46]. The potential increase in resting-state brain criticality may be even more relevant, as CLIS patients appear to show a general decrease in brain criticality, as indicated by changes in our selected criticality-related measures [14].

## CONCLUSION

We present here results indicating a relationship between resting-state brain criticality and subsequent performance with a visual P300-based BCI. Increasing brain criticality in resting-state brain activity, as indicated by criticality-related measures, appears to be beneficial for the subsequent use of a P300-based BCI in healthy participants.

## REFERENCES

- [1] Won K, Kwon M, Jang S, Ahn M, & Jun SC. P300 speller performance predictor based on RSVP multi-feature. *Frontiers in human neuroscience*. 2019; 13: 261
- [2] Kübler A, Holz EM, Riccio A, Zickler C, Kaufmann T, Kleih SC, et al. The user-centered design as novel perspective for evaluating the usability of BCI-controlled applications. *PloS one*. 2014; 9(12): e112392
- [3] Halder S, Hammer EM, Kleih SC, Bogdan M, Rosenstiel W, Birbaumer N, et al. Prediction of auditory and visual p300 brain-computer interface aptitude. *PloS one*. 2013; 8(2): e53513
- [4] Kleih SC, Nijboer F, Halder S, Kübler AJCN. Motivation modulates the P300 amplitude during brain-computer interface use. *Clinical Neurophysiology*. 2010; 121(7): 1023-1031
- [5] Mak JN, McFarland DJ, Vaughan TM, McCane LM, Tsui PZ, Zeitlin DJ, et al. EEG correlates of P300-based brain-computer interface (BCI) performance in people with amyotrophic lateral sclerosis. *Journal of neural engineering*. 2012; 9(2): 026014
- [6] Halder S, Ruf CA, Furdea A, Pasqualotto E, De Massari D, van der Heiden L, et al. Prediction of P300 BCI aptitude in severe motor impairment. *PloS one*. 2013; 8(10): e76148
- [7] Settgast T, Kübler A. Brain Criticality and Its Relation to BCI Operation, in 2024 12<sup>th</sup> International Winter Conference on Brain Computer Interface (BCI), IEEE, in press
- [8] Ahn M, Cho H, Ahn S, Jun SC. High theta and low alpha powers may be indicative of BCI-illiteracy in motor imagery. *PloS one*. 2013; 8(11): e80886
- [9] Lee M, Yoon JG, Lee SW. Predicting motor imagery performance from resting-state EEG using dynamic causal modeling. *Frontiers in human neuroscience*. 2020; 14: 321
- [10] Wang K, Tian F, Xu M, Zhang S, Xu L, Ming D. Resting-State EEG in Alpha Rhythm May Be Indicative of the Performance of Motor Imagery-Based Brain-Computer Interface. *Entropy*. 2022; 24(11): 1556
- [11] Zhang R, Xu P, Chen R, Li F, Guo L, Li P, et al. Predicting inter-session performance of SMR-based brain-computer interface using the spectral entropy of resting-state EEG. *Brain topography*. 2015; 28: 680-690
- [12] Shin GH, Lee M, Kim HJ, Lee SW. Prediction of event related potential speller performance using resting-state EEG, in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, 2973-2976
- [13] Settgast T, Zilio F, Kübler A, Northoff G. Correlation between Neurophysiological Measures of Consciousness and BCI Performance in a Locked-in Patient, in 2023 11th International Winter Conference on Brain-Computer Interface (BCI), IEEE, 2023, 1-6
- [14] Zilio F, Gomez-Pilar J, Chaudhary U, Fogel S, Fomina T, Synofzik M, et al. Altered brain dynamics index levels of arousal in complete locked-in syndrome. *Communications Biology*. 2023; 6(1): 757
- [15] Northoff G, Huang Z. How do the brain's time and space mediate consciousness and its different dimensions? Temporo-spatial theory of consciousness (TTC). *Neuroscience & Biobehavioral Reviews*. 2017; 80: 630-645
- [16] He BJ, Zempel JM, Snyder AZ, Raichle ME. The temporal structures and functional significance of scale-free brain activity. *Neuron*. 2010; 66(3): 353-369
- [17] O'Byrne J, Jerbi K. How critical is brain criticality?. *Trends in Neurosciences*. 2022; 45(11): 820-837
- [18] Maschke C, O'Byrne J, Colombo MA, Boly M, Gosseries O, Laureys S, et al. Critical dynamics in spontaneous EEG predicts anesthetic-induced loss of consciousness and perturbational complexity. *Communications Biology*. 2024; 7(1): 946
- [19] Höhn C, Hahn MA, Lendner JD, Hoedlmoser K. Spectral slope and Lempel-Ziv complexity as robust markers of brain states during sleep and wakefulness. *bioRxiv*. 2022
- [20] Schartner M, Seth A, Noirhomme Q, Boly M, Bruno MA, Laureys S, et al. Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. *PloS one*.

- 2015; 10(8): e0133532
- [21] Mateos DM, Guevara Erra R, Wennberg R, Perez Velazquez JL. Measures of entropy and complexity in altered states of consciousness. *Cognitive neurodynamics*. 2018; 12: 73-84
- [22] Carhart-Harris RL. The entropic brain-revisited. *Neuropharmacology*. 2018; 142: 167-178
- [23] Northoff G, Zilio F. Temporo-spatial Theory of Consciousness (TTC)–Bridging the gap of neuronal activity and phenomenal states. *Behavioural brain research*. 2022; 424: 113788
- [24] Alnes SL, Bächlin LZ, Schindler K, Tzovara A. Neural complexity and the spectral slope characterise auditory processing in wakefulness and sleep. *European Journal of Neuroscience*. 2023; 1-20
- [25] Medel V, Irani M, Crossley N, Ossandón T, Boncompte G. Complexity and 1/f slope jointly reflect brain states. *Scientific Reports*. 2023; 13(1): 21700
- [26] Gyurkovics M, Clements GM, Low KA, Fabiani M, Gratton G. Stimulus-induced changes in 1/f-like background activity in EEG. *Journal of Neuroscience*. 2022; 42(37): 7144-7151
- [27] Irmischer M, Poil SS, Mansvelder HD, Intra FS, Linkenkaer-Hansen K. Strong long-range temporal correlations of beta/gamma oscillations are associated with poor sustained visual attention performance. *European Journal of Neuroscience*. 2018; 48(8): 2674-2683
- [28] Bojorges-Valdez E, Yanez-Suarez O. Association between EEG spectral power dynamics and event related potential amplitude on a P300 speller. *Biomedical Physics & Engineering Express*. 2018; 4(2): 025028
- [29] Gervais C, Boucher LP, Villar GM, Lee U, Duclos C. A scoping review for building a criticality-based conceptual framework of altered states of consciousness. *Frontiers in Systems Neuroscience*. 2023; 17: 1085902
- [30] Herzog ND, Steinfath TP, Tarrasch R. Critical dynamics in spontaneous resting-state oscillations are associated with the attention-related P300 ERP in a Go/Nogo task. *Frontiers in neuroscience*. 2021; 15: 632922
- [31] Lee H, Golkowski D, Jordan D, Berger S, Ilg R, Lee J, et al. Relationship of critical dynamics, functional connectivity, and states of consciousness in large-scale human brain networks. *Neuroimage*. 2019; 188: 228-238
- [32] Heiney K, Huse Ramstad O, Fiskum V, Christiansen N, Sandvig A, Nichele S, et al. Criticality, connectivity, and neural disorder: a multifaceted approach to neural computation. *Frontiers in computational neuroscience*. 2021; 15: 611183
- [33] Li F, Liu T, Wang F, Li H, Gong D, Zhang R, et al. Relationships between the resting-state network and the P3: Evidence from a scalp EEG study. *Scientific Reports*. 2015; 5(1): 15129
- [34] Li F, Tao Q, Peng W, Zhang T, Si Y, Zhang Y, et al. Inter-subject P300 variability relates to the efficiency of brain networks reconfigured from resting-to task-state: evidence from a simultaneous event-related EEG-fMRI study. *NeuroImage*. 2020; 205: 116285
- [35] Won K, Kwon M, Ahn M, Jun SC. EEG dataset for RSVP and P300 speller brain-computer interfaces. *Scientific Data*. 2022; 9(1): 388
- [36] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on biomedical engineering*. 2004; 51(6): 1034-1043
- [37] Kübler A, Neumann N, Wilhelm B, Hinterberger T, Birbaumer N. Predictability of brain-computer communication. *Journal of Psychophysiology*. 2004; 18(2/3): 121-129
- [38] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*. 2004; 134(1): 9-21
- [39] Winkler I, Debener S, Müller KR, Tangermann, M. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP, in 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2015, 4101-4105
- [40] Winkler I, Haufe S, Tangermann M. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behavioral and brain functions*. 2011; 7: 1-15
- [41] Zhang XS, Roy RJ. Derived fuzzy knowledge model for estimating the depth of anesthesia. *IEEE Transactions on Biomedical Engineering*. 2001; 48(3): 312-323
- [42] Nagarajan R. Quantifying physiological data with Lempel-Ziv complexity-certain issues. *IEEE Transactions on Biomedical Engineering*. 2002; 49(11): 1371-1373
- [43] Aboy M, Hornero R, Abásolo D, Álvarez D. Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE transactions on biomedical engineering*. 2006; 53(11): 2282-2288
- [44] Samek W, Blythe DA, Curio G, Müller KR, Blankertz B, Nikulin VV. Multiscale temporal neural dynamics predict performance in a complex sensorimotor task. *Neuroimage*. 2016; 141: 291-303
- [45] Bettencourt R, Castelo-Branco M, Gonçalves E, Nunes UJ, Pires G. Comparing Several P300-Based Visuo-Auditory Brain-Computer Interfaces for a Completely Locked-in ALS Patient: A Longitudinal Case Study. *Applied Sciences*. 2024; 14(8), 3464
- [46] Kübler A. The history of BCI: From a vision for the future to real support for personhood in people with locked-in syndrome. *Neuroethics*. 2020; 13(2), 163-180

## BIDIRECTIONAL NEUROFEEDBACK: A CONTROL CONDITION COMPLEMENTARY TO SHAM?

Pierrieau, E.<sup>1</sup>, Pillette, L.<sup>2</sup>, Dussard, C.<sup>3</sup>, George, N.<sup>3\*</sup>, Jeunet-Kelway, C.<sup>1\*</sup>

<sup>1</sup>Aquitaine Institute for Cognitive and Integrative Neuroscience (INCIA), University of Bordeaux,  
Bordeaux, France

<sup>2</sup>Research Institute in Informatics and Random Systems (IRISA), Rennes, France

<sup>3</sup>Paris Brain Institute (ICM), Sorbonne University, Paris, France

\*co-last author

E-mail: emeline.pierrieau@u-bordeaux.fr

**ABSTRACT:** Neurofeedback (NF) is increasingly used for experimental and therapeutic purposes. However, the lack of proper control about the specificity of NF effects is criticized and hinders the development of reliable and efficient NF procedures. Bidirectional NF is based on the self-regulation of the targeted brain activity in opposite directions and might be better suited than other typical control conditions (e.g., sham) for assessing the link between modulations of brain activity and behavior. The present study aimed to determine if bidirectional regulation of a specific pattern of brain activity, namely motor beta power, can be achieved within a single session. Thirty participants performed several NF trials aiming to either down- or up-regulate their motor beta power. Results showed that participants significantly modulated their motor beta power in opposite directions with bidirectional NF. This modulation was constrained in space (central electrodes) and frequency (alpha/beta band). Overall, bidirectional NF appears as a valid method to probe brain-behavior relationships within a single session.

### INTRODUCTION

Neurofeedback (NF) consists of a brain-computer interface in which brain activity is measured and presented online to the participant through a sensory stimulus, for the purpose of enabling self-regulation of specific patterns of brain activity [1]. NF can be used as an experimental tool for studying brain-behavior relationships and as a potential therapeutic strategy for a variety of neurological diseases [2]. Yet, the lack of appropriate control condition in NF studies has been criticized and precludes its implementation as a valid experimental and clinical procedure [2–4]. One of the most commonly employed control condition in NF studies is a sham NF, which consists of presenting a NF whose features are independent from the participant's targeted brain activity (e.g., replay of prerecordings). Sham NF is well suited for assessing the specificity of NF on the effects found on the targeted brain activity and behavior because participants are supposedly

not aware of this sham condition, so that they should apply similar mental strategies when presented with the sham NF as with the real NF. However, using sham NF is not adequate for determining whether the changes in brain activity induced by NF are causally involved in the reported behavioral effects. Indeed, sham NF can lead to modulations of the targeted brain activity that do not differ significantly from the ones observed with real NF (e.g., [5]). Establishing the causality of brain-behavior relationships requires at least two NF conditions inducing significantly different changes in the targeted brain activity, such that it can be used as an independent variable. This can be achieved using bidirectional NF, that is implementing two NF conditions aiming to train participants to regulate the targeted brain activity in opposite directions (i.e., down- and up-regulate the targeted brain activity). Thus, opposite patterns of behavioral effects are expected in the two NF conditions. An additional sham-passive condition can be implemented to determine if each active NF condition led to significant change in brain activity in comparison to a "baseline" level without real NF nor mental strategy applied. Assessing whether behavioral effects remain significant when comparing active NF conditions to the sham-passive condition can then allow us to determine whether NF led to significant behavioral change compared to a "baseline" condition, without active modulation of brain activity. Overall, bidirectional NF comprehensively addresses numerous confounding factors associated with NF paradigms, including the ones that are purposely controlled with a sham NF such as placebo effects and global, spatially non-specific effects [6], while avoiding the ethical and time issues associated with sham NF. Indeed, a sham NF condition requires either a fairly high number of participants to ensure sufficient statistical power in the analyses when using a between-group experimental design (i.e., two separate groups of participants, real vs sham NF, that are compared to each other), or a long-lasting experiment if one chooses a within-group experimental design, because each participant will have to perform several sessions of each condition (real and sham NF). In each case, the ex-

perimental protocol will most likely be time-consuming. In addition to this, the use of a sham NF can bring some ethical issues in studies conducted on patients (e.g., only the group receiving the real NF will see their symptoms improving). Yet, there remains some gray areas regarding the feasibility of bidirectional NF, especially about the required time for learning to modulate a specific pattern of brain activity in opposite directions. Indeed, alternating up- and down-regulation can induce carry-over effects that might impair learning [7].

The present study aimed to determine whether efficient bidirectional regulation of a specific pattern of brain activity, namely the power of brain oscillations in the beta-band ( $\beta$ ; 13-30 Hz) over the motor cortex, can be achieved within a single experimental session using NF. Motor cortical  $\beta$  power was chosen as the target for the bidirectional NF protocol in the present study because it has been associated with changes in movement initiation and execution speed in non-invasive neurostimulation studies [8], as well as in a recent NF study using a sham control [5]. Motor cortical  $\beta$  power has also been widely used as a neural marker for decoding movement intention in the field of brain-computer interfaces. Therefore, the present study could provide replication of the results from the studies cited above, while assessing the feasibility of bidirectional NF within a single session. The frequential (frequency band) and spatial (electrodes) specificity of the effects on brain activity induced by the NF were also assessed.

## MATERIALS AND METHODS

**Participants:** Thirty participants (15 females, 15 males; age (mean  $\pm$  standard deviation (SD)):  $22 \pm 3$  years old) were recruited for the experiment. All participants were right-handed (mean Edinburgh score (mean  $\pm$  SD):  $94 \pm 5\%$ ), had normal or corrected-to-normal vision and were free of any known neurological or psychiatric condition. All subjects gave their written informed consent before participation in the study, which had been approved by the French committee for the protection of individuals (CPP) number 18-INSB-01. This study conformed to the standards set by the latest version of the Declaration of Helsinki. One participant was removed from the analyses because of excessive noise in the EEG data.

**Experimental design:** The experiment consisted of a bidirectional NF training coupled with a force task. Each participant performed a total of 125 trials, including 5 familiarization trials that were not considered in the analyses. Each trial started with the appearance of a fixation cross on a screen for 3 s. The cross was then replaced by a gauge, representing the NF based on the recorded online changes in motor  $\beta$  activity (see NF section below) for 2 to 10 s. This NF phase was followed by a hand grip task, subjective effort rating on an analog scale and written feedback about the performance at the hand grip task (Fig. 1C). Force data from the hand grip task and effort

ratings from the analog scale were not analyzed in the present article, as it focuses on modulations of  $\beta$  activity across NF conditions.

**Data acquisition and calibration:** At the beginning of the experiment, participants were comfortably seated in front of a screen (60 x 34 cm) with their right hand holding a dynamometer (K-Force Grip, Kinvent). They were asked to not exert any pressure on the dynamometer with their hand unless receiving an explicit instruction to do so. A 32-channel EEG cap (EEGo sports, ANTneuro) was placed on the participants' head and the EEG signal was recorded continuously for the duration of the experiment with a 500 Hz acquisition rate. Stimuli presented on the screen were synchronized with EEG and force recordings through an open-source software toolkit (lab streaming layer; LSL). The ground and reference electrodes were AFz and CPz respectively. The first EEG recording of the experiment was performed in a resting state to calibrate NF parameters in an individualized fashion. A white fixation cross was presented at the center of the screen and participants were asked to maintain their gaze on the cross and remain still as long as the cross was on. The cross was displayed for 1 min but only the last 30 s were recorded to ensure that participants were already in a resting state when starting EEG recordings. The distribution of beta power values during those recordings was used to set individualized NF thresholds (see NF section). The participants' maximal force for hand grip was also determined individually at the beginning and the middle of the experiment.

**NF:** The NF was implemented using OpenViBE (3.3) and Unity C# (2020.3.17f1). It was represented on the screen as a gauge (vertical rectangle cut by a horizontal midline) whose level varied as a function of motor  $\beta$  power. More precisely, the level of the gauge was refreshed every 250 ms based on the mean  $\beta$  power (squared amplitude of the EEG signal between 15 and 25 Hz, computed using a 4th order Butterworth filter) recorded at C3 electrode during the last 500 ms (i.e., sliding window of 500 ms per 250 ms step). This averaging procedure has been used by previous studies to avoid difficulties in reading the gauge because of excessive flickering. Spatial noise was attenuated using a Laplacian filter including 6 neighboring electrodes of C3 (FC5, FC1, F3, CP5, CP1, P3). Participants were trained to down-regulate and upregulate their motor  $\beta$  power in two separate experimental conditions:  $\beta$ -down and  $\beta$ -up respectively. Participants were encouraged to fill up the gauge in both conditions. In  $\beta$ -down, the lower the recorded  $\beta$  power, the higher the level of the gauge, whereas in  $\beta$ -up, the higher the recorded  $\beta$  power, the higher the level of the gauge. The gauge was calibrated according to  $\beta$  power detected at C3 electrode per 250 ms windows of the 30 s preliminary resting-state recording (120 values in total). The inferior boundary of the gauge corresponded to the median of resting-state  $\beta$  power in the two conditions. The gauge started to fill up if the recorded  $\beta$  power was below that median in  $\beta$ -down and above that

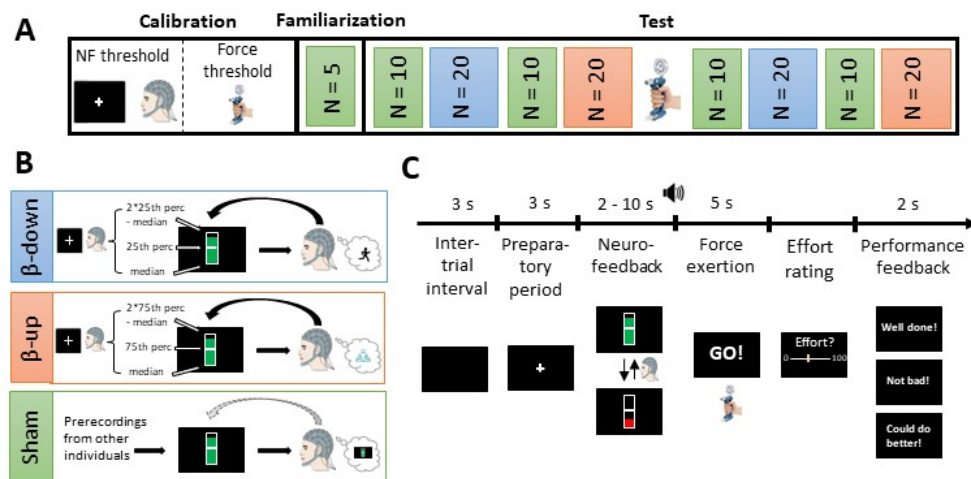


Figure 1: Methods. A. Overview of the experimental design. Blue, orange and green boxes indicate  $\beta$ -down,  $\beta$ -up, and sham-passive blocks respectively. N refers to the number of trials included in each block. The dynamometer's picture indicates the moments when the maximal force was (re)calculated. B. Neurofeedback conditions. The individualized values that were used for calibrating the level of the gauge are indicated on the left of the gauge. The images inside of the thought bubbles illustrate mental strategies that participants were advised to use in each condition. C. Trial timeline. Depiction of the name, duration and visual stimuli associated with each phase of a trial, in a chronological order.

median in  $\beta$ -up. As explained in Fig. 1, the gauge was half-filled/completely filled when  $\beta$  power reached the 25<sup>th</sup> percentile/2\*25<sup>th</sup> percentile-median of the resting-state distribution in  $\beta$ -down, or the 75<sup>th</sup> percentile/2\*75<sup>th</sup> percentile-median in  $\beta$ -up. The filling color of the gauge was red if its level was below midline and turned green if its level reached above midline. If the gauge remained green for 2 consecutive seconds (i.e.,  $\beta$  power was inferior to the 25<sup>th</sup> percentile in  $\beta$ -down/superior to the 75<sup>th</sup> percentile in  $\beta$ -up for 2 s), the NF stopped and the gauge was replaced by a go cue. In cases where this criterion was not met, the NF was automatically replaced by a go cue after 10 s.

To improve self-regulation of motor  $\beta$  power, participants were given mental strategies before starting the trials. In  $\beta$ -down, participants were advised to perform motor imagery, that is mentally representing themselves performing movements without executing actual movements. Indeed, motor imagery has been shown to reduce  $\beta$  power to an extent that can equal or even exceed the  $\beta$  event-related desynchronization observed for actual movements when combined with NF [9]. Conversely, relaxation strategies (e.g., relaxing body parts, conscious breathing, task-unrelated thinking) have been shown to increase  $\beta$  power [10], and were thus suggested to the participants in  $\beta$ -up. Those mental strategies were not mandatory, participants were told that they were free to use any mental strategy they found efficient for filling up the gauge. A reminder of the recommended mental strategy (motor imagery or relaxation) was presented on the screen for 5 s before the beginning of each block.

A sham passive condition was also implemented as a control condition with similar sensory inputs as in  $\beta$ -down and  $\beta$ -up (i.e., gauge) but without any mental strategy nor congruent NF. In this condition, participants were asked

to keep their gaze on the gauge but not to try controlling it. The level of the gauge varied according to fluctuations in  $\beta$  power in prerecordings from other subjects, such that the gauge was not informative of the actual online changes in  $\beta$  power of the tested participants. Blocks of 10 sham passive trials were interspersed with blocks of 20  $\beta$ -down or  $\beta$ -up trials (Fig. 1A). The presentation order of  $\beta$ -down and  $\beta$ -up blocks was counterbalanced across participants.

**Data processing:** Data processing was conducted on Matlab (R2018B, MathWorks).  $\beta$  power values computed in OpenViBE (squared amplitude of EEG signal comprised between 15 and 25 Hz at C3 electrode after applying a Laplacian filter, see NF section above for details) for updating the NF were averaged during the last 2 s of the NF phase for the analyses. Resting-state median  $\beta$  power was subtracted from average  $\beta$  power at the end of the NF phase individually to account for interindividual variability in baseline  $\beta$  power. Outliers were calculated separately for each participant and condition and were defined as values inferior to median-3\*absolute deviations around the median (MADs; [11]) or superior to median+3\*MADs, and were removed from the analyses.

**Statistical analyses:** A within-subject experimental design was conducted. A repeated-measures ANOVA was computed for comparing  $\beta$  power (value used for the NF in OpenViBE) across NF conditions (3 levels:  $\beta$ -down,  $\beta$ -up and sham-passive). An additional ANOVA including NF conditions and block number (2 levels: 1, 2) was performed to test if the effects differed across blocks. Greenhouse-Geisser's correction was applied to p-values when sphericity assumption was violated (Mauchly's test p-value < 0.05). Pairwise comparisons were conducted with paired Student's t-tests when observations were normally distributed (p-value from Shapiro-

Wilk test  $\geq 0.05$ ), and Wilcoxon's rank tests when normality assumption was violated (p-value from Shapiro-Wilk test  $< 0.05$ ). Bonferroni's correction was applied on p-values for multiple comparisons. Effect sizes were reported as partial eta squared for ANOVAs ( $\eta_p^2$ ), Cohen's d (d) for Student's t-tests and rank biserial correlation (r) for Wilcoxon's rank tests. Cluster-based permutation tests were performed using Fieldtrip to identify cluster of electrodes that were behaving differently between  $\beta$ -down and  $\beta$ -up [12]. Clusters were defined as adjacent electrode/time pairs whose test statistic exceeded the threshold for statistical significance ( $\alpha = 0.05$ , two-tailed paired t-tests). In the present analysis, a cluster was composed of at least two electrodes showing statistically significant t values within a radius of 4 cm. These tests were conducted from -2 to 0 s before the presentation of the go cue, in the theta (4-7 Hz), alpha (8-12 Hz), beta (15-25 Hz), and low-gamma (30-49 Hz) frequency bands separately.

## RESULTS

$\beta$  power during NF was significantly decreased as compared to resting-state  $\beta$  power in  $\beta$ -down ( $W(28) = 82$ ,  $p = 0.003$ ,  $r = -0.62$ , mean difference = -19.5%). Conversely, a trend toward an increase in  $\beta$  power during NF in comparison to resting-state  $\beta$  power was observed in  $\beta$ -up ( $W(28) = 301$ ,  $p = 0.072$ ,  $r = 0.38$ , mean difference = +9.3%).  $\beta$  power in the sham passive NF condition was not significantly different from resting-state  $\beta$  power ( $W(28) = 156$ ,  $p = 0.190$ ,  $r = -0.28$ , mean difference = -3.5%). Direct comparison of average  $\beta$  power across NF conditions showed a significant difference ( $F(1.4, 39.5) = 16.2$ ,  $p = 10^{-4}$ ,  $\eta_p^2 = 0.37$ ). Post-hoc analysis revealed that  $\beta$  power was significantly decreased in  $\beta$ -down in comparison to  $\beta$ -up ( $W(28) = 33$ ,  $p = 10^{-5}$ ,  $r = -0.85$ ) and sham-passive ( $W(28) = 51$ ,  $p = 10^{-4}$ ,  $r = -0.77$ ), whereas it was significantly increased in  $\beta$ -up as compared to sham-passive ( $W(28) = 372$ ,  $p = 0.001$ ,  $r = 0.71$ ; Fig. 2, top left panel). Single-subject analysis showed that 79% (23/29) of participants significantly decreased their  $\beta$  power in  $\beta$ -down as compared to  $\beta$ -up, 14% (4/29) did not significantly modulate their  $\beta$  power and 7% (2/29) inversely modulated their  $\beta$  power (i.e., increased their  $\beta$  power in  $\beta$ -down in comparison to  $\beta$ -up). When comparing  $\beta$ -down to sham-passive, results showed that 72% (21/29) of participants significantly decreased their  $\beta$  power in  $\beta$ -down as compared to sham-passive, 14% (4/29) did not significantly modulate their  $\beta$  power and 14% (4/29) inversely modulated their  $\beta$  power (i.e., increased their  $\beta$  power in  $\beta$ -down in comparison to sham-passive). Finally, 69% (20/29) of participants significantly increased their  $\beta$  power in  $\beta$ -up as compared to sham-passive, 14% (4/29) did not significantly modulate their  $\beta$  power and 17% (5/29) inversely modulated their  $\beta$  power (i.e., decreased their  $\beta$  power in  $\beta$ -up in comparison to sham-passive).

A two-way ANOVA for repeated measures showed a sig-

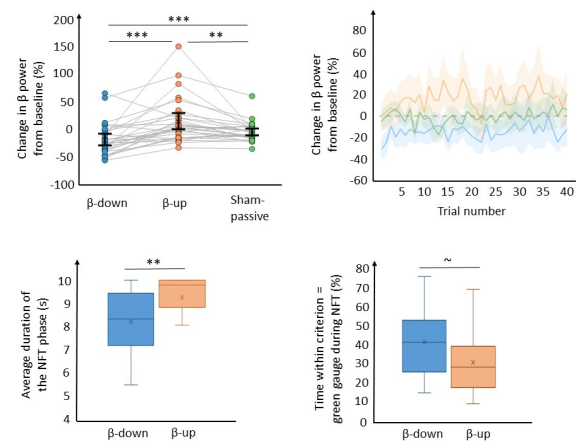


Figure 2: Effects of bidirectional NF on motor cortical  $\beta$  power. Top, left:  $\beta$  power change from baseline according to NF condition. Colored dots indicate individual data. Light gray lines connect dots pertaining to the same participant. Black error bars illustrate 95% confidence intervals around the mean. Top, right:  $\beta$  power change from baseline according to NF condition. Data is averaged across participants. Trials are represented in their chronological order of presentation during the experiment. Blue, orange, and green lines illustrate the  $\beta$ -down,  $\beta$ -up and sham-passive conditions. Shaded areas represent 95% confidence intervals around the mean. The zero-line (indicating no change from baseline  $\beta$  power) is highlighted with a dotted gray line. Bottom, left to right: average duration of the NF phase and time within criterion (i.e., when NF threshold is reached/the gauge is green) in  $\beta$ -down (blue) and  $\beta$ -up (orange). Median and mean are represented as horizontal line and cross respectively. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , ~  $p = 0.05-0.1$

nificant effect of NF condition ( $\beta$ -down,  $\beta$ -up and sham-passive;  $F(1.4, 39.7) = 16.0$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.36$ ) but no significant effect of block (1 and 2;  $F(1, 28) = 0.49$ ,  $p = 0.489$ ,  $\eta_p^2 = 0.02$ ) nor interaction between NF condition and block ( $F(1.5, 43.3) = 0.15$ ,  $p = 0.804$ ,  $\eta_p^2 = 0.01$ ) on  $\beta$  power. This suggests that NF performance did not improve throughout the experiment as  $\beta$  power did not significantly change with time across NF conditions (Fig. 2, top right panel). NF performance was also measured by the percentage of time spent within criterion (i.e.,  $< 25^{th}$  percentile of resting-state  $\beta$  power for  $\beta$ -down and  $> 75^{th}$  percentile of resting-state  $\beta$  power for  $\beta$ -up) during the NF phase, and the duration of the NF phase. There was a trend toward longer time spent within criterion in  $\beta$ -down as compared to  $\beta$ -up ( $t(28) = 1.83$ ,  $p = 0.078$ ,  $d = 0.34$ ; Fig. 2, bottom right panel). Additionally, the duration of the NF phase was significantly shorter in  $\beta$ -down than  $\beta$ -up ( $t(28) = -2.79$ ,  $p = 0.009$ ,  $d = -0.52$ ; Fig. 2, bottom left panel). This suggests that NF performance was significantly better in  $\beta$ -down than  $\beta$ -up as participants reached the criterion for 2 s more often, leading to shorter trials.

The last step of the analysis consisted of assessing the spatial and frequential specificity of the changes in brain activity induced by NF. To do so, the power of the EEG signal was computed in different frequency bands and across all electrodes by means of time-frequency anal-

ysis (see Methods for details). Signal power was first averaged in different frequency bands at C3 electrode to determine the frequential specificity of NF. Theta (4-7 Hz;  $F(2,54) = 1.21$ ,  $p = 0.306$ ,  $\eta_p^2 = 0.04$ ), alpha (8-12 Hz;  $F(1.57,42.34) = 2.32$ ,  $p = 0.122$ ,  $\eta_p^2 = 0.08$ ), and low-gamma (31-49 Hz;  $F(1.08, 29.11) = 1.05$ ,  $p = 0.319$ ,  $\eta_p^2 = 0.04$ ) power did not appear significantly affected by NF, suggesting that, over the contralateral motor cortex,  $\beta$  power was the only frequency band that was significantly modulated by NF. Spatial specificity was then quantified using cluster-based permutation tests. These tests highlighted a significant cluster of contralateral and midline fronto-central electrodes when comparing  $\beta$  power in  $\beta$ -down and  $\beta$ -up (Fig. 3, top panel). This cluster was centered on C3 and showed significant reduction of  $\beta$  power in  $\beta$ -down as compared to  $\beta$ -up. A significant negative cluster was also found when comparing alpha power in  $\beta$ -down and  $\beta$ -up, though in contrast with  $\beta$  power, alpha power appeared mostly attenuated over ipsilateral fronto-central electrodes (Fig. 3, top panel). No significant cluster was detected when comparing theta nor gamma power in  $\beta$ -down and  $\beta$ -up. Time-frequency maps of the activity recorded at C3 showed that baseline-corrected  $\beta$  power appears decreased in  $\beta$ -down and  $\beta$ -up, but the magnitude of this decrease appeared greater in  $\beta$ -down than in  $\beta$ -up.

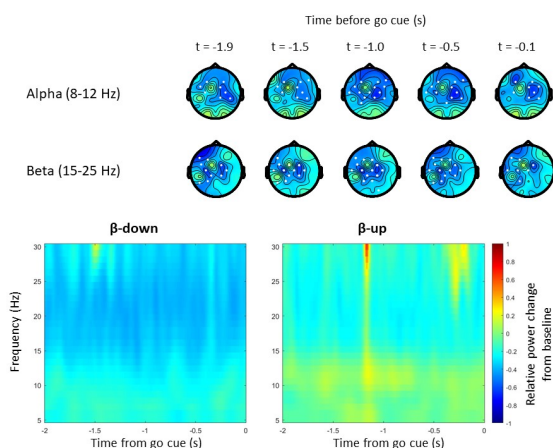


Figure 3: Spatial and frequential specificity of NF. Top: topographical maps representing changes in alpha (top) and beta (bottom) power between  $\beta$ -down and  $\beta$ -up, at different times before go cue onset (t). White dots highlight significant negative cluster of electrodes. Bottom: time-frequency maps of the signal recorded at C3 in  $\beta$ -down (left) and  $\beta$ -up (right).

## DISCUSSION

Results show that the bidirectional NF paradigm used in the present single-session experimental design effectively led to opposite modulations of motor cortical  $\beta$  power. Most participants (79%) significantly decreased their  $\beta$  power in  $\beta$ -down as compared to  $\beta$ -up. Additionally, participants significantly decreased and increased their  $\beta$  power in  $\beta$ -down and  $\beta$ -up respectively, as compared to

in sham-passive. The changes in brain oscillations induced by NF were not specific to the  $\beta$  frequency as a significant decrease in alpha power was also observed in  $\beta$ -down in comparison to  $\beta$ -up. However, this modulation of alpha power was mostly localized over the ipsilateral motor cortex, whereas changes in  $\beta$  power were more on the contralateral side. Taken as a whole, these results suggest that single-session bidirectional NF, including blocks belonging to each direction of regulation presented in an alternated order, constitutes an appropriate method to probe the nature of the relationship between motor cortical  $\beta$  power and behavior.

To our knowledge, the present results represent the first evidence of the possibility to use bidirectional NF for volitionally modulating motor cortical  $\beta$  power within a single session, and studying its effects on motor behavior over several trials using a within-subject experimental design. Two studies [13, 14] already demonstrated that bidirectional NF based on individualized task-specific features of sensorimotor rhythms can significantly affect motor performance. However, targeted features for NF varied across participants (e.g., an alpha and a  $\beta$  power components were included in 4 out of 8 participants, whereas only a  $\beta$  power component was included in the rest of the participants), thereby complicating the identification of the specific EEG patterns that were responsible for the behavioral changes observed. The experimental design presented herein allows the assessment of the specific influence of self-regulating motor cortical  $\beta$  power on motor behavior, and takes into account inter-individual variability in baseline EEG activity by calibrating  $\beta$  power values used for the NF individually. Yet, these  $\beta$  power values are calculated identically for all participants (median, 25<sup>th</sup> and 75<sup>th</sup> percentiles of the distribution of resting-state  $\beta$  power). This standardized procedure supposedly enables reliable conclusions about the relationship between a specific EEG pattern and behavior, though it may be inadequate for some participants (e.g., highly variable baseline activity, weak amplitude of movement-related  $\beta$  power changes). This does not appear to be the case in the present experiment as 79% of participants successfully decreased their motor cortical  $\beta$  power in  $\beta$ -down in comparison to  $\beta$ -up. The proportion of non-responders (21%) is in accordance with the numbers reported by previous studies (15 to 30%; [15]).

In addition to successful opposite regulation of motor cortical  $\beta$  power within a single session, the present results also showed that participants significantly decreased and increased their  $\beta$  power as compared to sham-passive in  $\beta$ -down and in  $\beta$ -up respectively. This demonstrates that the present experimental design is particularly well suited to assess brain-behavior relationships, considering that the influence of modulating motor cortical  $\beta$  power on motor behavior can be determined in each direction of regulation (down and up) and compared to a "baseline" level of  $\beta$  power and behavior. This design would enable to extend the results from previous studies that have shown significant effect of regulation of  $\beta$  power with NF

on motor behavior without control comparison or compared to a sham-passive NF only ([5, 16, 17]). Some limitations associated with the proposed experimental design should be underlined. First, self-regulation of  $\beta$  power with NF did not significantly improve across trials, suggesting that the single-session design, consisting of blocks of  $\beta$ -down and  $\beta$ -up trials presented in an alternated order, might have interfered with the learning process of self-regulation. An alternative option could be to perform each NF condition (i.e.,  $\beta$ -down and  $\beta$ -up) in separate sessions. However, in addition to significantly increase the duration of the experiment, splitting the experiment into distinct sessions brings other issues, such as potential differences in the positioning of the EEG cap, as well as differences in baseline  $\beta$  power values used for setting the NF threshold. Second, NF performance appeared slightly better in  $\beta$ -down than  $\beta$ -up, which could influence the following motor performance (i.e., filling up the gauge faster and more easily could boost motivation and thus lead to better motor performance). Possibilities for remediating to such an effect include modifying mental strategies for increasing  $\beta$  power in  $\beta$ -up, considering that the one used in the present experiment was fairly close to participants' resting state (i.e., relaxation strategy), or setting NF thresholds based on pre-movement  $\beta$  power instead of resting-state, as it should be closer to the baseline activity observed during the NF phase before the motor task.

## CONCLUSION

The present study provides evidence that bidirectional NF can be used to determine the effects of modulating brain oscillatory activity with NF, such as up- and down-regulation of motor  $\beta$  power, on behavior. Opposite modulations of brain oscillatory activity was achieved within a single session for most individuals. Therefore, bidirectional NF represents a relevant method for controlled, simple and rapid experimental designs aiming to study the association between a specific pattern of brain activity and behavior.

## REFERENCES

- [1] Sitaram R *et al.* Closed-loop brain training: The science of neurofeedback. *Nature Reviews Neuroscience*. 2017;18(2):86–100.
- [2] Jeunet C, Glize B, McGonigal A, Batail JM, Micoulaud-Franchi JA. Using eeg-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: Theoretical background, applications and prospects. *Neurophysiologie Clinique*. 2019;49(2):125–136.
- [3] Thibault RT, Raz A. When can neurofeedback join the clinical armamentarium? *The Lancet Psychiatry*. 2016;3(6):497–498.
- [4] Thibault RT, Lifshitz M, Raz A. Neurofeedback or neuroplacebo? *Brain*. 2017;140(4):862–864.
- [5] He S, Everest-Phillips C, Clouter A, Brown P, Tan H. Neurofeedback-linked suppression of cortical  $\beta$  bursts speeds up movement initiation in healthy motor control: A double-blind sham-controlled study. *Journal of Neuroscience*. 2020;40(20):4021–4032.
- [6] Sorger B, Scharnowski F, Linden DE, Hampson M, Young KD. Control freaks: Towards optimal selection of control conditions for fmri neurofeedback studies. *Neuroimage*. 2019;186:256–265.
- [7] Cortese A, Amano K, Koizumi A, Lau H, Kawato M. Decoded fmri neurofeedback can induce bidirectional confidence changes within single participants. *NeuroImage*. 2017;149:323–337.
- [8] Pogosyan A, Gaynor LD, Eusebio A, Brown P. Boosting cortical activity at beta-band frequencies slows movement in humans. *Current biology*. 2009;19(19):1637–1641.
- [9] Miller KJ, Schalk G, Fetz EE, Den Nijs M, Ojemann JG, Rao RP. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*. 2010;107(9):4430–4435.
- [10] Autenrieth M, Kober SE, Neuper C, Wood G. How much do strategy reports tell about the outcomes of neurofeedback training? a study on the voluntary up-regulation of the sensorimotor rhythm. *Frontiers in human neuroscience*. 2020;14:218.
- [11] Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*. 2013;49(4):764–766.
- [12] Maris E, Oostenveld R. Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*. 2007;164(1):177–190.
- [13] Boulay C, Sarnacki W, Wolpaw J, McFarland D. Trained modulation of sensorimotor rhythms can affect reaction time. *Clinical neurophysiology*. 2011;122(9):1820–1826.
- [14] McFarland DJ, Sarnacki WA, Wolpaw JR. Effects of training pre-movement sensorimotor rhythms on behavioral performance. *Journal of neural engineering*. 2015;12(6):066021.
- [15] Vidaurre C, Blankertz B. Towards a cure for bci illiteracy. *Brain topography*. 2010;23:194–198.
- [16] He S *et al.* Subthalamic beta-targeted neurofeedback speeds up movement initiation but increases tremor in parkinsonian patients. *Elife*. 2020;9:e60979.
- [17] Bichsel O, Stieglitz LH, Oertel MF, Baumann CR, Gassert R, Imbach LL. Deep brain electrical neurofeedback allows parkinson patients to control pathological oscillations and quicken movements. *Scientific Reports*. 2021;11(1):7973.

# DYNAMIC BRAIN NETWORKS IN MOTOR IMAGERY-BASED BCI

T. Venot<sup>1</sup>, A. Desbois<sup>1</sup>, F. De Vico Fallani<sup>1</sup>

<sup>1</sup>Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

E-mail: tristan.venot@icm-institute.org

**ABSTRACT:** Using the interactions between brain regions has great potential as new features to discriminate between mental tasks for brain computer interface (BCI). Network approaches applied to electroencephalographic (EEG)-derived functional connectivity has been recently used to identify discriminating brain organizational features in offline classification scenarios. However how those network properties temporally vary during the task, is still poorly understood. *A contrario*, the dynamics of event related desynchronization/synchronization resulting from local power spectra is widely known and used for online motor imagery-based BCIs. Here, we explored the offline time-frequency properties of dynamic brain networks in two subjects performing three sessions of MI-BCI for the control of a robotic arm. Results were compared to standard time-frequency power spectra and discussed in light of future implementation for online scenarios.

## INTRODUCTION

When dealing with motor imagery (MI) BCI, one of the central challenges is finding features both discriminant and interpretable from a neurophysiological perspective [1, 2]. More over, they often depend on the correct execution of the mental task. Performing motor imagery is tricky, mental strategies may vary from one subject to the next and features need to be robust to such variabilities [3]. When performing motor imagery, the main information that can be transferred into a command is the event related desynchronization (ERD) responsible for power spectral density variations in the frequency bands associated to motor task,  $\alpha$  (8-12 Hz) and  $\beta$  (13-30 Hz) [4]. However, this information is local, and the brain is a dynamic system whose regions constantly interact together. To capture those interactions, network approaches applied to functional connectivity [5] shows to be relevant as it studies the spectral correlation between electrodes (in the sensor space); the use of connectivity allows to apply metrics coming from network science such as node strength (NS) [6], which captures the amount of connections a node (in our case an electrode) has with the others. The interactive information coming from connectivity has been explored to differentiate MI from rest [7] however its use using network metrics as feature for classification is still not employed despite having the potential

to complete the local information of the PSD. Here, we want to investigate how the node strength evolves over time during a motor imagery task with respect to resting state and how the parameters used to estimate the functional connectivity will have an impact on the performances of a classifier, Figure 1 gives a representation of the offline approach. To this end, we used EEG data recorded from two subjects during a MI-BCI control of a robotic arm in three different sessions [8]. Our preliminary results points towards different conclusions. First, as it is the case for ERD/ERS, there is a need to average over time points to obtain subsequent performances to temper effects of variability coming from the spectral/coherence estimation. Second, adding the information of node strength as a complementary feature for the classification tend to improve the performances. Third, the temporal dynamics of node strength shows to fluctuate more than ERD/ERS on short windows which makes them more difficult to interpret.

## MATERIALS AND METHODS

### *Experimentation:*

The two subjects (2 F), aged  $24.5 \pm 1.5$  years, right-handed, provided informed consent and participated voluntarily in the protocol. The protocol was approved by Inria's national ethical committee as part of the BCIPRO protocol (authorization number 2021-35 - ref SICOERLE n°179). Experiments took place in the controlled environment of the EEG/MEG center within the neuroimaging core facility of the Paris Brain Institute.

A robotic arm facing the subjects reaches for objects on an augmented table used to show visual stimuli and neurofeedback (this table consists of a screen lying under a plexiglas, that displays visual cues directly underneath the objects to grasp). Subjects gaze towards a target to make the robot reach it and perform motor imagery for the robot to grasp the target. Each subject performs the control over the robot during three sessions where the robot moves before, during or after the subjects perform MI or rest. More details on the protocol can be found in the Braccio protocol [8].

In this study, we focus on motor imagery of the right hand closing and resting state trials lasting for three seconds. The acquisition uses BrainAmp 64 EEG cap, 500 Hz sampling frequency with TP9 and TP10 as reference

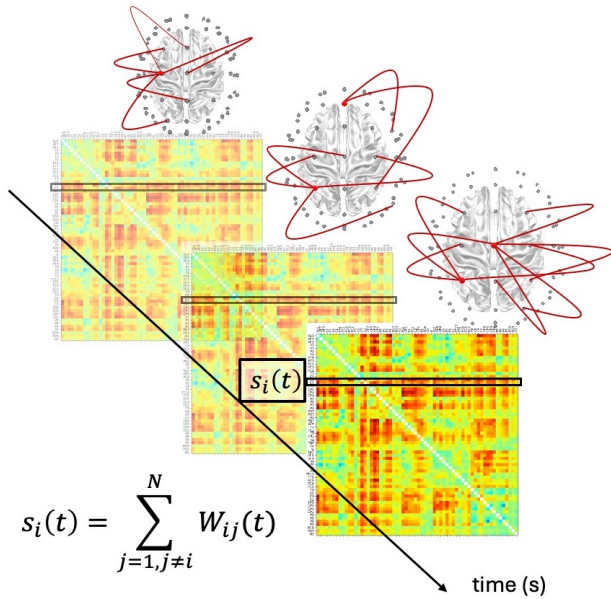


Figure 1: Representative view of the brain network dynamics through functional connectivity captured by node strength over time  $s(t)$ .  $w_{ij}(t)$  corresponds to the imaginary coherence (IC) between two electrodes calculated for each instant  $t$  by  $IC_{ij}[f](t) = \frac{|S(P_{ij}[f])|}{(P_i[f] \cdot P_j[f])^{1/2}}$ . with  $P$  the power and  $f$  the frequency,  $i$  and  $j$  are the couples of electrodes.

and ground. During the experimentation, we train an Linear discriminant Analysis (LDA) classification algorithm on PSD features selected from the  $R^2$  between MI and Rest trials for each subjects. These features are spectral amplitudes averages over trial for specific electrodes and frequency bins. To determine which features to select, we evaluated the highest  $R^2$  values in the sensorimotor cortex (electrodes of lines C and CP) in the  $\alpha$  and  $\beta$ . We selected 3 features for those two subjects :C3,CP3,C1 for subject 1 and C3,CP3,CP1 for subject 2. Each session is composed of two phases: first, 3 runs of control over the robot (Phase 1), then based on a training over the features of the 3 runs, a second phase of 2 runs (Phase 2), each run consists of 10 MI/10 Rest trials. Motor imagery and resting state trials lasted for 4 seconds, however only the last 3 seconds were kept to take into account the reaction time of subjects.

**Network metrics estimation:** To estimate the functional connectivity, we use imaginary coherence as it is more robust to volume conduction compared to spectral coherence[9]. Spectral properties were computed using Burg autoregressive (AR) method with a model order of 19, a frequency resolution of 0.5 Hz and make the parameters of windowing set to 0.33 s and overlap 53% to have a number of time points arbitrarily set to 18 points for 3 seconds of trial. The motivation of the choice of the filter order is based on two preliminary studies i) where we could identify a certain stability of the subject's patterns when we made the AR filter vary from 19 to 30 ii) using a particle swarm algorithm to optimize difference between

MI and rest for each subject, we identify 19 as the average filter. This operations were done using HappyFeat software (Inria) [10]. Based on the connectivity matrix, we use a local network metric called node strength (NS)[6], i.e. the average of all connection over each electrode in each condition. We then compute the average of the temporal node strength over the trials in Fig 2.

**Power spectrum estimation:** To estimate power spectral density, we also use Burg autoregressive method, set to a model of 19, a frequency resolution of 0.5 Hz and the window 0.25 s and 38% of overlap. For power spectrum only, a common average reference (CAR) was applied.

**Classification:** The classification algorithm used is a 2 class LDA, Phase 1 is used for the training and Phase 2 as a validation test. Brain features are electrodes at certain frequency bins used for the different sessions, they are selected after computing the  $R^2$  statistical test between trials of MI and resting state and with neurophysiological relevance - in the motor cortex in the  $\alpha$  or  $\beta$  band). We compare performance obtained with two different training approaches. A first method consists in using each estimated spectral window for each trial (both in PSD and NS) as a feature. It means that for 30 trials lasting for 3 seconds with 18 points, we trains the algorithm on  $30 \times 18 \times N \times M$  features per class (N being the number of electrodes and M of frequency bins). The second approach consists in averaging over time windows the features, meaning that the algorithm will be trained on  $30 \times N \times M$  features per class. In a first step, we use the same features for PSD and NS, then we select specific features for NS corresponding to its specific  $R^2$  map, and finally we combine the information coming from the two sources of information (NS and PSD).

## RESULTS

**Average Temporal dynamics over trials:** In a first step, we want to compare the trial-averaged evolution of NS compared to the PSD. As expected the separation in  $\alpha$  and  $\beta$  bands for PSD is clear and the amplitude corresponding to each condition is stable in time. Results are however more peculiar for NS, indeed, even though the evolution is averaged across trials, we still notice some strong oscillatory patterns that makes the separation between tasks more complex. From this, two different hypotheses can be made: first, the AR method used for the coherence estimation is more sensitive to noise in the context of short windows which forbids from using it to study the resulting node strength dynamics. The other hypothesis is that NS possesses properties different from the PSD on its temporal dynamic, ERD/ERS producing a stable pattern during MI/Rest task whereas node strength is intrinsically more oscillating.

We observe that Imaginary Coherence Node strength (NS-ImCoh) seems to follow same patterns as ERD/ERS as shown in Fig 2. Indeed we observe a decrease of Node strength from resting state to motor imagery mainly centered on 10-12Hz which is the expected behaviour. The

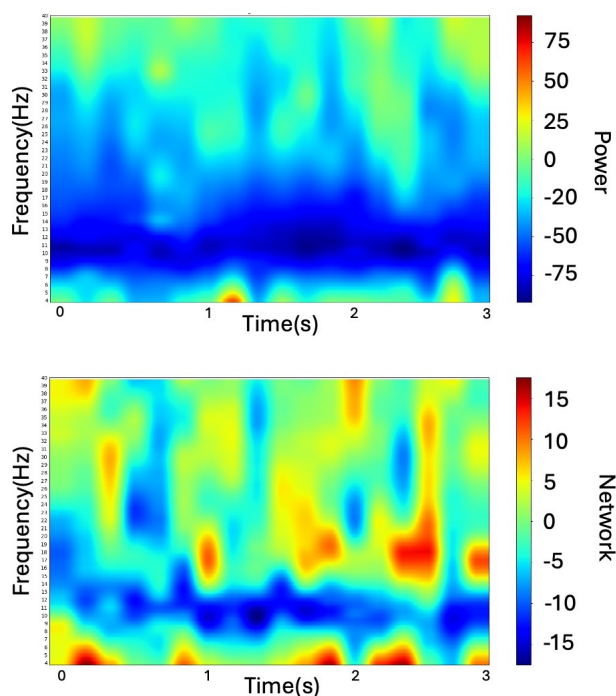


Figure 2: *Time-frequency maps of power and network-related dynamics*: Top: Relative difference of average spectral estimation across trials between motor imagery and resting state ( $\frac{MI-Rest}{Rest}$ ) for subject 2, session 3, electrode C3 selected. Bottom: Relative difference of average imaginary coherence estimation across trials between motor imagery and resting state ( $\frac{MI-Rest}{Rest}$ ) for subject 2, session 3, electrode C3 selected.

most interesting detail is that the most intense moment of desynchronization (in PSD) corresponds to the peak of the NS difference. Also, the dynamic seems to separate more the tasks in the  $\alpha$  band than in the  $\beta$  band. Finally, it is necessary to mention that the difference is far superior with PSD than with NS-ImCoh, however, the subtle changes of dynamics might be more easily captured even though the noise in the computation limits our interpretation. It is to note that the PSD and the NS do not use the same parameters of windowing and overlap hence intrinsic differences, when the coherence is computed using the same parameters as PSD, patterns are even less visible. This is due to the fact that coherence is more sensitive to noise and requires more information hence wider windowing for it to reveal relevant information.

*Investigating inter trials and temporal variability through classification*: If we evaluate each time point as a single feature (of the pre-selected electrodes at a specific frequency bin) and train the algorithm on all the time points and compare it to the average feature over all the time points and train the algorithm on the average features, it appears clear that the performances favour the average features. In both cases (PSD and NS-imCOH), averaging tends to increase class differentiation. Quite surprisingly the difference of intensity between NS and PSD does not seem to have an effect on performances: indeed, PSD, while slightly superior in average, is not necessarily better than node strength when training on each

separate time point. Even though, the  $R^2$  statistical test between MI and resting state trials shows higher scores for PSD than for NS. Altogether, linear machine learning algorithm (such as LDA) are sensitive to the noisy time points in PSD and NS, which stresses the importance for averaging along the trial to obtain good levels of discrimination.

*Combining information*: Two separate elements are to mention regarding the features and their resulting accuracies as shown in Fig 3. First, the node strength and PSD do not carry the same information: indeed if the same choices of features are made for both modalities (NS and PSD) and we base ourself on PSD, NS accuracies are lower. However, if other features are selected based on the specific NS  $R^2$ , NS based algorithm is showing accuracy improvement. Second, The interesting result we reveal is that if features of network and power are combined, performance always increase (in both training over each time windows and on average over trial). This tends to indicate the complementary nature of the two approaches, indeed while the local PSD information provides the majority of the information, the distributed information given by the network via the node strength has a role to play.

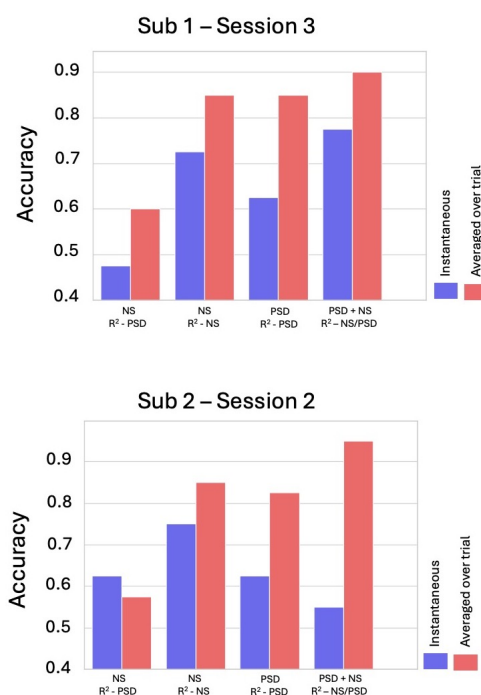


Figure 3: *Accuracy Comparison* LDA trained on motor imagery vs resting state trials of phase 1 and tested on phase 2 based on relevant neurophysiological features selected using the  $R^2$  in NS and PSD either with all the time points as separate features as if it was instantaneous (blue) or the average features over the time points (red). Top: Subject 1 on session 3. Bottom: Subject 2 on session 2. NS:Node Strength, PSD:Power Spectral Density

## DISCUSSION

*What does network dynamics reveal?:* Studying networks behaviour to discriminate between mental tasks is relatively new and yet to be used in online BCI paradigm[7, 11]. Even though connectivity using imaginary coherence is known to decrease during a motor imagery task with respect to a resting state in certain frequency bands [5], the temporal dynamics associated are still poorly studied. Here we propose to explore with preliminary results what are those dynamics in order to know how they could be integrated to the BCI context. We find that node strength follows a similar trend as ERD/ERS even though the data is more sensitive to noise which limits the amount of interpretations regarding the neurophysiological process.

*Auto regressive method needs fine tuning to overcome its limitations:* Spectral estimation is always subjected to the rule of its estimator and the use of any estimator over short time windows is a challenge. AR method (Burg) has been studied so far in the domain of power spectral density [12, 13] but its use for computing coherence has remained marginal compared to welch or multitaper which are not suited for short time window estimation[14]. Our first results show that auto regressive method with short windows and overlap are far more sensitive to noise with coherence compared to PSD which limits the ways we can use such method. This is especially revealed by the important decrease of performance when taking each time point as a feature. To contrast this effect, two approaches could be used, if it is for offline analysis, averaging over trials could filter those noises even though the estimation could still be erroneous. Using larger time windows could be the definitive solution to limit this noisy computational phenomenon.

*Can we use continuous MI BCI within this framework?:* One of the many problematic regarding BCI is the use of discrete or continuous feedback, which produce different effects on subjects[15, 16]. Continuous feedback can be used if the features they rely on are estimated on short time windows, the intrinsic noise of EEG makes. The logical follow up of our endeavour on the temporal dynamics of node strength is to interrogate the use of this feature in a feedback context. Here, our few results tend to demonstrate that continuous BCI which could use node strength as features will be highly impacted by the spectral estimation noise. Even though connectivity measures seems to improve performance when added to PSD, it is necessary to stress the use of an averaging over time series which means doing pseudo-continuous or discrete feedback.

## CONCLUSION

In this contribution, we investigate new forms of features that could be used for Motor imagery BCI relying on a network approach to EEG. We use data coming from subjects who perform during three sessions MI BCI to control the seizing of object with a robotic arm. We evaluate using imaginary coherence their network dynamics dur-

ing the MI task. We find out patterns similar to ERD/ERS but with some more subtle phenomenon which might be hidden due to the spectral estimator used. These results have to be tempered by the low amount of subjects and will require more of them to strengthen the conclusions. Nevertheless, based on those preliminary result, we advocate for the use of features averaged over time to maximize the differences that could be spotted and to use PSD and NS combined as they seem to be complimentary in the information they provide.

## ACKNOWLEDGEMENTS

FDVF acknowledges support from the European Research Council (ERC), Grant Agreement No. 864729 and from the Agence Innovation Defense DGA. The authors declare that they have no competing interests.

## REFERENCES

- [1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol.* 2002;113(6):767–791.
- [2] Lebedev MA, Nicolelis MAL. Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation. *Physiological Reviews.* 2017;97(2):767–837.
- [3] Guillot A, Collet C. Contribution from neurophysiological and psychological methods to the study of motor imagery. *Brain Research Reviews.* 2005;50(2):387–397.
- [4] Pfurtscheller G, Silva FH Lopes da. Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin Neurophysiol.* 1999;110(11):1842–1857.
- [5] Cattai T, Colonnese S, Corsi MC, Bassett DS, Scarano G, De Vico Fallani F. Phase/Amplitude Synchronization of Brain Signals During Motor Imagery BCI Tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* 2021;29:1168–1177.
- [6] Gonzalez-Astudillo J, Cattai T, Bassignana G, Corsi MC, Fallani FDV. Network-based brain-computer interfaces: Principles and applications. *J. Neural Eng.* 2021;18(1):011001.
- [7] Leeuwis N, Yoon S, Alimardani M. Functional Connectivity Analysis in Motor-Imagery Brain Computer Interfaces. *Frontiers in Human Neuroscience.* 2021;15.
- [8] Venot T, Desbois A, Corsi MC, Hugueville L, Saint-Bauzel L, Fallani FDV. Intentional binding enhances hybrid BCI control. *arXiv:2309.12195 [q-bio].* Oct. 2023. [Online]. Available: <http://arxiv.org/abs/2309.12195> (visited on 02/26/2024).
- [9] Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S, Hallett M. Identifying true brain interaction from EEG data using the imaginary part of coherence. *Clinical Neurophysiology.* 2004;115(10):2292–2307.

- [10] Desbois A, Venot T, De Vico Fallani F, Corsi MC. *HappyFeat*—An interactive and efficient BCI framework for clinical applications. *Software Impacts*. 2024;19:100610.
- [11] Corsi MC, Chevallier S, Fallani FDV, Yger F. Functional Connectivity Ensemble Method to Enhance BCI Performance (FUCONE). *IEEE Trans Biomed Eng*. 2022;69(9):2826–2838.
- [12] Bufalari S, Mattia D, Babiloni F, Mattiocco M, Marciani MG, Cincotti F. Autoregressive spectral analysis in Brain Computer Interface context. *Conf Proc IEEE Eng Med Biol Soc*. 2006;2006:3736–3739.
- [13] Krusienski DJ, McFarland DJ, Wolpaw JR. An evaluation of autoregressive spectral estimation model order for brain-computer interface applications. *Conf Proc IEEE Eng Med Biol Soc*. 2006;2006:1323–1326.
- [14] Diez P, Laciár E, Mut V, Avila E, Torres A. A Comparative Study of the Performance of Different Spectral Estimation Methods for Classification of Mental Tasks. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*. 2008;2008:1155–8.
- [15] Carabalona R. Attitude of the subject towards feedback and its implications for BCI use in neurorehabilitation. In: 2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010). Nov. 2010, 1–4.
- [16] Carino-Escobar RI, Rodríguez-García ME, Carrillo-Mora P, Valdés-Cristerna R, Cantillo-Negrete J. Continuous versus discrete robotic feedback for brain-computer interfaces aimed for neurorehabilitation. *Front Neurobot*. 2023;17:1015464.

# WHICH FACTORS AFFECT THE ACCEPTABILITY OF BCIS FOR FUNCTIONAL REHABILITATION AFTER STROKE AMONG PATIENTS?

E. Grevet<sup>1</sup>, M. Izac<sup>1</sup>, F. Amadiou<sup>2</sup>, J. Py<sup>2</sup>, D. Gasq<sup>\*3</sup>, C. Jeunet-Kelway<sup>\*1</sup> (\*co-last)

<sup>1</sup>Univ. Bordeaux, CNRS, EPHE, INCIA, UMR5287 F-33000 Bordeaux, France

<sup>2</sup>CLLE, Université de Toulouse, CNRS, Toulouse, France

<sup>3</sup>ToNIC, Université de Toulouse, INSERM, Toulouse, France

E-mail: elise.grevet@u-bordeaux.fr

**ABSTRACT:** Although motor imagery-based BCIs have been demonstrated to be relevant for improving motor recovery after stroke, they remain barely used in rehabilitation services. We hypothesise that *acceptability* (which is assessed in terms of perceived usefulness, ease of use and intention to use) could serve as a lever for fostering the adoption of BCIs through the improvement of their efficacy. More precisely, we suggest that improving the acceptability of BCIs could alleviate post-stroke patients' anxiety, stimulate their motivation and engagement in the BCI process, and thereby, favour skill acquisition (here self-regulation abilities), which will ultimately have positive effects on motor recovery. We created a model of acceptability of BCIs specifically for functional rehabilitation after stroke, and designed an associated questionnaire that was used to empirically assess the weight each factor of the model had on acceptability. Hereinafter, we introduce the methods and results obtained based on the responses received from 140 patients, and compare them with data collected in the general public (N=753). In a nutshell, for both the general public and patients perceived usefulness, scientific relevance and ease of learning emerge as the most influential factors.

## INTRODUCTION

BCI-based functional rehabilitation procedures have demonstrated their efficacy to improve post-stroke patients' motor and cognitive abilities [1, 2]. In the coming years, they are expected to substantially improve the quality of life of those patients [2].

In classical functional rehabilitation procedures, when subjects have no residual movement, i.e., when they cannot voluntarily move their affected limb, physical practice is impossible and both subjects and therapists must mainly rely on mental practice alone. Mental practice includes motor imagery (MI) as well as attempted movements. In this context, BCIs are very relevant as they enable the detection of MI / attempted movements of the impaired limb, which are underlain by modulations of the so-called sensori-motor rhythms (SMRs)—as defined in the BCI field by a large band covering mu ( $\mu$ ) and beta ( $\beta$ ) rhythms (8–30 Hz) [3]—, and provide the patient with a synchronised neurofeedback (NF), for instance using functional

electrical stimulation that triggers an arm muscle contraction, or visual feedback (movement of a virtual hand on a screen [4]). Such a NF training enables the participants to train to voluntarily self-regulate their SMRs in a closed loop process, which should favour synaptic plasticity and motor recovery [5].

While this is encouraging, BCI efficiency is still far from the level required to achieve the clinical breakthrough expected by both clinicians and patients. Thus, BCIs remain barely used in clinical practice, outside laboratories [6]. BCI efficiency is known to be modulated by several factors. Many researchers are working on improving this efficiency either from a “technical” point of view (e.g., signal processing [7]), or from the human learning standpoint [8, 9]. Nonetheless, it might not be sufficient for those technologies to be actually used in a clinical setting: fully optimised BCIs (in terms of sensors, signal processing, and training procedures) are pointless if patients and clinicians are not able or do not want to use them, i.e., if BCIs are not accepted [10].

The concepts of acceptability and acceptance were introduced in order to understand what led users to adopt or not a new system [11]. The adoption of a technology refers to a use that is maintained over time, i.e., without abandonment. Acceptability and acceptance differ by the moment they are measured at: acceptability concerns the user's standpoint before any interaction with the system, while acceptance comes after at least one first use.

Misconceptions that patients and their entourage have regarding BCIs may have a detrimental effect on the acceptance of these technologies. For instance, BCI procedures are not often adapted to the general clinical guidelines and practices (e.g., organisational constraints, lack of training time), so caregivers are not engaged to use them [12]. BCI acceptance could also be altered by the fact that most stroke patients experience depression, and therefore high anxiety levels [13] that have detrimental effects on BCI acceptance and learning [14]. Thus, BCI acceptance is likely to have a major impact on patients' learning processes and therefore on the efficiency of BCI-based stroke rehabilitation procedures.

Among this clinical context, this article focuses on patients. We hypothesise that identifying acceptability and accep-

tance factors will help us overcome these misconceptions and personalise the rehabilitation procedures, which will in turn result in reduced anxiety, and increased motivation and engagement levels for the patients. This should favour their learning and, ultimately, motor recovery. In other words, we expect that improving the acceptance levels of BCIs, through the design of personalised rehabilitation procedures, will result in an increased efficiency of these technologies and therefore be one step closer to their democratisation.

Yet, using acceptance to optimise BCI efficiency remains an aspect that has been little studied to date. To the best of our knowledge, only [15] for BCI-based stroke rehabilitation procedures, [16] with BCI training for elderly and [10, 17, 18] with BCIs for Amyotrophic Lateral Sclerosis patients assessed BCI in terms of acceptance. In addition, in the BCI field, acceptability is mostly assessed as an attribute of the user's satisfaction, itself being a dimension of user experience [6, 18]. It is possible that the reduced number of studies stems from the lack of proven methods to measure acceptability and acceptance (e.g., dedicated questionnaire or model). This is what we hope to remedy through our research.

To do this, we designed a general theoretical model of BCI acceptability [19] (under review) and a second one focused on BCI for functional rehabilitation after stroke. They are based on the *Technology acceptance model 3* (TAM3) [20], the *Unified theory of acceptance and use of technology 2* (UTAUT2) [21], and the *Components of user experience* (CUE) model [22]. In these existing models, acceptability measure is an evaluation of the user's behavioral intention (BI) i.e., their intention to use the studied technology. The main determinants of BI are perceived usefulness (PU) and perceived ease of use (PEOU). PU is the personal feeling about utility of the system, and PEOU the degree of belief to which using the system will require little or no effort. On the basis of our model, a questionnaire to assess the acceptability of BCI-based functional rehabilitation procedures among the general public was created and validated (N=753).

The aim of our paper is to study the acceptability of BCI-based functional rehabilitation procedures among post-stroke patients, in order to determine their most important acceptability factors and to compare the results to those of the general public. This paper details our methodology, then the results are presented, in addition to the data collected from the general public. Finally, the discussion includes a comparison of these two populations.

## MATERIALS AND METHODS

In order to study the acceptability factors among patients, we used a questionnaire methodology — as we previously did for the general public.

### *Experimental paradigm:*

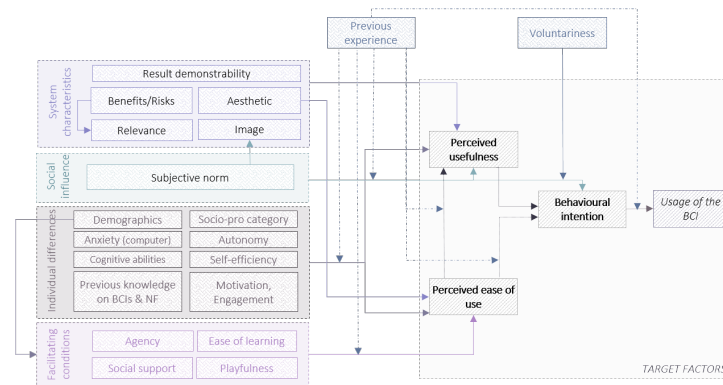
**Questionnaire:** The questionnaire is inspired from a previous questionnaire developed to identify and weigh the

factors influencing BCI acceptability for functional rehabilitation after stroke among the general public [23]. Detailed explanations regarding the design of this model are provided in [23]. In a nutshell, the BCI acceptability model for functional rehabilitation after stroke comprises four categories of factors: **(i) System characteristics** is a category related to the mental representation developed by the user to judge what the use of a technology can bring them in relation to their objective(s) (relevance of the system, perceived quality, etc.) [24]. **(ii) Social influence** is the influence of an individual's relatives and social group on their choice of whether or not to adopt a system. **(iii) Individual differences** is a category which groups the user personal characteristics (socio-demographic information, cognitive traits, etc.). Finally, **(iv) Facilitating conditions** brings together the factors related to the material, organisational and/or human conditions that facilitate the use of a technology [25] (Fig. 1).

We used the same questionnaire as the ones for the general public except that we added three factors into the **individual differences** category: *memory*, *attention* and *engagement in rehabilitation*. *Memory* [26] and *attention* [27] are both commonly affected after a stroke, and essential to learn to self-regulate brain patterns using a BCI (e.g., for memorising instructions and being able to stay focused on mental tasks [28]). The third factor was introduced in order to assess if the attitude towards BCI rehabilitation correlates with motivation in rehabilitation in general. All the questions are on the same scale in the patient and general public questionnaires.

The questionnaire was created on the Qualtrics tool, it was fully anonymous, and therefore not subject to the general data protection regulation (GDPR). It took between 20 min and 30 min to be completed, depending on the patients, and consisted of four parts: **(i)** Informed consent form; **(ii)** Questions regarding the participants' previous experience with BCIs; **(iii)** Questions related to each factor of the model (3-5 questions per factor). For example, for *subjective norm*, one of the question was 'People who are important to me would support the use of Brain-Computer Interfaces in post-stroke rehabilitation'. The scale used was a visual analogue scale from 0 to 10 ("strongly disagree" to "strongly agree") for quantitative factors and a checkbox question for categorical factors. Two explanatory videos were also included in the questionnaire: one explaining BCIs in general (video 1) and the second more specific to BCI-based stroke rehabilitation procedures, presenting EEG-based BCIs with motor imagery tasks (video 2). **(iv)** Socio-demographic data (for each item of this category, the participants could choose the option "I do not wish to answer").

With regard to the factors, some were assessed before and others after the second video. PU and BI were measured twice (before video 2: PU1/BI1; after video 2: PU2/BI2), the questions being the same for both times. The aim was to observe whether respondents' scores were influenced by the information given in the video. The factors following video 2 required a more detailed view of these new reha-



**FIGURE 1: Representation of the model of acceptability of BCIs for functional rehabilitation after stroke.** On the right are the target factors namely, PU, PEOU and BI. On the left are the four categories of factors that may influence the target factors. Finally, on top, two moderators are represented in blue. Those factors moderate the effect of the different categories on the target factors.

bilitation procedures (*result demonstrability*, *benefits/risks ratio* and *scientific relevance*).

For the added factors, we created the questions for *memory* and *attention*. Regarding *engagement in rehabilitation*, they come from the Treatment Self-Regulation Questionnaire (TSRQ) [29].

**Distribution:** The distribution of the questionnaire was done with two main methods: **(i) In hospital** (University hospitals of Bordeaux and Toulouse, France), with patients (N=40): experimenters visited stroke rehabilitation departments and helped patients to fill in the questionnaires. The experimenter read out the question and the patient was asked to answer orally or on a slate. The experimenters were medical students and a research engineer. **(ii) In autonomy, at home:** When the experimenters met patients in hospital, they always asked them whether they wished to complete the questionnaire on their own or accompanied by someone else. If the patient wished to take part independently, they gave their email address to the experimenter, and the anonymous link to the questionnaire was sent to them (N=52). The questionnaire link was also shared on social networks with the help of a patient who had a large online community concerned by stroke (N=48).

The exclusion criteria were people without experience of stroke and minors. The experimental protocol was carried out in accordance with the Declaration of Helsinki and was approved by the Research Ethics Committee of Bordeaux University (CER-BDX-AP-2022-14).

#### Data analysis:

We analysed the results in two stages. In both, the patient data was compared with the results for the general public.

**Descriptive analysis:** We measured the means and standard deviations (SD) of each quantitative factor, and the percentage distribution for the categorical factors. For the comparison, we used Welch t-tests (quantitative factors) and Chi<sup>2</sup> tests (categorical factors).

**Quantitative analysis:** We wanted to do observations that

do not depend on the architecture of our proposed model. Three linear regressions were implemented to find the most important determinants of BI, PU and PEOU. To predict BI and PU, we used all the acceptability factors in our questionnaire (regression for BI included PU and PEOU, regression for PU included PEOU, as the arrows in Fig. 1). For PEOU, *result demonstrability*, *benefits/risks ratio* and *relevance* (i.e., factors after video 2) were not included, nor were BI and PU.

It is a different type of regression than the one used for the general public (random forest regressions) as the number of respondents was lower. The categorical data from the questionnaire were formatted in order to enable their inclusion in the regression analyses. These regressions were implemented on R.

## RESULTS

### Participants:

A set of N = 140 respondents was obtained to the questionnaire, all of them were post-stroke patients in France. Of these, 60% were men and 40% women. The age group most represented was 55-65. The questionnaire was completed in hospital, with an experimenter, for 40 patients, while the other 100 completed it independently on their computer, at home. The socio-demographic details are provided in Table 1.

### Descriptive analysis:

In Table 2, are presented the mean scores of each quantitative factor, and the percentages for categorical factors. None of the factors was associated with a score below 5/10, which reflects globally positive feelings and well-perceived BCIs among the respondents. Indeed, regarding the target factors, for the patients, BI2 had a mean of 8.48/10 (SD = 2.03), for PU2 it was 8.34/10 (SD = 2.13) and for PEOU the mean was 6.43/10 (SD = 2.41).

These analyses show that certain factors differ significantly between the patients and the general public. Among target factors, only *PEOU* is significantly lower in the patient population. For *system characteristics*, patients have a significantly higher *benefit/risk ratio* score and a lower

		Number	%
AGE	18-24	2	1.43
	25-34	11	7.86
	35-44	31	22.14
	45-54	30	21.43
	55-65	32	22.86
	65-74	21	15.00
	74+	7	5.00
	Not know	6	4.29
GENDER	Male	84	60.00
	Female	56	40.00
POST-STROKE PERIOD	Acute (< 15 days)	2	1.43
	Subacute (15 days - 6 months)	21	14.00
	Chronic (> 6 months)	87	62.14
	No answer	30	21.43
SOCIO-PRO. CATEGORY	Students	2	1.43
	Craftsmen/shopkeepers	8	5.71
	Executives/Higher intellectual prof.	24	17.14
	Intermediate occupations	26	18.57
	Employees	32	22.86
	Manual workers	2	1.43
	Not in employment	7	5.00
	I do not wish to answer	3	2.14
	No answer	36	25.71
Questionnaire administration	With experimenter, at the hospital	40	28.57
	At home, recruited in hospital	52	37.14
	At home, recruited from social media	48	34.29

**TABLE 1: Respondents' socio-demographic information.**  
Number: number of respondents, %: percentage of respondents.

image score. Regarding *social influence*, the *subjective norm* score is significantly higher among patients. In *individual differences*, we found five significantly different factors. Perceived *autonomy* is higher among patients, and *computer anxiety* is lower (scores are inverted to be all negative to positive; a higher score means lower anxiety). Among the categorical factors, *self-efficacy*, *social support* and the *knowledge of BCIs* were significantly influenced by the group. For example, preferring to use a BCI with a human guidance is highly represented among patients.

#### Regressions:

Three linear regressions were performed in order to explain the main determinants of the three target factors: BI, PU and PEOU. Table 3 presents the variables with the most impact on these latter. The variables with significant p-values ( $p \leq 0.05$ ) were:

**For BI2:** *PU2*, *computer anxiety*, *socio-professional category* ("Students" and "No answer" categories), *autonomy*, *self-efficacy* ("I prefer to use the BCI alone, in autonomy" category) and *subjective norm*. Their coefficients were positive, except for *socio-professional category*. The quality of the prediction was good (adjusted  $R^2$ : 0.749, p-value < 2.2e-16). **For PU2:** *Relevance* and *computer anxiety*. The quality of the prediction was medium (adjusted  $R^2$ : 0.648, p-value < 2.2e-16). **For PEOU:** *Ease of learning* and *Playfulness*, but the prediction had a lower quality (adjusted  $R^2$ : 0.513, p-value = 3.178e-11).

## DISCUSSION

We created a model of acceptability of BCIs specifically for functional rehabilitation after stroke, and designed an associated questionnaire. We collected responses from 140 post-stroke patients and compare them with data previously obtained from the general public (N=753).

FACTORS	MEAN PATIENTS	MEAN GEN. PUBLIC
Scale from 0 to 10		
SYSTEM CHARACTERISTICS		
Result demonstrability	6.60 ± 1.93	6.84 ± 1.68
Benefits/Risks ***	7.80 ± 1.62	7.27 ± 1.51
Scientific relevance	8.09 ± 1.83	8.04 ± 1.48
Image *	5.54 ± 3.24	6.10 ± 2.18
Visual aesthetic	6.30 ± 2.47	6.62 ± 1.89
SOCIAL INFLUENCE		
Subjective norm **	7.80 ± 1.87	7.39 ± 1.71
INDIVIDUAL DIFFERENCES		
Engagement in rehabilitation	9.11 ± 1.52	/
Autonomy ***	7.99 ± 1.79	7.40 ± 1.46
General anxiety	5.23 ± 2.18	5.49 ± 1.87
Computer anxiety ***	7.40 ± 2.73	6.35 ± 2.51
Attention	6.16 ± 2.55	/
Memory	6.84 ± 2.52	/
BCI knowledge ***	Categorical variables <sup>Chi² residual</sup>	
No	70.7% <sup>0.231</sup>	68.7% <sup>-0.100</sup>
Yes (never used)	18.6% <sup>-1.697</sup>	27.1% <sup>0.732</sup>
Yes (already used)	10.7% <sup>2.991</sup>	4.0% <sup>-1.290</sup>
Self-efficacy ***		
Alone, in autonomy.	18.6% <sup>-0.931</sup>	23.1% <sup>0.401</sup>
Alone with a support function.	16.4% <sup>-3.487</sup>	36.7% <sup>1.504</sup>
Only with human guidance.	47.9% <sup>3.457</sup>	28.3% <sup>-1.491</sup>
Alone, if used similar technology before.	17.1% <sup>1.449</sup>	12.0% <sup>-0.625</sup>
Social support ***		
Independently at home.	22.1% <sup>-1.906</sup>	32.8% <sup>-1.906</sup>
With a healthcare professional.	67.1% <sup>2.727</sup>	47.7% <sup>-1.176</sup>
Alone, but in a healthcare establishment.	10.7% <sup>-2.063</sup>	19.5% <sup>0.890</sup>
FACILITATING CONDITIONS		
Playfulness	7.02 ± 2.41	6.90 ± 1.80
Ease of learning	6.06 ± 2.16	5.96 ± 1.62
Agency	6.25 ± 2.50	6.29 ± 1.65
TARGET FACTORS		
PEOU ***	6.43 ± 2.41	7.17 ± 1.57
PU	7.83 ± 2.00	7.87 ± 1.63
BI	8.11 ± 2.05	7.88 ± 1.73
PU2	8.34 ± 2.13	8.28 ± 1.57
BI2	8.48 ± 2.03	8.23 ± 1.69

**TABLE 2: Results from post-stroke patients' questionnaire in comparison to the general public's questionnaire.**

When a question was negative, the score was inverted (i.e., a high *general anxiety* score is in fact a low anxiety level).

For each factor, **Welch t-tests** (quantitative) and **Chi² tests** (categorical) were made. Factors in violet highlight significant differences between the two groups. '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05

Patients showed high acceptability levels, similarly to the general population (*behavioural intention*: 8.48/10 and 8.23/10, respectively and *perceived usefulness*: 8.34/10 and 8.28/10, respectively). Only the *perceived ease of use* was significantly lower in patients (6.43/10) than in the general public (7.17/10).

In addition to these target acceptability factors, descriptive analyses showed other significant differences between the two groups. Patients have a significantly higher *benefit/risk ratio* score (7.8/10), the advantages of BCI in functional rehabilitation seem greater to them than the disadvantages. Regarding *subjective norm* (7.8/10), compared with the general public, patients seem to consider that their close relatives and people who are important to them will have a more positive view of rehabilitation with BCI and will be more favourable to this type of rehabilitation. Nevertheless, they have a significantly lower *image* score (5.54/10): they think that the public image and the social status of people using BCI in rehabilitation will be less positive than respondents in the general population tend to expect. *Self-efficacy* and *social support* showed that the majority of pa-

PATIENTS - N=140					GENERAL PUBLIC - N=753	
Linear regressions					Random forest regressions	
	Estimate	Std. Error	t value	Pr(> t )	Importance values	
BEHAVIOURAL INTENTION						
Residual Std. error: 1.016 - R <sup>2</sup> : 0.821 - Adjusted R <sup>2</sup> : 0.749					% Variance explained: 86.09	
(Intercept)	-0.739	1.134	-0.652	0.516		
Perceived usefulness	0.451	0.080	5.617	1.79 x 10 <sup>-7</sup> ***	Perceived usefulness	100
Computer anxiety	0.125	0.046	2.739	0.007 **	Scientific relevance	37.44
Socio-pro 1 ("Students")	-2.543	0.952	-2.671	0.009 **	Benefits/Risks ratio	30.25
Socio-pro 3 ("No answer")	-1.992	0.748	-2.665	0.009 **	Subjective norm	29.56
Autonomy	0.139	0.056	2.505	0.014 *	Result demonstrability	28.03
Self-efficacy 1 ("Prefer to use BCI alone, at home")	0.758	0.339	2.234	0.028 *	Playfulness	27.67
Subjective norm	0.117	0.059	1.987	0.050 *	Perceived ease of use	24.94
PERCEIVED USEFULNESS						
Residual Std. error: 1.265 - R <sup>2</sup> : 0.746 - Adjusted R <sup>2</sup> : 0.648					% Variance explained: 79.64	
(Intercept)	-0.306	1.413	-0.217	0.829		
Scientific relevance	0.773	0.102	7.571	1.88 x 10 <sup>-11</sup> ***	Scientific relevance	100
Computer anxiety	0.125	0.055	2.250	0.027 *	Perceived ease of use	33.54
PERCEIVED EASE OF USE						
Residual Std. error: 1.682 - R <sup>2</sup> : 0.636 - Adjusted R <sup>2</sup> : 0.513					% Variance explained: 57.76	
(Intercept)	1.947	1.799	1.082	0.282		
Ease of learning	0.522	0.091	5.716	1.05 x 10 <sup>-7</sup> ***	Ease of learning	100
Playfulness	0.285	0.101	2.809	0.006 ***	Playfulness	83.21
					Subjective norm	80.86

**TABLE 3: Regression results for the target factors (BI2, PU2, PEOU).** For the patients, only the factors with significant p-value are displayed.

**Pr(>|t|):** probability of observing any value equal or larger than t (corresponds to p-value). **Estimate:** corresponds to the slope of the equation ("b" value). "\*\*\*\*" 0.001 "\*\*\*" 0.01 "\*\*" 0.05. For the general public, random forest regressions (500 trees and 5-fold cross-validation), were used. Importance values are the mean decrease accuracy (%IncMSE), scaled from 0 to 100.

tients want human guidance and expect to be accompanied by a caregiver when using a BCI, unlike the general public, who prefer a help system integrated into the computer. It is also interesting to note that patients have significantly higher perceived *autonomy* and lower *computer anxiety* than the general public. This shows that therapists or patients' relatives should not consider these factors as obstacles to offering patients a new type of rehabilitation technology such as BCI (i.e. not thinking that it is a bad idea for patients because they would have a fear of technologies, for example).

Regression analyses revealed that the intention to use BCIs was mainly motivated by the *perceived usefulness* of the system, itself mainly influenced by *scientific relevance* of BCIs in functional rehabilitation. *Subjective norm* likewise had a small but significant influence on acceptability. These factors were also important for the general public.

In line with those of the descriptive analyses, these results highlight the importance of scientific evidence and scientific communication not only to patients, but also to clinicians and the general public, as social norms (i.e., *subjective norm* and *image*) play an important role among patients.

In addition, within patients, *individual differences* showed a significant impact on the intention to use BCIs (it was not the case in the general public): the weight of psychological variables is greater in people who have suffered a stroke, as shown by the importance of *computer anxiety* and *autonomy*. It appears also that patients who prefer to use the BCI alone, at home (i.e. higher level of *self-efficacy*), are more likely to want to use a BCI in their rehabilitation, which is coherent with existing recommendations [14]. Reducing patients' anxiety and taking into account their perceived autonomy and self-efficacy is also something that could be achieved by personalising BCI protocols. For example,

by proposing training sessions where the degree of support provided by a therapist can be modulated.

As for the general public, among the patients, *ease of learning* and *playfulness* were the main determinants of *perceived ease of use*. Thus, with regard to the lower patients' *perceived ease of use* score, the aim is to improve this target factor by making it easier for patients to learn how to use the BCI. One way can be to personalise BCI protocols depending on patients' profiles, with the aim of making learning easier and more enjoyable for them. For instance, instructions must be clear and the feedback motivating. These findings are consistent with the guidelines for successful MI-BCI training [9] and with studies on gamified rehabilitation processes [30, 31].

## CONCLUSION

This study provides insights on how to foster BCI acceptability, notably by better informing the patients and the general public on the scientific evidence related to BCIs and by personalising rehabilitation procedures to facilitate learning. One next step will consist in adopting the same approach with clinicians in order to understand the conditions for a high acceptability of BCIs, be they related to scientific, technical or organisational aspects.

## ACKNOWLEDGEMENTS

We thank Jeremy Narayaninsamy and Emma Salgues for participating in the administration of the questionnaire at Toulouse University Hospital during their internship.

## REFERENCES

- [1] Cervera MA *et al.* Brain-computer interfaces for post-stroke motor rehabilitation: A meta-analysis. *Annals of clinical and translational neurology*. 2018;5(5):651–663.
- [2] Nojima I, Sugata H, Takeuchi H, Mima T. Brain-computer interface training based on brain activity can induce motor recovery in patients with stroke: A meta-analysis. *Neurorehabilitation and Neural Repair*. 2022;36(2):83–96.
- [3] Pfurtscheller G, Guger C, Müller G, Krausz G, Neuper C. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience letters*. 2000;292(3):211–214.
- [4] Pichiorri F *et al.* Brain-computer interface boosts motor imagery practice during stroke recovery. *Annals of neurology*. 2015;77(5):851–865.
- [5] Jeunet C, Glize B, McGonigal A, Batail JM, Micoulaud-Franchi JA. Using eeg-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: Theoretical background, applications and prospects. *Neurophysiologie Clinique*. 2019;49(2):125–136.
- [6] Kübler A *et al.* The user-centered design as novel perspective for evaluating the usability of bci-controlled applications. *PloS one*. 2014;9(12):e112392.
- [7] Lotte F *et al.* A review of classification algorithms for eeg-based brain-computer interfaces: A 10 year update. *Journal of neural engineering*. 2018;15(3):031005.
- [8] Pillette L, Jeunet C, Mansencal B, N'kambou R, N'Kaoua B, Lotte F. A physical learning companion for mental-imagery bci user training. *International Journal of Human-Computer Studies*. 2020;136:102380.
- [9] Roc A *et al.* A review of user training methods in brain computer interfaces based on mental tasks. *Journal of Neural Engineering*. 2021;18(1):011002.
- [10] Blain-Moraes S, Schaff R, Gruis KL, Huggins JE, Wren PA. Barriers to and mediators of brain-computer interface user acceptance: Focus group findings. *Ergonomics*. 2012;55(5):516–525.
- [11] Alexandre B, Reynaud E, Osiurak F, Navarro J. Acceptance and acceptability criteria: A literature review. *Cognition, Technology and Work*. 2018.
- [12] Seguin PRR, Maby E, Mattout J. What is missing for physicians to prescribe a bci? In: 8th Graz BCI Conference 2019. 2023.
- [13] Burton CAC, Murray J, Holmes J, Astin F, Greenwood D, Knapp P. Frequency of anxiety after stroke: A systematic review and meta-analysis of observational studies. *International Journal of Stroke*. 2013;8(7):545–559.
- [14] Jeunet C, N'Kaoua B, Lotte F. Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates. *Prog. Brain Res.*. 2016;228:3–35.
- [15] Morone G *et al.* Proof of principle of a brain-computer interface approach to support poststroke arm rehabilitation in hospitalized patients: Design, acceptability, and usability. *Archives of physical medicine and rehabilitation*. 2015;96(3):S71–S78.
- [16] Lee TS *et al.* A brain-computer interface based cognitive training system for healthy elderly: A randomized control pilot study for usability and preliminary efficacy. *PloS one*. 2013;8(11):e79419.
- [17] Huggins JE, Wren PA, Gruis KL. What would brain-computer interface users want? opinions and priorities of potential users with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*. 2011;12(5):318–324.
- [18] Nijboer F. Technology transfer of brain-computer interfaces as assistive technology: Barriers and opportunities. *Annals of physical and rehabilitation medicine*. 2015;58(1):35–38.
- [19] Pillette L *et al.* The acceptability of BCIs and neurofeedback: Presenting a systematic review, a field-specific model and an online tool to facilitate assessment. Will be submitted soon. in preparation.
- [20] Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*. 2008;39(2):273–315.
- [21] Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS quarterly*. 2012:157–178.
- [22] Thüring M, Mahlke S. Usability, aesthetics and emotions in human-technology interaction. *International journal of psychology*. 2007;42(4):253–264.
- [23] Grevet E *et al.* Modeling the acceptability of bcis for motor rehabilitation after stroke: A large scale study on the general public. *Frontiers in Neuroergonomics*. 2023;3:1082901.
- [24] Terrade F, Pasquier H, Juliette R, Guingouain G, Somat A. L'acceptabilité sociale: La prise en compte des déterminants sociaux dans l'analyse de l'acceptabilité des systèmes technologiques. *Le travail humain*. 2009;72.
- [25] Février F. Vers un modèle intégrateur "expérience-acceptation". Ph.D. dissertation. Université Rennes 2; Université Européenne de Bretagne. 2011.
- [26] Lugtmeijer S, Lammers NA, Haan EH de, Leeuw FE de, Kessels RP. Post-stroke working memory dysfunction: A meta-analysis and systematic review. *Neuropsychology review*. 2021;31:202–219.
- [27] Loetscher T, Potter KJ, Wong D, Nair R das. Cognitive rehabilitation for attention deficits following stroke. *Cochrane Database of Systematic Reviews*. 2019;(11).
- [28] Foong R *et al.* Assessment of the efficacy of eeg-based mi-bci with visual feedback and eeg correlates of mental fatigue for upper-limb stroke rehabilitation. *IEEE Transactions on Biomedical Engineering*. 2019;67(3):786–795.
- [29] Ryan RM, Connell JP. Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of personality and social psychology*. 1989;57(5):749.
- [30] Burke JW, McNeill M, Charles DK, Morrow PJ, Crosbie JH, McDonough SM. Optimising engagement for stroke rehabilitation using serious games. *The Visual Computer*. 2009;25(12):1085–1099.
- [31] Lopes S *et al.* Games used with serious purposes: A systematic review of interventions in patients with cerebral palsy. *Frontiers in psychology*. 2018;9:1712.

# BREAKING OUT OF THE FEEDBACK LOOP: TRANSFERRING MASTERY OF SELF-REGULATION DURING NEUROFEEDBACK TO OTHER CONTEXTS

S.E. Kober<sup>1</sup>, G. Wood<sup>1</sup>

<sup>1</sup> Department of Psychology, University of Graz, Graz, Austria

E-mail: [silvia.kober@uni-graz.at](mailto:silvia.kober@uni-graz.at)

**ABSTRACT:** One important question in neurofeedback (NF) research is the mastery of self-regulation and the generalizability of the NF training results. Here, we investigated whether NF users can voluntarily increase the Sensorimotor Rhythm (SMR, 12-15 Hz) activity during repeated NF training sessions while receiving visual feedback and if they can also increase SMR during subsequent transfer sessions without any feedback. We also assessed the used mental strategies during the sessions. Nine healthy adults received real feedback, nine received sham feedback. Only the real feedback group was able to linearly increase SMR within the six NF training sessions. However, they could not increase SMR during the transfer sessions. Participants reported multiple different mental strategies during NF training as well as during transfer sessions with different success rates. These results indicate that participants were not able to transfer successful mental strategies to other situations after six sessions of SMR-based NF training.

## INTRODUCTION

In neurofeedback (NF) applications, users can learn to voluntarily modulate brain signals, in most cases the electrical brain activation recorded with the electroencephalogram (EEG), in a desired direction through real-time feedback. The aim is to intentionally reinforce EEG patterns that are associated with optimal cognitive or motor processes. Successful NF training can consequently lead to cognitive, motor, or affective improvements [1]. However, long-term effects of NF training or transfer effects are hardly investigated [2]. Users have varying degrees of success in regulating their own brain activation during NF training. Up to 30% of NF users are so-called non-responders and the exact reasons for this inability are still open [3]. The mental strategies used appear to be an important predictor of the success of NF training. In SMR-(12-15 Hz) based NF training studies, it turned out that participants report many different mental strategies during NF training with different success rates [4-6]. The use of no specific mental strategy seems to be advantageous to upregulate the SMR over central brain areas [6]. But there are also other successful mental strategies reported to increase SMR activity during NF training [4-6]. In this context, the question arises as to whether NF users can transfer mental strategies that they use during NF training to increase SMR to other situations where they

do not receive real-time feedback via NF training. In theory, learning how to up-regulate SMR at a given time, which should lead for instance to improved cognitive performance [1, 7], should be transferable to other contexts (e.g., school, work) without real-time feedback of one's own brain activity [8]. In the present study, we investigated the ability of NF users to increase SMR during NF training receiving visual feedback of SMR changes as well as subsequent transfer sessions without any feedback of one's own brain activity. Additionally, we assessed the used mental strategies during NF training and the transfer sessions to see whether the same strategies lead to an increase in SMR in both NF and transfer sessions, or not.

Gruzelier [2, pp. 18] mentioned in his review article that the ultimate goal of NF training is the *mastery of self-regulation* and that this can be evaluated using transfer trials where the participants do not receive any feedback or reward. Such transfer trials are generally included in slow cortical potentials (SCP) NF training (e.g., [9]). However, a differential process analysis is lacking. For instance, Gevensleben et al. [10] included transfer trials without contingent feedback in their NF protocol and also gave home-work. NF users were required to practice their focused mental state, which they should achieve during NF training, at home. However, NF training results or changes in EEG activity during transfer trials were not reported [10]. In a SCP-based NF study by Barth et al. [11], EEG activity during transfer trials were reported. However, as in the other SCP NF studies using transfer trials, no feedback was presented during the transfer trials but participants received reinforcement following the transfer trials in case they had regulated in the desired direction often in form of a smiley. The authors used the transfer trials to categorize the NF users in learners and non-learners. Changes in SCP during NF and transfer trials were not directly comparable [11]. Kleih-Dahms et al. [12] reported SCP results of single subjects for NF and transfer trials. It turned out that in some NF users, SCP changes were not present in transfer trials while in other users, SCP changes were even stronger during transfer trials than during NF trials. In this study, mental strategies during SCP regulation were assessed as well. However, they did not differentiate between mental strategies used during NF and transfer trials [12].

In the present study, we investigated changes in SMR activity during repeated NF training sessions as well as

in subsequent transfer sessions. We also assessed the used mental strategies during NF training and transfer trials. To rule out possible placebo or unspecific effects [13,14], we also included sham control groups receiving fake feedback. We expect that the real feedback group should be able to increase SMR activity during NF training, while the sham group should not show linear increases in SMR activity. In line with previous NF studies, NF users should report many different mental strategies with different success rates during NF training [4-6]. According to previous SCP-NF studies [11,12], there might be differences in changes in EEG activity during the NF and transfer trials.

## MATERIALS AND METHODS

*Participants:* Eighteen participants were randomly assigned to two groups. One group received real feedback of their own SMR activity ( $N = 9$ , 5 females, average age = 25.7 years,  $SD = 3.02$ ), one group received sham feedback by receiving feedback of another participants' EEG recording ( $N = 9$ , 5 females, average age = 22.8 years,  $SD = 2.96$ ). All participants signed a consent form. The study was approved by the ethics committee of the University of Graz, Austria (GZ. 39/9/63 ex 2019/20).

*Design and procedure:* All participants performed in sum 9 sessions on different days within 3 weeks. The first session was an instructional session where participants should be relaxed but mentally focused but did not get any visual feedback of their brain activity. In session 2 to 7, NF training was performed. Participants of the real feedback group received feedback of their own SMR activity (12-15 Hz) over Cz via visual feedback. The sham group saw the same visual feedback screen but the movement of the visual feedback was not related to their own brain activity. It showed changes in SMR activity recorded in another participant. The last two sessions were transfer sessions. In these sessions, participants should try to reach the mental state they had during the NF training sessions without getting any visual feedback. Session 8 was performed directly after the last NF training session, session 9 one week later.

*Neurofeedback training:* During NF training, participants received visual feedback. Changes in target EEG activity were depicted by vertically moving bars. Three bars were depicted on a conventional computer screen. The bar in the middle of the screen depicted changes in SMR (12-15 Hz) power. The bar on the left side of the screen showed changes in theta (4-7 Hz) power (to control for eye movement artifacts) and the bar on the right side of the screen depicted high beta (21-35 Hz) power (to prevent the participants from producing too many muscle artifacts). Per NF training session, 7 three-minute runs were performed. The first three-minute run was a baseline run. Here, participants were instructed to relax and watch the moving bars without trying to control them. This baseline run was used to define individual threshold values per participant (median SMR value for middle bar, median + 1  $SD$  for theta and beta

bars). The subsequent six runs were feedback runs where participants were instructed to increase the size of the middle bar while keeping the bars on the left and right as small as possible. An increase in EEG power led to an increase in size of the bar and vice versa. When the SMR bar exceeded its threshold and theta and beta bars were below their thresholds, the bars turned green, and a reward counter increased. Otherwise, the bars turned red. Participants were instructed to be physically relaxed and mentally focused and concentrated to increase SMR activity. This NF protocol has been successfully used in previous NF studies to increase SMR while controlling artifact activity (e.g., blinking, muscle activity) [5-7].

*EEG recording and analysis:* 12 EEG electrodes were recorded (F3, Fz, F4, C3, C1, Cz, C2, C4, CPz, P3, Pz, P4) using a g.USBamp 16 channels standard amplifier (g.tec, Austria). A linked mastoid reference was used, the ground was placed at FPz. Vertical and horizontal EOGs were placed on the outer canthi of the eyes and superior to the nasion. Impedances were kept below 5 kOhms for the EEG electrodes and below 10 kOhms for the EOGs. EEG signals were digitized at 256 Hz and filtered with a 0.5 Hz high-pass and a 60 Hz low-pass filter. To analyze the EEG data, the Brain Vision Analyzer software (version 2.2, Brain Products GmbH, Germany) was used. Ocular artifacts such as eye blinks were corrected using an automatic ocular correction method (Gratton & Coles), followed by a semi-automatic artifact rejection (criteria for rejection:  $>50.00 \mu V$  voltage step per sampling point, absolute voltage value  $\geq \pm 150.00 \mu V$ , lowest allowed activity in 100 ms intervals:  $0.5 \mu V$ , maximal allowed difference of values in 200 ms intervals:  $200 \mu V$ ). All data points with artifacts were excluded from further EEG analysis. Absolute SMR power values recorded over electrode position Cz were extracted by means of complex demodulation (Brain Products GmbH, 2009). Power values were averaged per run.

*Mental strategies:* To assess the mental strategies the participants used to modulate SMR in the desired direction during NF training as well as during the transfer session, participants had to write down the mental strategies used after the first and the last NF training session as well as after the last two transfer sessions. The verbal descriptions were classified by two independent raters in different categories in accordance with prior studies [4-6]. The categories were: Visual (focusing on visual things, e.g., visual properties of the feedback screen), Cheering the feedback bars on, Breath (concentration on one's own breathing), Auditory (thinking of auditory stimuli), Concentration, Body (focusing on one's own body or bodily sensations), Relax, Cognitive (e.g., mental calculations), No Strategy (reporting to have no specific strategy, to do nothing in particular), Other Strategies. Inter-rater reliability was sufficiently high (across all categories: Kappa = 0.6). Participants reported generally multiple strategies during the sessions.

**Statistical analysis:** To define successful NF training performance, we analyzed changes in SMR power within NF training sessions across feedback runs. This is in line with prior studies showing changes in SMR power within NF training sessions but not between training sessions (e.g., [7]). A linear increase in SMR power across NF runs within a training session is an indicator for successful voluntary up-regulation of SMR activity at a given time and voluntary phasic EEG changes, which does not necessarily be related to changes in tonic or background EEG measures indicated by SMR changes across sessions [1,7,15,16].

Hence, to quantify NF training performance, we performed linear regression analysis with SMR power as dependent variable and NF run number (averaged across all NF training sessions) as predictor variable for each participant. The resulting regression slope was used as indicator for NF training performance. A positive slope is a sign of a linear increase in SMR power across NF training runs, while a negative slope is a sign of unsuccessful training. These regression slopes were then compared to zero using *t*-tests against zero per group. Alpha (0.05) levels were adjusted using Bonferroni correction, normal distribution was given.

Furthermore, *t*-tests were used comparing average SMR power across all NF sessions with average SMR power of the transfer sessions (session 8 and 9).

To analyze the mental strategies descriptively, we calculated the percentage of participants reporting a specific strategy per session, averaged over the classification results of both raters. Then, we calculated the average regression slope (changes in SMR power over the runs within a session) per reported mental strategy to determine which mental strategy was associated with a successful or unsuccessful SMR increase and if successful mental strategies during NF training were also successful during the transfer sessions. Note that we only report on the mental strategies used by the real feedback group.

## RESULTS

In a first step, we analyzed changes in SMR power within NF training sessions. The real feedback group could successfully increase their SMR power over the NF runs within the NF training sessions. Comparing the slopes of SMR power across the feedback runs against zero showed a significant difference for the real feedback group ( $t(8) = 3.20$ ,  $p = 0.01$ ) but not for the sham feedback group ( $t(8) = 0.32$ ,  $p = 0.76$ ) (Fig. 1).

During the transfer sessions (Fig. 2), neither the real feedback group ( $t(8) = -0.90$ ,  $p = 0.40$ ) nor the sham feedback group ( $t(8) = 0.45$ ,  $p = 0.67$ ) showed a significant linear increase in SMR power across runs. Absolute SMR power across the NF sessions was numerically higher in the real feedback group than in the sham feedback group and higher during the NF training sessions than during the transfer sessions (Fig. 3), but there were no statistically significant differences.

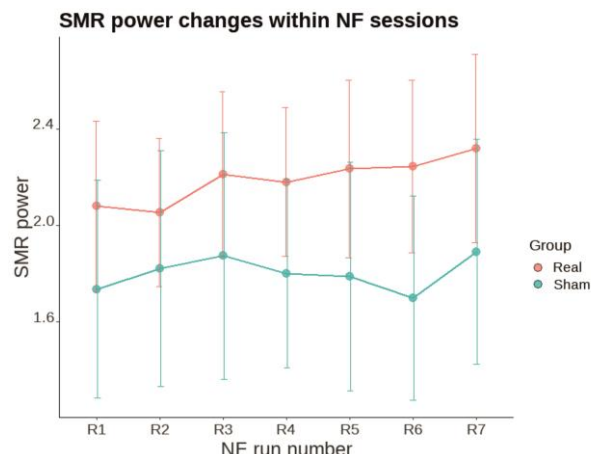


Figure 1: Mean changes in SMR power [ $\mu V^2$ ] per group across feedback runs within NF training sessions. Error bars show *SE*.

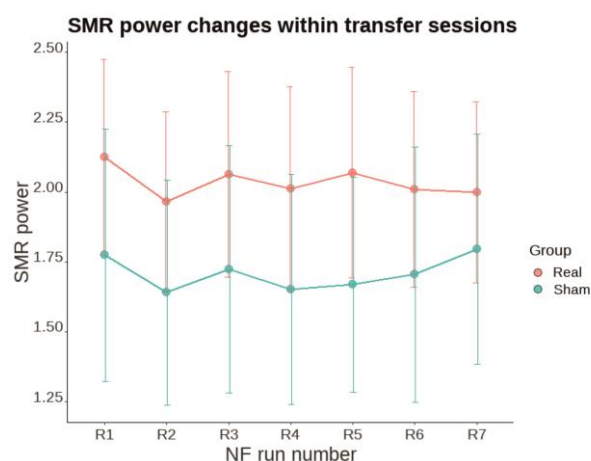


Figure 2: Mean changes in SMR power [ $\mu V^2$ ] per group across runs within transfer sessions. Error bars show *SE*.

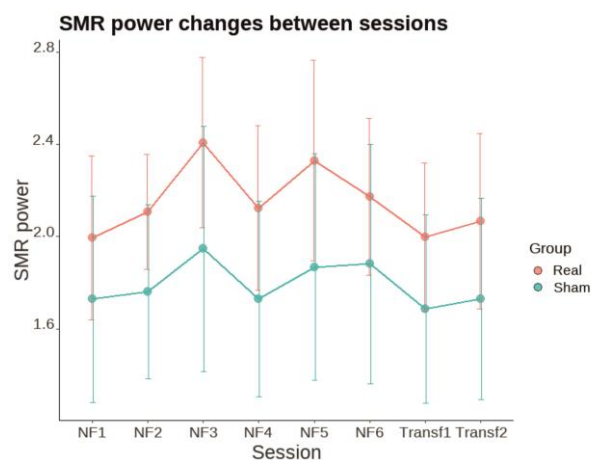


Figure 3: Mean changes in SMR power [ $\mu V^2$ ] per group across sessions. Error bars show *SE*.

The analysis of the mental strategies revealed that participants used many different strategies during the NF training sessions as well as during the transfer sessions (Fig. 4). Across all sessions, Concentration was mentioned most often, followed by Cognitive and Visual strategies. The frequency of the usage of a specific strategy changed over sessions, but in many cases, the strategies were more frequently used during the NF sessions than during the transfer sessions (e.g., Visual, Cheering, Auditory, Concentration, Relax). The breathing strategy was more frequently used during the transfer sessions compared to the NF sessions. This mental strategy turned out to be the most successful one during NF training indicated by positive regression slopes. However, this strategy was not successful during the transfer session as shown by negative slopes (Fig. 5). The Visual strategy seemed to be as successful during the first transfer session as during the NF training sessions. No Strategy was the most successful one during the transfer sessions.

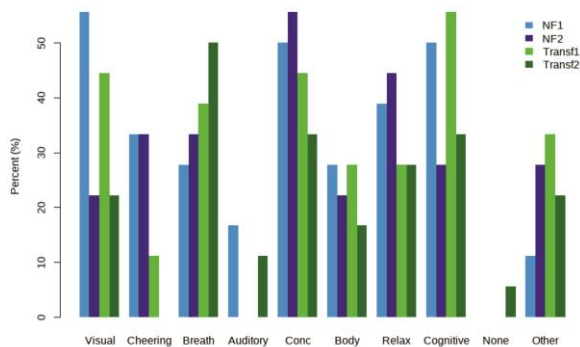


Figure 4: Percentage of participants reporting a specific mental strategy during the first (NF1) and the last NF session (NF2) and during the last two transfer sessions (Transf 1 & 2).

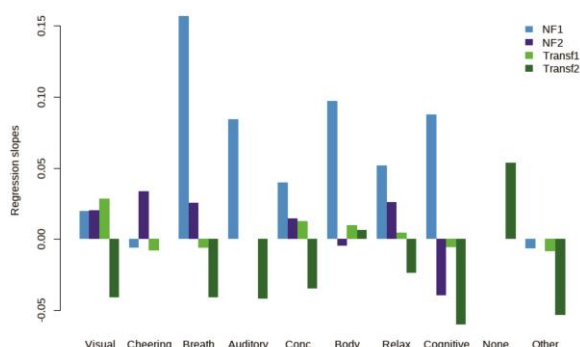


Figure 5: Slopes of linear changes in SMR power across runs within training sessions for the different mental strategies during the first (NF1) and the last NF session (NF2) and during the last two transfer sessions (Transf 1 & 2).

## DISCUSSION

Here we focused on the question of mastery of self-regulation of SMR activity during NF training and the generalizability of the NF training results. We were interested in the ability of NF users to increase SMR during transfer trials without any feedback and the used mental strategies during NF training as well as transfer trials.

In line with previous SMR-based NF training studies, participants in the real feedback group were able to linearly increase SMR activity within NF training sessions, while the sham group showed no linear SMR increases [1,7,15,16]. This indicates that the real feedback group was able to increase the target EEG feedback frequency at a given time and that NF learning happened when receiving real-time visual feedback in the real feedback group.

Although the real feedback group showed some form of learning during the NF training sessions, participants were not able to transfer these mental states to a situation without real-time feedback of one's own brain activity, at least after six sessions of NF training. Hence, they did not show a linear increase in SMR activity during the transfer sessions as in the NF sessions. So far, transfer trials were mainly used in SCP-based NF training protocols [8-12]. Most of these prior SCP studies did not report on SCP changes during the transfer trials. Studies reporting changes in SCPs during transfer trials reported heterogeneous results [11,12]. Some participants managed to transfer successful self-regulatory processes to situations without feedback, others did not [12]. Kleih-Dahms et al. [12] defined the start of the transfer trials individually depending on the SCP control during NF training. When participants successfully controlled their SCP, transfer trials were included in the NF training. This took between 15 and 17 sessions [12]. In this study, we analyzed group data rather than individual data. Transfer sessions started after the sixth NF training sessions. It may be beneficial to customize the start of transfer sessions based on NF training performance also in SMR-based NF training protocols.

Participants tried multiple different mental strategies during NF training as well as during the transfer sessions. This is in line with prior SMR-based NF training studies that also analyzed the used mental strategies [4-6]. These prior studies also consistently revealed that the strategy "Concentration" is one of the most frequently reported mental strategy during SMR-based NF training [4-6], which might be caused by the instruction of being physically relaxed but mentally focused and concentrated during NF training. "Visual" and "Cognitive" strategies are also often mentioned by participants to increase SMR during NF training [4-6]. But the most frequently mentioned mental strategies are not necessarily the most successful. As in previous studies, breathing strategies are rarely used, but are among the most successful strategies during NF training [4-6]. Surprisingly, these breathing strategies were no longer successful in the transfer trials, although even more participants reported

using them. The “Visual” strategy seemed to be as successful during the first transfer session as during the NF training sessions. In the transfer sessions, participants mentioned here to visualize the moving bars which they have seen in the previous NF sessions. This strategy successfully led to an increase in SMR power over the runs of the first transfer session, which was performed directly after the last NF training session, but not of the second transfer session, which was performed one week after the first transfer session. Having no specific mental strategy (“None”) turned out to be the most successful one during transfer trials. Previous SMR-based NF training studies also showed that NF users that learned to increase SMR successfully also stopped to use any specific mental strategy [6]. An uncontrolled attempt to use too many mental strategies at once or alternately could overload cognitive resources and could be detrimental to the mental state needed to produce SMR.

## CONCLUSION

We could not show that NF users are able to transfer mastery of self-regulation of SMR activity achieved during NF training to other situations without visual feedback. Also, successful mental strategies used during NF training could not be transferred. An individual adjustment of the start of the transfer sessions depending on the NF training success could be useful and should be investigated in future studies.

## REFERENCES

- [1] Gruzelier JH. EEG-neurofeedback for optimising performance. I: A review of cognitive and affective outcome in healthy participants. *Neurosci. Biobehav. Rev.* 2014;44:124–41.
- [2] Gruzelier JH. EEG-neurofeedback for optimising performance. III: A review of methodological and theoretical considerations. *Neurosci. Biobehav. Rev.* 2014;44:159–82.
- [3] Allison B, Neuper C. Could Anyone Use a BCI? In: Tan D, Nijholt A, editors. *Brain-Computer Interfaces: Human-Computer Interaction Series*. London: Springer-Verlag; 2010, p. 35–54.
- [4] Autenrieth M, Kober SE, Neuper C, Wood G. How Much Do Strategy Reports Tell About the Outcomes of Neurofeedback Training? A Study on the Voluntary Up-Regulation of the Sensorimotor Rhythm. *Front. Hum. Neurosci* 2020;14:218.
- [5] Kober SE, Witte M, Ninaus M, Koschutnig K, Wiesen D, Zaiser G et al. Ability to Gain Control Over One’s Own Brain Activity and its Relation to Spiritual Practice: A Multimodal Imaging Study. *Front. Hum. Neurosci* 2017;11:1–12.
- [6] Kober SE, Witte M, Ninaus M, Neuper C, Wood G. Learning to modulate one’s own brain activity: the effect of spontaneous mental strategies. *Front. Hum. Neurosci* 2013;7:1–12.
- [7] Kober SE, Witte M, Stangl M, Valjamae A, Neuper C, Wood G. Shutting down sensorimotor interference unblocks the networks for stimulus processing: An SMR neurofeedback training study. *Clin Neurophysiol* 2015;126(1):82–95.
- [8] Gevensleben H, Holl B, Albrecht B, Schlamp D, Kratz O, Studer P et al. Neurofeedback training in children with ADHD: 6-month follow-up of a randomised controlled trial. *Eur Child Adolesc Psychiatry* 2010;19(9):715–24.
- [9] Strehl U, Leins U, Goth G, Klinger C, Hinterberger T, Birbaumer N. Self-regulation of slow cortical potentials: a new treatment for children with attention-deficit/hyperactivity disorder. *Pediatrics* 2006;118(5):e1530–40.
- [10] Gevensleben H, Holl B, Albrecht B, Schlamp D, Kratz O, Studer P et al. Neurofeedback training in children with ADHD: 6-month follow-up of a randomised controlled trial. *European child & adolescent psychiatry* 2010;19(9):715–24.
- [11] Barth B, Mayer-Carius K, Strehl U, Wyckoff SN, Haeussinger FB, Fallgatter AJ et al. A randomized-controlled neurofeedback trial in adult attention-deficit/hyperactivity disorder. *Scientific reports* 2021;11(1):16873.
- [12] Kleih-Dahms SC, Botrel L. Neurofeedback therapy to improve cognitive function in patients with chronic post-stroke attention deficits: a within-subjects comparison. *Frontiers in Human Neuroscience* 2023;17:1155584.
- [13] Ros T, Enriquez-Geppert S, Zotev V, Young KD, Wood G, Whitfield-Gabrieli S et al. Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). *Brain* 2020;143(6):1674–85.
- [14] Thibault RT, Lifshitz M, Raz A. Neurofeedback or neuroplacebo? *Brain a journal of neurology* 2017;140(4):862–4.
- [15] Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev* 1999;29(2-3):169–95.
- [16] Dempster T, Vernon D. Identifying Indices of Learning for Alpha Neurofeedback Training. *Appl Psychophysiol Biofeedback* 2009;34(4):309–18.

# INTEGRATING CORTEC BRAIN INTERCHANGE DEVICE AND BCI2000 WITH A CLOUD INTERFACE

Filip Mivalt<sup>1,3</sup>, Max A. Van den Boom<sup>2</sup>, Frederik Lampert<sup>2</sup>, Jiwon Kim<sup>1</sup>, Andrea Duque Lopez<sup>1</sup>,  
Will Engelhardt<sup>4</sup>, Inyong Kim<sup>1</sup>, Su-youn Chang<sup>2</sup>, Dora Hermes<sup>1,5</sup>, Peter Brunner<sup>4</sup>, Vaclav  
Kremen<sup>1,5</sup>, Nuri Ince<sup>2,5</sup>, Gerwin Schalk<sup>6</sup>, Gregory A. Worrell<sup>1,5</sup>, Kai J. Miller<sup>2,5</sup>

<sup>1</sup>Department of Neurology, Mayo Clinic, MN, USA

<sup>2</sup>Department of Neurosurgery, Mayo Clinic, MN, USA

<sup>3</sup>Biomedical Engineering, Brno University of Technology, Brno, Czechia

<sup>4</sup>Department of Neurosurgery, Washington University School of Medicine, St Louis, MO, USA

<sup>5</sup>Department of Biomedical Engineering, Mayo Clinic, MN, USA

<sup>6</sup>Chen Frontier Lab, Tianqiao and Chrissy Chen Institute, Shanghai, China

E-mail: mivalt.filip@mayo.edu

**ABSTRACT:** Emerging brain-computer interface (BCI) systems may aim to develop invasive implantable systems to restore functionality in people with paralytic disabilities and to deliver adaptive brain stimulation (ABS) to treat severe neurological disorders. A key characteristic of next-generation implantable systems will be their capability to record extended periods of local field potential (LFP) data. Timely transfer of the recorded LFPs to the clinical team is crucial for monitoring the implanted system's reliability, safety and to dynamically enhance BCI and ABS applications in response to changing brain states. Our team is developing a comprehensive therapeutic BCI ecosystem that combines the Cortec BrainInterchange hardware with the BCI2000 software environment. We have designed an architecture that seamlessly integrates recorded neural signals with device performance metrics, delivering these insights to the care team through a cloud-based interface. In order for future centers-of-excellence to be able to deliver care with clinical BCIs, closed-loop algorithms will need to be able to be dynamically updated without physically interacting with the patient for each adjustment. Our BCI ecosystem is currently being tested with canine subjects, and this manuscript describes how device function (impedance measures) and brain data (LFP signals) were measured daily for an 8 week period following implantation through the cloud interface. Cloud based data synchronization for implantable brain technologies is essential for dynamic re-calibration of reliable and safe BCI and ABS therapies in the clinical setting.

## INTRODUCTION

Recent advancements in the field of implantable neurotechnology have enabled continuous streaming of local field potential data (LFP) spanning years, mainly thanks to rechargeable batteries. Some applications of such devices focus on seizure monitoring and optimizing

epilepsy treatment [1–5], while others focus on brain-computer interface (BCI) applications for people with paralytic disabilities [6].

The CorTec Brain Inter-Change (BIC) device aims to develop an advanced ecosystem with BCI2000 to facilitate chronic data recording to restore functionality in people with paralytic disabilities [6] as well as to deliver adaptive brain stimulation (ABS) as a treatment for severe neurological disorders [7]. Initial developments have demonstrated the utility of such a system and its capability to record clinically relevant data [7, 8].

Worrell and his team have demonstrated the utility of an implantable system for continuous LFP streaming to track primary markers of epilepsy and its comorbidities, including interictal epileptiform discharges, seizures, sleep, cognition, and mood [3, 5, 9–11].

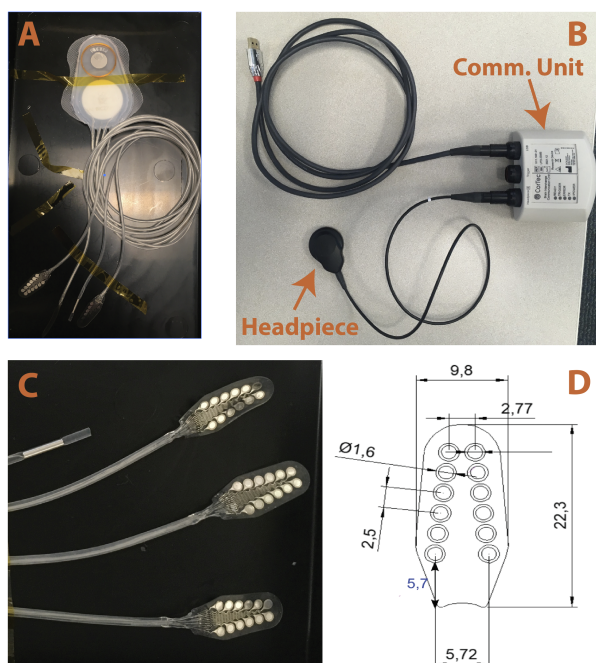
The transfer of LFP data recorded from implantable neurostimulators to the clinical team is a critical task enabling remote supervision of autonomous implantable systems and facilitating the improvement of algorithms for BCI and ABS applications [9].

This work describes progress in the development of an ecosystem for BCI and adaptive neuromodulation in humans by documenting efforts to integrate the CorTec BrainInterchange (BIC) device with BCI2000 [12] and a cloud interface to transfer LFP data from a computer interacting with the implanted device into a research storage compliant with BIDS data structure [13].

## MATERIALS AND METHODS

*Recording System:* The Brain Interchange (BIC) unit (CorTec GmbH) is an implantable battery-less research system with customizable leads, offering up to 32 channels and inductive powering. BIC is capable of simultaneous recording from all 32 channels with a 1 kHz sampling rate and generating stimulation pulses with a maximum frequency of 200 Hz, an amplitude of 6.12 mA, and

a pulse width of up to 2.5 ms. The BIC unit consists of three pieces (Fig. 1): the implanted device, the headpiece for inductive powering of the implanted device, and the communication unit [8].

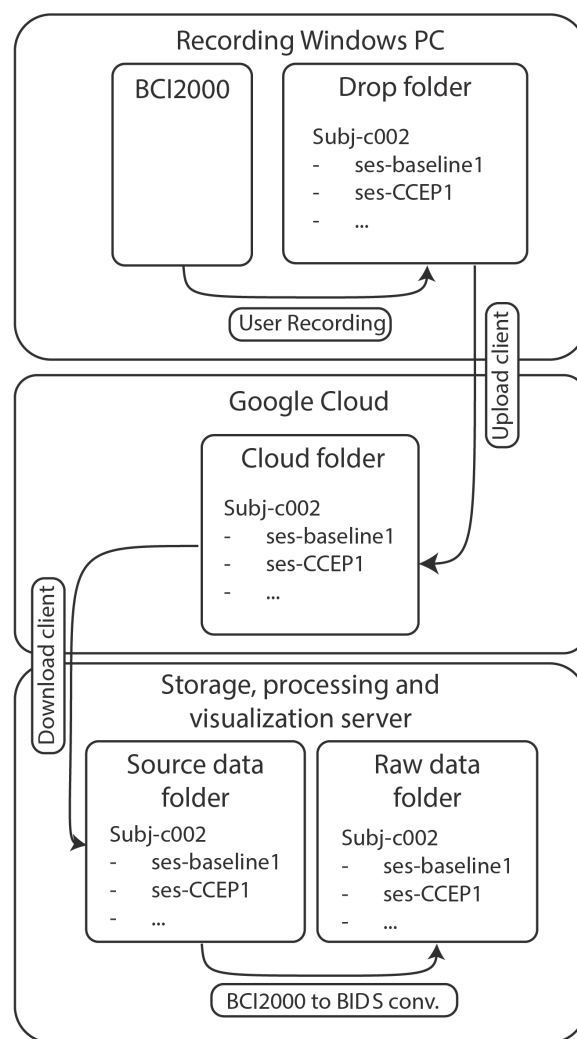


**Figure 1: Experimental CorTec Brain Interchange implantable system.** A) The Brain Interchange device in 3 grid electrode configurations with an additional ground electrode. B) The headpiece connects to the device magnetically and inductively powers the brain implant. The communication unit is plugged into a Windows computer and facilitates wireless communication and real-time data streaming from the implanted device. C) The configuration of the implanted electrodes. D) Electrode dimensions for the electrode with 12 contacts.

**Cloud Synchronization and Data Storage:** Electrophysiology data recorded using the implanted BIC device and BCI2000 ecosystem [12] are manually stored on the Recording Windows Computer in the “Drop directory”. This data is already organized in the subject- and session-oriented file scheme compliantly with the BIDS data structure [13]. Data is automatically transferred over to a cloud storage platform and further synchronized to its final destination, which in our implementation is a research Linux server. The data is automatically converted into MEF3 format and stored in BIDS format with corresponding annotation files (Figure 2). The original raw data are preserved as well in the *sourcedata* folder.

**Ethical Statement:** This research was conducted under Mayo Clinic IACUC protocol A00001713. According to the State of Minnesota statute 135A.191, the canines can be made available for adoption if for any reason the research were to be discontinued. The intent of this animal research is to test and develop a platform for novel human therapeutics.

**Subject:** Canines present a promising translational model for human implantable systems [1, 10]. Dogs



**Figure 2: Data transfer flowchart.** The developed data synchronization system automatically transfers and converts files generated by BCI2000 on the Recording Windows Personal Computer (PC). Brain signals are recorded using the BCI2000 platform connected to the implanted CorTec BIC system. The recorded data is stored in a “Drop Folder” in a subject- and session-oriented storage scheme designed for seamless conversion into the BIDS data format [13]. The data is automatically transferred to a Google Cloud storage as a data transfer platform and subsequently converted into MEF3 format and stored compliantly with the BIDS data organization scheme. The example in this figure represents a scheme for one subject “c002” with two sessions called “baseline1” and “CCEP1”.

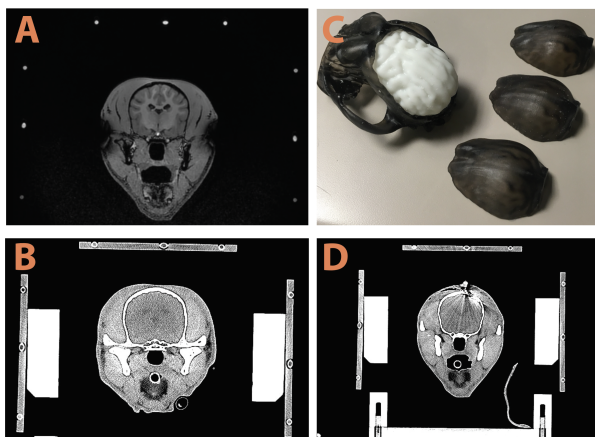
share an evolutionary history with humans and are a promising model for studying behavior, sleep, and neurological disorders [14, 15]. Moreover, canines are large enough to accommodate human-sized electrodes and devices, and canine neurological disorders share features with humans [9, 10].

One adult female intact beagle was housed on a 12/12 light cycle and fed approximately two cups of Lab Diet 5L18, with water provided ad libitum. The animal was housed in temperature-controlled rooms with elevated floors that met all size, material, and sanitation require-

ments according to the Guide for the Care and Use of Laboratory Animals and the Animal Welfare Act. The animal was provided with mats and daily enrichment through assorted treats, chew toys, and human interaction. Animals were socially housed and were assessed daily by a team of veterinarians.

**Implant Surgery:** Pre-surgical magnetic resonance imaging (MRI) - T1-MPRAGE and computed tomography (CT) were utilized to segment brain and skull anatomy. MRI imaging was coregistered to an existing stereotactic atlas of a canine brain [16]. A full-scale replica was 3D-printed using the MRI and CT scans (Figure 3). The 3D brain model aided in pre-surgical planning. The canine was implanted with a research 3-lead CorTec BIC device. The surgery was performed using a stereotactic targeting software, BrainLab, with a custom-made stereotactic frame [10, 17]. Grid electrodes were placed targeting primarily the Sensorimotor, Occipital, and Temporal cortex (Figure 4).

LFP signals were recorded during the surgery under anesthesia and while the canine was waking up to assess the signal quality. Post-surgical CT was utilized to detect individual electrodes.



**Figure 3: Preoperative imaging and 3D print.** **A)** Pre-operative Magnetic Resonance Imaging (MRI) T1-MPRAGE Sequence used for developing a 3D brain model and stereotactic targeting. **B)** Pre-operative Computed Tomography (CT) used for developing a 3D brain model and stereotactic targeting. **C)** A full-scale 3D-printed model aiding in surgery planning. **D)** Post-operative CT with distinct metal artifacts caused by the implanted electrodes.

**Long-term Recording Protocol:** A recording protocol was designed to validate the recording stability of the implanted BIC system. The impedance of all electrodes was measured daily along with a set of three 3-minute recordings, each with a different reference electrode.

**Reproducible Research - Data & Code Sharing:** The authors are committed to sharing data and code to facilitate reproducible research. All codes utilized and developed within this project beyond the BCI2000 ecosystem are publicly available on GitHub

as Python software packages: *Behavioral State Analysis Toolbox (BEST)* ([https://github.com/bnelair/best\\_toolbox](https://github.com/bnelair/best_toolbox)) and *Mef Tools* ([https://github.com/bnelair/mef\\_tools](https://github.com/bnelair/mef_tools)). The data were published as a dataset called *Intracranial recordings using BCI2000 and the CorTec BrainInterchange* on OpenNeuro [18].

## RESULTS

We integrated the CorTec BIC system with BCI2000 with a cloud data synchronization system [1, 12]. The developed data synchronization system automatically transfers LFP recordings from the acquisition computer running BCI2000 to a cloud environment and further to the hospital. We demonstrated that such a system can serve for long-term monitoring of the technical parameters of the implanted system, as well as the LFP recordings.

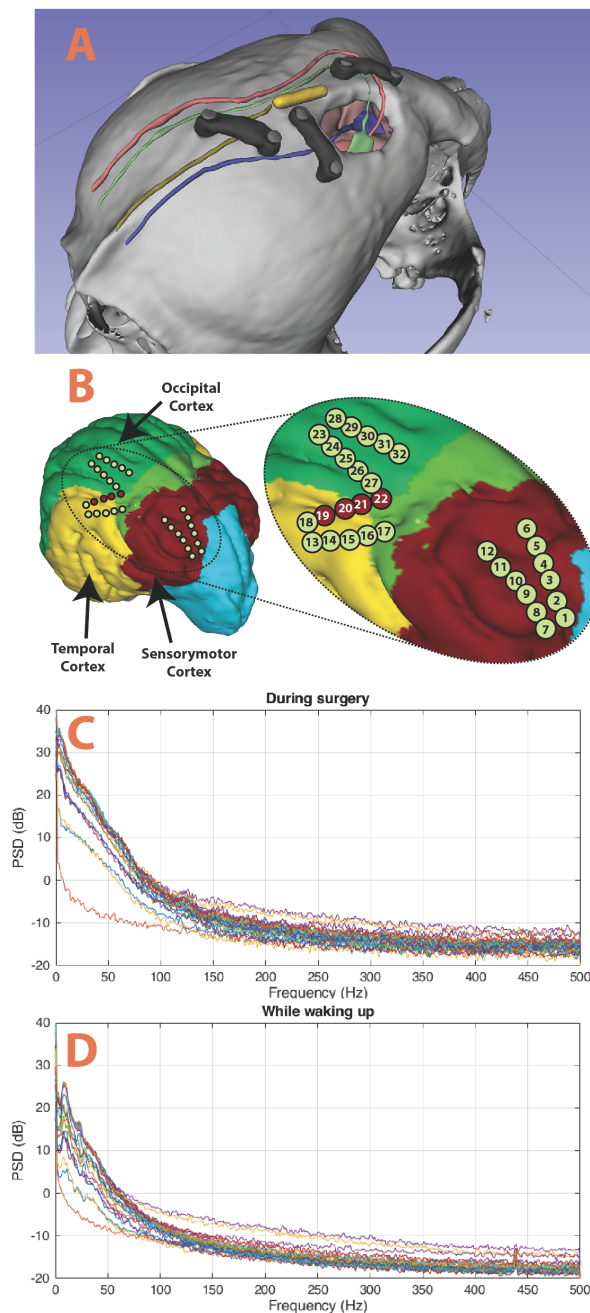
The implanted BIC system has demonstrated sensitivity to LFP changes related to different behavioral states (anesthesia during surgery vs. wakefulness). LFP changes were dominant in lower frequencies below 25 Hz (anesthesia vs. wakefulness: > 30 dB vs. 20 dB) (Fig. 4). Subsequently, we monitored the dog for 40 days after the implant surgery using 3-minute LFP recordings every day along with collecting impedance measurements. The electrical impedance and LFP power stabilized after 20 recording days, which corresponded to an increase in the signal power (Fig. 5). We also identified 4 channels with either permanently high impedance (>5 k $\Omega$ ) or sudden impedance changes. These channels were excluded from subsequent analysis.

## DISCUSSION

Multiple chronic implantable BCI systems are currently being developed with the aim of restoring functionality in paralyzed people and providing ABS for neurological conditions such as epilepsy, movement disorders, stroke, and others [1, 7].

LFPs recorded by an implantable system are a key data source in assessing the system performance and reliability while patients live their lives in a home environment. The performance of any implantable system in BCI or adaptive DBS applications might be significantly impacted by medication-induced LFP changes or high electrode impedance caused by compromised channel integrity. These factors can jeopardize the performance of detection algorithms crucial in BCI and responsive DBS applications.

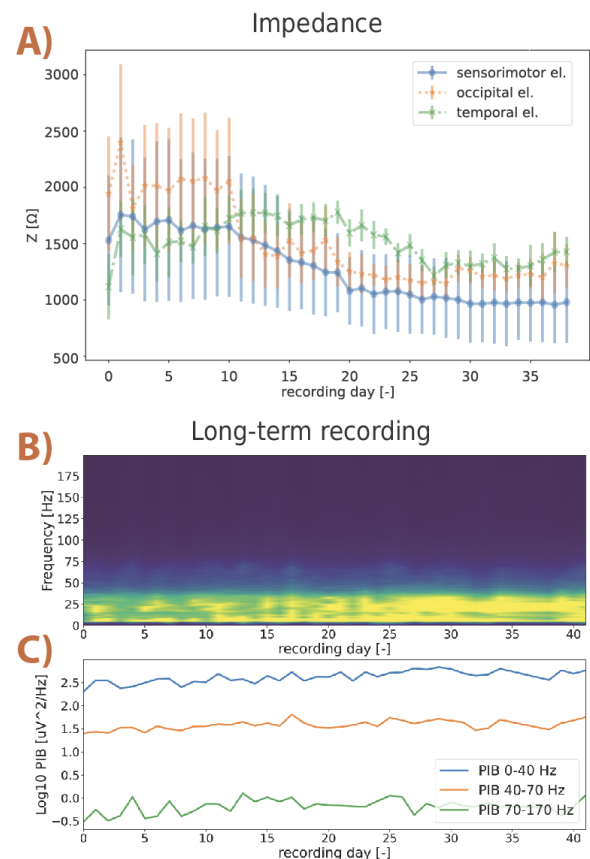
Timely delivery of LFP signals and technical health information from the implanted systems from a patient home environment to a clinical team with minimal patient and physician effort is, therefore, an important task to ensure continuous safety and efficacy of the



**Figure 4: Post-surgical rendering and LFP recording.** **A)** A surface rendering of the leads secured to the skull. **B)** A visualization of all 32 channels distributed among 3 leads placed over Sensorimotor Cortex, Occipital Cortex and Temporal Cortex [16]. Red color symbolizes electrodes with impedance  $> 5 \text{ k}\Omega$  at any time of chronic recording. **C)** A power spectrum density (PSD) of the iEEG signals recorded during deep anesthesia. **D)** iEEG PSD for the canine waking up from the anesthesia manifests different power in low frequencies compared to deep anesthesia.

implantable system. Data of interest can include LFP recordings, electrode impedance, battery level, amount of delivered stimulation, and any other data that might be beneficial to monitor [1, 2].

Here, we integrated an experimental BCI system (CorTec



**Figure 5: Impedance and recording demonstrated over the first 40 recording days.** **A)** Impedance average calculated over all channels for each implanted lead (sensorimotor cortex, occipital cortex, and temporal cortex) decreases and stabilizes over the first 10-20 days after the implant. **B)** Power spectrum density mildly increases over time corresponding to the impedance decrease. **C)** Gradual power increase demonstrated using power in the band (PIB) visualization for bands 0-40 Hz, 40-70 Hz, and 70-170 Hz.

BIC) with an open-source software platform for non-commercial use, BCI2000, which integrates multiple recording and stimulating systems to record, process data, and stimulate using a single platform. Recorded data is automatically synchronized over a cloud interface into a BIDS structured data storage. We demonstrated the feasibility of such a system in one canine implanted with CorTec BIC that was monitored daily for 40 days after the implant surgery. The utility of the proposed system and the need for the collection of technical and LFP data from implanted devices were demonstrated by revealing and identifying faulty channels and monitoring changes in the signal power related to impedance stabilization after the implant.

The cloud-based interface allowed us, in simulation of a remote interaction, to identify sources of packet loss, including A) sensitivity of the Communication Unit orientation towards the implant with a data dropout rate ranging from 1-20%, depending on orientation; B) Interference with surrounding wireless devices. Both of these

shortcomings can be overcome by careful experiment design and/or by implementing a multi-directional antenna in the Communication Unit. Initial experiments show automatic reconnecting of the BCI2000 to the implanted BIC can significantly improve data transmission. The BIC is powered inductively, and the headpiece magnet can be a point of connection loss during extended periods of time, exacerbated by animal movement. Notably, the canine represents a particularly challenging subject due to the range and frequency of movements. The system can easily be improved but will likely be adequate for applications in people with restricted movement, e.g., the locked-in syndrome [6].

## CONCLUSION

BCI systems have the potential to improve the quality of life for many patients with neurological disease, but therapeutic interventions will require scalability beyond the attention of technical experts directly interacting with the patient. Cloud-based interfaces like the novel one that we demonstrate here, will enable trained technicians to interact with patient devices from a distance. We demonstrate regular interaction with our nascent ecosystem for eight consecutive weeks post implant, without disruption.

## ACKNOWLEDGEMENTS

We are grateful to be able to work with this noble animal Belka, and for the care and attention provided by the Veterinary staff at Mayo Clinic. Belka will be available for adoption at the conclusion of the research in accordance with the Minnesota state Beagle Freedom Bill. This work was supported by the NIH U01-NS128612 (KJM, GAW, PB). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Our founder played no role in data collection and analysis, study design, decision to publish, or manuscript preparation.

## REFERENCES

- [1] Sladky V *et al.* Integrated human-machine interface for closed-loop stimulation using implanted and wearable devices. 2022;1730–1735.
- [2] Sladky V *et al.* Integrated human-machine interface for closed-loop stimulation using implanted and wearable devices. IEEE, Oct. 2022, 1730–1735.
- [3] Mivalt F *et al.* Electrical brain stimulation and continuous behavioral state tracking in ambulatory humans. *Journal of Neural Engineering*. 2022;19:016019.
- [4] Mivalt F *et al.* Impedance rhythms in human limbic system. *The Journal of Neuroscience*. 2023;43:6653–6666.
- [5] Gregg NM *et al.* Thalamic deep brain stimulation modulates cycles of seizure risk in epilepsy. *Scientific Reports*. 2021;11.
- [6] Vansteensel MJ *et al.* Fully implanted brain-computer interface in a locked-in patient with ALS. *New England Journal of Medicine*. 2016;375:2060–2066.
- [7] Schalk G *et al.* Toward a fully implantable ecosystem for adaptive neuromodulation in humans: Preliminary experience with the cortex braininterchange device in a canine model. *Frontiers in Neuroscience*. 2022;16:932782.
- [8] Ayyoubi AH *et al.* Benchmarking signal quality and spatiotemporal distribution of interictal spikes in prolonged human EEG recordings using cortex wireless brain interchange. *Scientific Reports*. 2024;14(1).
- [9] Sladky V *et al.* Distributed brain co-processor for tracking spikes, seizures and behaviour during electrical brain stimulation. *Brain Communications*. 2022;4.
- [10] Kremen V *et al.* Integrating brain implants with local and distributed computing devices: A next generation epilepsy management system. *IEEE Journal of Translational Engineering in Health and Medicine*. 2018;6:1–12.
- [11] Balzekas I *et al.* Invasive electrophysiology for circuit discovery and study of comorbid psychiatric disorders in patients with epilepsy: Challenges, opportunities, and novel technologies. *Frontiers in Human Neuroscience*. 2021;15.
- [12] Schalk G, McFarland D, Hinterberger T, Birbaumer N, Wolpaw J. Bci2000: A general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*. 2004;51(6):1034–1043.
- [13] Gorgolewski KJ *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*. 2016;3:160044.
- [14] Berendt M, Høgenhaven H, Flagstad A, Dam M. Electroencephalography in dogs with epilepsy: Similarities between human and canine findings. *Acta Neurologica Scandinavica*. 1999;99:276–283.
- [15] Hare B, Brown M, Williamson C, Tomasello M. The domestication of social cognition in dogs. *Science*. 2002;298:1634–1636.
- [16] Johnson PJ *et al.* Stereotactic cortical atlas of the domestic canine brain. *Scientific Reports*. 2020;10(1).
- [17] Mivalt F *et al.* Automated sleep classification with chronic neural implants in freely behaving canines. *Journal of Neural Engineering*. 2023;20:046025.
- [18] Mivalt F *et al.* Intracranial recordings using bci2000 and the cortex braininterchange. *OpenNeuro*. 2023.

## SPATIAL AND SPECTRAL CHANGES IN CORTICAL POTENTIALS DURING PINCHING VERSUS THUMB AND INDEX FINGER FLEXION

Panagiotis Kerezoudis<sup>1,2</sup>; Michael A Jensen<sup>1,3</sup>; Harvey Huang<sup>3</sup>; Jeffrey G. Ojemann<sup>4</sup>;  
Bryan T. Klassen<sup>5</sup>; Nuri F. Ince<sup>2,6</sup>; Dora Hermes<sup>6</sup>; Kai J Miller<sup>2,6</sup>

<sup>1</sup>Division of Neuroscience, Mayo Graduate School of Biomedical Sciences, Rochester, MN, USA

<sup>2</sup>Department of Neurologic Surgery, Mayo Clinic, Rochester, MN, USA

<sup>3</sup>Medical Scientist Training Program, Mayo Clinic, Rochester, MN, USA

<sup>4</sup>Department of Neurosurgery, Seattle Children's Hospital, Seattle, Washington

<sup>5</sup>Department of Neurology, Mayo Clinic, Rochester, MN, USA

<sup>6</sup>Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN, USA.

E-mail: kerezoudis.panagiotis@mayo.edu

**ABSTRACT:** We analyzed the electrocorticographic signals of 3 patients implanted with subdural electrode arrays for identification of seizure foci. Patients performed screen cue-based flexion movement of the thumb or index finger, or a pinch movement of both. Broadband power changes with each movement were estimated. Topological maps for each type of movement were created using co-registered brain renderings, and the overlap in spatial extent was quantified using a resampling metric. Activation during pinching was compared with thumb flex, index flex and three composite metrics: the maximum, the geometric mean and a modified geometric mean. Significant increase in broadband power was observed in all three patients when pinch (max signed  $r^2$  0.64-0.80-0.89), thumb flexion (0.63-0.72-0.89) or index flexion (0.66-0.67-0.83) was performed compared to rest. Spatial overlap was highest between pinching and index flexion (69-87%, all  $p < .001$ ), followed by the modified geometric mean (61-96%, all  $p < .001$ ), while thumb flexion had the lowest overlap of all analyzed metrics (27-77%, significant in 2/3).

### INTRODUCTION

The hallmark of human motor behavior is the ability to perform fine individual finger movements. Dexterous hand gestures are essential to carrying out activities of daily living and can be considered as a combination of individual finger movements with varying degrees of freedom, particularly flexion and extension. Interestingly, certain moves, including reaching, grasping (palmar or pincer), transporting, and placing have been previously proposed as motor primitives that represent building blocks to execute an overall goal-directed action, such as a monkey reaching for a seed, grasping it, bringing it towards the mouth and finally eating it [1].

Whether the neural representation of composite moves is equal to the linear summation of their constituents

(e.g. palmar grasp vs concurrent all-finger flexion) or non-linear and therefore, not easily predictable, remains unknown and it is an area of ongoing electrophysiology and imaging research [2,3]. Recordings of the human brain cortical surface the last 30-40 years have revealed movement-related signals over the motor and sensory regions with high accuracy [4-7]. Specifically, electrocorticography (ECoG) experiments have demonstrated that hand and tongue movement are associated with spatially focal but spectrally broadband changes in the frontoparietal cortex [6,8]. Distinct individual finger movements have been identified with electrode spacing resolution as low as 1cm using subdural recordings [9].

The spectral and spatial relationship between individual finger movement and synergistic action has not yet been fully elucidated, and the extent to which grasping movements represent sums of the individual finger moves or activate different brain areas, e.g. premotor cortex, (posterior) parietal cortex etc. Herein, we report our findings in 3 young patients with medically refractory epilepsy that underwent invasive ECoG recordings as part of their seizure workup. We demonstrate the spatially focal spectral changes that occur with thumb flexion, index flexion and pinching and compare these three movement types in their spatial overlap.

### MATERIALS AND METHODS

*Ethics statement:* All patients participated in a purely voluntary manner and provided informed consent, based on protocols approved by the Institutional Review Board of the University of Washington, Seattle.

*Patient population:* We recruited 3 patients (2 females, aged 18, 21 and 19 years) that underwent invasive clinical EEG monitoring for localization of seizure foci as part of their workup for medically refractory epilepsy using subdural ECoG grids.

*Electrical recordings:* The platinum ECoG arrays were configured as a combination of grid (8x8 or 4x8) arrays and strip arrays, numbering a total of 32 to 81 contacts (i.e. channels). The diameter of the electrode contacts was 4mm and the inter-electrode distance was 10mm. These arrays were surgically placed by one of the senior authors (J.G.O). The ECoG signals were split into two identical sets, one towards the clinical EEG system (XLTEK, Oakville, Ontario, Canada) and the research set was recorded using the Synamps2 (Neuroscan, El Paso, TX) biosignal amplifiers at 1 kHz with a bandpass-filter at 0.3-200 Hz. Finger position was recorded using a sensor dataglove (5DT, Irvine, CA). The general purpose BCI2000 software was used for stimulus presentation and ECoG signal data collection [10].

*Cortical rendering and electrode localization:* The cortical surface from a preoperative MRI was rendered using either Freesurfer or Spm5 software in order to determine the relationship between the gyral anatomy and electrode position. The electrode positions were calculated after co-registering the post-operative computed tomography (CT) to the pre-operative MRI using the CTMR package by Hermes et al., 2010, which has been shown to accurately localize the electrode positions within a ~4mm error [11].

*Movement tasks:* Subjects were presented with a word cue displayed on a bedside monitor to perform the following self-paced movement tasks, each on a separate run: 1) pinching move between the thumb and the index finger or 2) individual thumb or index finger flexion. A 2-second rest trial (blank screen) followed each movement trial. There were up to 30 cues for each movement type. Segmentation of movement vs rest period was performed based on visual inspection of the data glove trace and marking the initiation and termination of movement. A total of 10-15 movement/rest trials were completed for each task.

*Signal pre-processing:* The potential measured at each electrode was re-referenced with respect to the common average of all electrodes. Electrodes significant for artifact or epileptiform activity were visually inspected, removed from re-referencing and omitted prior to further analysis.

*Power spectral snapshots:* Power spectral snapshots using Welch's method and a Hann window of 1s were generated for the entire experiment in each electrode as well as separately for movement and rest blocks. In addition, we mean-normalized the log power spectra by dividing the power spectrum for each trial with the mean electrode power spectrum. The frequency range of interest included 1-200 Hz.

*Dynamic power spectrum:* We generated time-frequency approximations after convolving the voltage time-series data with a Morlet wavelet (10 cycles) in order to estimate the amplitude and phase of the signal across the frequency range for every point in time.

*Power spectrum decoupling:* We performed decoupling of the power spectra using a previously established method by Miller et al, 2009 [12]. In

summary, we performed Principal Component Analysis (i.e. eigendecomposition) of the power spectral density-covariance matrices (freq [f], chan [k], trials [q]). The resultant eigenvectors (principal spectral components-PSCs) reveal which frequencies vary in power together and are ordered according to the value of the corresponding eigenvalue. It has been shown that the 1<sup>st</sup> PSC corresponds to broadband spectra (i.e. spanning all frequency ranges), whereas the 2<sup>nd</sup>-4<sup>th</sup> PSCs typically capture rhythmic power spectral phenomena (i.e. mu rhythm, alpha rhythm etc). In addition, the projection of the dynamic spectrum onto the 1<sup>st</sup> PSC will yield the logarithm of the time course of the power spectrum power law coefficient  $P(f, t) = A(t) \cdot f^\alpha$  [13]. This logarithm was subsequently smoothed with a Gaussian window (250 ms), z-scored and re-exponentiated in order to obtain the broadband trace plotted in Fig. 1. This broadband power time course has been shown to be a robust estimate of behaviorally relevant local cortical activity as well as predictive of the finger movement time course, as measured by the data glove [5]. Finally, we calculated the signed  $r^2$  cross-correlation values from broadband activity, between each movement modality (pinching, thumb flexion, or finger flexion) and rest. Movement trials were compared with the rest trials that followed the same movement type [14]. Thus,  $r^2$  represents the percentage in variance in the joint distribution of movement-rest blocks, that can be explained by a difference in the movement and rest trial means, respectively.

$$r^2 = \frac{(\bar{m} - \bar{r})^2}{|\bar{m} - \bar{r}| \sigma_{mUr}^2} \frac{N_m \cdot N_r}{N_{mUr}^2}$$

*Quantifying spatial overlap:* The spatial extent of overlap and degree of change in broadband power activity was quantified and compared using a resampling metric. In summary, for measures of type  $X_n$  and  $Y_n$  (n denoting the electrode), the true overlap ( $O^{T_{XY}}$ ) was calculated from the pairwise dot product of the broadband  $r^2$  value. The spatial overlap metric was calculated by dividing the true overlap with the maximum possible overlap ( $O^{M_{XY}}$ ), by computing the dot product of distributions  $X_n$  and  $Y_n$  assorted in ascending order. The electrodes were subsequently scrambled  $10^6$  times ( $n \rightarrow m$ ) and a surrogate overlap was obtained from each permutation ( $O^{S_{XY}}$ ). Finally, p-value for statistical significance was estimated as the percentage of surrogate distributions  $O^{S_{XY}}$  that were greater than  $O^{T_{XY}}$  (or the percentage that were less in case  $O^{T_{XY}} < 0$ ). The methodology is graphically displayed in Fig. 2. The  $r^2$  electrode activation maps of the following conditions were compared with pinching [summarized in Table 1]: finger flexion, thumb flexion, their maximum value per channel, their geometric mean and modified geometric mean.

Table 1. Summary of metrics used for the analysis of spatial overlap with pinching.

Condition	Rationale
Thumb flexion $r_{thb}^2$	Constituents of pinching movement.
Index flexion $r_{idx}^2$	
Maximum value of thumb flex, index flex $\max\{r_{thb}^2, r_{idx}^2\}$	Activation reflects the maximal activity evoked by each individual finger movement (i.e. the union of two activation sets).
Geometric mean of thumb flex and index flex $\sqrt{r_{thb}^2 * r_{idx}^2}$	Activation reflects the average activity evoked by each individual finger movement (i.e. the intersection of two activation sets).
Modified geometric mean of thumb flex and index flex $1 - \sqrt{(1 - r_{thb}^2) * (1 - r_{idx}^2)}$	Activation reflects the average activity evoked by each individual finger movement (i.e. the complement of the intersection of the non-activation sets).

## RESULTS

*Spectral changes related to movement:* Across all 3 subjects, we observed significant increase in broadband power in the higher frequencies with movement and decrease in power in narrowband, lower-frequency oscillations (Fig. 1). Specifically, the maximum channel  $r^2$  values associated with movement were 0.63/0.72/0.89 for thumb flexion, 0.66/0.67/0.83 for index finger flexion and 0.64/0.80/0.89 for pinching. Broadband power increases following each movement type were more spatially focused, whereas motor beta rhythm (12-20 Hz) power decrease was more widely distributed over the cortical surface. Following decoupling of the power spectrum as described above, we projected the dynamic spectrum to the first eigenvector and recreated the log(broadband) timeseries. Following smoothing with a Gaussian window, z-scoring and re-exponentiation, we found that the time course of broadband activation matched finger movement with high accuracy.

*Somatotopy of movement-related spectral changes:* The  $r^2$  activation values corresponding to each channel in the subdural grid were plotted on the 3D brain renderings for each individual finger movement type as well as their composite metrics (Fig. 3). In all three patients, spatial overlap between pinch and index flexion (69-87%, all  $p < .001$ ) was considerably higher compared to pinch and thumb flexion (27-77%; significant in 2/3 patients). Furthermore, the modified

geometric mean metric performed similarly or better compared to the index flex (61-96%), while the max (39-89%) and geometric mean (68-81%) metrics ranked in between.

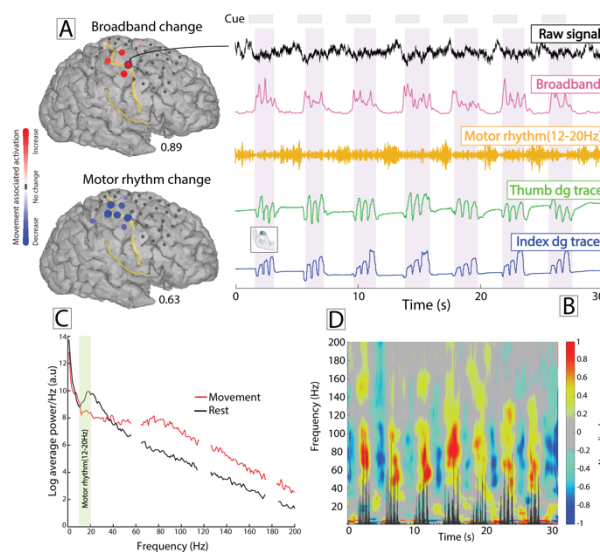


Figure 1. Electroencephalographic (EECoG) activity in the right frontoparietal cortex during pinch movement and rest (subject 3). (A) Topological maps demonstrating focal broadband power increase and more widespread motor rhythm (12-20 Hz) power decrease. (B) Spectral changes demonstrate excellent correlation with finger movement. (C) Power spectral density plot from the entire behavioral task showing decrease in low-frequency oscillations and broadband increase across the rest of the frequency range, during movement (red) compared to rest (black). (D) Time-varying power spectral density plot during the same period of recording. The black trace inset represents the 1st derivative of thumb movement.

## DISCUSSION

In the present study, we analyzed the cortical activity in three patients implanted with subdural electrode arrays during three types of finger movement, i.e. thumb flexion, index flexion and pinching, and compared their spatial overlap in broadband activation. Overall, we observed the spatial overlap between pinching and index flexion to be significantly higher compared to thumb flexion, reflecting a considerably smaller digit movement of the latter during the combined motor behavior. This aligns with the results by Cramer et al., who demonstrated that the degree of muscle activation, such as force of squeezing, correlates with the volume of activated cortical tissue as measured by fMRI bold signal [15].

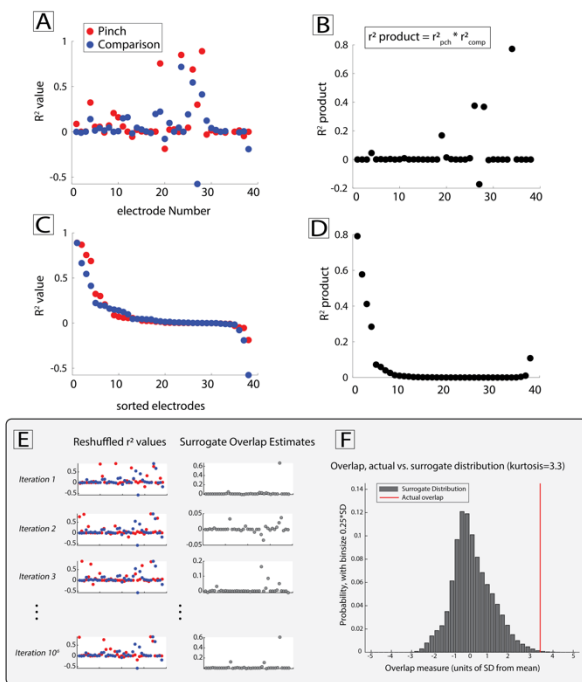


Figure 2. Demonstration of the method for quantifying the overlap in spatial distribution. (A) The  $r^2$  value, as a metric of activation during movement vs rest, is computed and plotted for each electrode, during pinching versus thumb flexion, index flexion or one of the other metrics we used (collectively labeled here as ‘comparison’). (B) The overlap can be quantified as the dot product across electrodes, thereby denoting that activities are paired together at each electrode. (C) We can probe all of the possible configurations of activation values by reshuffling the electrode positions (left panels) for one of the conditions (pinch) and recalculating the dot products (right panels). This is done  $10^6$  times, generating the distribution of surrogate dot products. (D) The generated surrogate distribution can be used to estimate a p-value for the significance of the overlap.

To date, the majority of work investigating neuronal representation of multi-task computations is derived from simulated models, non-human primate recordings or human fMRI experiments [16,17]. Evarts’ pioneering work demonstrated linear relation of single-unit neuronal activity with forces and torques about the wrist [18]. Later studies refuted this notion, supporting altered gain of this relation depending on the size of inertial load and the anticipated range of force.[19,20] For example, Shah et al. recently reported their findings from recorded motor activity from high-density cortical arrays in two patients with tetraplegia that performed single, pairwise and higher order finger movements [2]. Given the subjects’ medical history, kinematics was captured and analyzed using an animated hand (Unity Software, Unity Technologies, San Francisco). They observed that neural activity associated with multiple finger movements aligned with a ‘pseudo-linear’ summation of individual finger activities, i.e. had lower

magnitude, which authors attributed to vector unit-normalization.

Chestek et al. classified different hand posture ECoG signals with subdural micro- and macro-electrodes implanted over the primary sensorimotor areas [21]. The authors examined 4 finger movements and 4 distinct isometric hand postures (pinch, fist, splay and point). Similar to our findings, they observed significant co-modulation of the thumb electrode (up to 93% likelihood) for index, and observed index finger flexion and pinching to be similar movements that could be decoded at 79% correct based on ‘gamma-band’ power (66-114Hz) [21]. Our study build further upon that work by delving into the topological representation of composite vs individual finger movements.

Recent work in monkeys has revealed similar findings. Naufel et al. investigated the relation between primary motor cortex (M1) and muscle activity (as captured by electromyography, EMG) during three wrist movement tasks of varying dynamical load and forces [3]. The authors observed that a greater proportion of EMG changes were explained by a linear model of M1 activity supplemented by a condition-dependent gain. In our analysis, the non-linear metric of modified geometric mean explained a greater proportion of spatial overlap compared to maximum value and geometric mean. Nonetheless, the association between biological underpinnings of more complex limb movements and nonlinear dynamics remains to be determined; potential explanatory phenomena may include modulation of cortical activity (M1, premotor etc) by subcortical structures and differential inhibitory interneuron activity.

*Implications for future research:* These findings have implications for brain-computer interface applications. One common limitation of motor-controlled brain-computer interface (BCI) in non-human primates is the suboptimal coupling between neural activity and movement output [22]. Subsequent iterations of cortically-based BCI designs should take into consideration the strength and the spatiotemporal patterns of neural signals and how these relate to muscular activity. Future research can further expand on our work to investigate how individual finger actions relate to and overlap with other complex gestures, such as palmar grasping, squeezing and pointing.

*Limitations:* First, our sample consisted only of 3 subjects. Second, we employed standard-scale clinical ECoG, rather than high-density research grids; the latter have been shown to have superior spatial performance in 6 elementary movements in the alpha, beta and gamma frequency bands [21]. Third, we were unable to incorporate kinetics and/or kinematics in our analysis to correlate with the ECoG signal. The data glove captures mean deformation of an electrically conductive elastomer with piezoresistive properties but does not allow for detailed analysis of force of contraction or degree of displacement/range of motion in the interphalangeal joints.

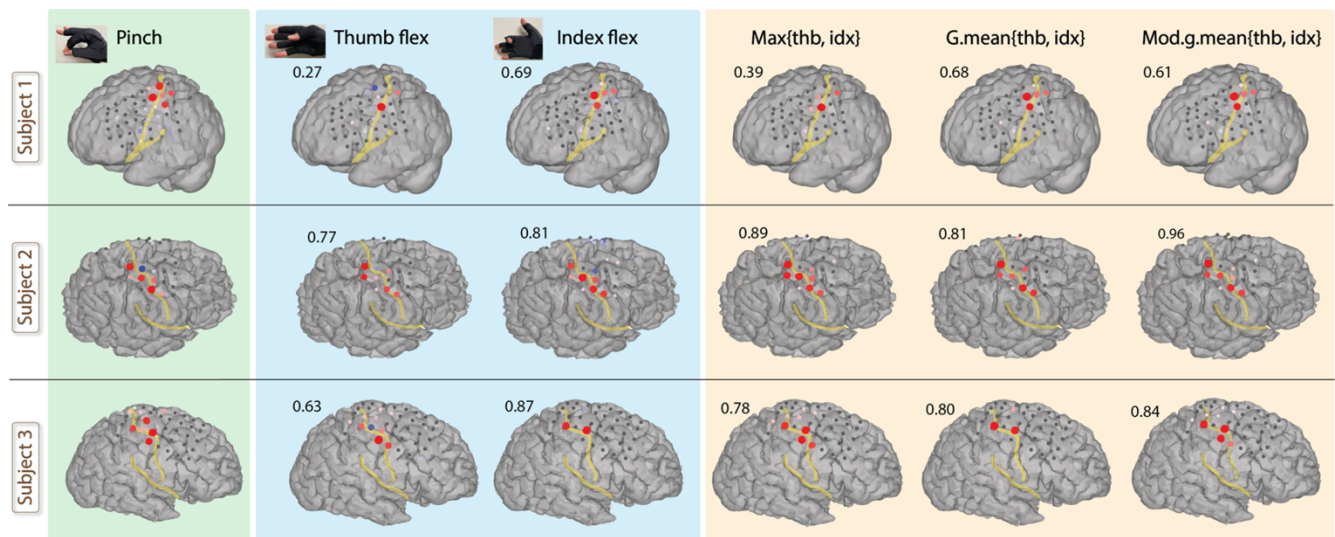


Figure 3. Topological maps in each subject demonstrating changes in broadband power at different cortical sites for pinching (green panel), thumb and index finger flexion (blue panel) and three combined  $r^2$  metrics (orange panel): the maximum, the geometric mean and the modified geometric mean; please refer to Methods for detailed description. The spatial overlap of each condition with pinching is shown on the top left corner of the brain renderings. Overall, it can be appreciated that index flexion, followed by the modified geometric mean, had the highest overlap with pinching.

## CONCLUSION

For pinch, the ECoG signal seems to comprise a combination of the signals from the individual thumb and index movements, with considerably larger overlap with the latter. This analysis may provide insight into the tuning of the motor cortex toward specific types of motor behaviors and help refine currently available BCI control algorithms and their therapeutic efficacy.

## ACKNOWLEDGEMENTS

All patients participated in a purely voluntary manner, after providing informed written consent, under experimental protocols approved by the Institutional Review Board of the University of Washington (#12193). All patient data was anonymized according to IRB protocol, in accordance with HIPAA mandate. It was made available through the library described in “A Library of Human Electroencephalographic Data and Analyses” by Kai Miller [Miller, Kai J. “A library of human electroencephalographic data and analyses.” *Nature human behaviour* 3.11 (2019): 1225-1235]. This work was also supported by the National Institutes of Health (NIH) NINDS U01-NS1286.

## REFERENCES

- [1] Hemeren PE, Thill S. Deriving motor primitives through action segmentation. *Front Psychol*. 2010;1:243.
- [2] Shah NP, Avansino D, Kamdar F, Nicolas C, Kapitonava A, Vargas-Irwin C, et al. Pseudo-linear Summation explains Neural Geometry of Multi-finger Movements in Human Premotor Cortex. *bioRxiv* [Internet]. 2023 Oct 12; Available from: <http://dx.doi.org/10.1101/2023.10.11.561982>
- [3] Naufel S, Glaser JJ, Kording KP, Perreault EJ, Miller LE. A muscle-activity-dependent gain between motor cortex and EMG. *J Neurophysiol*. 2019 Jan 1;121(1):61–73.
- [4] Crone NE, Miglioretti DL, Gordon B, Sieracki JM, Wilson MT, Uematsu S, et al. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization. *Brain*. 1998 Dec;121 ( Pt 12):2271–99.
- [5] Miller KJ, Hermes D, Honey CJ, Hebb AO, Ramsey NF, Knight RT, et al. Human motor cortical activity is selectively phase-entrained on underlying rhythms. *PLoS Comput Biol*. 2012 Sep 6;8(9):e1002655.
- [6] Miller KJ, Honey CJ, Hermes D, Rao RPN, denNijs M, Ojemann JG. Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *Neuroimage*. 2014 Jan 15;85 Pt 2:711–20.
- [7] Crone NE, Miglioretti DL, Gordon B, Lesser RP. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain*. 1998 Dec;121 ( Pt 12):2301–15.
- [8] Miller KJ, Leuthardt EC, Schalk G, Rao RPN, Anderson NR, Moran DW, et al. Spectral changes in cortical surface potentials during motor movement. *J Neurosci*. 2007 Feb 28;27(9):2424–32.
- [9] Schalk G, Miller KJ, Anderson NR, Wilson JA, Smyth MD, Ojemann JG, et al. Two-dimensional movement control using electrocorticographic signals in humans. *J Neural Eng*. 2008 Mar;5(1):75–84.
- [10] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-

- purpose brain-computer interface (BCI) system. *IEEE Trans Biomed Eng.* 2004 Jun;51(6):1034–43.
- [11] Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF. Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J Neurosci Methods.* 2010 Jan 15;185(2):293–8.
- [12] Miller KJ, Zanos S, Fetz EE, den Nijs M, Ojemann JG. Decoupling the cortical power spectrum reveals real-time representation of individual finger movements in humans. *J Neurosci.* 2009 Mar 11;29(10):3132–7.
- [13] Miller KJ, Sorensen LB, Ojemann JG, den Nijs M. Power-law scaling in the brain surface electric potential. *PLoS Comput Biol.* 2009 Dec;5(12):e1000609.
- [14] Jensen MA, Huang H, Valencia GO, Klassen BT, van den Boom MA, Kaufmann TJ, et al. A motor association area in the depths of the central sulcus. *Nat Neurosci.* 2023 Jul;26(7):1165–9.
- [15] Cramer SC, Weisskoff RM, Schaechter JD, Nelles G, Foley M, Finklestein SP, et al. Motor cortex activation is related to force of squeezing. *Hum Brain Mapp.* 2002 Aug;16(4):197–205.
- [16] Driscoll L, Shenoy K, Sussillo D. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs [Internet]. bioRxiv. 2022 [cited 2024 Jan 15]. p. 2022.08.15.503870. Available from: <https://www.biorxiv.org/content/10.1101/2022.08.15.503870>
- [17] Ito T, Murray JD. Multitask representations in the human cortex transform along a sensory-to-motor hierarchy. *Nat Neurosci.* 2023 Feb;26(2):306–15.
- [18] Evarts EV. Relation of pyramidal tract activity to force exerted during voluntary movement. *J Neurophysiol.* 1968 Jan;31(1):14–27.
- [19] Humphrey DR. Relating motor cortex spike trains to measures of motor performance. *Brain Res.* 1972 May 12;40(1):7–18.
- [20] Hepp-Reymond M, Kirkpatrick-Tanner M, Gabernet L, Qi HX, Weber B. Context-dependent force coding in motor and premotor cortical areas. *Exp Brain Res.* 1999 Sep;128(1-2):123–33.
- [21] Chestek CA, Gilja V, Blabe CH, Foster BL, Shenoy KV, Parvizi J, et al. Hand posture classification using electrocorticography signals in the gamma band over human sensorimotor brain areas. *J Neural Eng.* 2013 Apr;10(2):026002.
- [22] Kuo CH, Blakely TM, Wander JD, Sarma D, Wu J, Casimo K, et al. Context-dependent relationship in high-resolution micro-ECoG studies during finger movements. *J Neurosurg.* 2019 Apr 26;132(5):1358–66.

# PASSIVE OLFACTORY BRAIN-COMPUTER INTERFACE PARADIGM FOR AWARENESS LEVEL PREDICTION

Tomasz M. Rutkowski<sup>1,2,3</sup>, Hubert Kasprzak<sup>2</sup>, Nina Niewińska<sup>2</sup>,  
Mihoko Otake-Matsuura<sup>1</sup>, Tomasz Komendziński<sup>2</sup>

<sup>1</sup>RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan

<sup>2</sup>Department of Cognitive Science, Nicolaus Copernicus University, Toruń, Poland

<sup>3</sup>Graduate School of Education, The University of Tokyo, Tokyo, Japan

E-mail: tomasz.rutkowski@riken.jp & tkomen@umk.pl

**ABSTRACT:** The sense of smell, also known as olfaction, can improve the usability of brain-computer interfaces (BCIs) and support passive modalities for monitoring cognitive states. In reactive BCI, users can assign specific scents to commands for natural interaction, while a passive application can monitor cognition. However, some challenges still need to be addressed, such as the need for accurate odor delivery systems and robust algorithms for detecting and interpreting brain activity patterns. We propose combining electroencephalogram (EEG) and electrobulbogram (EBG) in an olfactory modality oddball paradigm to predict a user's awareness level. Our pilot study indicates promising results for a new passive olfactory BCI modality combining CSP filtration and awareness level classification.

## INTRODUCTION

Recent reports suggest a link between COVID-19 and loss of smell, also known as olfactory dysfunction [1]. Mounting evidence suggests that this condition may be an early symptom of Alzheimer's or Parkinson's syndromes [2, 3]. A new method of objectively measuring olfactory bulb (OB) activity named electrobulbogram (EBG) has been proposed [4], utilizing standard EEG amplifiers by placing electrodes above eyebrows to evaluate olfaction-related brain activity.

According to a recent study, olfactory sleep stimulation may improve cognitive and memory performance in the elderly, providing a potential intervention to protect against Alzheimer's syndrome [5]. As olfactory neuroscience and applied neurotechnology gain interest, their potential for use in BCI becomes more appealing [6]. The state-of-the-art visual [7], auditory [8], and tactile [9] modalities have been successfully implemented in BCI, and the olfactory modality could represent the next frontier.

Sensory awareness focuses on a specific sensory detail rather than simply responding to stimuli. In the context of olfactory awareness, it refers to the ability to distinguish a target odor in an oddball stimulus presentation. When evaluating the awareness level, we predict the subject's



Figure 1: An experimental setting with a subject wearing an EEG cap in front of an olfactometer delivering scent stimuli.

ability to differentiate between odors in a binary setting, i.e., whether they are above or below the chance level of half (three out of six) target stimulants.

The authors recently conducted a pilot study to test a new form of olfactory stimulation. The study recruited ten healthy and BCI-naïve participants and used wearable neurotechnology to capture their brainwave (EEG) and OB sensory activity (EBG). During the EEG and EBG preprocessing stages, we used the typical spatial pattern (CSP) filtering technique [10, 11] to extract features from the signals. Specifically, we focused on the gamma frequency band (which ranges from 35 to 100 Hz). This frequency band is known to carry the most meaningful olfactory responses, as reported in previous studies [4]. We then compared the classification accuracy of EEG-only, EBG-only, and combined EEG+EBG CSP features using various classifiers.

The paper is organized as follows: First, we will introduce the experimental conditions, signal preprocessing with CSP feature extractions, and classification techniques in the materials and methods section. Then, we will present preliminary results from pilot experiments

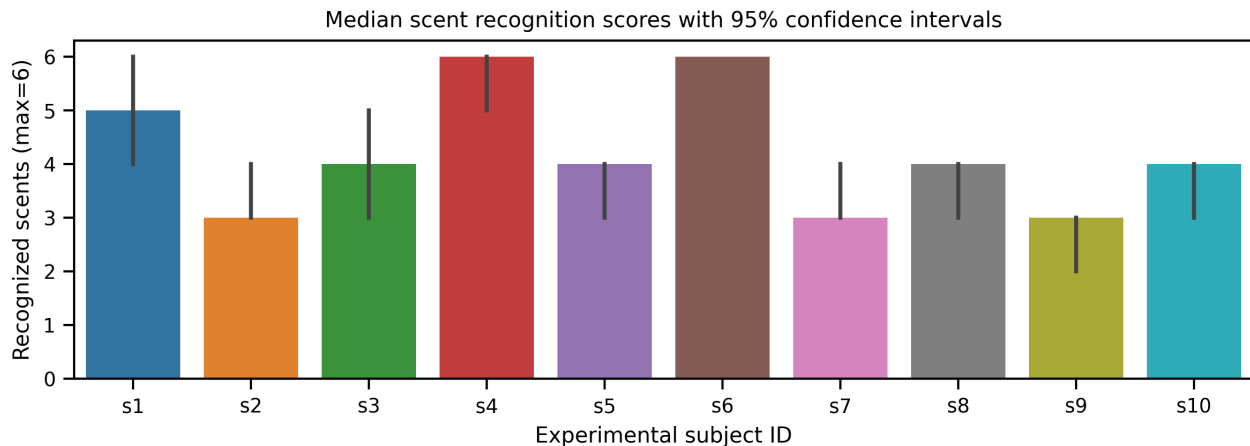


Figure 2: The study involved scent identification trials that included six targets each. The median results we represented as bar plots in the above figure, along with 95% confidence intervals. Out of all the subjects, only three of them were able to achieve the maximum score at once. The scores were then divided into two categories: lower awareness (scores ranging from 0 to 3) and high awareness (scores ranging from 4 to 6).

involving ten young, healthy olfactory BCI users. Finally, we will discuss these results and draw our conclusions.

## MATERIALS AND METHODS

The olfactory BCI has the potential to revolutionize human-computer interaction. It can also passively monitor olfactory cognition and age-related changes, making it helpful in supporting the elderly. Our recent feasibility study involved ten technology novices to identify any potential usability and related unresolved issues. During the study, the users were asked to perform classical oddball-style olfactory BCI tasks while their brain activity was monitored using EEG and EBG electrodes. The users were also asked to report on the number of target scents they could identify in each trial. We utilized data from scents delivered through an oddball paradigm to train a machine-learning model. The model predicts levels of sensory awareness based on olfactory stimulation in a passive brain-computer interface (BCI) application. This application estimates the user's mental state instead of generating commands.

### *Olfactory oddball BCI paradigm:*

In 2024, during winter, we conducted a pilot study on adult volunteers at Nicolaus Copernicus University in Toruń, Poland. The study aimed to record EEG and EBG using an olfactory passive BCI paradigm. The Institute of Psychology UNC Ethical Committee for Experiments with Human Subjects approved the investigation under the ethical principles of The Declaration of Helsinki. The study involved nine females and one male, with an average age of  $20.4 \pm 1.71$  years. The report presents the study's findings.

This pilot project presents the findings of a study that employs the ETT Olfactometer<sup>25</sup> to deliver odors in an oddball BCI paradigm. Users are asked to identify and report the number of instructed target scenes in each session,

comprising six odors. In each trial, one odor becomes a target accompanied by five randomly presented distractors. Our study uses classical reactive BCI settings to determine how well users can identify a specific target odor amidst other distracting odors. However, the actual application passively monitors the user's cognitive states. We used an olfactometer to ensure the odors were delivered uniformly in all trials. This system includes an airflow delivery unit, an odorant carrier, tubes, a nose applicator, and ETT Direct Control software. We have developed a Python script that communicates with the olfactometer's original software. This software connects to a laptop through a USB cable. The experiment setup can be seen in Figure 1. During the experiment, a participant wore a wireless Unicorn EEG wearable cap, manufactured by g.tec medical engineering, Austria, which was connected to a laptop running our in-house developed EEG with EBG recording and stimulus presentation software. This software controlled the delivery of olfactory stimuli from a pipe near the participant's nose. We used an ETT Olfactometer 2<sup>5</sup> manufactured by Emerging Tech Trans, LLC., USA, to deliver the odor stimulus. The olfactometer can be identified as a blue box on the left side of the photograph in Figure 1, with a blue-orange pipe on the participant's right side. In our passive olfactory BCI paradigm experiments, we used an olfactometer with six scents to control odorant stimulation. Each scent stimulus lasts four seconds, followed by a four-second break without odor delivery. To ensure accurate results, we conducted a pilot study that included natural odors such as rose, cinnamon, lavender, orange, lemon, and vanilla. The subject is visually instructed on a computer screen to breathe at the beginning of each new odor presentation during the experiment. After each session of six oddball trials, the subject reported the number of correctly identified scents, with six being a perfect score and three a chance level. The behavioral results are summarized as

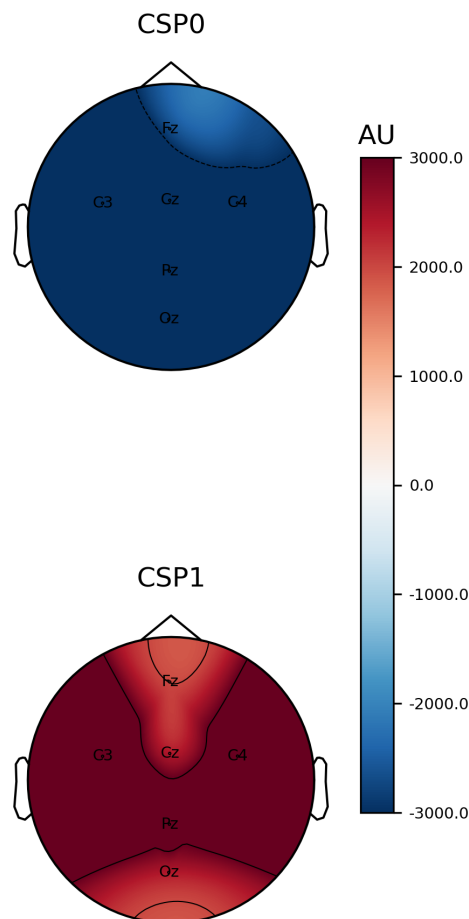


Figure 3: The averaged over all subjects and sessions head plot images show the CSP filter coefficients overlaid on a topographic map of the gamma ( $\gamma$ ) frequency band. The EEG data was collected from six electrodes, namely *Fz*, *C3*, *Cz*, *C4*, *Pz*, and *Oz*. The *CSP0* filter represents the low awareness, while *CSP1* represents the high awareness signals in the passive olfactory BCI paradigm.

medians with 95% error bars in Figure 2. The results of this study provide valuable insights into odor-based BCI paradigms and their potential applications in medicine, psychology, and neuroscience.

#### EEG and EBG recording:

The reported pilot research study collected EEG and EBG data using the Unicorn EEG headset from g.tec medical engineering, based in Austria. Our previous studies have demonstrated the reliability of this device compared to other available wearables [12, 13]. We used six EEG channels in the pilot investigation: *Fz*, *C3*, *Cz*, *C4*, *Pz*, and *Oz*. We also placed two EBG sensors on a 10/5 international standard EEG cap, approximately above each eyebrow at *AFp9h* and *AFp10h*.

In the first preprocessing stage, we converted six EEG and two EBG streams into a digital format with a sampling frequency of 250 Hz. After that, we removed any baseline shifts and high-frequency noise outside the frequency range of 7 Hz and 100 Hz by applying a bandpass filter. To eliminate any power line interference at 50 Hz, we used a notch filter. We segmented the EEG and EBG

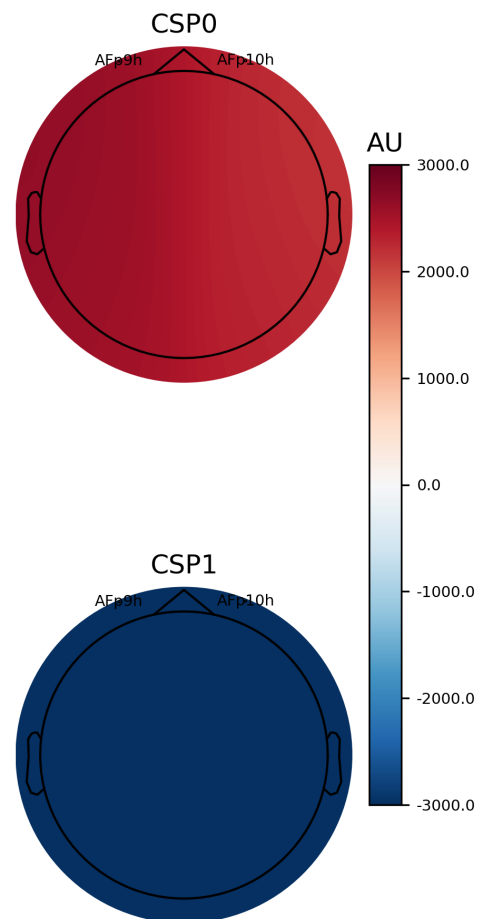


Figure 4: The averaged over all subjects and sessions head plot images show the CSP filter coefficients overlaid on a topographic map of the gamma ( $\gamma$ ) frequency band. The EBG electrodes, *AFp9h* and *AFp10h*, were employed. The gamma-band filter *CSP0* represent low awareness, while *CSP1* the high awareness in the passive olfactory BCI paradigm..

signals into eight-second sections with four seconds of odor stimulation and four-second flush breaks, using experimental triggers recorded together in oddball, target, and non-target recognition tasks.

To ensure the accuracy of our data, we utilized the empirical mode decomposition (EMD) technique to remove distortions caused by eye blinks or muscle movements in the EEGs and EBGs. We applied this method separately to each channel, which allowed us to identify and eliminate artifacts effectively, thus improving the overall quality of the data, as previously proposed in [9, 13].

Using a CSP method, we transformed EEG and EBG signals to increase the variance in one class and decrease it in the other, as described in [10, 11]. We used the CSP method to create distinguishable patterns across space for the olfactory stimulus-induced potential in the gamma sub-bands. These patterns were generated for two different levels of awareness. The first level had behavioral median scores of three and below (three subjects in our study). The second level had median behavioral scores of four and above (seven subjects in our study). We have illustrated all possible user behavioral responses in Fig-

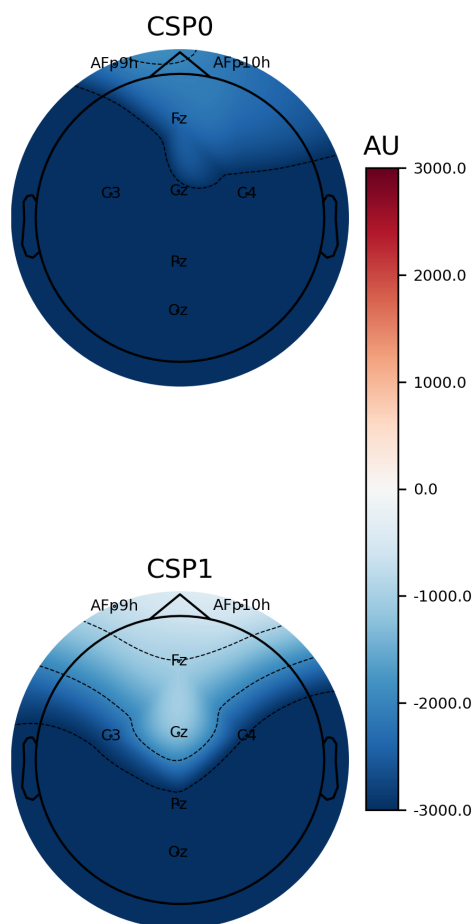


Figure 5: The averaged over all subjects and sessions head plot images display the CSP filter coefficients superimposed on a topographic map of the gamma ( $\gamma$ ) frequency band. The EEG data was obtained from six electrodes, namely *Fz*, *C3*, *Cz*, *C4*, *Pz*, and *Oz*, and two EBG electrodes, *AFp9h* and *AFp10h*, were employed. The gamma-band filters *CSP0* represent low awareness, while *CSP1* represents high awareness in the passive olfactory BCI paradigm.

ure 2. The patterns that could be distinguished were visualized in Figures 3, 4, and 5 for EEG-only, EBG-only, and combined EEG+EBG channels, respectively. We used the same scaling in arbitrary units (AU) [14]. The differences in CSP filter patterns that we observed were due to varying levels of EEG and EBG signals.

To assess the potential separability of a CSP feature, we used a supervised clustering technique called uniform manifold approximation and projection (UMAP) [15]. We merged CSP features derived from band-pass filtered EEG and EBG signals in the gamma band. This enabled us to obtain distinct clusters in a two-dimensional feature space. The CSP filters we used were four-dimensional.

As part of our research, we conducted a preliminary trial to test the accuracy of an offline passive olfactory BCI application. The trial involved three recording sessions, each comprising six single oddball trials. There were 36 EEG and EBG responses, each lasting eight seconds, with six targets and thirty non-target responses. We transformed these responses into two-dimensional CSP fea-

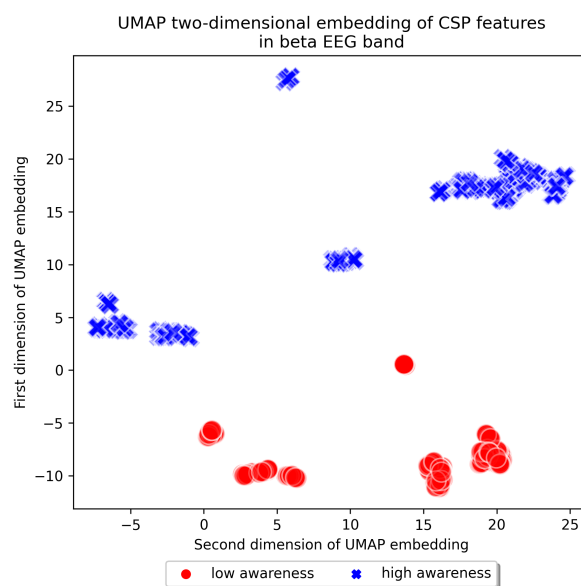


Figure 6: A scatter plot using two-dimensional UMAP results for EEG+EBG combined gamma bands.

tures in gamma EEG and EBG frequency bands, based on previous research that indicated higher frequency as the best carriers of olfactory information in the brain-waves [2, 3].

We used five different machine learning models for our experiment: linear SVM, random forest classifier (RCT), decision tree classifier (DTC), linear discriminant analysis (LDA), and deep fully connected neural network (DFNN). The DFNN had five hidden layers with 128, 64, 32, 4, and 2 RELU units each. We performed ten-fold cross-validation using these models, available in scikit-learn v1.4.0 [16]. We did not observe overfitting of the machine learning models despite the unequal distribution of low and high awareness cases, as shown in Figure 2.

## RESULTS

A pilot study was conducted using ten healthy subjects to fit spatial filters and analyze patterns in the EEG and EBG within the gamma frequency band. The study showed promising outcomes, summarized in Figures 3, 4, and 5, for EEG-only, EBG-only, and EEG+EBG combined, respectively. The outputs of CSP filtering also formed separable clusters of low versus high awareness mental states, as shown in Figure 6 using UMAP applied to EEG+EBG combined features. The results of the initial classification trials, which used ten-fold-cross-validation, are presented in Figure 7. The figure shows median balanced accuracies and percentile ranges. The evaluation used different classifiers, including linear SVM, LDA, RFC, DTC, and DFNN. All the results were significantly above the balanced accuracy chance level of 50%, are reported in Figure 7. The results also indicate that using combined EEG+EBG electrodes led to statistically significant balanced accuracy results (at  $p < 0.05$ ) for RFC, DTC,

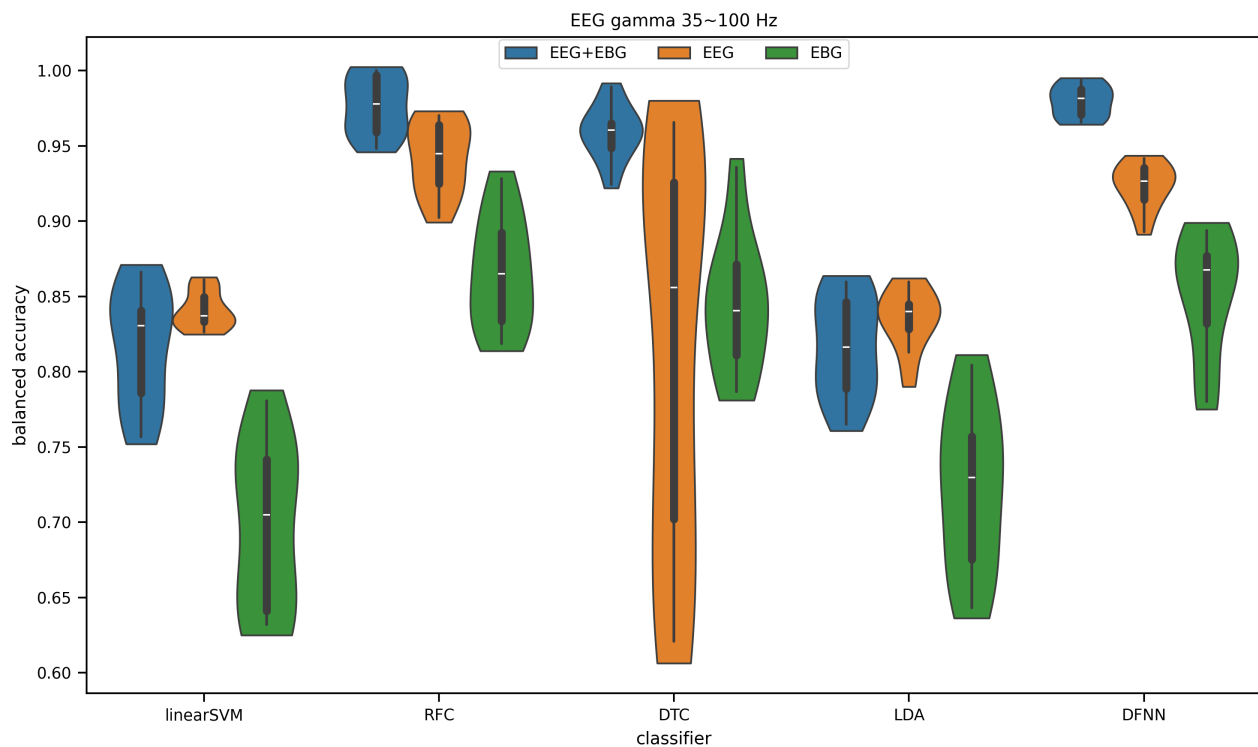


Figure 7: Distribution plots comparing EEG+EBG (blue), EEG-only (orange), and EBG-only (green) classification results for the evaluated classifiers in the reported study.

and DFNN classifiers, compared to EEG-only and EBG-only modality trials. The significance was tested using Wilcoxon tests. The RFC, DTC, and DFNN classifiers achieved median balanced accuracy scores above 95% for combined EEG+EBG electrode cases. Therefore, combining both electrode modalities creates a promising possibility in the olfactory modality for subject awareness estimation.

## DISCUSSION

We conducted a trial study on a passive olfactory BCI, a non-invasive method that uses smell to track brain activity in order to estimate user awareness at a later stage. Our study has shown that olfactory stimuli can be processed quickly without causing attention overload. This makes it a promising area for BCI research and quantifying the mental state of users. The study discovered that using a combination of EEG and EBG electrodes with classifiers such as RFC, DTC, and DFNN resulted in almost perfect classification outcomes. This supports the hypothesis that recording brainwaves through multiple modes can lead to improved results. However, simpler linear classifiers such as SVM and LDA showed a different level of improvement and had lower median accuracies than the classifiers mentioned earlier.

## CONCLUSIONS

The use of the olfactory modality in neurotechnology

has recently become popular due to the positive results of the proposed passive olfactory Brain-Computer Interface (BCI). These results are based on a preliminary pilot study with young and healthy subjects. However, our team faces specific challenges that need to be addressed to improve our approach. These challenges include developing more reliable odor delivery systems and implementing more robust algorithms to detect brain activity patterns. To refine and validate our approach, we plan to conduct a more extensive study with elderly subjects who may have reduced awareness due to mind wandering or daydreaming.

The study aimed to explore the potential of a passive olfactory BCI that could be used to track dementia or COVID-19-related olfactory impairment, as well as interventions. A passive modality could be used in the future for this purpose. The study findings confirm the feasibility of using a passive olfactory BCI, which could be a tool in non-invasive monitoring of brain activity and associated disorders based on user awareness tracking in various cognitive tasks.

These findings have implications for developing non-invasive techniques for monitoring brain activity and associated disorders. The discovery of a practical passive olfactory BCI could help track olfactory impairment related to dementia or COVID-19, which would help develop early interventions. This could be particularly useful in the COVID-19 pandemic, where olfactory impairment has been identified as a common symptom.

The study concludes that a passive olfactory BCI option

can be an effective way to track awareness, which may help in early interventions. The development of this technology can have significant implications for non-invasive monitoring of brain activity and related disorders.

#### AUTHOR CONTRIBUTIONS

TMR, HK, NN, and TK, and worked together to develop the concept of an olfactory brain-computer interface (BCI). TMR suggested using CSP to extract features from combined EEG and EBG time-series data and clustering/classification using supervised machine-learning techniques. HK and TK recruited subjects for the study. TMR designed and programmed the experimental stimulus presentation and the acquisition and analysis of EEG and EBG data. TMR and HK conducted the data analysis, while NN, TK, and MOM examined the results. TMR and HK wrote the manuscript.

#### FUNDING

MOM and TMR were partly supported by the Japan Science and Technology Agency AIP Trilateral AI Research Grant No. JPMJCR20G1 from the Japan Science and Technology Agency.

TMR received support from Nicolaus Copernicus University in Toruń, Poland, through the 2022 and 2024 Mobility Grants.

TK received support from Nicolaus Copernicus University's Emerging Field projects "Cognition and Language 2022" and "Culture, Development & Wellbeing" in Toruń, Poland.

#### ACKNOWLEDGMENTS

We want to thank the committed student volunteers from the Department of Cognitive Science at Nicolaus Copernicus University in Toruń, Poland, for their invaluable contribution to the olfactory BCI pilot study project. Their participation has significantly advanced the EEG and EBG-based brainwave modalities utilized in the study.

#### REFERENCES

[1] Xydakis MS *et al.* Post-viral effects of covid-19 in the olfactory system and their implications. *The Lancet Neurology*. 2021;20(9):753–761.  
[2] Murphy C. Olfactory and other sensory impairments in alzheimer disease. *Nature Reviews Neurology*. 2019;15(1):11–24.  
[3] Iravani B, Arshamian A, Schaefer M, Svenningsson P, Lundström JN. A non-invasive olfactory bulb measure dissociates parkinson's patients from healthy controls and discloses disease duration. *npj Parkinson's Disease*. 2021;7(1):75.

[4] Iravani B, Arshamian A, Ohla K, Wilson DA, Lundström JN. Non-invasive recording from the human olfactory bulb. *Nature communications*. 2020;11(1):648.  
[5] Woo CC, Miranda B, Sathishkumar M, Dehkordi-Vakil F, Yassa MA, Leon M. Overnight olfactory enrichment using an odorant diffuser improves memory and modifies the uncinate fasciculus in older adults. *Frontiers in Neuroscience*. 2023;17:1200448.  
[6] Kato M *et al.* Spatiotemporal dynamics of odor representations in the human brain revealed by eeg decoding. *Proceedings of the National Academy of Sciences*. 2022;119(21):e2114966119.  
[7] Pereira Junior J, Teixeira C, Rutkowski TM. "Visual motion onset brain-computer interface." *arXiv. Tech. Rep*. 2016. [Online]. Available: <http://arxiv.org/abs/1607.02695>.  
[8] Chang M *et al.* Comparison of P300 responses in auditory, visual and audiovisual spatial speller BCI paradigms. In: *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013*. Graz University of Technology Publishing House, Austria: Asilomar Conference Center, Pacific Grove, CA USA, 2013, Article ID: 156 (1–2).  
[9] Rutkowski TM, Mori H. Tactile and bone-conduction auditory brain computer interface for vision and hearing impaired users. *Journal of Neuroscience Methods*. 2015;244:45–51.  
[10] Dornhege G, Blankertz B, Curio G. Speeding up classification of multi-channel brain-computer interfaces: Common spatial patterns for slow cortical potentials. In: *First International IEEE EMBS Conference on Neural Engineering*, 2003. Conference Proceedings. 2003, 595–598.  
[11] Zhao Q, Rutkowski TM, Cichocki A, Zhang L. High resolution common spatial frequency filters for classifying multi-class EEG. In: *Advances in Cognitive Neurodynamics (II)*. Springer Netherlands, 2011, 683–688.  
[12] Rutkowski TM, Narebski S, Bekier P, Komendzinski T, Sugimoto H, Otake-Matsuura M. Cross-cultural evaluation of dementia passive bci neuro-biomarker candidates. In: *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS & ISIS)*. 2022, 1–5.  
[13] Rutkowski TM, Abe MS, Komendzinski T, Sugimoto H, Narebski S, Otake-Matsuura M. Machine learning approach for early onset dementia neurobiomarker using eeg network topology features. *Frontiers in Human Neuroscience*. 2023;17.  
[14] Gramfort A *et al.* MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*. 2013;7(267):1–13.  
[15] McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. 2018.  
[16] Pedregosa F *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825–2830.

# SSVEP-BASED COVERT COMMUNICATION USING HYPERSCANNING

L.H. Reintsema<sup>1</sup>, C.M. Sweeney-Reed<sup>2</sup>, S. Dürschmid<sup>3,4</sup>, H. Hinrichs<sup>5</sup>, C. Reichert<sup>3</sup>

<sup>1</sup> Faculty of Electrical Engineering and Information Technology, Otto von Guericke University, Magdeburg, Germany

<sup>2</sup> Neurocybernetics and Rehabilitation, Department of Neurology, Otto von Guericke University, Magdeburg, Germany

<sup>3</sup> Department of Behavioral Neurology, Leibniz Institute for Neurobiology, Magdeburg, Germany

<sup>4</sup> Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, CA, USA

<sup>5</sup> Department of Neurology, Otto von Guericke University, Magdeburg, Germany

E-mail: christoph.reichert@lin-magdeburg.de

**ABSTRACT:** Communication by means of evoked brain signals is one of the main applications of brain-computer interfaces (BCIs). Commonly, in BCI applications the user's intention is directly fed back and openly perceivable. Here we used hyperscanning to investigate a communication approach, in which two users can covertly communicate by brain signal modulation. To achieve this, we artificially generated synchronous and asynchronous oscillatory brain activity by presenting a choice of two flickering stimuli inducing steady-state visual evoked potentials (SSVEPs) and provided feedback that indicated the synchronicity of the brain signals of participant pairs. We used different approaches to determine synchronicity. When we used broadband activity, the accuracy varied considerably between participant pairs, which could be attributed to individual differences in the timing and the amplitudes of SSVEPs. However, when we involved features reflecting the stimulus frequencies, the predictions were highly reliable. Beyond demonstrating the feasibility of our approach, our findings have the potential to identify challenges in studying social interaction using hyperscanning.

## INTRODUCTION

Hyperscanning refers to recording brain data from more than one person simultaneously. It is increasingly being used to investigate neuronal correlations during social interaction [1,2] and learning [3]. A common approach in hyperscanning is to determine the degree of brain-to-brain synchrony, i.e., the synchronicity, of two users. Portable electroencephalography (EEG) headsets allow involvement of multiple individuals in real-world environments, e.g., a classroom [4]. Simultaneous EEG recordings from multiple users have also been utilized in brain-computer interface (BCI) research to incorporate collaborative/competitive BCI control in video games [5] and to increase the decoding accuracy in single trial classification of visual evoked potentials (VEPs) [6].

Collaborative BCIs have the potential to increase the signal-to-noise ratio by combining activity from multiple brains. For communication purposes, collaborative control would not have a practical benefit, since users would be required to know their mutual intention. In contrast, a brain-to-brain communication, in which one user sends a message and the other user infers the message from comparing BCI-generated feedback and their own intention, could enable a covert communication only perceivable by the users involved. Brain-to-brain interface control using brain stimulation techniques applied to the receiver has been demonstrated in rats [7] and humans [8,9]. Here we introduce a noninvasive approach for implementing indirect brain-to-brain communication by simultaneously recording EEG signals from both the sender and the receiver of a communication. The content of the communicated message is inferred from feedback indicating the degree of brain synchrony. We artificially induce brain synchronicity and asynchronicity by presenting flicker stimuli. While the approach of decoding the steady-state visually evoked potentials (SSVEPs) induced by these types of stimuli is a common approach for controlling BCIs [10,11], decoding the synchronicity of brain signals from two brains induced by flicker stimuli has not yet been performed.

## MATERIALS AND METHODS

*Subjects and Task:* Fourteen participants (mean age 26.4±4.2 years, 8 male) were recruited to participate in the hyperscanning BCI experiment, resulting in seven participant pairs, or dyads (6 male/female, 1 male/male). They provided informed consent and received 25 € for participation. The study was approved by the Ethics Committee of the Otto von Guericke University, Magdeburg, Germany.

Two participants were seated next to each other, each in front of a custom stimulation device, which consisted of

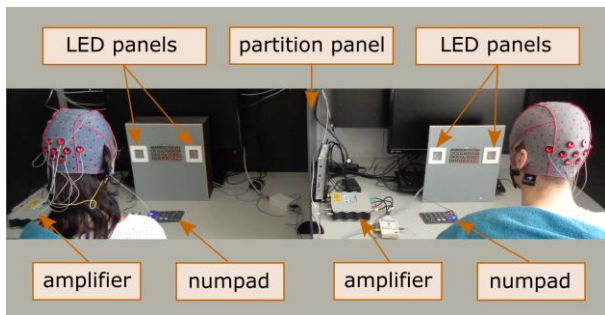


Figure 1: Technical setup for the hyperscanning BCI.

two small light-emitting diode (LED) panels (35×35 mm<sup>2</sup>), 18cm apart. A partition panel on the desk prevented the participants from being distracted by the stimuli presented to the partner (see Fig. 1 for the technical setup). The task of the participants was to silently communicate yes/no responses by focusing on particular flickering stimuli. One participant was asked to send a yes/no response (Sender role) and the other was asked to infer the response from the feedback generated by the BCI (Receiver role). The roles were changed between participants halfway through the experiment. The experimenter provided verbal instructions to guide the participants throughout the experiment. A trial started by verbally cueing the participants to focus on their next intended response. Subsequently, the presentation of the stimuli started. The left LED panel flickered with a frequency of 9.09 Hz (110 ms stimulus onset asynchrony) and was associated with the response “yes”. The right LED panel flickered with a frequency of 11.11 Hz (90 ms stimulus onset asynchrony) and was associated with the response “no”. We identified these flicker frequencies as reliable following pilot testing with individual test subjects using 9.09 Hz, 11.11 Hz, 12.5 Hz and 15.0 Hz stimuli.

The occipital brain oscillatory activity of the participant classically synchronizes with the flicker frequency of the panel on which the participant focuses their gaze, forming the basis of a classical SSVEP-based BCI. The stimuli for the two participants were synchronized, i.e., the LEDs in front of them were on and off at the same time. After five seconds, the stimulus stopped and the synchronicity feedback was presented by a computer voice saying “equal” if the BCI detected synchrony between the EEG signals of the participant pair, and “different”, if the BCI detected asynchronous EEG signals. Afterwards, the Receiver combined their own internal response with the synchronicity feedback, inferring what the Sender had intended to answer and pressed the corresponding button on a keypad. The inferred Sender response was presented as inference feedback (“yes” or “no”). Finally, the Sender assessed whether the inferred response was correct or not by pressing the corresponding button on a keypad, yielding a third feedback item, which could be “correct” or “wrong”. This final feedback provided the Receiver with the ground truth. Note that for silent communication only, the synchronicity feedback is required but no button press. The button presses and the additional feedback

were only necessary for evaluation purposes.

A session started with 20 training trials in which all combinations of responses were cued an equal number of times (yes/yes, no/no, yes/no, no/yes) and no feedback was presented. These trials were for initial classifier training. Afterwards, the feedback mode started, and the experimenter asked subjective questions, e.g. “Do you play an instrument?”. After each trial, the participant’s intentions were determined from their responses and the classifier was retrained. We also performed trials in which the participants were asked to communicate one of 32 items. A table was presented, showing 32 numbers and letters, or 32 cards of a card deck. Each item shown in the tables was assigned a binary color code, where five bits were coded with green and red bars located under each item. The participants associated green bits with “yes” and red bits with “no”. This approach required five trials, with the concomitant five binary decisions, to communicate one item. Trials in which participants made mistakes according to their own statement (e.g., wrong button press) were immediately excluded by the experimenter from the further analysis. Therefore, the number of questions asked varied, resulting in a total number of 118.7±6.5 trials on average, including training trials.

**Recordings:** EEG data were recorded with a sampling frequency of 512 Hz from ten electrode sites (OI1h, OI2h, PO7, O1, Oz, O2, PO8, POO1, POO2, Fz) and referenced against the right ear lobe using two g.tec gUSBamp devices. Synchronization of EEG recordings was achieved by sending a trigger signal simultaneously to both amplifiers and correcting for time shifts in the received data buffers during online processing. The 5 s data segments were notch filtered to remove 50 Hz line noise and bandpass filtered between 5 and 30 Hz. Finally, we resampled the data to 256 Hz sampling rate to reduce computational demands in further processing.

**Decoding approach:** Different decoding approaches were applied in the online and offline analyses described in this section. All approaches used canonical correlation analysis (CCA), a statistical method that maximizes correlation between two variable sets  $X$  and  $Y$ :

$$(U, V) = \underset{A, B}{\operatorname{argmax}} \operatorname{corr}(XA, YB) \quad (1)$$

where the canonical coefficients in  $A$  and  $B$  linearly combine  $X$  and  $Y$ , such that the correlation is maximal in the first variables of matrices  $U = XA$  and  $V = YB$ , and decreases with increasing component ranking. The approach suggested by Lin et al. [12] has become established in BCIs as a reliable way of detecting SSVEPs and uses  $X$  as the time varying brain signals (EEG) and  $Y$  as a set of sine and cosine functions with frequencies equal to the stimulus frequencies and their harmonics. We used this approach to determine classifier features, using the two stimulation frequencies, 9.09 and 11.11 Hz, and their first harmonics, resulting in four features per stimulation frequency. We used only the first two canonical correlation coefficients per stimulation frequency and participant as features, resulting in four

features per participant and eight features per dyad. We refer to this feature set as  $R_{SSVEP}$ . This feature set specifically captures the brain activity associated with the flicker stimuli. To investigate whether the brain synchronicity, which we artificially induced through the stimulus, can also be predicted when we do not include information about the stimulus, we calculated another feature set by setting  $X$  as the EEG signal of one participant and  $Y$  as the EEG signal of the other participant. With this approach, we investigated whether we can determine synchronous brain signals from broadband (here limited to 5-30 Hz) brain activity. Here we used only the first two canonical correlation coefficients and refer to this feature space as  $R_B$  and to the canonical components as  $U_B$  and  $V_B$ , respectively.

During online decoding, we used the  $R_{SSVEP}$  feature set and trained a k-nearest-neighbor classifier (kNN) using trials with equal stimulation frequencies as one class and trials with different stimulation frequencies as the other class to decode the synchronicity and present feedback accordingly. We also predicted the synchronicity directly from this feature set in a leave-one-out cross-validation (LOOCV) using a nonlinear support vector machine (SVM) classifier and radial basis function (RBF) as kernel. Finally, the  $R_{SSVEP}$  feature set was used to classify the SSVEP response in single participants as in conventional SSVEP-based BCIs, using LOOCV and the RBF SVM classifier. To test whether the direct classification of synchronicity is advantageous compared to classifying the SSVEP of the participants separately and subsequently determining the synchronicity indirectly by comparing the predictions of both participants, we calculated the indirect decoding accuracy from the predictions obtained by conventional SSVEP decoding.

In a final approach, we used LOOCV and RBF SVM to decode the synchronicity of brain activity based on broadband EEG signals, using the  $R_B$  feature set as described above. To compensate for potential shifts in individual latencies of VEPs, we shifted the signals in steps of single sample points against each other before applying CCA and compared the maximum accuracy of the time-shifted analysis with the accuracy achieved with the non-shifted signals.

We performed permutation testing for all reported classification approaches by permuting the labels that indicate the focused stimulus frequency and repeating the LOOCV 1000 times. This procedure resulted in a distribution of chance accuracies from which we determined the mean chance level and the 95% confidence intervals.

## RESULTS

The intention of participants with Sender role was correctly determined on the basis of the feedback regarding the synchronicity of their brain activity with that of the Receiver in  $\mu=93.6\%$  ( $\sigma=10.7\%$ ) on average during the BCI hyperscanning. The LOOCV using the same approach (direct classification of synchronicity)

yielded an average accuracy of  $\mu=94.3\%$  ( $\sigma=9.4\%$ ), which was not statistically significantly different to that attained using online decoding. Using the approach of indirectly classifying synchronicity resulted in an average accuracy of  $\mu=94.7\%$  ( $\sigma=9.2\%$ ) and was neither different from online decoding nor from direct classification of synchronicity. Conventional decoding of the stimulus frequency from SSVEPs using  $R_{SSVEP}$  features resulted in an average decoding accuracy of  $\mu=97.4\%$  ( $\sigma=6.8\%$ ) across all 14 participants. These decoding accuracies are shown in Fig. 2 for each participant pairs. It can be seen that the classification of synchronicity is bounded by the accuracy in detection of the focused stimulus frequency of the less well-performing participant.

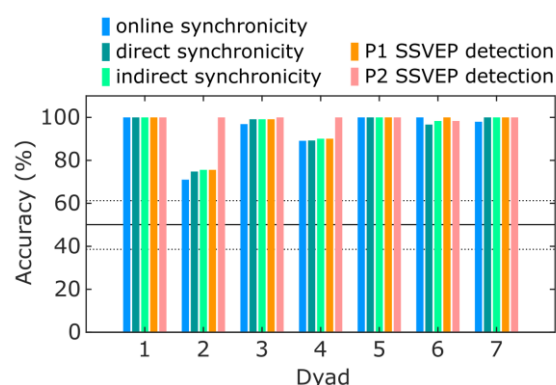


Figure 2: Decoding accuracies achieved with different decoding approaches using the  $R_{SSVEP}$  feature space. All bars show accuracies obtained with LOOCV except the blue bar, which shows online accuracy. Solid and dashed black lines indicate the mean and upper 95% confidence interval of the chance level obtained by permutation testing.

In a next step, we did not include information about the stimulus frequencies but rather calculated canonical correlation coefficients using the broadband signals from all channels of both participants as variable sets to use them as features for classification of the brains' synchronicity. This analysis showed strong variability between dyads, ranging from 50.5% to 97.2% decoding accuracy ( $\mu=72.5\%$   $\sigma=18.1\%$ ). The decoding accuracy could be improved by shifting the time series of either participant to compensate for potential individual differences in visual processing latencies. Selecting the maximum accuracy from the latency shifts, decoding accuracy ranged from 62.2% to 97.2% ( $\mu=78.2\%$   $\sigma=14.5\%$ ). Latency shifts resulting in these improved accuracies ranged from 0ms to 23.4 ms ( $\mu=10.0$  ms,  $\sigma=9.0$  ms).

To investigate the reason for the large inter-dyad differences when using the  $R_B$  features, we calculated the spectra of the first component in  $U_B$ , obtained by CCA, of each trial and participant and averaged across trials where the visual stimuli were identical in both participants. For each stimulus frequency and their first harmonics, we calculated the Pearson correlation coefficient  $\rho$  between amplitudes (obtained by the spectra of the first component in  $U_B$ ) and decoding accuracies

(achieved with the broadband synchronicity classification approach) across dyads (see Fig. 3). We found a significant correlation ( $p < 0.05$ ) for frequencies 9.09 Hz ( $\rho = 0.93$ ), 11.11 Hz ( $\rho = 0.86$ ) and 18.18 Hz ( $\rho = 0.73$ ) but not for 22.22 Hz ( $\rho = 0.34$ ). Therefore, the accuracy of broadband synchronicity classification is influenced by the magnitude of SSVEP amplitudes.

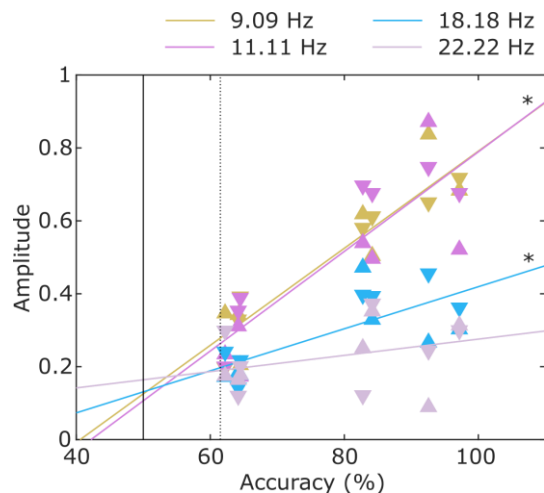


Figure 3: Decoding accuracies of broadband synchronicity classification according to signal amplitudes at different stimulus frequencies and their 1<sup>st</sup> harmonics. Regression lines visualize the correlation, asterisks indicate significant correlation ( $p < 0.05$ ). Upward and downward triangles indicate first and second participant of a dyad. Solid and dashed black lines indicate mean and upper 95% confidence interval of the chance level obtained by permutation testing.

## DISCUSSION

The study demonstrates that covert communication can be performed using a noninvasive hyperscanning BCI. We induced synchrony of brain signals by presenting synchronous and asynchronous visual stimuli and used the degree of synchrony as a feedback signal. Brain-to-brain synchrony was decoded with an accuracy close to 100% in five dyads. This was only constrained by the lowest single subject SSVEP detection accuracy within a dyad in two instances using features that incorporate prior knowledge about the flicker frequency of the stimuli. Although direct classification could potentially exploit multivariate relationships between individual  $R_{SSVEP}$  features, decoding accuracy was not significantly different from the indirect approach that compared independently predicted SSVEPs of each participant. Thus, both approaches are well suited to implementing the proposed BCI for covert communication.

A secondary aim of the study was to investigate the feasibility of classifying brain-to-brain synchrony in general, without taking knowledge about stimulus frequencies into account. When we used broadband activity to calculate CCA features, decoding accuracy strongly depended on the SSVEP amplitudes, i.e., on

how much the stimulus was reflected in the brain signals. Two of the dyads were highly reliably decoded ( $>90\%$ ), two were moderately reliably decoded ( $>80\%$ ) and three achieved accuracies slightly above chance level ( $>61.5\%$ ).

Flicker frequencies in the alpha band have been shown to have an impact on attentional processing, which is presumed to result from interactions between the flicker rhythm and the endogenous alpha rhythm [13]. Note that although the stimulus frequencies we used both lie within the alpha band, common CCA-based decoding resulted in high accuracy. The interaction between flicker stimuli and the endogenous alpha rhythm is, however, highly variable, with dependency on the individual alpha frequency and also the particular source of the endogenous alpha rhythm [14]. Future work should include exploration of potential relationships between synchronicity decoding accuracies and both individual alpha activity and flicker frequencies within and outside the alpha band.

Our approach provides an alternative strategy for hyperscanning experiments, given the general discussion on interpretability of coherence measures applied for investigating social interaction with this technique [15,16]. Mainly, however, our findings uncover important challenges in hyperscanning, namely that individual differences in EEG signals can lead to quite different degrees of brain-to-brain synchrony detection. This variability is illustrated by our findings when we classified synchronicity using  $R_{SSVEP}$  and  $R_B$  features. The SSVEP amplitude was sufficiently high to enable decoding of the focused frequency at close to 100% in 12 of 14 subjects, and the synchrony decoding based on these features was related to these accuracies. However, the SSVEPs were not sufficiently represented in the EEGs of some dyads to enable comparably reliable synchrony decoding from broadband activity. Furthermore, latency shifts of evoked potentials might be a limiting factor for determining brain-to-brain synchrony, as not only suggested by our analyses but also known from the literature.

In future work, features based on coherence and information theory could be investigated for suitability of synchronicity detection for comparison with our broadband CCA approach. A limitation of our approach to performing covert communication is that it can only be performed in binary mode, by inferring binary responses from feedback indicating the degree of synchrony. However, using color-coded items, the participants also could communicate letters in five steps of binary decisions and thus, the proposed BCI could potentially be used to communicate whole sentences. Another limitation is that SSVEPs depend on eye movements, and only sensory processing is decoded rather than higher cognitive functions. However, gaze-independent paradigms communicating binary decisions using attention processes exist [17], which would be suitable for decoding brain-to-brain synchrony as well, and might be the next step towards a gaze-independent, covert communication BCI.

## CONCLUSION

SSVEPs can be used to stimulate brain-to-brain synchrony, which was decoded with high accuracy using stimulus-based features and enabled dyads of participants to reliably communicate binary messages, not perceivable by external observers. Synchrony features obtained from broadband signals reflected the synchronous stimulating signals, resulting in reliable decoding only in some dyads. These findings may have implications for other neuroscientific hyperscanning studies investigating social interaction.

## DATA AVAILABILITY

The datasets used for this study are available to download from <https://zenodo.org/doi/10.5281/zenodo.10809098>.

## ACKNOWLEDGEMENTS

This study was partially supported by the German Research Foundation - DFG grant 521761873 and DFG SFB-1436, TPA03. We also would like to acknowledge the intramural funding of the Leibniz Institute for Neurobiology - LINSeeds Grant.

## REFERENCES

- [1] Czeszumski A, Eustergerling S, Lang A, Menrath D, Gerstenberger M, Schuberth S, et al. Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Front Hum Neurosci.* 2020;14(39):1–17.
- [2] Babiloni F, Astolfi L. Social neuroscience and hyperscanning techniques: Past, present and future. *Neurosci Biobehav Rev.* 2014;44:76–93.
- [3] Tan SHJ, Wong JN, Teo WP. Is neuroimaging ready for the classroom? A systematic review of hyperscanning studies in learning. *Neuroimage.* 2023;281(August):120367.
- [4] Dikker S, Wan L, Davidesco I, Kaggen L, Oostrik M, McClintock J, et al. Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Curr Biol.* 2017;27(9):1375–80.
- [5] Bonnet L, Lotte F, Lécuyer A. Two brains, one game: Design and evaluation of a multiuser bci video game based on motor imagery. *IEEE Trans Comput Intell AI Games.* 2013;5(2):185–98.
- [6] Yuan P, Wang Y, Wu W, Xu H, Gao X, Gao S. Study on an online collaborative BCI to accelerate response to visual targets. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. San Diego: IEEE; 2012. p. 1736–9.
- [7] Pais-Vieira M, Lebedev M, Kunicki C, Wang J, Nicolelis MAL. A brain-to-brain interface for real-time sharing of sensorimotor information. *Sci Rep.* 2013;3:1–10.
- [8] Rao RPN, Stocco A, Bryan M, Sarma D, Youngquist TM, Wu J, et al. A direct brain-to-brain interface in humans. *PLoS One.* 2014;9(11):e111332.
- [9] Jiang L, Stocco A, Losey DM, Abernethy JA, Prat CS, Rao RPN. BrainNet: A Multi-Person Brain-to-Brain Interface for Direct Collaboration Between Brains. *Sci Rep.* 2019;9(1):1–11.
- [10] Vialatte FB, Maurice M, Dauwels J, Cichocki A. Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Prog Neurobiol.* 2010;90(4):418–38.
- [11] Li M, He D, Li C, Qi S. Brain-computer interface speller based on steady-state visual evoked potential: A review focusing on the stimulus paradigm and performance. *Brain Sci.* 2021;11(4):450.
- [12] Lin Z, Zhang C, Wu W, Gao X. Frequency recognition based on canonical correlation analysis for SSVEP-Based BCIs. *IEEE Trans Biomed Eng.* 2006;53(12):2610–4.
- [13] Gulbinaite R, Van Viegen T, Wieling M, Cohen MX, Vanrullen R. Individual alpha peak frequency predicts 10 hz flicker effects on selective attention. *J Neurosci.* 2017;37(42):10173–84.
- [14] Nuttall R, Jäger C, Zimmermann J, Archila-Melendez ME, Preibisch C, Taylor P, et al. Evoked responses to rhythmic visual stimulation vary across sources of intrinsic alpha activity in humans. *Sci Rep.* 2022;12(1):1–15.
- [15] Burgess AP. On the interpretation of synchronization in EEG hyperscanning studies: A cautionary note. *Front Hum Neurosci.* 2013;7(881):1–17.
- [16] Hamilton AFC. Hyperscanning: Beyond the Hype. *Neuron.* 2021;109(3):404–7.
- [17] Reichert C, Dürschmid S, Sweeney-Reed CM, Hinrichs H. Visual spatial attention shifts decoded from the electroencephalogram enable sending of binary messages. In: 2022 IEEE Workshop on Complexity in Engineering (COMPENG). Florence; 2022. p. 1–4.

# EEG-BASED PERFORMANCE ESTIMATION DURING A REALISTIC DRONE PILOTING TASK

Marcel F. Hinss<sup>1</sup>, Vincenzo Maria Vitale<sup>1</sup>, Anke M. Brock<sup>1</sup>, Raphaëlle N. Roy<sup>1</sup>

<sup>1</sup> Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, France.

E-mail: marcel.hinss@isae-supero.fr

## ABSTRACT

Passive brain-computer interfaces (pBCIs) developed within the neuroergonomic field usually aim to improve safety by augmenting human-machine interaction. To accomplish said goal, many pBCIs classify mental states such as mental workload or mental fatigue. An alternative is to forego mental states and aim to predict performance. Despite its drawbacks, we argue that performance estimation is a more goal-oriented approach than mental state estimation. In a realistic experiment, 25 participants had to control an uncrewed aerial system for two hours, continuously switching between target search and navigation. EEG classification accuracies based on mental states and performance were compared. With a Tangent Space Logistic Regression, we could predict an increased likelihood of lapses in the form of missing instructions with an above-chance level accuracy of 62.09 %.

## INTRODUCTION

Passive Brain-Computer Interfaces (pBCI), i.e., BCIs that observe brain activity that is not influenced by the presence of a BCI, are a valuable component of neuroergonomics [1]. They promise to provide complex systems, such as cockpits, with valuable information on their user and the Human-Machine Interaction. A machine can then use that information employing adaptation or feedback to improve said interaction [2].

To do so, pBCIs are often trained to detect specific mental states such as mental workload [3] or mental fatigue [4]. The underlying argument for detecting said mental states is their correlation with erroneous or sub-optimal behaviour by the operator. Thus, by detecting, e.g. a high mental workload, the system may adapt itself to reduce workload and increase safety [5]. This approach has certain drawbacks. Mental states as constructs are not observable and vary across definitions [6, 7]. Furthermore, mental states depend on the current context and tasks. Differing task instructions can result in differing brain activity [8]. A high mental workload during an N-Back task may not be comparable to a high workload during a Stroop task [9, 10]. Finally, mental states are not always strong predictors of performance. In the case of mental fatigue, evidence suggests that participants, using e.g. compensatory strategies, can uphold performance despite

fatigue [11–13]. So, all these aspects must be accounted for when constructing a mental state-based pBCI to be used in an open-loop adaptation (feedback) or a closed-loop adaptation (interface change) with a complex system [2].

Alternatively, a pBCI could forego the mental state aspect and try to predict a participant's behaviour directly. Performance estimation has been proven to work in several contexts [14–17] and does not suffer from any of the aforementioned issues. It works by assigning labels to the physiological data, using the recorded behavioural data of participants, such as reaction time, accuracies and misses. Performance prediction allows direct observation of the variable we want to optimize with a pBCI, but also faces challenges. To predict performance, we first need to define good and bad performance. Many tasks, such as the Stroop or N-back tasks, involve some measure of correctness and reaction time [9, 10]. Reaction time or accuracy may be considered a valid performance measure in these cases, but only the combination into a global score will provide a complete picture of performance. Combining scores, on the other hand, raises questions about how to weigh each metric. Here, it needs to be considered that these measures are not orthogonal [8]. As mentioned above, the issue gets more complex as we move away from very controlled tasks and move towards more ecologically valid measures that may include several different reaction times and accuracies.

A related challenge is then how these cases should be labelled. Imagine, for example, an experiment where participants continuously perform a task for one hour. A global score of performance is assigned to each minute of the task. The value is continuous from 0 (bad) to 1 (good). How can the data now be divided into a 2-class problem? The 10 worst minutes of performance versus the remaining 50 minutes, the 10 best minutes versus the 10 worst minutes, the good half versus the bad half, or values exceeding a threshold (e.g. >0.8) or subceeding another (e.g.<0.2) are all present plausible approaches. The metric calculation and label assignment issue is further complicated when algorithms are tested to classify the data. Does a chance level classification accuracy mean the algorithm doesn't work or that the label assignment is sub-optimal?

In many cases, predicting any change in performance may be helpful, whether it is the likelihood of committing

an error, missing a trial or the speed at which a participant responds. Still, trying out unlimited combinations of labels may also create a global performance score not because of its usefulness but because of its ability to be classified.

To the best of our knowledge, no study has yet evaluated performance prediction using a pBCI applied to a prolonged realistic drone task. Hence, this experiment tested whether an EEG-based pBCI can predict meaningful performance metrics from participants performing a complex Uncrewed Aerial System (UAS, drone) piloting task. Moreover, extended mission duration makes UAS pilots vulnerable to mental fatigue and the associated risks [18]. Participants were asked to switch between a search and a navigation task during the experiment for two hours. This work expands on a previous protocol that focused only on a visual search task –and which did not yield above chance level performance estimations– [19] by adding a second navigation task and making the overall performance more difficult and longer. The long duration allows for comparing performance-based labelling and more traditional labelling based on subjective fatigue scores and Time-On-Task (TOT). Our goal is to illustrate how label assignment impacts classification accuracies, particularly in the absence of an absolute performance definition, for such a realistic task.

## METHODS

### *Participants:*

25 Participants (7 female, mean age 23.54 years (std 2.7), 11 English speaking & 14 French speaking) were recruited and completed the experiment. From the subsequent analysis, one participant had to be dropped due to inconsistencies and missing data in the recordings.

**Procedure:** Participants who agreed to participate signed the informed consent forms and were equipped with the EEG sensor. Next, participants completed a battery of questionnaires. They then completed a training phase of 16 minutes before a five-minute resting state was recorded (30-second intervals of eyes open and closed). They then started the 120-minute main phase of the experiment. After completing the main phase, participants filled in another battery of RSME, KSS, SPS, and VAS questionnaires. The ethical committee of Toulouse (Comité éthique de l'Université de Toulouse) approved the experimental protocol (Dossier 2022-501).

### *Materials:*

**Task:** The UASOS task (Fig. 1) combines some of the fundamental aspects of UAS operations with a Task-Switching protocol to allow the investigation of cognitive flexibility [20]. The task requires participants to alternate between tasks on a trail-based system. On average, every 7 seconds (with a  $\pm 1000$ ms jitter,  $\sim 1020$  trials during the main condition), written instructions appear on a widget in the middle of the participant's visual field to indicate the current task. To ensure adequate performance, small pretests were conducted to calibrate these param-



Figure 1: Experimental setup. Left Screen: Search Task. Right Screen: Navigation Task. Center Top: Flight director with task information. (Note: The text in the centre of the screens is feedback only displayed during the training phase.)

eters. Participants work on two main tasks, with two modes each. The Navigation task (NAV) requires the participant to navigate the UAV either using headings (heading mode, HDG) or waypoints (Waypoint mode, WPY). The design was balanced with an equivalent number of trials in all tasks and modes.

In the heading mode of the NAV task, participants receive a heading instruction (e.g. 350) in each trial. Using a joystick, they then turn the UAS in said direction. For the WPY mode, they receive a waypoint consisting of a letter and a number (e.g. F13). They must choose the corresponding waypoint using a trackball on a grid overlaying the navigational display. The other task is the SRC task. This task was adapted from previous work [19], and integrated into the overall protocol. Participants see a 3x3 grid of black-and-white images that visual filters may further distort. They are instructed to search either People or Vehicles. If they detect a target on one picture, they select the corresponding picture using a numpad. For all tasks, reaction times and the correctness of responses are recorded. The instructions on the flight director widget tell the participant which mode to perform at the onset of each trial.

The task was coded in Python and presented on two identical computer screens. A detailed description of the experimental environment can be found in [21].

**Questionnaires:** Participants answered 5 questionnaires at varying moments. At the beginning of the experiment, participants completed the demographics questionnaire, and their handedness was also assessed using the shortened Edinburgh handedness questionnaire [22]. Next, the Karolinska Sleepiness Scale (KSS), a 9-point Likert and the Samn-Perelli Fatigue (SPF) 7-point Likert scale were used to assess fatigue [23, 24]. Participants also filled in the RSME scale [25] that evaluates participants' mental effort invested in the task. The versions in which all items are labelled were used [26]. The translated versions of the KSS and SPS questionnaires originate from the ICAO [27]. Participants also responded to two Visual Analogue Scales (VAS) scales: cognitive fatigue (VAS-

F) and drowsiness (VAS-D). The entire battery was presented a second time following the completion of the experiment. The VAS scales were also filled in at 19-minute intervals during the main experimental phase.

**EEG:** Using an active AG-AgCl electrode system with an ActiCHamp amplifier (Brain Products, GmbH), EEG data was recorded from 64 electrodes. The international 10-20 system was used for electrode placement [28]. Data were recorded at 500 Hz, and impedances were kept below 50 k $\Omega$ . Data was streamed and synchronized using the Lab-StreamingLayer (LSL) [29].

As part of the data validation, we performed a frequency analysis of the EEG data independently from the mental state prediction. For this, a zero-padded channel was added to the EEG data before an average referencing, with a subsequent removal of the zero-padded channel. Extreme values were clamped following the method proposed by [30]. The data was then cut into 5-second non-overlapping epochs. The power of each frequency band was calculated by band-pass filtering the signal and calculating the root mean square for each electrode. Using the parameters suggested in [31], the power of the theta (4-8Hz) alpha (8-12Hz) bands were extracted. For the statistical analysis the data was then averaged into 10-minute epochs over three clusters of electrodes (i) Frontal: F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4; (ii) Central: C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4; and (iii) Parieto-Occipital: P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2.

#### *Performance and Mental State classification:*

**EEG preprocessing:** The EEG data for each participant was cut into 5-second non-overlapping epochs, to allow for robust covariance matrix estimation, and to be independent of the task trials. The epoched data was then referenced and filtered between 2-36 Hz using the *mne.filter()* function. Data points of each channel that exceeded 20 std of the robustly scaled data were clamped to the value equal to  $\pm 20$  std; for a detailed explanation of this method, see [30]. Finally, the data was resampled to 125Hz.

**Label creation:** The performance metrics used misses (Miss), reaction times (RT) and accuracies (Acc) of all subtasks. Across subtasks, all values were first normalized to give equal importance to each subtask. Next, averages of misses, RT and accuracies were calculated. For each value, the best and worst 33% were used to assign labels—the global performance (OVR) score combined misses, RT, and accuracy. Three different mental state-based labels were created. The time-on-task used the first and last 33% of each recording, respectively. The VAS scores were used for the other two approaches. Using the drowsiness and cognitive fatigue scales, the blocks corresponding to the most extreme values of each scale were used to label the data. Adjusted chance levels were calculated for each label-type based on [32].

**Classification:** The data from each participant was divided into an 80/20 split for training/testing datasets. Next, the covariances were computed using OAS or LWF,

and the data was projected to the target space. We compared performances of logistic regression (Log Reg), Support Vector Machine (SVM) and Random Forest (RF) Classifiers. Hyper-parameters for each classifier were optimized using 5-fold cross-validation using Bayesian search.

#### *Statistical Analysis:*

The general inference criterion is a p-value of  $p < .05$ . In multiple comparisons, we adjusted that criterion according to the adjusted Bonferroni method. Assumptions for each statistical test were checked and accounted for if not satisfied. Outlier detection was performed based on the interquartile range criterion. This was done for trials grouped by condition.

**Subjective:** To analyse the subjective results, we compared the SPS and KSS scores from the beginning to the end. For this, we used a paired samples T-Test. We also performed a one-way repeated measures ANOVA for both VAS scales. For one participant, the questionnaires at the end of the experiment were not recorded.

**Behavioral:** The behavioural analysis was divided into three sections for the (i) Search Task, (ii) Navigation Task - Heading Mode, and (iii) Navigation Task - Waypoint mode, respectively. An overall analysis was not possible due to the differences between tasks. Due to the randomized order of the tasks, missing data occurred in some blocks as single participants did not engage in a task in a given block. In this case, the missing values were replaced with the list-wise mean. This occurred in 0.55 % of the behavioural data.

For the search task, reaction times, F1 score and misses were used as dependent variables in repeated measures within-subjects ANOVA with Task (searching humans or searching vehicles) as an independent variable and time on task (19-minute blocks) for repeated measures.

To analyse the heading task reaction times, turning direction misses and deviation were the dependent variables of repeated measures within-subjects ANOVA with TOT (19-minute blocks) for repeated measures.

The Waypoint mode was evaluated using reaction time, correct choices, and misses as dependent variables, again TOT was the independent variable for repeated measures.

**EEG Frequency:** The extracted powerbands were compared across blocks in a repeated measures ANOVA for each cluster.

**Classification:** To analyze the classification results, a 2-way ANOVA with factors Classifier and Label-type was performed on the dependent variable of accuracy.

## RESULTS

*Subjective Data:* Scores for the KSS, RSME and SPS measures all showed significant increases in values comparing the beginning and the end of the experiment KSS:  $t(23) = -6.912, p < .001, d = -1.411$ ; SPS:  $Z = -4.000, p < .001, r = -0.933$ ; RSME:  $t(23) = -6.380, p < .001, d = -1.302$ . The assumption of normality was violated for the SPS test (Shapiro-Wilk

$W = 0.875$ ,  $p = 0.007$ ); therefore, the Wilcoxon result is reported. VAS scores on both cognitive fatigue and drowsiness showed linear increases over time (Cognitive:  $F(2.374, 49.849) = 25.979$ ,  $p < .001$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.553$  and  $\eta_p^2 = 0.553$ ; Drowsiness:  $F(4.294, 90.184) = 12.159$ ,  $p < .001$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.367$  and  $\eta_p^2 = 0.367$ , see Figure 2 a-d).

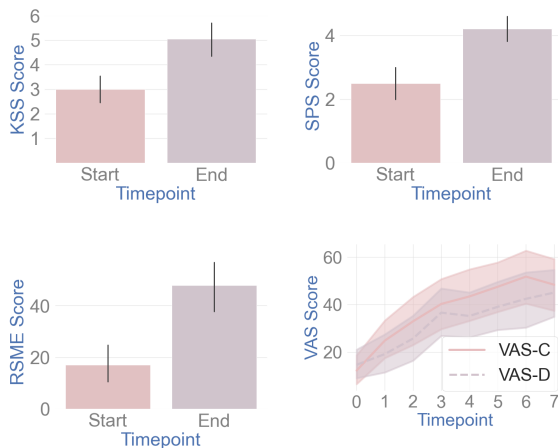


Figure 2: Subjective results: a) KSS Scores comparing the beginning and end of the experiment. b) SPS Scores comparing the beginning and end of the experiment. c) RSME Scores comparing the beginning and end of the experiment. d) VAS scores throughout the experiment for both the Cognitive Fatigue and the Drowsiness scale. Timepoint 0 is before the start and before the training, and then each subsequent point occurred every 19 minutes into the experiment. The last point occurred after the completion of the experimental phase.

#### Behavioral Data:

**SRC Task:** Searching humans resulted in significantly larger reaction times and more misses (RT:  $F(1, 24) = 214.872$ ,  $p < .001$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.900$  Misses:  $F(1, 24) = 34.308$ ,  $p < .001$ , Greenhouse-Geisser corrected,  $\eta^2 = .588$ ). Surprisingly, the F1 score was slightly higher for searching humans (F1:  $F(1, 24) = 108.011$ ,  $p < .001$ , Greenhouse-Geisser corrected,  $\eta^2 = .818$ ). Time did not significantly affect performance on any metric (RT:  $F(4.117, 120) = .381$ ,  $p = .827$ , Greenhouse-Geisser corrected,  $\eta^2 = .0160$  F1:  $F(4.056, 120) = 2.268$ ,  $p < 0.066$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.086$  Miss:  $F(2.024, ) = .751$ ,  $p = .479$ , Greenhouse-Geisser corrected,  $\eta^2 = .030$ ), see Figure 3 a-d.

**NAV task HDG mode:** RT, correct turn and misses were all influenced by time (RT:  $F(4.1, 60) = 41.711$ ,  $p < .005$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.635$ , correct turn:  $F(4.207, 60) = 3356$ ,  $p = .011$ , Greenhouse-Geisser corrected,  $\eta^2 = .123$ , miss:  $F(3.327, 60) = 4.185$ ,  $p = .007$ , Greenhouse-Geisser corrected,  $\eta^2 = .148$ ). Contrast analysis revealed significant cubic effects for RT, correct turn and misses (RT:  $F(5, 60) = 99.569$ ,  $p < .005$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.806$ ,

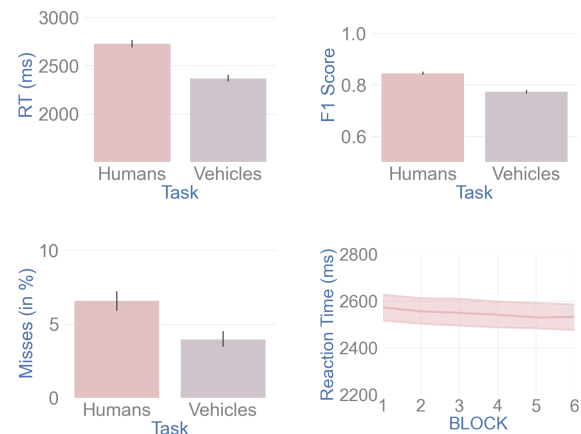


Figure 3: Behavioral Results of the SRC task: a) Reaction time by Mode. b) F1 scores by Mode. c) Misses by mode. d) Reaction Times over time

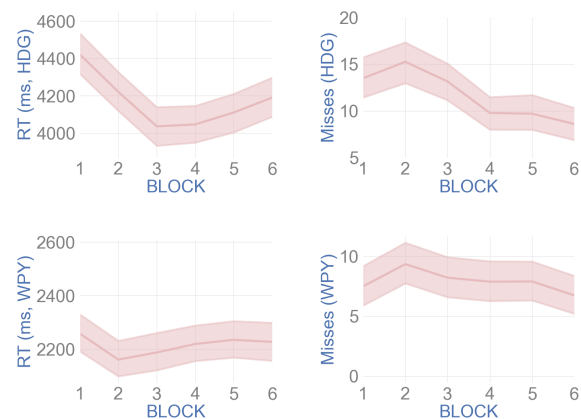


Figure 4: Behavioral Results of the NAV task: a) HDG: Reaction times over time b) HDG: Misses over time. c) WPY: Reaction times over time. d) WPY: Misses over time.

correct turn:  $F(5, 60) = 7.627$ ,  $p = .011$ , Greenhouse-Geisser corrected,  $\eta^2 = .231$ , miss:  $F(5, 60) = 5.253$ ,  $p = .031$ , Greenhouse-Geisser corrected,  $\eta^2 = .180$ ). See Figure 4 a & b.

**SRC task WPY mode:** Reaction times were also influenced by time (RT:  $F(3.078, 60) = 2.707$ ,  $p = .05$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.101$ ). Contrast tests revealed a significant cubic effect (RT:  $F(1, 60) = 10.395$ ,  $p = .004$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.302$ ). See Figure 4 c & d.

**EEG:** The alpha frequency band showed significant effects throughout the experiment. In all three clusters, an increase in alpha power was observed (Frontal :  $(F(1, 4.6) = 3.184$ ,  $p = .012$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.117$ ; Central:  $(F(1, 6.051) = 3.995$ ,  $p < .001$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.143$ ; Parieto-Occipital:  $(F(1, 4.899) = 2.508$ ,  $p = .035$ , Greenhouse-Geisser corrected,  $\eta^2 = 0.095$ ). Theta power did not show any significant change over time (see Figure 5 a-d).

**Classification:** The 2-way ANOVA showed a signifi-

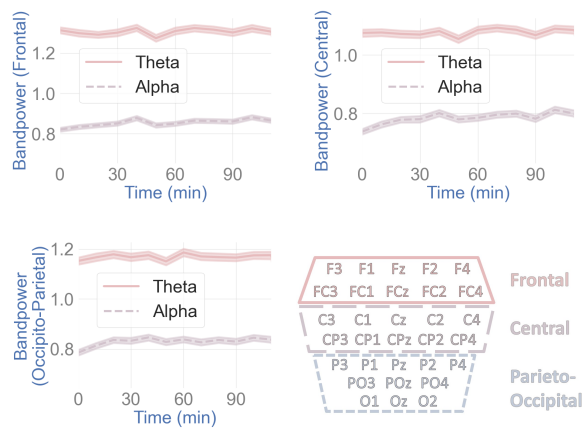


Figure 5: Frequency power in the theta and alpha power bands over time for the a) Frontal Cluster; b) Central Cluster; c) Parieto-Occipital Cluster. d) Topology of the 3 clusters.

cant effect of Label type ( $F(6,24) = 342.12$ ,  $p < .001$ ,  $\eta^2 = 0.81$ ). There was no significant effect of classifier ( $F(7,24) = 2.398$ ,  $p > .09$ ,  $\eta^2 = 0.01$ ). The highest classification accuracy was obtained using the TOT labels, with an average accuracy of 94.86% across all classifiers. However, both VAS scales, the performance labelling based on misses and accuracy, also performed above their respective chance levels. This resulted in a 62.09 % accuracy for detecting misses using the tangent space logistic regression (see Figure 6).

## DISCUSSION

To use pBCIs in complex environments, the output of a pBCI needs to have some predictive value. The results presented here compare mental-state estimation and performance estimation using EEG data. The subjective and EEG analyses both point toward increased mental fatigue over time. Yet, while some behavioural metrics, such as reaction times in the navigation task, seem to show a similar trend, variability in overall performance is not best explained by TOT.

While mental state estimation using Time-On-Task or subjective metrics as ground truth performs considerably above chance level, our algorithms could also predict misses with an above chance level likelihood. The algorithm's success with TOT metrics may be attributable to the observed alpha power increase often associated with mental fatigue. [31, 33]. It may also be due to slow drifts and the non-stationarity of the EEG signal [34]. Subjective fatigue scores increased over time, creating similarities between the TOT and VAS labels. The absence of stronger effects in the spectral analysis may have been attenuated due to the complexity of the task [35]. The reaction time-based performance estimation had the lowest accuracy. One possible explanation is that longer reaction times may reflect several processes that are then mixed up. Slow reaction times may be due to fatigue [36] or a speed-accuracy tradeoff [37]. The moderate success of the performance estimation based on misses suggests that

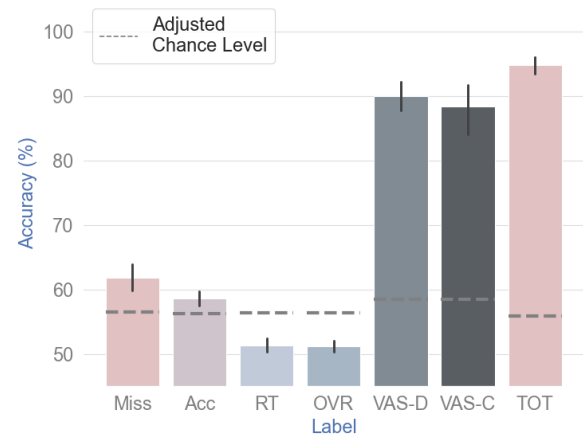


Figure 6: Classification results: Accuracy by Label Type with adjusted chance levels.

spectral EEG features, especially theta power, are sensitive to lapses [38]. Future work could evaluate incorporating Bayesian updating, which may further improve the performance estimation

## CONCLUSION

This study highlights the challenges and possibilities of EEG-based performance estimation. The differences between definitions of performance highlight the importance of label assignment. In our opinion, performance scores should (i) be defined a priori, (ii) be explainable, and (iii) provide real-world value.

## REFERENCES

- [1] Lotte F, Roy RN. Brain-computer interface contributions to neuroergonomics. In: Neuroergonomics. Elsevier, 2019, 43–48.
- [2] Krol, LR., Zander, TO. Passive Bci-Based Neuroadaptive Systems. 2017.
- [3] Aricò P, Borghini G, Di Flumeri G, Sciaraffa N, Colosimo A, Babiloni F. Passive BCI in Operational Environments: Insights, Recent Advances, and Future Trends. IEEE Transactions on Biomedical Engineering. 2017;64(7):1431–1436.
- [4] Trejo LJ, Kubitz K, Rosipal R, Kochavi RL, Montgomery LD. EEG-Based Estimation and Classification of Mental Fatigue. Psychology. 2015;06(05):572–589.
- [5] Zander TO *et al.* Automated Task Load Detection with Electroencephalography: Towards Passive Brain-Computer Interfacing in Robotic Surgery. Journal of Medical Robotics Research. 2017;02(01):1750003.
- [6] Young MS, Brookhuis KA, Wickens CD, Hancock PA. State of science: Mental workload in ergonomics. Ergonomics. 2015;58(1):1–17.
- [7] Phillips RO. A review of definitions of fatigue – And a step towards a whole definition. Transportation Research Part F: Traffic Psychology and Behaviour. 2015;29:48–56.

- [8] Heitz RP. The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*. 2014;8.
- [9] Kirchner WK. Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*. 1958;55(4):352.
- [10] Stroop JR. Studies of interference in serial verbal reactions. *Journal of experimental psychology*. 1935;18(6):643.
- [11] Hamann A, Carstengerdes N. Assessing the development of mental fatigue during simulated flights with concurrent EEG-fNIRS measurement. *Scientific Reports*. 2023;13(1):4738.
- [12] Linden D van der. The urge to stop: The cognitive and biological nature of acute mental fatigue. In: *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications*. APA: Washington, DC, US, 2011, 149–164.
- [13] Robert J. Hockey G. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*. 1997;45(1):73–93.
- [14] Müller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B. Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*. 2008;167(1):82–90.
- [15] Ko LW *et al*. Single channel wireless EEG device for real-time fatigue level detection. In: *IJCNN*. Jul. 2015.
- [16] Blaha LM, Fisher CR, Walsh MM, Veksler BZ, Gunzelmann G. Real-Time Fatigue Monitoring with Computational Cognitive Models. In: *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*. Springer International Publishing: Cham, 2016, 299–310.
- [17] Alhazmi S, Saini MK, El Saddik A. Multimedia Fatigue Detection for Adaptive Infotainment User Interface. In: *Human-Computer-Media Communication*. ACM: New York, NY, USA, Oct. 2015, 15–24.
- [18] Hinss MF, Brock AM, Roy RN. Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces. *Frontiers in Neuroergonomics*. 2022;3.
- [19] Hinss MF, Jahanpour ES, Brock AM, Roy RN. Labeling mental fatigue for passive BCI applications: Accuracy vs applicability tradeoff. *Bruxelles*, 2023.
- [20] Hinss MF, Brock AM, Roy RN. The double task-switching protocol: An investigation into the effects of similarity and conflict on cognitive flexibility in the context of mental fatigue. *PLOS ONE*. 2023;18(2):e0279021.
- [21] Hinss MF, Vitale VM, Phan NT, Roy RN, Brock AM. UASOS: An Experimental Environment For Assessing Mental Fatigue & Cognitive Flexibility During Drone Operations. ACM: Boulder Colorado, USA, 2024.
- [22] Veale JF. Edinburgh Handedness Inventory - Short Form: A revised version based on confirmatory factor analysis. *Laterality*. 2014;19(2):164–177.
- [23] Samn SW, Perelli LP. “Estimating Aircrew Fatigue: A Technique with Application to Airlift Operations.” Brooks AFB, TX: USAF School of Aerospace Medicine. Technical Report SAM-TR-82-21. 1982.
- [24] Åkerstedt T, Gillberg M. Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*. 1990;52(1-2):29–37.
- [25] Ghanbary A, Ashnagar M, Habibi E, Sadeghi S. Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses. *Journal of Occupational Health and Epidemiology*. 2016;5(4):211–217.
- [26] Miley A, Kecklund G, Åkerstedt T. Comparing two versions of the Karolinska Sleepiness Scale (KSS). *Sleep and Biological Rhythms*. 2016;14(3):257–260.
- [27] ICAO. “Manual for the Oversight of fatigue Management Approaches.” Tech. Rep. Doc 9966. 2016.
- [28] Jasper H. The ten-twenty electrode system of the international federation \textbar CiNii Research. *Electroencephalogr. Clin. Neurophysio*. 1958;10, 370–375.
- [29] Kothe C. Scnn/labstreaminglayer. original-date: 2018-02-28T10:50:12Z. 2015. (visited on 01/11/2023).
- [30] Défossez A, Caucheteux C, Rapin J, Kaveli O, King JR. Decoding speech perception from non-invasive brain recordings. *Nature Machine Intelligence*. 2023;5(10):1097–1107.
- [31] Wascher E *et al*. Frontal theta activity reflects distinct aspects of mental fatigue. *Biological Psychology*. 2014;96:57–65.
- [32] Mueller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: A closer look on BCI results. *International Journal of Bioelectromagnetism*. 2008.
- [33] Boksem MA, Meijman TF, Lorist MM. Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research*. 2005;25(1):107–116.
- [34] Urigüen JA, Garcia-Zapirain B. EEG artifact removal—state-of-the-art and guidelines. *Journal of Neural Engineering*. 2015;12(3):031001.
- [35] Kamzanova AT, Matthews G, Kustubayeva AM, Jakupov SM. EEG Indices to Time-On-Task Effects and to a Workload Manipulation (Cueing). *International Journal of Psychological and Behavioral Sciences*. 2011;5(8):928–931.
- [36] Csathó, Linden D vd, Hernádi I, Buzás P, Kalmár G. Effects of mental fatigue on the capacity limits of visual attention. *Journal of Cognitive Psychology*. 2012;24(5):511–524.
- [37] Wood CC, Jennings JR. Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. *Perception & Psychophysics*. 1976;19(1):92–102.
- [38] Peiris MTR, Jones RD, Davidson PR, Carroll GJ, Bones PJ. Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects. *Journal of Sleep Research*. 2006;15(3):291–300.

# LESS IS MORE: ADVANCING EEG-BASED ONLINE CONTINUOUS MACHINE ERROR DETECTION WITH THE LIGHTWEIGHT MAX-MIN AMPLITUDE NOISE FILTERING TECHNIQUE

Y. Pan, L. Rabe, M. Klug

Chair of Neuroadaptive Human-Computer Interaction, Brandenburg University of Technology  
Cottbus-Senftenberg, Cottbus, Germany  
Young Investigator Group – Intuitive XR, Brandenburg University of Technology Cottbus-  
Senftenberg, Cottbus, Germany

E-mail: yanzhao.pan06@gmail.com

**ABSTRACT:** To apply synchronous laboratory passive brain-computer interface (pBCI) systems to dynamic real-world scenarios, it is essential to develop asynchronous, event-independent pBCIs that can continuously interpret brain activity. Minimizing false alarms (FAs) caused by artifacts in continuous online sessions without compromising the hit rate is one of the primary challenges in EEG-based brain activity classification tasks. To address this challenge, this study introduces the Max-Min Amplitude Noise Filtering (MANF) technique, which is designed to reduce FAs in the online EEG-based machine error detection task. To achieve this, we pre-trained a classifier on labeled data and then tested the performance of the technique on a simulated continuous online classification. The MANF technique, using a predetermined noise threshold, simplifies the noise filtering process by comparing the difference between the maximal and minimal amplitude of incoming EEG data against this threshold, substantially reducing FAs while maintaining high hit rate. This technique outperforms the unfiltered condition and those using the Artifact Subspace Reconstruction technique, achieving an optimal balance between sensitivity and specificity with medium and conservative thresholds. Highlighting the "less is more" principle, the MANF technique proves highly suitable for continuous online pBCI applications. This development contributes to the ongoing efforts in creating more user-friendly and reliable pBCIs for dynamic real-world use.

## INTRODUCTION

Passive brain-computer interfaces (pBCIs) derive the output from ongoing brain activity, enriching the human-machine interaction by integrating implicit information on the actual user's intentions and emotional states into technical systems [1]. In recent years, electroencephalogram (EEG)-based pBCIs have been used in various scenarios [2–5], particularly notable in the use of error-related potentials (ErrPs) as implicit feedback. This approach has been validated as feasible in various use cases, e.g. BCI speller [6–8],

cursor control [2,9], and improvement of robot control [10–12]. However, most paradigms involved are lab-based and time-locked to specific events. To bridge the gap between synchronous laboratory pBCI systems and dynamic real-world usage, it is essential to develop asynchronous, event-independent pBCIs that can continuously interpret brain activity, ensuring more natural and seamless human-machine interactions. A few studies have investigated continuous, asynchronous error detection [13–15], with the aim of reducing false alarms (FAs) caused by artifacts in continuous online sessions while maintaining precise error detection.

When exploring the real-time classification of EEG signals in dynamic environments or during intense physical activities, artifact interference is a primary obstacle [16,17]. Research in Mobile Brain/Body Imaging [18,19] has highlighted the efficacy of offline artifact correction techniques, such as Independent Component Analysis (ICA), for cleaner signal analysis [20]. Nonetheless, the adaptability of ICA for online application is limited. Novel approaches, including Artifact Subspace Reconstruction (ASR) [21] and Online Recursive ICA [22], show potential in certain online scenarios. However, their performance in continuous, asynchronous online classification, especially considering computational demands, remains unexplored. Therefore, further research is necessary to evaluate the performance of these algorithms or explore new techniques to reduce the effects of artifacts in continuous online application.

This study aims to fill this gap by designing and evaluating a continuous online classification approach that incorporates a novel noise filtering technique. In the context of tactile-based machine error detection, we have implemented a comprehensive methodology that includes feature extraction, class re-balancing, Support Vector Machine (SVM) classifier training, and a simple noise filtering technique. The results indicate that our methodology achieves good performance in the simulation of online continuous classification. Furthermore, through comparative analyses of different noise filtering conditions, we have demonstrated that our simplistic noise filtering technique outperforms both unfiltered conditions and those using the ASR

technique. This study highlights the “less is more” principle in controlled continuous online classification scenarios. Therefore, our contributions are twofold: we not only validate the feasibility of a minimalist approach for online machine error detection but also offer insights for future research and applications in EEG-based continuous online classification sessions.

## MATERIALS AND METHODS

*Materials overview:* This study is based on an open-source dataset (<https://zenodo.org/records/8345429>) [23], which was used for the IJCAI 2023 competition (<https://ijcai-23.dfki-bremen.de/competitions/inter-hri/>), supported by German Research Center for Artificial Intelligence, Robotics Innovation Center. The dataset contains recordings of the EEG data from eight subjects who were assisted in moving their right arm by an active orthosis. For each participant, 8 labeled single-trial training sets and 2 unlabeled test sets are included. The training sets contain EEG data and all the event markers across the whole experimental sessions, while the test sets contain only continuous EEG data streams and markers indicating the onset of introduced errors.

*Participants:* The experiment involved 8 healthy right-handed volunteers (4 males and 4 females) with an average age of 21.8 years. Before the experiment, they attended a short session in the lab for an introduction and preliminary tests, which included fitting the orthosis and determining the EEG cap size according to head circumference. All participants were informed of their rights including voluntary withdrawal. The experiment lasted an average of 4.9 hours (SD = 0.6 hours), and participants were remunerated at a rate of 10 euros per hour.

*Experimental setup and procedure:* Participants were equipped with a 64-channel EEG system and wore an active orthosis on their right arm while holding an air-filled ball in their left hand. The orthosis facilitated the participant's arm movements through a sequence of trials that included both flexion and extension. Certain errors were induced for a short duration of time during these movements. An error was defined as a short-term (250 ms) alteration in the direction of the orthosis's movement. For example, if the orthosis was amid executing a flexion movement, an error would cause a temporary switch to extension before resuming the original flexion path. Similarly, during an extension, it would momentarily change to flexion. The participants' primary task was to identify these errors in the orthosis's operation. The experiment's initial run aimed to establish a baseline with 30 movements without any errors. This was followed by a training session to familiarize the participants with the sensation of the error and the corresponding response — squeezing the ball in the left hand. During the experiment, 6 errors were randomly introduced among 30 movement trials across 10 runs, with the sequence of errors varied in each run. To reduce the artifacts in the data, participants were asked to maintain specific postures and gaze

directions. The experiment was designed to elicit a total of 480 error detection responses, calculated from 6 errors in each of the 10 runs, across all 8 participants. The timing of the stimulus onset for the trials without errors was determined by averaging the onset times from the trials that contained errors.

*Data acquisition:* EEG data collection was performed using the 64-channel LiveAmp system paired with the ActiCap slim electrode setup, adhering to an extended 10-20 layout, both supplied by Brain Products GmbH. The reference and ground electrodes were placed at FCz and AFz, respectively. The impedance for all 64 electrodes was consistently kept below 5 k $\Omega$ . Data sampling was at 500 Hz using Brain Products GmbH's Recorder software (version 1.25.0001), which applied hardware filters to limit the data's frequency range to 0.0 Hz to 131.0 Hz. The acquired EEG data was organized in the BrainVision Core Data Format 1.0, comprising three essential files: a binary data file (.eeg), a header file (.vhdr), and a marker file (.vmrk). Within each participant's EEG data folder, the marker files (.vmrk) recorded all critical events during the experiments. Markers for the start of flexion and extension movements were coded as S64 and S32, respectively. Error-free trials are indicated by S48 markers, placed around the calculated mean onset of errors from the trials that contain errors. The introduction of an error in a trial was marked by S96, while the participant's action of squeezing the ball was recorded as an S80 event in the marker file.

*Task description and evaluation metrics:* For 10 runs from each participant, 8 serve as labeled single-trial training sets for training machine learning models capable of detecting the onset of the deliberately introduced errors in the data, while 2 serve as unlabeled test sets without event markers for model evaluation. During the evaluation phase, a buffer-like sliding window moves through the EEG data along the temporal dimension to simulate continuous online data acquisition. A binary classification of event types (error or non-error) is performed at each position of the sliding window. The predicted error onsets are then compared with true error onsets for evaluation. A predicted error onset that occurs within a 1000 ms window following the true error onset is defined as a hit or true positive. Conversely, any predicted error onset outside this 1000 ms window is defined as a FA or false positive. It is important to note that within the 1000 ms window following a true error onset, multiple error predictions are collectively counted as a single hit, while each FA is included in the cumulative count of FAs. The evaluation metrics consist of the total number of hits in the 2 test sets for all 8 participants (with a maximum of 96), the average number of FAs across the participants, and the average FA rate across the participants. The average FA rate is calculated as the total number of FAs divided by the total number of non-error epochs across the 2 test sessions, averaged across participants.

*Preprocessing:* For the training sets, EEG data was preprocessed by re-referencing to an average reference

and applying a zero-phase, non-causal Hamming windowed-sinc FIR highpass (0.1 Hz passband edge, 0.1 Hz transition bandwidth, 0.05 Hz cutoff frequency (-6db)) and lowpass filter (15 Hz passband edge, 3.75 Hz transition bandwidth, 16.875 Hz cutoff frequency (-6db)) in succession, using the EEGLAB `pop_eegfiltnew` function. The data was then segmented into epochs of interest, ranging from 100 ms before to 800 ms after the stimulus onset. Baseline correction was applied to each epoch, using the interval from 100 ms before the stimulus to its onset. The error epochs were organized into a [64, 400, 48] matrix indicating the number of channels, time points per epoch, and total error epochs, respectively, derived from 8 training sets per participant, with each set containing 6 error epochs. Similarly, the error-free epochs were organized into a [64, 400, 192] matrix, derived from 8 training sets per participant, with each set containing 24 error-free epochs.

**Feature extraction and classifier training:** Features were extracted using 50 ms non-overlapping moving windows within the [0-800] ms post-stimulus period across all 64 channels. In each window, the mean value was calculated, resulting in 1024 features per epoch for classification. To address the class imbalance issue, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to enhance the representation of the minority class (error) by generating synthetic samples through interpolation with neighboring instances, thereby equalizing the number of epochs in both classes within the feature matrix [24]. The resulting balanced feature set contained 384 epochs (192 error and 192 non-error epochs), each characterized by a 1024-element feature vector. This feature set served as the input for training the SVM classifier [25]. The SVM model was configured with a linear kernel and a regularization parameter set to 1.0. To validate the model's performance, a 10-fold cross-validation was carried out individually for each participant.

**Continuous online classification simulation:** During the online classification simulation phase, a buffer-like sliding window moved through the test EEG data along the temporal dimension, simulating the process of continuous real-time data acquisition. Spanning 900 ms, this window covered the same time range used during the offline training phase for feature extraction (including 100 ms baseline range). With a step size of 20 ms, the window continuously “fetched” EEG data, ensuring a seamless and overlapping coverage of the incoming EEG data. The preprocessing approach and feature set chosen for classification in this phase were consistent with those applied in the offline training phase. To reduce the FAs in the continuous classification, we implemented a lightweight noise-filtering technique, termed the “Max-Min Amplitude Noise Filtering (MANF) Technique”. The noise level in each window was evaluated and compared to a predetermined threshold. Epochs with a noise level above this threshold were considered “noisy” and directly classified as non-error (0). Finally, the trained

#### Max-Min Amplitude Noise Filtering (MANF) Technique Process

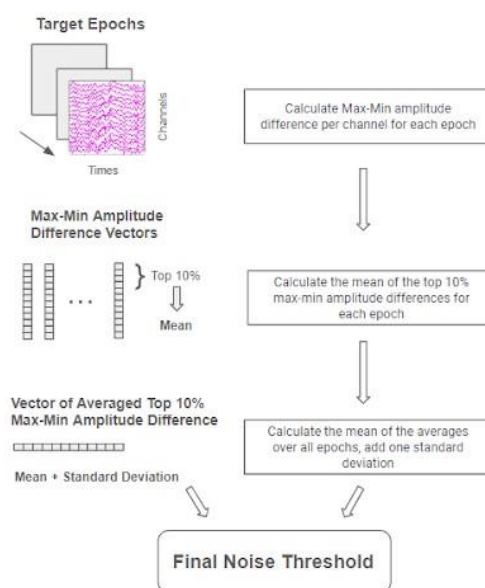


Figure 1: Max-Min Amplitude Noise Filtering (MANF) technique process. This process diagram illustrates the process of determining the noise threshold using target epochs within training sets. First, the max-min amplitude difference within each target epoch is calculated for each channel. Then, the average of the highest 10% of these differences is calculated for each epoch. The final noise threshold is derived by averaging these means over all target epochs and adding a variable number of standard deviations to this average.

SVM classifier continuously provided predictions for each position of the sliding window, where a prediction of 1 indicated the detection of an error, and 0 indicated its absence. Furthermore, we replicated the online simulation using the ASR technique [21] with the ‘clean\_asr’ function in EEGLAB. The evaluation of both the MANF and ASR techniques included three conditions: aggressive, medium, and conservative, each reflecting a different level of noise tolerance. The aggressive strategy used a lower threshold, resulting in more stringent noise filtering, while the conservative strategy adopted a higher threshold, allowing for less stringent filtering. The medium strategy maintained a balance between these two extremes. To implement these strategies, specific hyperparameters were tuned within each technique. For the MANF technique, noise thresholds were set at one, two, and three standard deviations above the mean noise level. For ASR, deviation cutoffs of 5, 20, and 30, relative to calibration data, defined the respective aggressive, medium, and conservative conditions. In addition, a baseline condition without any noise filtering was also evaluated to provide a comparative benchmark.

**MANF technique:** As shown in Figure 1, a threshold is calculated by examining max-min amplitude differences using the target epochs in the training sets. To this end, first, the max-min amplitude difference

Table 1: Online classification simulation results across different noise filtering conditions.

Condition	Total Hits	Average FAs (Mean $\pm$ SD)	Average FA Rate (Mean $\pm$ SD)
Unfiltered (baseline)	88	957.3 $\pm$ 308.6	0.037 $\pm$ 0.011
MANF aggressive	80	416.6 $\pm$ 158.9	0.016 $\pm$ 0.006
MANF medium	86	539.5 $\pm$ 186.2	0.021 $\pm$ 0.007
MANF conservative	87	620.8 $\pm$ 207.5	0.024 $\pm$ 0.008
ASR aggressive	26	27.8 $\pm$ 22.1	0.001 $\pm$ 0.001
ASR medium	81	521.1 $\pm$ 192.0	0.020 $\pm$ 0.007
ASR conservative	86	653.6 $\pm$ 169.5	0.025 $\pm$ 0.006

within each target epoch is calculated for each channel. Then, the mean of the top 10% max-min amplitude differences (6 channels/values) is calculated for each target epoch, and last, the mean of those means across all target epochs, plus a variable number of standard deviations, is set as the final noise threshold. In the online simulation phase, epochs that contain mean max-min amplitude differences in their top 10% channels above this threshold are considered “noisy” and excluded from being classified as error epochs.

## RESULTS

The pre-trained classifier's performance was evaluated in a 10-fold cross-validation. The average balanced accuracy across all participants was  $91.3\% \pm 4.5\%$  (mean  $\pm$  SD), which is significantly above the chance level of 0.5 (significance with  $\alpha = 0.001$  would have been reached with 73.68% correct classification, see [26]). The average true positive rate was  $84.9\% \pm 7.8\%$  (mean  $\pm$  SD), and the average true negative rate was  $97.7\% \pm 1.4\%$  (mean  $\pm$  SD).

In the continuous online classification simulation phase, 16 test sets, with 2 for each of the 8 participants, were used for online simulation. The average count of non-error epochs across all participants was  $26027.0 \pm 1733.5$  (mean  $\pm$  SD). The unfiltered condition served as a baseline, yielding the highest total hits and average FAs, as shown in Table 1. When implementing noise filtering using the MANF and ASR techniques with varying hyperparameters, there was a notable decrease in the average FAs. However, this improvement was accompanied by a corresponding reduction in total hits. Figure 2 illustrates the percentage changes in hit rate and average FA rate when applying both noise filtering techniques across three noise tolerance levels. Both techniques substantially reduced the FA rate, with only a relatively minor decrease in the hit rate. The noise filtering strategy that was most aggressive in its approach was the most effective in reducing FAs.

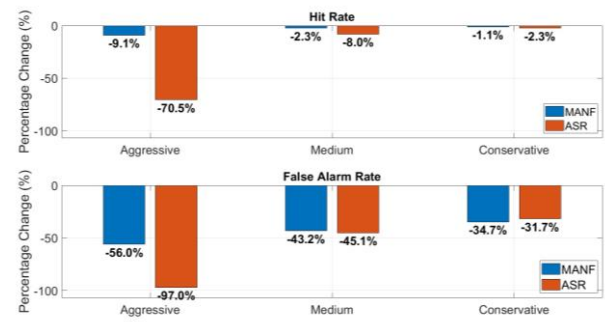


Figure 2: Impact of noise filtering techniques on hit rate and FA rate. This figure compares the percentage changes in hit rate and average FA rate resulting from the application of two noise filtering techniques (MANF and ASR) across three levels of noise tolerance: aggressive, medium, and conservative.

However, it also resulted in the largest decline in hit rate. This was particularly noticeable when using the ASR technique in its aggressive application, where there was a 97.0% reduction in FAs but an unacceptable 70.5% drop in hit rate. The medium and conservative conditions produced a more balanced outcome, with both techniques achieving similar reductions in FAs (43.2% and 34.7% for MANF, and 45.1% and 31.7% for ASR, under medium and conservative conditions, respectively). Notably, the MANF technique outperformed the ASR in preserving the hit rate, showing only a 2.3% reduction compared to 8.0% under the medium condition, and a 1.1% reduction compared to 2.3% under the conservative condition. To demonstrate the process of continuous error prediction in the online simulation, we display the error predictions and true error onsets over time using a test set from one participant as an illustrative example (see Figure 3). This visualization is presented for both the baseline condition and the MANF condition, across three different noise threshold levels.

## DISCUSSION

This study investigates the issue of artifact interference in continuous online classification of EEG signals, with a focus on machine error detection. We introduced and evaluated a simple noise filtering technique and found it to be superior to the unfiltered condition and those using the ASR technique in terms of reducing FAs while maintaining the hit rate. Specifically, our noise filtering technique achieved a 43.2% reduction in FA rate with only a 2.3% decrease in hit rate using a medium noise threshold, and a 34.7% reduction in FA rate with a minimal 1.1% decrease in hit rate using a conservative noise threshold. The comparative analysis of different noise filtering conditions highlights an essential consideration in the design of noise filtering strategies: the trade-off between sensitivity and specificity. While aggressive noise filtering effectively minimizes FAs, it may also inadvertently filter out genuine signals, leading to missed detections. Especially when using an

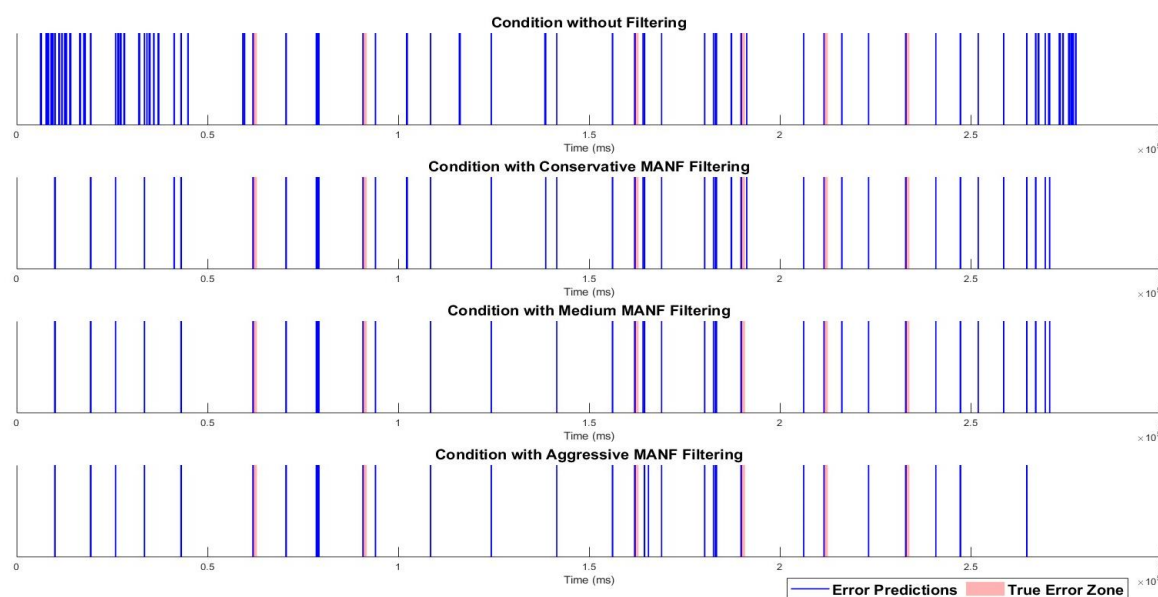


Figure 3: Visualization of continuous error prediction in online simulation. This figure displays the temporal distribution of error predictions (indicated by blue vertical lines) and true error onsets (highlighted with light red shadows) for both the baseline and MANF-filtered conditions across three different noise threshold levels. The data presented here is specific to participant “AQ59D”, using test set 6 as an illustrative example.

aggressive cutoff with the ASR technique, this approach tends to remove a significant portion of valuable signal features. As a result, the signals become flattened and lose their distinctive characteristics. Conversely, less stringent filtering preserves more hits but at the risk of higher FA rates. Therefore, for practical applications, it is recommended to use the medium or conservative strategies of the MANF technique. These strategies effectively minimize FAs while preserving a satisfactory hit rate, making them a preferable choice over the ASR technique or no filtering. Besides, the simplicity of the MANF technique offers an additional advantage. This technique simplifies the noise detection process by calculating the difference between the maximum and minimum EEG amplitudes for new incoming data streams in online sessions, and then comparing this calculated difference with a predetermined noise threshold value. This computational efficiency makes the MANF technique particularly suitable for continuous online applications, providing a computationally lightweight and effective solution for enhancing pBCI systems. In essence, this approach perfectly illustrates the “less is more” principle.

The implications of our findings extend beyond the context of tactile-based machine error detection and shed light on the broader domain of asynchronous, continuous online EEG-based classification tasks. The results suggest that the MANF technique shows potential in developing more intuitive and user-friendly pBCI systems, such as assistive devices. By effectively reducing FAs, it promises smoother and more reliable human-machine interactions. However, it is important to note that the MANF technique primarily focuses on excluding noisy epochs rather than correcting artifacts,

which might limit its effectiveness in dynamic environments with substantial movement. Advancing artifact correction techniques in such environments is still crucial. Furthermore, the testing phase of our study was conducted through an online simulation. Therefore, future research should evaluate the performance of the MANF technique in real online sessions and across diverse EEG-based applications, especially in complex or unpredictable environments. Additionally, further investigation into the optimization of noise filtering parameters, possibly through machine learning algorithms or adaptive filtering techniques, could yield even more effective and flexible solutions.

## CONCLUSION

This study demonstrates that the MANF noise filtering technique can substantially reduce FAs while maintaining a high hit rate, outperforming both unfiltered conditions and those using the ASR technique in continuous online EEG signal classification. This approach, emphasizing the “less can be more” principle, offers a computationally efficient solution for enhancing pBCI systems, particularly in applications requiring continuous, real-time interaction. Our work contributes to the field by validating a minimalist yet effective strategy for online machine error detection and providing a foundation for future research in EEG-based continuous online classification. Looking forward, it is essential to explore the generalizability of the MANF technique in dynamic environments and further refine noise filtering parameters to broaden its applicability and effectiveness.

## REFERENCES

- [1] T. O. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general," *J. Neural Eng.*, vol. 8, no. 2, p. 025005, Apr. 2011.
- [2] T. O. Zander, L. R. Krol, N. P. Birbaumer, and K. Gramann, "Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 52, pp. 14898–14903, Dec. 2016.
- [3] F. Dehais et al., "Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions," *Sensors*, vol. 19, no. 6, Mar. 2019, doi: 10.3390/s19061324.
- [4] A. Mora-Sánchez, A.-A. Pulini, A. Gaume, G. Dreyfus, and F.-B. Vialatte, "A brain-computer interface for the continuous, real-time monitoring of working memory load in real-world environments," *Cogn. Neurodyn.*, vol. 14, no. 3, pp. 301–321, Jun. 2020.
- [5] T. O. Zander and S. Jatzev, "Context-aware brain-computer interfaces: exploring the information space of user, technical system and environment," *J. Neural Eng.*, vol. 9, no. 1, p. 016003, Dec. 2011.
- [6] M. Spüler, M. Bensch, S. Kleih, W. Rosenstiel, M. Bogdan, and A. Kübler, "Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI," *Clin. Neurophysiol.*, vol. 123, no. 7, pp. 1328–1337, Jul. 2012.
- [7] M. Bevilacqua, S. Perdakis, and J. D. R. Millan, "On Error-Related Potentials During Sensorimotor-Based Brain-Computer Interface: Explorations With a Pseudo-Online Brain-Controlled Speller," *IEEE Open J Eng Med Biol*, vol. 1, pp. 17–22, Feb. 2020.
- [8] P. Margaux, M. Emmanuel, D. Sébastien, B. Olivier, and M. Jérémie, "Objective and Subjective Evaluation of Online Error Correction during P300-Based Spelling," *Advances in Human-Computer Interaction*, vol. 2012, Dec. 2012, doi: 10.1155/2012/578295.
- [9] O. E. Krigolson, C. B. Holroyd, G. Van Gyn, and M. Heath, "Electroencephalographic correlates of target and outcome errors," *Exp. Brain Res.*, vol. 190, no. 4, pp. 401–411, Oct. 2008.
- [10] S. K. Kim, E. A. Kirchner, A. Stefes, and F. Kirchner, "Intrinsic interactive reinforcement learning – Using error-related potentials for real world human-robot interaction," *Sci. Rep.*, vol. 7, no. 1, pp. 1–16, Dec. 2017.
- [11] S. K. Ehrlich and G. Cheng, "A Feasibility Study for Validating Robot Actions Using EEG-Based Error-Related Potentials," *International Journal of Social Robotics*, vol. 11, no. 2, pp. 271–283, Apr. 2019.
- [12] A. F. Salazar-Gomez, J. DelPreto, S. Gil, and D. Rus, "Correcting robot mistakes in real time using EEG signals," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, unknown, May 2017, pp. 6570–6577.
- [13] C. Lopes-Dias et al., "Online asynchronous detection of error-related potentials in participants with a spinal cord injury using a generic classifier," *J. Neural Eng.*, vol. 18, no. 4, p. 046022, Mar. 2021.
- [14] J. Omedes, I. Iturrate, and L. Montesano, "Asynchronous detection of error potentials," in *Proceedings of the 6th Brain-Computer Interface Conference 2014*, 2014.
- [15] M. Spüler and C. Niethammer, "Error-related potentials during continuous feedback: using EEG to detect errors of different type and severity," *Front. Hum. Neurosci.*, vol. 9, p. 155, Mar. 2015.
- [16] A. Tandle, N. Jog, P. D'cunha, and M. Chheta, "Classification of artefacts in EEG signal recordings and EOG artefact removal using EOG subtraction," *Commun. Appl. Electron.*, vol. 4, no. 1, pp. 12–19, Jan. 2016.
- [17] D. Gorjan, K. Gramann, K. De Pauw, and U. Marusic, "Removal of movement-induced EEG artifacts: current state of the art and guidelines," *J. Neural Eng.*, vol. 19, no. 1, Feb. 2022, doi: 10.1088/1741-2552/ac542c.
- [18] S. Makeig, K. Gramann, T.-P. Jung, T. J. Sejnowski, and H. Poizner, "Linking brain, mind and behavior," *Int. J. Psychophysiol.*, vol. 73, no. 2, pp. 95–100, Aug. 2009.
- [19] K. Gramann et al., "Cognition in action: imaging brain/body dynamics in mobile humans," *Rev. Neurosci.*, vol. 22, no. 6, pp. 593–608, Nov. 2011.
- [20] T. P. Jung et al., "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, Mar. 2000.
- [21] C. A. E. Kothe and T.-P. Jung, "Artifact removal techniques with signal reconstruction," 20160113587:A1, Apr. 28, 2016 Accessed: Mar. 06, 2024. [Online]. Available: <https://patentimages.storage.googleapis.com/c3/6b/dc/7dadfae33c0062/US20160113587A1.pdf>
- [22] S.-H. Hsu, T. R. Mullen, T.-P. Jung, and G. Cauwenberghs, "Real-Time Adaptive EEG Source Separation Using Online Recursive Independent Component Analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 3, pp. 309–319, Mar. 2016.
- [23] N. Kueper et al., "EEG and EMG dataset for the detection of errors introduced by an active orthosis device," *arXiv [cs.HC]*, May 19, 2023. [Online]. Available: <http://arxiv.org/abs/2305.11996>
- [24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *jair*, vol. 16, pp. 321–357, Jun. 2002.
- [25] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [26] G. R. Mueller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller, "Better than Random? A closer look on BCI results," *Int. J. Bioelectromagn.*, vol. 10, no. 1, pp. 52–55, Jan. 2008.

# ANA-E: A NOVEL APPROACH FOR PRE-TRAINED ERROR DETECTION MODELS IN BRAIN-COMPUTER INTERFACES

Alexandros Christopoulos<sup>1</sup>, Matias Valdenegro-Toro<sup>1</sup>, Andreea Ioanna Sburlea<sup>1</sup>

<sup>1</sup> Department of Artificial Intelligence, Faculty of Science and Engineering, University of Groningen, Groningen, The Netherlands.

E-mail: a.i.sburlea@rug.nl

**ABSTRACT:** Error-related potentials hold the potential to enhance self-correcting behaviors in Brain-Computer Interfaces (BCIs), pivotal for human-machine interactions. However, integrating error detection mechanisms poses challenges, notably in lengthy calibration sessions required for different BCI modules. To address this, we propose a novel approach using Self-Supervised Learning (SSL) with an autoencoder architecture, called Ana-E, to develop pre-trained error detection pipelines. We recorded EEG data from participants navigating a game scenario imposed with errors. Offline analyses within and between participants were conducted for both pre-processed EEG trials and Ana-E features with two classifiers. Within-participants analysis showed comparable performance between Ana-E features and EEG trials. While in between-participants analysis, Ana-E exhibited an 8% performance improvement (72%) over the second-best pipeline (64%). Our study offers valuable insights into the future of pre-trained models for error detection in BCIs, providing a baseline for more complex architectures with the goal of significantly enhancing BCI usability and reducing dependency on calibration sessions, thereby improving user experience and applicability.

## INTRODUCTION

In the field of Brain-Computer Interfaces (BCIs), Error-related Potentials (ErrPs) have been utilized as error detection instruments to expand the usability of architectures and develop a smoother experience between the user and external device [1]. Such implementations have been applied both in offline [2] and online paradigms [3]. Although the integration of errors as corrective instruments could improve the usability of BCIs as assistive tools, an immediate challenge emerges. Different BCI modules would require multiple calibration sessions [4], making the existence of pretrained models a requirement for self-correcting BCI implementation to prove applicable. In this paper, we attempt to develop pretrained models by adapting an autoencoder architecture to conduct a Self-Supervised Learning (SSL) task.

SSL reflects a subset of unsupervised learning methods in which neural networks are trained with automatically generated labels (pretext task) and then tested on a supervised task (downstream task), where human annotations

are utilized to evaluate the performance of the model [5]. SSL relies on the premise that input information has distinguishable characteristics, and learned feature representations from the pretext task can be transferred to the downstream task [5]. SSL has been used successfully in visual feature learning tasks like image colorization [6], temporal order verification [7], and visual-audio correspondence verification [8]. SSL methods have further been deployed for time series data [9], where a common method is that of masked autoencoders, which randomly mask patches of the original time series data and learn temporal dynamics by recovering the masked patches [10]. Recently, SSL methods were implemented in EEG data for sleep stage recognition and pathology detection, outperforming purely supervised deep neural networks in low-labeled cases [11].

Autoencoders represent an unsupervised learning technique where the core idea is to conduct a representation learning task [12]. To do so, a deterministic encoder-decoder network pair is trained to learn a feature vector, often referred to as a 'bottleneck,' capable of encoding the underlying structural characteristics of the input samples. The learned feature vector could then be used by the decoder to fully reconstruct the input data samples [13].

In this paper, we used a 1D convolutional autoencoder architecture for EEG reconstruction that we coined as Ana-E. After training our model on EEG reconstructions, we extracted the encoder part and used it as a feature extractor of EEG trials, which we then fed to a classification head (CH). Our goal was to develop an architecture capable of deconstructing and reconstructing the input EEG as our pretext task. We then expect that the learned features from our encoder would be robust enough to classify errors in human participants in a downstream task. By doing so, we hope to address the issue of pre-trained BCI models. To investigate the novelty of our approach, we tested our architecture both within-participants and between-participants.

## MATERIALS AND METHODS

*Participants:* We recorded 10 participants (6 females) with a mean age of 22 years (SD = 2.3), each undergoing a single recording session. One participant was removed due to incomplete markers. Participants were recruited

via advertisement fliers and compensated at a rate of 8 euros per hour. They provided informed consent and were informed of their right to withdraw at any time during the experiment. Ethical approval was obtained from the Ethical Committee of the Faculty of Arts, University of Groningen, The Netherlands (ID 92123476).

**Procedure:** Upon arrival, participants were introduced to the laboratory, briefed on the study, and signed informed consent forms. EEG cap placement took approximately 30 minutes. Following this, participants underwent a brief training session (5-10 minutes) to familiarize themselves with the experimental paradigm, including game rules and controls. After training, participants explained the game rules to researchers and began playing, with the game duration lasting 60-90 minutes. Rest periods were provided between trials as needed. The total experiment duration ranged from 120 to 150 minutes.

**EEG Recordings:** Participants' EEG was recorded with antiCAP slim/snap 32 gel based active electrodes according to the 10-20 international system with a sampling rate of 500Hz using the LiveAmp BrainProducts amplifier. The measured EEG channels were: FP1, FPz, FP2, AF3, AFz, AF4, F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP5, CP1, CPz, CP2, CP6, P3, Pz, P4, O1, Oz, O2. Ground and reference electrodes were placed on the left and right mastoid, respectively. EEG was recorded with the Brainvision recorder. Finally, impedance of electrodes was kept below 20  $\Omega$  for all participants.

**Experimental Procedure:** The experiment was developed in the Unity game engine [14] and had the code name Honey Heist. The experiment consisted of two phases: the training phase (approximately 5 minutes) and the testing phase (approximately 60–90 minutes). In this game, participants had to control a 3D avatar (bear) using keyboard buttons (W, S, D, A, or arrow keys) to reach a target (acquiring the honey) and then escape from the predefined boundaries to reach the finish line (forest). After participants passed the starting line (fence), they were chased by an artificial agent (chicken) throughout the rest of the trial.

Each trial had two possible outcomes: either participants acquired the target (honey) and reached the finish line, resulting in "winning" the trial, or the agent caught them before reaching the finish line, resulting in "losing" the trial. Participants were instructed to complete the task as quickly as possible. This experiment consisted of 400 trials, divided into three experimental conditions: 1) Normal Trials: 280 trials, 2) Control Error: 60 trials, and 3) Environment Error: 60 trials. Each trial ranged from 9 to 13.5 seconds depending on the participants' performance. In the normal trials, the procedure was identical to what is described above. In the control error condition, after crossing a specified threshold, the player lost the ability to jump over fences (Fig. 1), resulting in the agent catching up with the player and subsequently "losing" the trial. The threshold was an invisible box that was randomly selected between 3 – 3.6 units on the Z axis, in the game

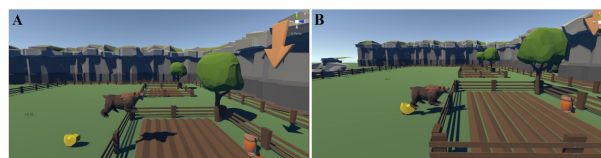


Figure 1: A) Normal trial where the participant is able to jump over the fence. B) Control error trial where the participant loses the ability to jump.



Figure 2: Environment Error Condition: The artificial agent is teleported onto the participant while enlarged, resulting in a jitter effect.

environment.

In the environment error conditions, after participants crossed the threshold, the agent was teleported inside the player's avatar while doubling its size, resulting in jitter effects and instant loss of the trial (Fig. 2).

Participants were not informed about the different error conditions, which comprised 30% (15% per error condition) of the total trials. The total number of trials was divided into 10 blocks with a uniform distribution to ensure that only six errors per condition occurred per block. Furthermore, the onset of error trials occurred after participants crossed the specified threshold, randomly selected in each trial.

**EEG Pre-processing:** EEG recordings were first band-pass filtered (FIR) between 1-30 Hz (filter length of 1651 samples) to remove slow drifts and power line noise. Then, to remove eye artifacts, an Extended Infomax Independent Component Analysis (ICA) [15] was computed with as many components as EEG electrodes (32). The ICA components were then visually inspected and scored by adaptive z-scoring based on the three frontal electrodes (FP1, FPz, FP2). After correcting for eye artifacts, the data were filtered again between 1-15 Hz (FIR) with a filter length of 1651 samples (3.302 sec) and epoched for each condition from 0 to 1 s, where 0 s was the onset of our markers. To ensure that no significant artifacts were maintained in our epochs, we dropped epochs based on maximum peak-to-peak signal amplitude (PTP) with a rejection threshold of  $100 \times 10^{-6}$  V. Epochs were then saved per participant to be later used for the training of our models.

**Data preparation:** The dataset consisted of epochs  $\times$  channels  $\times$  time-points. Additionally, we selected only the 11 central electrodes AFz, F1, Fz, F2, FC1, FCz, FC2,

C1, Cz, C2, CPz, based on the brain regions frequently associated with the encoding of error processing [16]. Furthermore, to prepare the dataset for our model, min-max normalization was computed per electrode, transforming the amplitude of the electrodes between 0 and 1. Finally, to account for the unbalanced dataset (as normal trials accounted for 60% of the trials), the number of epochs was equalized between the error conditions and the normal trials, with the aim of the remaining epochs occurring as close as possible in time. Thus, we removed those normal trials that fell further in time from the epochs of control and environment error conditions and maintained normal trials that fell closer. Finally, we combined the two error conditions (control, environment) into a single class, resulting in a binary classification task between normal and error trials

*Autoencoder Architecture Ana-E:* The architecture we developed, termed Ana-E, is a 1D convolutional autoencoder comprised of an encoder, decoder, and an intermediate dense layer for reshaping the encoded representation. The encoder consists of three layers, with each layer applying a 1D convolution with kernel sizes of 32, 64, and 128, respectively. The first layer's input size mirrors the 11 selected electrode, where a 1D convolution with a kernel size of 32 is applied over the 500 time points per electrode. This results in EEG epochs as input, with the first layer's output producing 22 filters. Subsequent layers double the number of filters, culminating in 88 in the final encoder layer. No padding is applied in any layer. For the first two layers, batch normalization and ReLU activation functions are applied after each convolutional layer, while the final encoder layer consists of a 1D convolution with a kernel size of 128, followed by flattening the output. This results in a high-dimensional tensor, which is then passed through a linear layer to reduce the dimensionality to 750.

Between the encoder and decoder, we integrated an intermediate dense layer with a linear transformation from 750 to  $88 \times 280$  dimensions, reshaping the flattened encoder output for decoding. The decoder network mirrors the encoder, excluding the final flatten and linear layers. Additionally, the decoder's output is passed through a sigmoid function to reconstruct the original EEG signal. For the EEG feature representations, we utilized the output of our encoder. For the classification task, our classification head (CH) consisted of five linear layers with dimensions 750, 500, 250, 125, 60, and 1, respectively. The first input layer corresponds to the output number of Ana-E's encoder. Each linear layer is followed by a ReLU activation function. The output of the final layer undergoes a sigmoid function for binary classification between normal and error trials.

*Training and evaluation:* We trained and evaluated Ana-E both within-participants and between participants. For the within-participants case, we split each participant's session into train/val/test splits of 0.6, 0.2, and 0.2, respectively. In the between-participants case, the model was trained on sessions from all participants except the

one being tested, resulting in 8 training sessions and 1 testing session. The 8 sessions, after being combined, were split into train and validation sets of 0.8 and 0.2, respectively. This process was conducted iteratively for each participant. The model was trained for 250 iterations with a batch size of 64.

We selected Adam as an optimizer with a learning rate of  $1 \times 10^{-3}$  and weight decay of  $1 \times 10^{-6}$ . For the loss function, we chose the Mean Squared Error (MSE) as we wanted our model to be fine-tuned based on the difference between the original input and the reconstructed output. The most optimal parameters for the models, such as training iterations, batch size, learning rate, weight decay, and the number of neurons of the final linear layer of the encoder, were selected based on GridSearch.

For our classification head (CH), we used 200 training iterations with a batch size of 10 using Adam with a learning rate of  $1 \times 10^{-3}$  and weight decay of  $1 \times 10^{-6}$ . We employed Binary Cross-Entropy (BCE) as the loss function.

*Ana-E: Error classification as downstream task:* To assess the effectiveness of our architecture in extracting reliable features for developing pre-trained error detection models, we compared the features extracted by Ana-E with the raw (preprocessed) EEG trials within and between participants. In each comparison, we employed two classifiers: our CH and Linear Discriminant Analysis (LDA) [17], resulting in four different pipelines: AnaE-LDA, AnaE-CH, RAW-LDA, and RAW-CH. In the RAW pipelines, we flattened the 3D EEG trials into 2D. Each epoch's input for LDA and CH in the RAW pipelines consisted of 11 electrodes multiplied by 500 time points. To meet the specified input size of 750 in the first layer for CH, we added an extra layer with an input of size  $11 \times 500$  and an output size of 750.

We evaluated the quality of each pipeline and its ability to differentiate between classes by examining accuracy scores, True Negative Rates (TNRs), and True Positive Rates (TPRs) for 2-class classification within and between participants.

## RESULTS

First, to assess the quality of our feature extractor in the within-participants case, we provide the grand average (GA) per pipeline, together with each participant's accuracy and TNRs and TPRs. We observe that the best GA is achieved by RAW-CH ( $M = 79\%$ ,  $SD = 0.08$ ). For the second-best performance, both AnaE-LDA ( $M = 78\%$ ,  $SD = 0.11$ ) and AnaE-CH ( $M = 78\%$ ,  $SD = 0.11$ ) performed equally well, while the worst accuracy was achieved from RAW-LDA ( $M = 73\%$ ,  $SD = 0.13$ ).

Further inspection of the accuracy per participant reveals that the top three pipelines performed equally well across participants, as each pipeline resulted in the best performance across three participants. Differences in GA are reflective of the variation of classifiers' performances within each participant. For example, our custom classi-

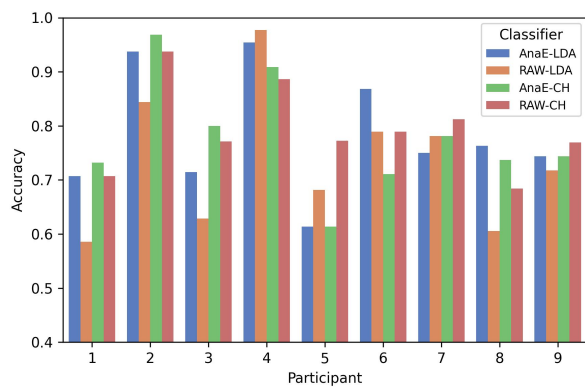


Figure 3: Within-Participants: Classification accuracy

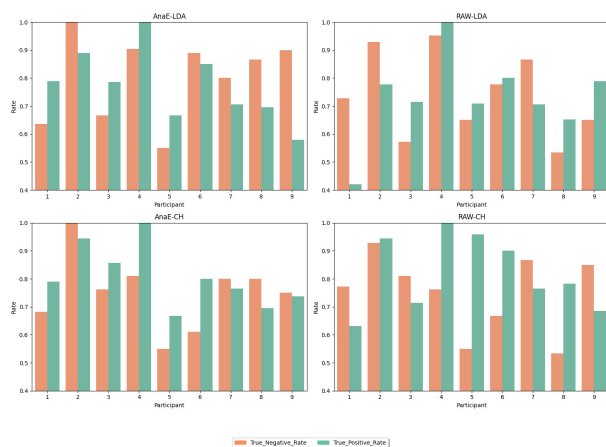


Figure 4: True Negative and True Positive Rates: Within-Participants

fication head seems to achieve the best GA due to the difference it has with the other classifiers in participant 6 and not due to being the most optimal classification method. Additionally, we notice that all pipelines across all participants scored higher than the binary classification chance level (50%), with the best performances achieved in participants 3 and 5 (>80%). Our top three pipelines consistently perform around the 70 mark for almost all participants. (Fig.3).

To gain a better understanding of the quality of the classification made by our tested pipelines, we further investigated the TNRs and TPRs (Fig. 4). We observed that the LDA pipelines predict both classes more equally, with the normal class being predicted slightly more frequently. In contrast, the Ana-E pipelines seem to predict the error condition more strongly, as evidenced by the average TNRs and TPRs (Fig. 5).

Similarly, to assess the quality of our approach between participants, we provide the GA for each pipeline together with the accuracy performances per participant and then TNRs and TPRs. The best GA is reached by AnaE-CH with a score of  $M = 72\%$  ( $SD = .07$ ), the second-best metric is achieved by both RAW-CH ( $M = 64\%$ ,  $SD = .14$ ), and AnaE-LDA ( $M = 64\%$ ,  $SD = .05$ ).

By investigating the accuracy metric per participant, we

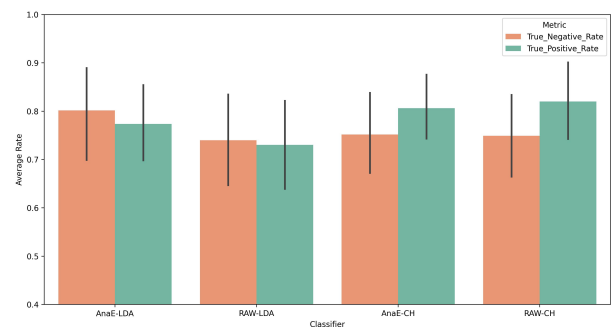


Figure 5: Average True Negative and True Positive Rates: Within-Participants

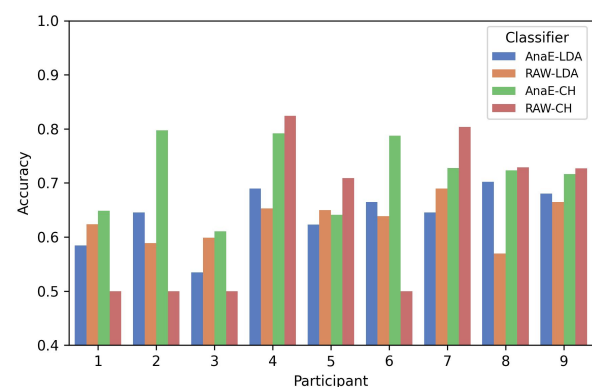


Figure 6: Between-Participants: Classification accuracy

observe that our pipeline (Ana-E) performed above average in all participants and achieved the best accuracy in 4 out of 9 participants. Additionally, we notice that our custom classifier coupled with the preprocessed data (epochs) performed the best in 5 out of the 9 participants but next to chance levels in the remaining participants (Fig. 6).

By further inspecting the average TNRs and TPRs per pipeline, we notice that pipelines utilizing our feature extractor perform the best in terms of error condition recognition. Although RAW-CH seems to achieve the best performances in 5 out of the 9 participants in terms of accuracy metrics, we now notice that the classifier mainly learns to predict the normal trials and performs poorly in terms of error detection.

Furthermore, by inspecting the average TNRs and TPRs, we can deduce that when classifiers were utilizing the features extracted by our encoder, they were firstly able to better predict error conditions while secondly maintaining more stable performances across the different participants (Fig. 8). Finally, by inspecting the TNRs and TPRs per participant we can observe the effects that are responsible for the below chance level of error condition predictions as in 4 out 9 participants the custom classifier trained on the preprocessed data predicts every class as normal trials (Fig.7).

## DISCUSSION

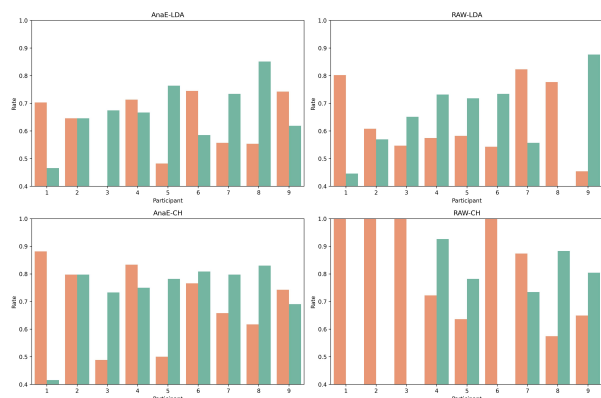


Figure 7: True Negative and True Positive Rates: Between-Participants

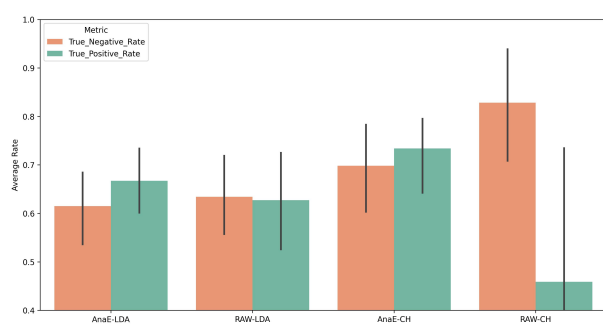


Figure 8: Average True Negative and True Positive Rates: Between-Participants

The study aimed to explore the development of pre-trained error detection models using an autoencoder in a semi-supervised learning (SSL) setup. Initially, a game simulation was designed in Unity, where participants navigated obstacles, with 30% of trials containing simulated errors. EEG data were recorded during the game. We adapted an autoencoder, termed AnaE, for SSL, training it on a reconstruction task and utilizing EEG features for classification. Our pipeline (AnaE) was compared against preprocessed EEG data (RAW) using two classifiers. Both within-participant and between-participant analyses were conducted. In the within-participant analysis, the top three approaches (RAW-CH, AnaE-LDA, AnaE-CH) performed equally. In the between-participant analysis, AnaE-CH outperformed RAW-CH and AnaE-LDA.

Our within-participant analysis offers insights into the performance of our architecture within a classical BCI framework, where models are trained and tested on the same participant session. Results suggest comparable performance among our top three pipelines (RAW-CH, AnaE-LDA, AnaE-CH), with minor variation. Further examination of TNRs and TPRs reveals a tendency for our custom architectures (RAW-CH and AnaE-CH) to exhibit stronger predictions for positive class instances (error conditions). Moreover, it is noteworthy that the sample size for training and testing within participants was significantly smaller compared to between-participant

analysis. Given the complexity of AnaE, the available data may not have been sufficient for our model to learn highly discriminable features. Potentially a solution could be implemented by integrating variational autoencoders (VAE) [18] or other generative models [19] in the processes and amplifying the total number of EEG samples [20]. Furthermore, while the within-participant analysis highlights the usability of our architecture as a "classical" BCI paradigm, the between-participant analysis will provide evidence regarding the feasibility of developing pre-trained error detection models.

The between-participants analysis provided insight into the potential of pre-trained models, as it underscores the capacity of an architecture to generalize to unseen participants while being trained on a sizable dataset comprising multiple individuals. However, in the context of BCIs, certain constraints, such as the non-stationarity of EEG recordings [21], hinder the application of classical training approaches similar to those used in image processing. In this study, we implemented an architecture designed to learn generalizable features from an unsupervised task (pretext) and subsequently transfer the learned EEG feature representations to a downstream classification task. Our architecture (AnaE-CH) achieved the highest GA, surpassing the second-best approach by 8%. Further examination of the TNRs and TPRs revealed that our classifiers were more reliable in predicting the error condition only when our encoder (AnaE) was used to extract features. Conversely, when preprocessed trials were utilized to train the classifiers, they primarily predicted normal trials and struggled to identify the error condition. Our approach provides support for the idea of generalizable features across participants, laying the foundation for pre-trained error detection models. Integration of such models into classical BCI scenarios could potentially reduce the need for calibration sessions.

In the current study, there were certain limitations that could have hindered the performance of our architecture, such as the lack of sufficient datasets in our within-participants analysis. The small sample size for training and testing within participants might have limited AnaE's robustness and generalizability in terms of feature learning. Insufficient data can impact the model's capacity to learn complex EEG signal patterns and features, potentially leading to suboptimal performance. A potential solution could involve the integration of VAE [18] or GANs [19] to amplify the number of trials within participants without increasing the duration of participants' sessions. By doing so, we could further examine the quality of our approach as a classical BCI pipeline. Furthermore it should be highlighted that the results of this study are based on an offline analysis where excessive attenuation of artifacts was possible.

Moreover, our study diverged from the classical autoencoders. Typically, autoencoders aim to reduce dimensions, but in our case, the number of dimensions increases with each layer. Existing literature suggests that autoencoders with large inter-layer dimensions may sim-

ply copy input to output without learning meaningful features [22]. Despite the increasing dimensions, our model achieved improved classification performance based on the learned features, indicating meaningful feature representations from our encoder. Further investigation could explore classical autoencoder principles to determine if the improvements seen in this study stem from our architectural choices.

## CONCLUSION

In conclusion, our study investigated the feasibility of developing pre-trained models for error detection using an autoencoder architecture in a semi-supervised learning (SSL) setting. We designed an experimental setup simulating a game scenario to collect EEG data, which were then employed to train our proposed approach. The results demonstrated comparable performance across different pipelines in the within-participant analysis and a notable enhancement in classification performance in the between-participant analysis when utilizing Ana-E. Particularly promising was our architecture's ability to generalize to unseen participants, indicating its potential utility in real-world applications. Moving forward, further research should explore alternative architectural modifications to enhance the adaptability and robustness of Ana-E. Overall, our study provides valuable insights into the opportunity of developing pre-trained models for error detection in BCI scenarios, laying the foundation for future advancements in the field.

## REFERENCES

- [1] Dias CL, Sburlea AI, Müller-Putz GR. Masked and unmasked error-related potentials during continuous control and feedback. *Journal of Neural Engineering*. 2018;15(3):036031.
- [2] Omedes J, Iturrate I, Minguez J, Montesano L. Analysis and asynchronous detection of gradually unfolding errors during monitoring tasks. *Journal of Neural Engineering*. 2015;12(5):056001.
- [3] Lopes-Dias C, Sburlea AI, Müller-Putz GR. Online asynchronous decoding of error-related potentials during the continuous control of a robot. *Scientific Reports*. 2019;9(1):17596.
- [4] Krauledat M, Tangermann M, Blankertz B, Müller KR. Reducing calibration time for brain-computer interfaces: A clustering approach. *Jan*. 2006, 753–760.
- [5] Jing L, Tian Y. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021;43(11):4037–4058.
- [6] Larsson G, Maire M, Shakhnarovich G. Colorization as a proxy task for visual understanding. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017, 840–849.
- [7] Misra I, Zitnick CL, Hebert M. Shuffle and learn: Unsupervised learning using temporal order verification. In: *Computer Vision – ECCV 2016*. Springer International Publishing: Cham, 2016, 527–544.
- [8] Korbar B, Tran D, Torresani L. Cooperative learning of audio and video models from self-supervised synchronization. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2018.
- [9] Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C. A transformer-based framework for multivariate time series representation learning. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery: Virtual Event, Singapore, 2021, 2114–2124.
- [10] Zha M, Wong S, Liu M, Zhang T, Chen K. Time series generation with masked autoencoder. 2022.
- [11] Banville H, Chehab O, Hyvärinen A, Engemann DA, Gramfort A. Uncovering the structure of clinical eeg signals with self-supervised learning. *Journal of Neural Engineering*. 2021;18(4):046020.
- [12] Tschannen M, Bachem O, Lucic M. Recent advances in autoencoder-based representation learning. 2018. arXiv: 1812.05069 [cs.LG].
- [13] Michelucci U. An introduction to autoencoders. 2022. arXiv: 2201.03898 [cs.LG].
- [14] Haas JK. “A history of the unity game engine.” Worcester. Tech. Rep. 2014. [Online]. Available: <https://digital.wpi.edu/show/2f75r821k>.
- [15] Lee TW, Girolami M, Sejnowski T. Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation*. 1999;11:417–441.
- [16] Rousseau S, Jutten C, Congedo M. The error-related potential and bcis. 2012.
- [17] Mika S, Ratsch G, Weston J, Scholkopf B, Mullers K. Fisher discriminant analysis with kernels. In: *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*. 1999, 41–48.
- [18] Kingma DP, Welling M. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*. 2019;12(4):307–392.
- [19] Ruthotto L, Haber E. An introduction to deep generative modeling. 2021.
- [20] Luo Y, Zhu LZ, Wan ZY, Lu BL. Data augmentation for enhancing eeg-based emotion recognition with deep generative models. 2020.
- [21] Krumpe T, Baumgärtner K, Rosenstiel W, Spüler M. Non-stationarity and inter-subject variability of eeg characteristics in the context of bci development. Sep. 2017.
- [22] Bourlard H, Kabil S. Autoencoders reloaded. *Biological Cybernetics*. 2022;116.

# ASSESSING CALIBRATION DURATIONS FOR C-VEP-BASED BCIS: INSIGHTS FROM NON-BINARY PATTERNS AND SPATIAL FREQUENCY VARIATIONS

V. Martínez-Cagigal<sup>1, 2</sup>, Á. Fernández-Rodríguez<sup>3</sup>, E. Santamaría-Vázquez<sup>1, 2</sup>, A.  
Martín-Fernández<sup>1</sup>, R. Hornero<sup>1, 2</sup>

<sup>1</sup>Biomedical Engineering Group, University of Valladolid, Valladolid, Spain

<sup>2</sup>Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina  
(CIBER-BBN), Valladolid, Spain

<sup>3</sup>UMA BCI Group, University of Malaga, Malaga, Spain

E-mail: [victor.martinez.cagigal@uva.es](mailto:victor.martinez.cagigal@uva.es)

**ABSTRACT:** BCIs using code-modulated visual evoked potentials (c-VEP) have become popular for their reliable, high-speed control of applications and devices. However, traditional circular shifting paradigms based on black & white stimuli can cause eyestrain for some users. We previously showed that adjusting the number of code events and spatial frequency can enhance user comfort. Despite c-VEP calibration being notably shorter than other BCIs, the optimal number of calibration cycles for effective system control remains unexplored. This study aims to investigate the impact of calibration duration on various c-VEP-based BCIs, with stimulus variations to improve user experience. We evaluated performance with different calibration cycles using five  $p$ -ary  $m$ -sequences encoded with shades of gray and eight spatial frequency variations of checkerboard-like stimuli. Results indicate that all conditions achieved over 90% accuracy and 80 bpm with calibration durations ranging between 6–70 seconds. These findings highlight the importance of selecting a configuration based on the functional requirements of the BCI.

## INTRODUCTION

Non-invasive brain-computer interface (BCI) systems enable users to control external applications or devices by processing their electroencephalographic (EEG) activity in real-time [1]. However, direct interpretation of users' intentions from EEG signals is not feasible, making necessary to rely on task-based paradigms that elicit specific control signals. These paradigms encompass strategies that induce measurable deflections in the EEG during cognitive tasks or the processing of external stimuli, such as visual flashes [1]. Among such approaches, code-modulated visual evoked potentials (c-VEPs) have gained popularity in recent years due to their ability to achieve high-performance BCIs with short calibration times. Traditional c-VEP-based BCIs utilize flickering stimuli generated by pseudorandom binary codes that shows perfect autocorrelation properties [2]. These time series en-

code selectable commands using temporally shifted versions of the same code. Calibration in this paradigm, known as circular shifting, typically requires extracting the brain response elicited by the original code over the primary visual cortex as a template. It is assumed that the response to subsequent commands corresponds to temporally shifted versions of this template according to each command's lag [2]. Thus, the main advantage of the circular shifting paradigm is that calibration is drastically reduced by estimating the brain response to a single sequence, independently of the total number of commands. Despite the excellent performance of high-contrast flickering produced by binary codes, which encode commands with black and white flashes, several studies have highlighted potential issues such as visual eyestrain and fatigue among certain users [3, 4]. One of the current areas of research in the state-of-the-art is focused on improving users' comfort without compromising performance. Previous studies have shown that this goal can be achieved through various methods, such as increasing the stimulation rate [5], employing customized codes that confine spectral density to high-frequency bands [6], using sequences with a high number of events allowing encoding with different shades of gray rather than high-contrast stimuli [5], or increasing the spatial frequency of checkerboard-like stimuli [7]. Specifically, we have previously demonstrated in studies by Martínez-Cagigal *et al.* (2023) [5] and Fernández-Rodríguez *et al.* (2023) [7] that these two latter approaches effectively enhance user comfort while maintaining similar levels of accuracy and information transfer rate (ITR).

In comparison with other systems, such as those based on P300 potentials or sensorimotor rhythms, c-VEP-based BCIs require notably fewer calibration trials. Nevertheless, some authors have proposed adaptive algorithms to further reduce or completely eliminate the need for calibration. For example, Spüler *et al.* (2013) [8] introduced an unsupervised clustering-based approach with two calibration targets, Thielen *et al.* (2021) [9] presented an adaptive version of "reconvolution" tailored for

zero-shot calibration contexts, and Stawicki & Volosyak (2022) [10] explored the potential of transfer learning to minimize recalibration across multiple sessions.

In spite of the increasing popularity of c-VEPs in non-invasive BCIs, to our knowledge, no study has specifically investigated the number of calibration cycles needed to effectively control the system. Consequently, this has not been explored for stimulus modifications aimed at enhancing user experience either. Therefore, the aim of this study is to analyze the impact of calibration duration in circular shifting c-VEP paradigms. Specifically, we examine typical binary black and white stimuli,  $p$ -ary m-sequences encoded with different shades of gray, and variations of spatial frequencies in checkerboard patterns.

## SUBJECTS

In this study, we utilized data from two previously recorded databases [5, 7]. In both databases, all aspects of BCI operation were managed using open-source applications of MEDUSA©, accessible at [medusabci.com](https://medusabci.com) [11]. Visual stimuli were presented on an LED FullHD @ 144 Hz monitor (model: KEEP OUT XGM24F+ 23.8") with a refresh rate of 120 Hz. EEG signals were registered using a g.USBamp device (g.Tec, Guger Technologies, Austria) from 16 active Ag/AgCl channels at positions F3, Fz, F4, C3, Cz, C4, CPz, P3, Pz, P4, PO7, POz, PO8, Oz, I1, and I2, according to the International System 10/5. The device was grounded at AFz and referenced to the right earlobe. All participants provided informed consent before participating [5, 7].

***P-ary m-sequences database:*** This dataset comprises 15 healthy participants (aged  $28.80 \pm 5.02$  years, 10 males, 5 females) [5] who engaged in BCI spelling tasks using the open-source “P-ary c-VEP Speller” application of MEDUSA©, accessible at [medusabci.com/market/pary\\_cvep](https://medusabci.com/market/pary_cvep) [11]. Participants completed a single session consisting of a calibration phase comprising 300 cycles and an online spelling task comprising 32 trials (with 10 cycles per trial) for each  $p$ -ary m-sequence. A total of five  $p$ -ary m-sequences were assessed: binary GF( $2^6$ ) with a base of 2, GF( $3^5$ ) with a base of 3, GF( $5^3$ ) with a base of 5, GF( $7^2$ ) with a base of 7, and GF( $11^2$ ) with a base of 11. Events were encoded using various shades of gray; e.g., GF( $2^6$ ) employed black and white flashes, while GF( $5^3$ ) used three equidistant grey tones in addition to black and white flashes, and so forth. Figure 1(A) illustrates all  $p$ -ary m-sequences, along with the arrangement of commands’ lags. Online selections were made using a  $4 \times 4$  command matrix (chance level of 6.25%), comprising alphabetic characters from A to P. For further details, please see Martínez-Cagigal *et al.* (2023) [5].

***Checkerboard database:*** This dataset comprises 16 healthy participants (aged  $29.63 \pm 4.06$  years, 11 males, 5 females) [7] who undertook BCI spelling tasks using a modified version of the open-source “c-VEP Speller” application provided by MEDUSA©, accessible at [medusabci.com/market/cvep\\_speller](https://medusabci.com/market/cvep_speller) [11].

Participants completed a single session, which included a calibration phase consisting of 300 cycles and an online spelling task comprising 18 trials (with 8 cycles per trial) for each spatial condition. All conditions employed a binary m-sequence of 63 bits encoded with black-background checkerboard (BB-CB) patterns, i.e., the encoding represented “1” with a checkerboard pattern and “0” with a black flash. Each condition assessed a specific spatial frequency of the stimuli, which refers to the size of the squares within a checkerboard-like stimulus, measured in cycles (pairs of squares of two alternative colors) per degree of visual angle ( $c^\circ$ ). A total of 8 conditions were examined: C001 ( $0 c^\circ$ ), C002 ( $0.15 c^\circ$ ), C004 ( $0.3 c^\circ$ ), C008 ( $0.6 c^\circ$ ), C016 ( $1.2 c^\circ$ ), C032 ( $2.4 c^\circ$ ), C064 ( $4.79 c^\circ$ ), and C128 ( $9.58 c^\circ$ ). The stimuli and the arrangement of lags are shown in Figure 1(B). Online selections were made using a  $3 \times 3$  command matrix (chance level of 11.11%). For further details, please see Fernández-Rodríguez *et al.* (2023) [7].

## METHODS

***Paradigm:*** In both datasets, the circular shifting paradigm was employed, which relies on using shifted versions of a pseudorandom sequence to encode individual commands, ensuring that only the original sequence needs to be calibrated. Consequently, it is imperative for the sequence to demonstrate minimal autocorrelation, thereby facilitating subsequent decoding [2]. Maximal length sequences (i.e., m-sequences), represent pseudorandom temporal series characterized by nearly optimal autocorrelation properties. These sequences are generated employing linear-feedback shift registers (LFSR). The characteristics of m-sequences are determined by three main factors: the base  $p$ , denoting the quantity of different events (e.g.,  $p = 2$  for binary m-sequences: events 0 and 1); the order  $r$ , indicating the number of LFSR taps; and the generator polynomial expressed within a Galois Field of  $p$  elements, GF( $p$ ), which defines the arrangement of the LFSR taps [5]. In addition to conforming to various mathematical constraints, the length of a m-sequence exactly equals  $N = p^r - 1$  bits, repeating cyclically [5]. The larger the m-sequence, the greater the number of commands that can be encoded with it.

***Signal processing:*** During the calibration stage, participants are instructed to focus on a single command encoded by the original m-sequence (without delay) for a duration covering  $k$  cycles (i.e., repetitions of the same m-sequence). Initially, the EEG signal undergoes pre-processing using a filter bank comprising three bandpass filters (ranging from 1 to 60 Hz, 12 to 60 Hz, and 30 to 60 Hz) and a notch filter set at 50 Hz [5, 7]. Two versions of the EEG response are subsequently computed for each signal: (1) concatenated epochs, denoted as  $\mathbf{A} \in \mathbb{R}^{[kN_s \times N_c]}$ ; and (2) epochs averaged over the  $k$  cycles, denoted as  $\mathbf{B} \in \mathbb{R}^{[N_s \times N_c]}$ . Here,  $N_s$  represents the number of samples per cycle, and  $N_c$  represents the number

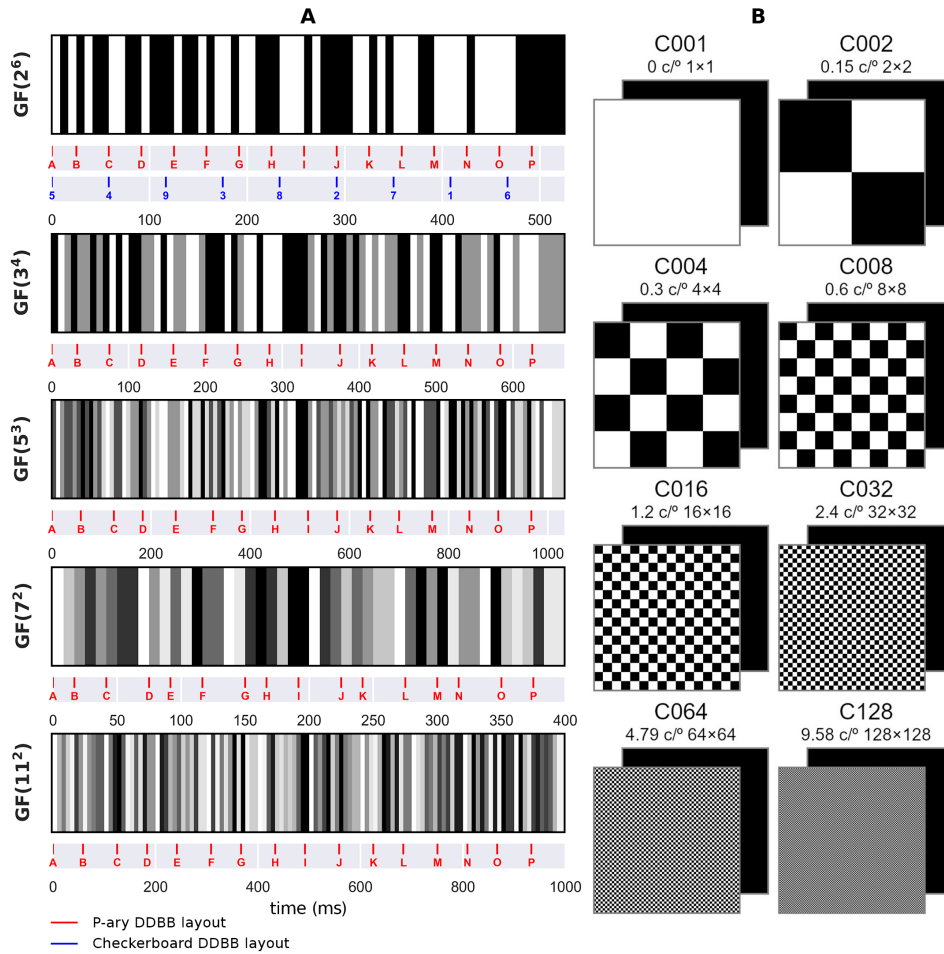


Figure 1: Stimuli details for both databases. (A) Gray encoding of each  $p$ -ary m-sequence over time, depicting associated lags for each command in the  $p$ -ary m-sequence database (shown in red) and checkerboard database (shown in blue). (B) Binary patterns of the black-background checkerboard (BB-CB) stimulus for the eight distinct spatial frequencies assessed in the checkerboard database. Note that all patterns (event 1) were coupled with a flickering monochromatic black square (event 0).

of channels. Subsequently, a canonical correlation analysis (CCA) is utilized to train the spatial filter  $\mathbf{w}_b$  that maximizes the correlation between the projected versions of  $\mathbf{A}$  and  $\mathbf{B}$ . In this process,  $\mathbf{B}$  is replicated  $k$  times to match the dimensions of  $\mathbf{A}$ . The main template (i.e., for the command without delay) is established by projecting the averaged signal using the spatial filter  $\mathbf{w}_b$ , resulting in  $\mathbf{x}_0 = \mathbf{B}\mathbf{w}_b$ . Templates for the other commands are then generated by cyclically shifting this main template based on their respective delays. Following this procedure,  $N_t \times 3$  templates, each for a command and filtered signal, are obtained, where  $N_t$  indicates the number of commands in the online stage. Therefore,  $N_t = 16$  for the  $p$ -ary m-sequences database, and  $N_t = 9$  for the checkerboard database. Calibration epochs with a standard deviation three times greater than the average standard deviation of all epochs were excluded before training the CCA [5, 7]. During the online mode, a similar approach is employed to determine the command the user is focusing on in real-time. The EEG signal undergoes preprocessing, and individual epochs are averaged and projected using the spatial filter  $\mathbf{w}_b$ . The correlation between the result-

ing projection and all templates is then computed, yielding  $\hat{\mathbf{p}} \in \mathbb{R}^{N_t \times 3}$  values. After averaging across the filtered signals,  $\mathbf{p} \in \mathbb{R}^{N_t}$  is obtained. The selected command corresponds to the one that produces the highest correlation value, identified as  $\arg \max_i(\mathbf{p})$  [5, 7].

## RESULTS

To understand how the duration of the calibration period affects the system's final performance, we (1) selected a specific number of calibration cycles  $k$ , (2) trained the model as outlined in the methods section, and (3) predict the outcome of the test trials and extract accuracy and ITR while varying the number of online cycles. The parameter  $k$  was systematically increased until the entire calibration dataset was utilized for each database, i.e.,  $k \in [1, 300]$ . Figures 2 and 3 depict the grand-averaged accuracy across subjects for each condition and dataset. This accuracy is presented as a function of the number of calibration cycles  $k$  and the number of online cycles. The amount of calibration data is expressed in both the number of cycles and duration in seconds. As depicted, a

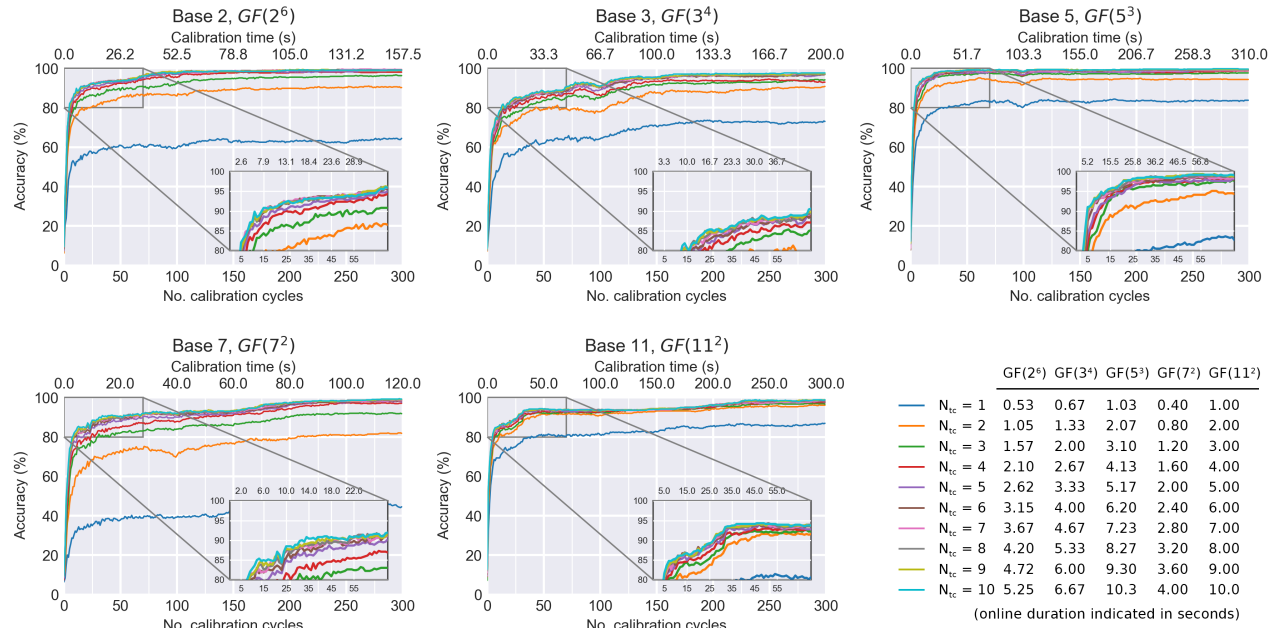


Figure 2: Grand-averaged accuracy across subjects for the  $p$ -ary m-sequences database as a function of the number of calibration cycles and the number of online cycles. The x-axis illustrates the quantity of calibration data utilized to train the model, expressed in both the number of cycles (bottom axis) and duration in seconds (top axis). Each curve represents a distinct number of online cycles (i.e., the length of the test epoch). Each plot corresponds to a different base. Chance level was 6.25%

Table 1: Analysis of plateau performance on the  $p$ -ary m-sequences database.

		Base 2, GF(2 <sup>6</sup> )			Base 3, GF(3 <sup>4</sup> )			Base 5, GF(5 <sup>3</sup> )			Base 7, GF(7 <sup>2</sup> )			Base 11, GF(11 <sup>2</sup> )		
	N <sub>oc</sub>	2	5	10	2	5	10	2	5	10	2	5	10	2	5	10
0.9M	cal. (s)	74.55	9.45	5.78	96.00	37.33	28.67	19.63	10.33	5.17	-	24.40	12.40	34.00	27.00	24.00
	acc. (%)	89.45	89.84	89.84	88.09	87.89	87.89	90.04	91.02	91.02	-	89.65	89.84	89.06	89.45	89.06
	ITR (bpm)	177.3	71.53	35.76	135.3	53.90	26.95	91.25	37.31	18.66	-	93.47	46.94	92.24	37.22	18.45
0.95M	cal. (s)	-	34.12	31.50	-	76.00	52.00	62.00	22.73	11.37	-	73.60	65.20	195.0	47.00	37.00
	acc. (%)	-	94.73	94.34	-	92.97	92.97	95.12	95.51	94.92	-	94.53	94.53	93.95	93.95	94.14
	ITR (bpm)	-	79.91	39.60	-	60.45	30.22	102.4	41.34	20.39	-	104.4	52.21	103.0	41.21	20.70
1M	cal. (s)	-	-	109.2	-	-	175.3	-	-	276.9	-	-	118.8	-	-	237.0
	acc. (%)	-	-	99.22	-	-	97.66	-	-	99.80	-	-	99.41	-	-	98.83
	ITR (bpm)	-	-	44.61	-	-	33.73	-	-	23.06	-	-	58.88	-	-	23.17

N<sub>oc</sub> indicates the number of online cycles,  $M$  the maximum accuracy for each base, “cal” the duration of the calibration in seconds, “acc.” the accuracy in %, and “ITR” the information transfer rate in bits per minute. This analysis unveils the minimum calibration duration necessary to achieve 90%, 95%, and 100% of the maximum accuracy for each base, provided that such accuracy can be attained with each number of online cycles.

performance plateau is generally observed across all conditions. Tables 1 and 2 summarize the minimum calibration duration required to achieve 90%, 95%, and 100% of the maximum accuracy for each condition and database. Results are provided for varying numbers of online cycles. If that accuracy cannot be achieved with a particular number of online cycles, it is denoted with a hyphen.

## DISCUSSION

As expected, the longer the calibration duration, the greater the accuracy achieved, regardless of the  $p$ -ary m-sequence or the spatial frequency of the BB-CB stimuli. While all conditions reached a performance plateau after a specific number of calibration cycles, independently of the number of online cycles, the slope (learning curve) appears to be dependent on the database or condition.

In the  $p$ -ary m-sequences database, all conditions achieved an average accuracy higher than 97% when using the maximum calibration duration. Overall, it is evident for all conditions that at least 2 online cycles are required to reach suitable performance. However, the performance plateau varied among them, indicating a trade-off between calibration duration and final performance. Specifically, GF(5<sup>3</sup>) and GF(2<sup>6</sup>) exhibited higher accuracy with less calibration duration, followed by GF(11<sup>2</sup>), GF(7<sup>2</sup>), and GF(3<sup>4</sup>). Considering an intermediate number of 5 online cycles, the 95th percentile of maximum accuracy (all above 92%) was achieved by using the following calibration durations, as shown in Table 1: 22.73 s for GF(5<sup>3</sup>), 34.12 s for GF(2<sup>6</sup>), 47.00 s for GF(11<sup>2</sup>), 65.20 s for GF(7<sup>2</sup>), and 76.00 s for GF(3<sup>4</sup>). These configurations also yielded ITRs above 40 bpm in all cases, i.e. maximum of 104.1 bpm for GF(7<sup>2</sup>), and minimum

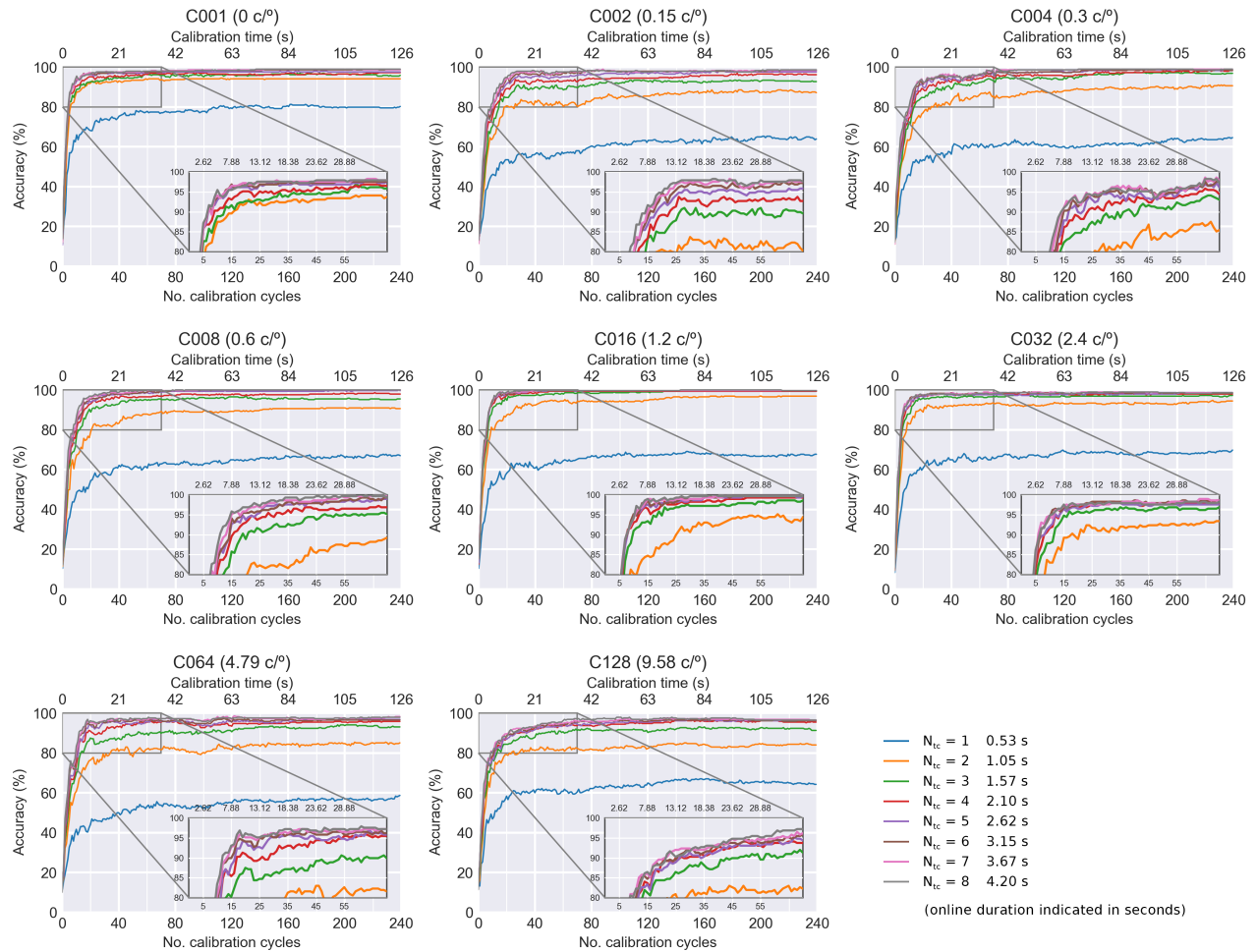


Figure 3: Grand-averaged accuracy across subjects for the checkerboard database as a function of the number of calibration cycles and the number of online cycles. The x-axis illustrates the quantity of calibration data utilized to train the model, expressed in both the number of cycles (bottom axis) and duration in seconds (top axis). Each curve represents a distinct number of online cycles (i.e., the length of the test epoch). Each plot corresponds to a different spatial frequency. Chance level was 11.11%.

Table 2: Analysis of plateau performance on the checkerboard database.

		C001 (0 c/°)		C002 (0.15 c/°)		C004 (0.3 c/°)		C008 (0.6 c/°)		C016 (1.2 c/°)		C032 (2.4 c/°)		C064 (4.79 c/°)		C128 (9.58 c/°)	
$N_{oc}$		5	8	5	8	5	8	5	8	5	8	5	8	5	8	5	8
0.9M	cal. (s)	3.67	3.67	9.45	6.83	8.93	7.35	7.88	5.78	4.20	4.20	3.67	3.67	7.88	6.30	12.07	9.45
	acc. (%)	90.97	91.67	89.93	89.58	90.97	89.93	92.71	90.28	90.63	90.97	89.58	91.67	88.19	90.62	88.89	88.19
	ITR (bpm)	73.37	46.58	71.67	44.44	73.37	44.79	76.31	45.14	72.80	45.86	71.11	46.58	68.91	45.50	70.00	43.07
0.95M	cal. (s)	6.83	5.25	13.12	9.45	14.70	12.07	12.60	7.88	5.25	5.78	6.30	5.25	9.97	7.88	23.10	18.90
	acc. (%)	94.10	95.49	94.10	94.10	94.79	94.44	95.14	95.83	95.14	97.22	95.14	94.79	94.10	93.06	93.06	93.06
	ITR (bpm)	78.76	50.83	78.76	49.23	80.03	49.62	80.68	51.25	80.68	52.98	80.68	50.02	78.76	48.07	76.91	48.07
1M	cal. (s)	-	71.92	-	44.10	-	68.25	124.95	31.50	-	19.43	-	71.92	-	32.02	-	71.92
	acc. (%)	-	98.96	-	98.61	-	98.96	100.00	100.00	-	100.00	-	98.96	-	97.92	-	97.92
	ITR (bpm)	-	55.37	-	54.86	-	55.37	91.43	57.14	-	57.14	-	55.37	-	53.89	-	53.89

$N_{oc}$  indicates the number of online cycles,  $M$  the maximum accuracy for each base, “cal.” the duration of the calibration in seconds, “acc.” the accuracy in %, and “ITR” the information transfer rate in bits per minute. This analysis unveils the minimum calibration duration necessary to achieve 90%, 95%, and 100% of the maximum accuracy for each base, provided that such accuracy can be attained with each number of online cycles.

of 41.21 bpm for GF(11<sup>2</sup>). Nevertheless, the tradeoff between calibration/online selection duration and performance makes it difficult to select any specific configuration. For instance, in GF(2<sup>6</sup>), it is feasible to calibrate with only 9.45 s, albeit at the cost of reducing the average accuracy to 89.84% and utilizing 2.62 s of online

selection duration (at 71.53 bpm). Conversely, an impressive ITR of 177.3 bpm (at 89.45%) can be attained with an online selection duration of only 1.05 s by employing a calibration duration of 74.55 s. To sum up, all  $p$ -ary m-sequences prove capable of achieving above 90% mean accuracy through various configurations of calibra-

tion and online durations.

Regarding the checkerboard database, all spatial frequency conditions displayed similar performing plateaus. Once again, two online cycles are necessary to attain a practical level of control over the system. Note that condition C001 (0 c°) is analogous to GF(2<sup>6</sup>), and its superior performance may stem from the fact that the checkerboard database comprises only 9 commands instead of 16 [5, 7]. For 5 online cycles (2.62 s), all conditions demonstrated more than 93% accuracy and 76 bpm. Condition C016 stands out by achieving 95.15% accuracy and 80.68 bpm using only 5.25 s of calibration. Extending the calibration duration to 19.43 s and the online cycles to 8 (4.20 s), C016 attains 100% accuracy and 57.14 bpm. While we emphasize C016, similar behaviors are observed for C008 and C032, as indicated in Table 2. Generally, calibrations lasting between 3.67–12.07 s are sufficient to achieve accuracies around 88%–90% for all conditions, where C128 likely exhibits the poorest results.

It is evident that there exists a four-variable tradeoff involving calibration duration, online selection duration, system performance, and user comfort. Although all conditions can provide high-speed and high-performance BCIs, the performance plateau varies among them. The total number of commands also plays a significant role in interpreting these results. In conclusion, performance levels around 95% accuracy and 80 bpm can be readily attained with calibration durations ranging of 6–35 s and online selections of 2–3 s for binary m-sequences. Higher bases (i.e.,  $p > 2$ ) would lead to increased user comfort, as suggested by Martínez-Cagigal *et al.* (2023) [5], albeit at the cost of longer calibration periods. Concerning spatial frequencies, C016, followed by C008, emerged as the BB-CB stimuli associated with higher comfort scores [7]. Ultimately, the choice of configuration would depend on the functional requirements of the BCI system.

## CONCLUSION

To our knowledge, this is the first study to examine the impact of the calibration phase on circular shifting paradigms for c-VEP-based BCIs. Regardless of the  $p$  base of the code or the spatial frequency of the BB-CB stimuli, all conditions can achieve high-speed and high-performance BCIs with sufficient calibration. We identified a tradeoff between calibration duration, online selection duration, performance, and user comfort. Performance levels nearing 95% accuracy and 80 bpm can be reached with calibrations lasting 6–70 s. While achieving over 95% accuracy with approximately 5 s of calibration and a selection duration of 2.62 s is possible with binary m-sequences, we stress the importance of selecting a configuration based on the requirements of the final BCI system, such as target performance and user comfort.

## ACKNOWLEDGMENTS

This research has been developed under the grants

TED2021-129915B-I00, PID2020-115468RB-I00 and PDC2021-120775-I00 funded by MCIN/AEI/10.13039/501100011033 and ERDF; under project ‘0124\_EUROAGE\_MAS\_4\_E’ (‘Cooperation Programme Interreg VI-A Spain-Portugal POCTEP 2021–2027’) funded by ‘European Commission’ and ERDF; and by CIBER-BBN through ‘Instituto de Salud Carlos III’ co-funded with ERDF funds.

## REFERENCES

- [1] Wolpaw J, Wolpaw EW. Brain-computer interfaces: principles and practice. OUP USA (2012).
- [2] Martínez-Cagigal V, Thielen J, Santamaría-Vázquez E, Pérez-Velasco S, Desain P, Hornero R. Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): a literature review. *Journal of Neural Engineering*. 2021;18(6):061002.
- [3] Ladouce S, Darmet L, Torre Tresols JJ, Velut S, Ferraro G, Dehais F. Improving user experience of SSVEP BCI through low amplitude depth and high frequency stimuli design. *Scientific Reports*. 2022;12(1):1–12.
- [4] Gembler FW, Rezeika A, Benda M, Volosyak I. Five Shades of Grey: Exploring Quintary m-Sequences for More User-Friendly c-VEP-Based BCIs. *Computational Intelligence and Neuroscience*. 2020;2020.
- [5] Martínez-Cagigal V, Santamaría-Vázquez E, Pérez-Velasco S, Marcos-Martínez D, Moreno-Calderón S, Hornero R. Non-binary m-sequences for more comfortable brain-computer interfaces based on c-VEPs. *Expert Systems with Applications*. 2023;232(June):120815.
- [6] Shirzhiyan Z *et al.* Introducing chaotic codes for the modulation of code modulated visual evoked potentials (c-VEP) in normal adults for visual fatigue reduction. *PLoS ONE*. 2019;14(3):1–29.
- [7] Fernández-Rodríguez Á, Martínez-Cagigal V, Santamaría-Vázquez E, Ron-Angevin R, Hornero R. Influence of spatial frequency in visual stimuli for cVEP-based BCIs: evaluation of performance and user experience. *Frontiers in Human Neuroscience*. 2023;17.
- [8] Spüler M, Rosenstiel W, Bogdan M. Unsupervised BCI Calibration as Possibility for Communication in CLIS Patients? In: *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013*. 2013, 10–12.
- [9] Thielen J, Marsman P, Farquhar J, Desain P. From full calibration to zero training for a code-modulated visual evoked potentials brain computer interface. *Journal of Neural Engineering*. 2021;18(5):56007.
- [10] Stawicki P, Volosyak I. cVEP Training Data Validation—Towards Optimal Training Set Composition from Multi-Day Data. *Brain Sciences*. 2022;12(2).
- [11] Santamaría-Vázquez E *et al.* MEDUSA©: A novel Python-based software ecosystem to accelerate brain-computer interface and cognitive neuroscience research. *Computer Methods and Programs in Biomedicine*. 2023;230(107357).

# IMPROVED MOTOR IMAGERY DECODING WITH SPATIOTEMPORAL FILTERING BASED ON BETA BURST KERNELS

S. Papadopoulos<sup>1,2,3</sup>, L. Darmet<sup>1,2,3</sup>, M.J. Szul<sup>1,3</sup>, M. Congedo<sup>4</sup>, J.J. Bonaiuto<sup>1,3,†</sup>, J. Mattout<sup>1,2,†</sup>

<sup>1</sup> University Lyon 1, Lyon, France

<sup>2</sup> Lyon Neuroscience Research Center, CRNL, INSERM, U1028, CNRS, UMR 5292, Lyon, France

<sup>3</sup> Institut des Sciences Cognitives Marc Jeannerod, CNRS, UMR 5229, Lyon, France

<sup>4</sup> GIPSA-lab, University Grenoble Alpes, CNRS, Grenoble-INP, Grenoble, France

† These authors contributed equally

E-mail: sotirios.papadopoulos@univ-lyon1.fr

**ABSTRACT:** The description of the event-related desynchronization and synchronization phenomena in the mu and beta frequency bands has to a significant extent shaped our understanding of motor-related brain processes. Accordingly, Brain-Computer Interface applications leveraging attempted or imagined movements usually depend on spatially- and band-limited power changes as the brain markers of interest. Yet, converging neuroscience evidence question the idea that signal power best describes the movement-related modulation of brain activity. On a single-trial level, beta band activity is characterized by short, transient and heterogeneous events termed bursts rather than sustained oscillations. In a recent study we demonstrated that a beta burst analysis of hand motor imagery binary classification tasks is often superior to beta power in terms of classification score. Here we expand upon this idea proposing a comparable to state-of-the-art algorithm. We confirm our previous results by using convolution kernels extracted from beta bursts. Moreover, we show that these kernels can effectively be used in inter-session transfer learning strategies.

## INTRODUCTION

Three decades ago, a number of seminal studies in motor neuroscience revealed for the first time, time-locked changes in induced power within specific frequency bands [1-3]. These studies described a relative-to-baseline gradual reduction in the power of brain signals recorded during an ongoing movement or motor imagery (MI) task in the mu (~8-12 Hz) [3-6] and beta (~13-30 Hz) [3, 5] frequency bands, termed event-related desynchronization (ERD). They, also, demonstrated a relative increase in power in the beta band shortly after the end of the task, known as event-related synchronization (ERS) [6-8]. The ERD is considered to be a high-level indication of brain processes pertaining to movement preparation and

execution, and is particularly prominent in the contralateral sensorimotor cortex [2, 9-12].

Because of this spatial and frequency specificity, ERD is the main marker of interest for motor-related and especially MI-based, non-invasive BCI applications [13, 14]. Typically, signals recorded during MI are transformed in the time-frequency domain (TF) [15-17], and are then spatially filtered using the common spatial pattern algorithm (CSP) [18-20]. This results in an increase of signal-to-noise ratio and extracts the signal power in specific time windows and frequency bands of interest, while also maximizing the spatial disparity among different MI classes (e.g. "left" or "right" hand).

The hypothesized reliability and reproducibility of these signal characteristics has also served as the basis for a range of transfer-learning attempts. Transfer learning refers to the exploitation of specific signal features extracted from past recording sessions, different subjects and/or experiments to guide decoding during future sessions [21].

Despite the fact that the ERD and ERS are widely observed, their nature is not well-understood and converging neurophysiology evidence puts these phenomena into question. The ERD and ERS are typically revealed by averaging signal power in the TF domain over multiple trials, especially in the beta frequency band [11, 22], under the assumption of sustained oscillations. However, this evidence points out that, on the contrary, on a single trial level, beta band activity occurs in short events, termed bursts [11, 22-26]. The rate of these beta bursts is more behaviorally relevant in motor processes [11, 24, 27-30] than averaged beta band power. Moreover, it has been shown that beta bursts comprise heterogeneous events [29] with different functions, alluded to by their differential modulation during different task conditions [31, 32] or phases [29, 30].

In a recent study we showed that the analysis of beta

bursts from channels C3 and C4 during hand motor imagery can be advantageous to beta power in terms of classification, confirming the hypothesis that on the single-trial level beta burst rate modulations are more behaviorally relevant than beta band power changes [31]. In this article we expand upon that study. We introduce an algorithm that exploits beta bursts in order to transform brain signals into measures of waveform-resolved burst rate. Moreover, this algorithm can take advantage of an arbitrary number of recorded signals while being computationally efficient, thus constructing decoding features that are comparable to state-of-the-art in BCI. We analyze the activity during “left” and “right” hand MI of three open EEG datasets and show that the use of beta bursts instead of beta band power can improve classification results. Finally, we adopt a transfer learning approach and show that beta bursts detected in one recorded session can be exploited to guide the decoding in other sessions.

## MATERIALS AND METHODS

**Datasets:** We analyzed three open EEG MI datasets: BNCI 2014-001 [13], BNCI 2014-004 [33] and Zhou 2016 [34], all available through the MOABB project [14] (Table 1). These datasets are composed of recordings of numerous subjects who were required to perform sustained MI following the appearance of a visual cue on a screen. For our analysis we only considered trials corresponding to the ‘left hand’ or ‘right hand’ classes even though two experimental paradigms consisted of more MI classes. A brief account of the tasks is described in [31].

**Pre-processing:** Each subject’s epoched recordings were loaded with the MOABB python package (v0.4.6) LeftRightImagery class, and were filtered with a low pass cutoff of 120 Hz (zero-phase FIR filter designed with the windowed approach and transition bandwidth of 25% of the low pass frequency). Then, we rejected trials using the autoreject python package [35] (v0.4.0, function `get_rejection_threshold`) (Table 1). We refer the reader to [31] for more details regarding the pre-processing.

**Burst-detection:** Following pre-processing, a subset of channels above the sensorimotor cortex was defined (‘C3’, ‘Cz’, ‘C4’, and ‘FC3’, ‘FCz’, ‘FC4’, ‘CP3’, ‘Cpz’, ‘CP4’ when available; as in [31]). The corresponding recordings were first transformed in the time-frequency (TF) domain from 1 to 43 Hz using the superlets algorithm [36] (parameters:  $\sigma_{\min} = 1$ ,  $\sigma_{\max} = 40$ ,  $c = 4$ ) with a frequency resolution of 0.5 Hz. Then, we identified bursts within the beta frequency range (15-30Hz) from each TF matrix using a previously published procedure that allowed us to extract the waveforms of the beta bursts within a fixed time-window (see [29] for more information on the algorithm and [31] on how it was specifically applied to these datasets).

**Kernel selection:** For each dataset and subject we randomly sampled 10% of the recording trials (or all

Table 1. Dataset attributes

Dataset	BNCI 2014-001	BNCI 2014-004	Zhou 2016
# Subjects	9	9	4
# Channels	3	22	64
# Sessions	2	5	3
Min - Max # Trials across subjects	288	680-760	290-319
Min - Max # Trials (after trial rejection)	217-287	269-621	114-280
Sampling freq. (Hz)	250	250	250
Trial duration (s)	4.0	4.5	5.0
Reference	[13]	[33]	[34]

trials corresponding to a recording session when assessing transfer learning) of each participant after trial rejection (Table 1) ensuring class balance, in order to create a large sample of beta burst waveforms while restricting the number of trials per participant we excluded from classification (see *Feature Selection* and *Classification*). We aggregated all detected bursts within these trials in a matrix (irrespective of the trial class, i.e. ‘left hand’ or ‘right hand’) after robust scaling (scikit-learn package [37], v1.0.2). Then, we used principal component analysis (PCA) [38] (scikit-learn package, v1.0.2) in order to reduce the time dimension of the waveforms. We defined an index of lateralized modulation of the average-per-axis PCA score  $I_m$  (a metric of the difference between any burst waveform and the average shape) from the baseline to the trial periods of the recordings as using electrodes C3 and C4:

$$I_m = |(u_{ipsi}^{C3} - u_{contr}^{C4}) - (u_{ipsi}^{C4} - u_{contr}^{C3})|, m \in \{2, \dots, 9\}$$

$$u = |\widehat{score}_{trial\ period} - \widehat{score}_{baseline}|$$

where *ipsi* (*contr*) refers to bursts recorded from channels C3 / C4 during a left / right (right / left) hand MI. Using this index, we identified the three PCA axes that maximized  $I_m$  among components 2 to 9. We did not consider the first PCA component as it simply describes the temporal skew of the burst waveforms [29, 31]. Finally, based on our previous study [31], we split the computed score range of each of the three selected axes in seven equally spaced groups. We considered the two groups that lie further away from the origin (score equal to 0), grouped similarly shaped bursts together and defined 2 kernels per PCA axis by computing the average waveform of all bursts within these two groups. As a result, we ended up with 6 kernels that describe the



Figure 1: Average decoding score (area under the curve of the receiver operator characteristic) and standard deviation per dataset, subject and classification feature.

burst waveforms that were expected to be maximally rate-modulated during the task, compared to baseline, for each participant of each dataset separately.

**Feature selection:** We used these kernels in order to transform the available EEG recordings of all initially available channels, not considering the subset of trials that were used for defining the kernels (the random trials sample or a given recording session). Specifically, a copy of each subject's epoched data was convolved with one kernel, resulting in a proxy of the waveform-resolved burst rate. Then, each of the temporally filtered epoched data was spatially filtered using the CSP algorithm (MNE package [39], v 1.5.1, function CSP, parameters: `n_components = 4`, `transform_into = "average_power"`).

We also created band-limited spatial features with a standard filtering technique. Specifically, we computed the envelope of the epoched data Hilbert transform (MNE package, v 1.5.1, function `apply_hilbert`) after independently applying a single filter in the beta frequency band (15-30 Hz) and a wider frequency range encompassing both the mu and beta bands (6-30 Hz). Then, these signals served as inputs to the CSP algorithm (we kept the parameters unchanged).

In order to assess whether the number of spatial features used for classification (24 for the beta burst kernel approach versus 4 for the Hilbert power approach)

affected the observed results, we also adopted a filter bank approach. We split either frequency range in non-overlapping filter banks of 3 Hz range; this resulted in 5 filters for the beta band (15-18 Hz, 18-21 Hz, 21-24 Hz, 24-27 Hz, 27-30 Hz) and 8 filters for the mu-beta band (6-9 Hz, 9-12 Hz, 12-15 Hz, 15-18 Hz, 18-21 Hz, 21-24 Hz, 24-27 Hz, 27-30 Hz) corresponding to 20 and 32 spatial features respectively.

We also estimated kernel-specific CSP filters on one session and used these filters in the remaining sessions. For comparison, we estimated the CSP filters of a session and used them to transform the signals of the other sessions for each of the previously described filtering techniques.

**Classification:** The resulting spatially filtered data were concatenated in a single matrix, once per subject. Using a repeated ( $n=10$ ), 5-fold cross validation procedure we shuffled the remaining trials (Table 1) and estimated the decoding score using LDA (scikit-learn, v1.0.2) as a classifier using the whole trial recordings. The decoding scores were based on the area under the curve (AUC) of the receiver operating characteristic (scikit-learn, v1.0.2). All numeric computations were based on the numpy python package (v1.21.6; [40]) and an environment running python (v3.10).

**Statistical analysis:** To estimate, at the population level, any statistical difference between methods, we compa-

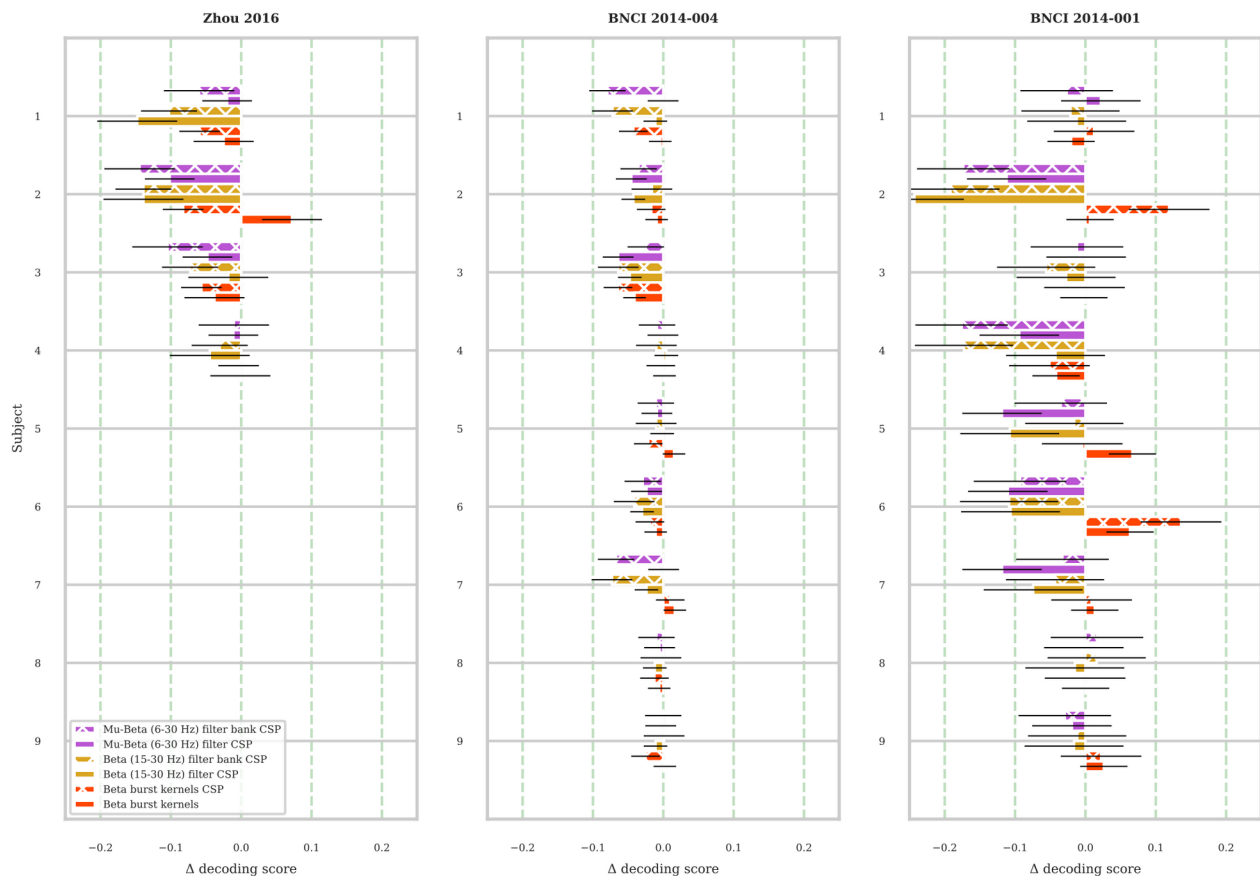


Figure 2: Average difference in decoding score (area under the curve of the receiver operator characteristic) between classification score of a transfer learning session and classification score obtained using all available trials per subject. A value of 0 indicates no difference in decoding between the two approaches. Positive values or negative values indicate a performance gain or loss respectively when adopting a transfer learning approach.

red classification results of the waveform-resolved burst features against multiple band-limited power features. We used a generalized linear mixed model with a binomial distribution and logit link function with across-trial average classification score as the dependent variable setting the number of trials as prior weights, the type of classification feature as a fixed effect, and the subject nested within the dataset as random intercepts. We also compared the across-session average difference in classification score between a transfer learning approach and an across-session, global classification, using a similar model but with a Gaussian distribution. Statistical analyses were conducted using R (v4.1.2) and lme4 (v1.1-31; [41]). Fixed effects were assessed using type II Wald X<sup>2</sup> tests using car (v3.1-1; [42]). Pairwise Tukey- or Sidak-corrected follow-up tests were carried out using estimated marginal means from the emmeans package (v.1.8.7; [43]).

## RESULTS

We estimated the across-session decoding score per subject of each dataset using the waveform-resolved approach. In order to assess the significance of the waveform-resolved burst rate features we also

computed classification features based on signal power and compared the different decoding results. We used a filter bank approach to assess whether the number of spatial features used for classification affected the decoding scores (Fig. 1).

Across all datasets, the waveform-resolved burst rate features resulted in statistically significant improvement in decoding performance compared to beta band power estimated using a single filter ( $X^2(5) = 97.081$ ,  $p < 0.001$ ), non-significant differences compared to the beta band filter bank and mu-beta band single filter approaches ( $p = 0.1047$  and  $p = 0.9987$  respectively), and significant decoding score decrease compared to mu-beta band filter bank approach ( $p = 0.004$ ).

Moreover, we assessed whether the beta burst kernels can be exploited in a transfer learning approach. To do so, after trial rejection we iteratively estimated the kernels based on the recordings of one session and used them for transforming the recordings of the remaining sessions. We also examined whether we can use kernel-specific CSP filters for transfer learning. We compared these approaches to transfer learning of CSP filters after applying the standard filtering techniques.

The difference between the results depicted in Fig 1. and the transfer learning approach results per subject of

each dataset are depicted in Fig. 2. Negative values indicate greater decoding performance when estimating the results based on all available recordings, whereas positive values indicate an improvement in decoding score using transfer learning. Without surprise, transfer learning based on the beta burst kernels slightly reduced the overall decoding score, but was, interestingly, the only method that had the potential to improve decoding for some of the subjects. The two kernel-based transfer learning approaches yielded statistically non-significant discrepancies in decoding score difference ( $X^2(6) = 28.8$ ,  $p = 0.93$ ). Transfer learning based only on the kernels yielded significantly smaller reduction in decoding score difference compared to all filtering methods ( $p = 0.013$  vs beta band filter;  $p < 0.001$  vs beta band filter bank;  $p = 0.0237$  vs mu-beta band filter;  $p = 0.0011$  vs mu-beta band filter bank). Transfer learning based on kernel-specific CSP patterns resulted in significantly smaller decoding score differences in all comparisons except for filtering in the mu-beta band using a single filter ( $p = 0.0392$  vs beta band filter;  $p = 0.0161$  vs beta band filter bank;  $p = 0.3061$  vs mu-beta band filter;  $p = 0.034$  vs mu-beta band filter bank).

## DISCUSSION

Recently, the fields of neurophysiology and systems neuroscience have been experiencing a surge in the development of novel methods for analyzing neural recordings. An increasing number of articles are concerned with unveiling traditionally disregarded signal characteristics [11, 22-26, 29, 30] with several important implications for BCI. Particularly, the description of beta bursts has put into question the importance of band-limited power modulations in motor or MI tasks.

BCI applications often rely on signal power under the assumption of sustained and/or oscillatory signals. However, the description of waveform-specific beta burst modulations opens up new possibilities and holds the potential of improving decoding [31]. In this article we verified that waveform-resolved burst rate features can be informative markers of the underlying brain activity during MI tasks. In line with previous results we showed that indeed waveform-resolved burst rate features can be more informative than beta band power alone. We also showed that the information content of these features is comparable to that of filtering within a wider frequency band encompassing the mu band, although the mu-beta filter bank approach still yielded the best decoding results. A possible explanation for this finding is that beta burst kernels also capture slower modulations of the underlying activity and, thus, by adapting the waveform-resolved burst rate features to the mu band characteristics [45, 46] we may be able to further improve decoding.

Additionally, we adopted an inter-session transfer learning approach, and showed that the waveform-resolved features are relatively stable over recording sessions. We demonstrated that on the dataset level

reduction in classification score can be minimal, and that these features can even contribute in improving decoding for subjects with low classification score. This finding is important because it proves that, despite the difficulty of exploiting spatial features learned during previous sessions, there is potential in improving how previously acquired data can be leveraged during offline calibration or even online decoding sessions. In the future, an assessment of the across-subject kernel similarity may open the path for more reliable inter-subject transfer learning.

## CONCLUSION

The waveform-resolved burst rate analysis is a promising, neurophysiology-grounded alternative to classic descriptions of beta band activity in motor and MI tasks. The results of this work reaffirm that classification features based on beta bursts can efficiently decode MI binary classification tasks and suggest that beta bursts kernels are stable across recording sessions, thus potentially serving as an interesting feature inter-session transfer learning.

## REFERENCES

- [1] Pfurtscheller G 1981 Central beta rhythm during sensorimotor activities in man *Electroencephalogr. Clin. Neurophysiol.* 51 253–64
- [2] Pfurtscheller G and Berghold A 1989 Patterns of cortical activation during planning of voluntary movement *Electroencephalogr. Clin. Neurophysiol.* 72 250–8
- [3] Pfurtscheller G and Da Silva F L 1999 Event related EEG/MEG synchronization and desynchronization: basic principles *Clin. Neurophysiol.* 110 1842–57
- [4] Pfurtscheller G, Brunner C, Schlögl A and Da Silva F L 2006 Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks *NeuroImage* 31 153–9
- [5] Pfurtscheller G, Neuper C, Flotzinger D and Pergenzer M 1997 EEG-based discrimination between imagination of right and left hand movement *Electroencephalogr. Clin. Neurophysiol.* 103 642–51
- [6] Neuper C, Wörtz M and Pfurtscheller G 2006 ERD/ERS patterns reflecting sensorimotor activation and deactivation *Progress in Brain Research* vol 159 ch 14, pp 211–22
- [7] Pfurtscheller G, Stancák A and Neuper C 1996 Post-movement beta synchronization. A correlate of an idling motor area? *Electroencephalogr. Clin. Neurophysiol.* 98 281–93
- [8] Alayrangues J, Torrecillos F, Jahani A and Malfait N 2019 Error-related modulations of the sensorimotor post-movement and foreperiod beta-band activities arise from distinct neural substrates and do not reflect efferent signal processing *NeuroImage* 184 10–24
- [9] Makeig S, Enghoff S, Jung T P and Sejnowski T J

- 2000 A natural basis for efficient brain-actuated control *IEEE Trans. Rehabil. Eng.* 8 208–11
- [10] Pfurtscheller G and Neuper C 1997 Motor imagery activates primary sensorimotor area in humans *Neurosci. Lett.* 239 65–68
- [11] Little S, Bonaiuto J, Barnes G and Bestmann S 2019 Human motor cortical beta bursts relate to movement planning and response errors *PLoS Biol.* 17 1–30
- [12] Zich C, Quinn A J, Bonaiuto J J, O'Neill G, Mardell L C, Ward N S et al. 2023 Spatiotemporal organization of human sensorimotor beta burst activity *elife* 12 e80160
- [13] Tangermann M et al 2012 Review of the BCI competition IV *Front. Neurosci.* 6 1–31
- [14] B. Aristimunya, “Mother of all BCI Benchmarks”. Zenodo, Oct. 23, 2023. doi: 10.5281/zenodo.10034224.
- [15] Bruns A 2004 Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *J. Neurosci. Methods* 137 321–32
- [16] Herman P, Prasad G, McGinnity T M and Coyle D 2008 Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification *IEEE Trans. Neural Syst. Rehabil. Eng.* 16 317–26
- [17] Brodu N, Lotte F and Lécuyer A 2011 Comparative study of band-power extraction techniques for motor imagery classification *IEEE SSCI 2011—Symp. Ser. Comput. Intell.—CCMB 2011* 2011 IEEE Symp. Comput. Intell. Cogn. Algorithms, Mind, Brain pp 95–100
- [18] Koles Z J 1991 The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG *Electroencephalogr. Clin. Neurophysiol.* 79 440–7
- [19] Blankertz B, Kawanabe M, Tomioka R, Hohlefeld F U, Nikulin V and Müller K R 2008 Invariant common spatial patterns: alleviating non-stationarities in brain-computer interfacing *Advances in Neural Information Processing Systems 20—Proc. 2007 Conf.* Pp 1–8
- [20] Müller-Gerking J, Pfurtscheller G and Flyvbjerg H 1999 Designing optimal spatial filters for single-trial EEG classification in a movement task *Clin. Neurophysiol.* 110 787–98
- [21] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A et al. 2018 A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update *J. Neural Eng.* 15 031005
- [22] Jones S R 2016 When brain rhythms aren't 'rhythmic': implication for their mechanisms and meaning *Curr. Opin. Neurobiol.* 40 72–80
- [23] Lundqvist M, Rose J, Herman P, Brincat S, Buschman T and Miller E 2016 Gamma and beta bursts underlie working memory *Neuron* 90 152–64
- [24] Wessel J R 2020  $\beta$ -bursts reveal the trial-to-trial dynamics of movement initiation and cancellation *J. Neurosci.* 40 411–23
- [25] Shin H, Law R, Tsutsui S, Moore C I and Jones S R 2017 The rate of transient beta frequency events predicts impaired function across tasks and species *elife* 6:e29086
- [26] Torrecillos F et al 2018 Modulation of beta bursts in the subthalamic nucleus predicts motor performance *J. Neurosci.* 38 8905–17
- [27] Hannah R, Muralidharan V, Sundby K K and Aron A R 2020 Temporally-precise disruption of prefrontal cortex informed by the timing of beta bursts impairs human action-stopping *NeuroImage* 222 117222
- [28] Enz N, Ruddy K L, Rueda-Delgado L M and Whelan R 2021 Volume of  $\beta$ -bursts, but not their rate, predicts successful response inhibition *J. Neurosci.* 41 5069–79
- [29] Szul M J, Papadopoulos S, Alavizadeh S, Daligaut S, Schwartz D, Mattout J et al. 2023 Diverse beta burst waveform motifs characterize movement-related cortical dynamics *Prog. Neurobiol.* 165187
- [30] Rayson H, Szul M J, El-Khoueiry P, Debnath R, Gautier-Martins M, Ferrari P F, et al. 2023 Bursting with potential: how sensorimotor beta bursts develop from infancy to adulthood *J. Neurosci.* 43 8487–503
- [31] Papadopoulos S, Szul M J, Congedo M, Bonaiuto J J and Mattout J, Beta bursts question the ruling power for brain-computer interfaces *J. Neural Eng.*, 2024, 21 (1), pp.016010. <https://doi.org/10.1088/1741-2552/ad19ea>.
- [32] Langford Z D, Procyk E and Wilson C R E 2023 Frontal oscillatory beta bursts have rhythmically distinct regimes with differing functional relevance *bioRxiv* 1–19
- [33] Leeb R, Lee F, Keinrath C, Scherer R, Bischof H and Pfurtscheller G 2007 Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment *IEEE Trans. Neural Syst. Rehabil. Eng.* A 15 473–82
- [33] Zhou B, Wu X, Lv Z, Zhang L and Guo X 2016 A fully automated trial selection method for optimization of motor imagery based brain-computer interface *PLoS One* 11 1–20
- [35] Jas M, Engemann D, Bekhti Y, Raimondo F and Gramfort A 2017 Autoreject: Automated artifact rejection for MEG and EEG data *NeuroImage* 159 417–129
- [36] Moca V V, Bârzan H, Nagy-Dăbâcan A and Muresan R C 2021 Time-frequency super-resolution with superlets *Nat. Commun.* 12 1–18
- [37] Pedregosa F et al 2011 Scikit-learn: machine learning in Python *Fabian J. Mach. Learn. Res.* 12 2825–30
- [38] Shlens J 2014 A tutorial on principal component analysis (arXiv:1404.1100)
- [39] Gramfort A, Luessi M, Larson E, Engemann D A, Strohmeier D, Brodbeck C, et al. 2013 MEG and EEG data analysis with MNE-Python *Front. Neurosci.*
- [40] Harris C R et al 2020 Array programming with NumPy *Nature* 585 357–62

- [41] Bates D, Mächler M, Bolker B M and Walker S C 2015 Fitting linear mixed-effects models using lme4 J. Stat. Softw. 67 1–48
- [42] Fox J and Weisberg S 2019 An R Companion to Applied Regression (Sage)
- [43] Lenth R V 2023 emmeans: estimated marginal means, aka least-squares means
- [44] Gerster M, Waterstraat G, Litvak V, Lehnertz K, Schnitzler A, Florin E, et al. 2022 Separating Neural Oscillations from Aperiodic 1/f Activity: Challenges and Recommendations Neuroinformatics 20 991–1012
- [45] Vigué-Guix I and Soto-Faraco S 2022 Using occipital  $\alpha$ -bursts to modulate behaviour in real-time BioRxiv
- [46] Chen Y Y, Lambert K J M, Madan C R and Singhal A 2021 Mu oscillations and motor imagery performance: a reflection of intra-individual success, not inter-individual ability Hum. Mov. Sci. 78 1–12

# EEG-BASED STIMULUS CLASSIFICATION IN A FULL-BODY MOVEMENT, VIRTUAL REALITY PARADIGM

L. Rabe, Y. Pan, M. Klug

Young Investigator Group, Institute of Medical Technology, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany

E-mail: rabelea@b-tu.de

**ABSTRACT:** The use of EEG brain-computer interfaces (BCI) during movement is inherently difficult due to motion artifacts interfering with measured brain signals. Thus, most BCI research utilizes rather immobile conditions, thereby decidedly limiting its range of use cases. We aim to overcome this restriction by introducing a novel virtual reality paradigm which allows full-body movement of participants in combination with a processing pipeline specifically designed to deal with motion artifacts. Stimulus discrimination (target versus distractor) upon fixation was tested in 32 participants. Results indicate that targets elicit a higher P300 amplitude than distractors. Comparing the performance of different classifiers, shrinkage linear discriminant analysis (sLDA), support vector machine (SVM), and EEGNet, yielded equally sized, above chance classification accuracies. Overall, the results suggest the feasibility of studying and applying BCI in full-body motion paradigms given refined data preprocessing. The authors conclude with suggestions for future BCI studies in motion.

## INTRODUCTION

After the first years of brain-computer interface (BCI) research, during which BCI was mainly investigated as a means to partially compensate lost motor functions in people with severe motor disabilities, research on applications of BCI expanded to its use in non-medical areas. Following Zander and Kothe [1], BCI will be most beneficial to a wide range of applications if it does not replace or compete with fundamental human interaction patterns, like using hands or speech. Rather, it should substantially add value to the human-computer interaction (HCI) without distracting the user in his/her task. Because this kind of BCI covertly, or rather *passively* boosts HCI, it's referred to as passive BCI (pBCI) which will be the object of this paper.

Promising fields for BCI usage beyond medical application are summarized by van Erp, Lotte, and Tangermann [2] as device control, user state monitoring, evaluation, training and education, gaming and entertainment, cognitive improvement, as well as safety and security.

Even though, there is a definite vision to apply pBCI in real-life contexts, most of the research has been conducted in highly controlled laboratory settings which

are limited in their ecological validity. A major challenge in investigating pBCI under highly realistic (simulated) or real scenarios lies in their inherent complexity comprised of artefacts, non-brain influences, and other mental states [3]. Some research has been done to fill this gap and promising results were obtained, yet most studies were conducted in seated scenarios, notably driving, aviation, and desktop gaming. Of all studies included in the review [3], only one investigated participants who were standing and moving rather freely while performing a surgical task [4]. We believe it is imperative to conduct more studies allowing for free full-body movement in order to make pBCI applicable universally in real-life scenarios and not only in seated conditions. To achieve this goal, we introduce a new paradigm which allows participants to move and interact freely in a relatively fast-paced game-like scenario. First promising results of electroencephalography (EEG) analysis will be presented underlining the feasibility to work with pBCIs in a rather movement-intensive environment.

To deal with motion artifacts, we adopted methods from another discipline which has emerged to understand brain and body in motion: Mobile Brain-Body Imaging (MoBI; [5]). For MoBI studies, it's common to apply joint measurements of EEG, muscular activity, motion capture, and eye tracking. This, in combination with elaborate data processing techniques, like independent component analysis (ICA) for artifact rejection or machine learning for pattern recognition, proved to be efficacious in studying EEG in moving participants [6,7]. We combined MoBI with virtual reality (VR), an emerging tool to investigate more realistic scenarios, permitting participants to move around freely, while at the same time being highly controllable [8].

To demonstrate the feasibility of our setup, we investigated two well-established findings of BCI research. First, we aimed at replicating the P300 response and second, we compared the most common BCI classification algorithms in terms of their performance in a visual categorization task. In both cases we investigated fixation-related potentials (FRP), i.e. the cortical patterns locked to the onset of a fixation.

*Replicating the P300 component:* The main goal of this first step is a proof of concept: investigating whether we can replicate a typical pattern of cortical activation, the P300 in response to targets vs. distractors, in such a movement-rich paradigm.

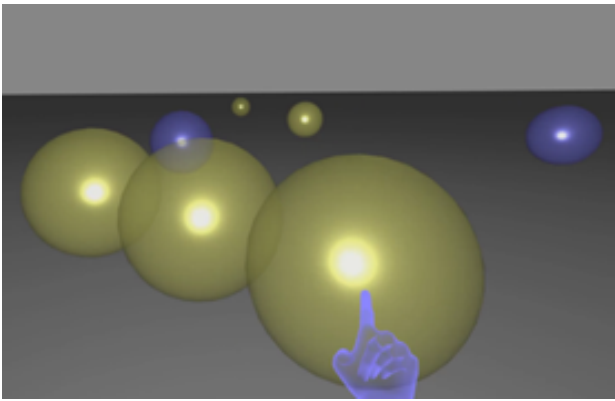


Figure 1: Stimulus presentation in virtual reality. Spheres flow towards the participant. The translucent blue hand represents positioning of the controller and bursts spheres upon touch. Depicted is a condition in which sphere color is permanently visible and colors are easy to distinguish.

While the P300 speller is probably the most popular application of this cortical potential in a BCI (first demonstrated by Farwell and Donchin, [9]), there are numerous examples of BCI studies investigating the P300 [10]. De Vos, Gandras, and Debener [11] investigated the P300 on walking participants while they performed an auditory oddball task. They were not only able to replicate the P300 component to rare targets, also single-trial classification in the P300 time window worked with a decent accuracy of 64%.

In our experiment, P300 was analyzed to visually presented targets and distractors. Since stimuli were appearing in a rather fast paced, free-viewing environment, some challenges had to be addressed to obtain a meaningful FRPs (for a detailed discussion of free-viewing paradigms see [12]). First, fixation durations were shorter than the following cognitive processes: a fixation usually lasts around 200-300ms [13], but some cognitive components, like P300, occur even later. Hence, components to subsequent fixations would overlap. This overlap can be controlled for mathematically with a linear regression [14,15]; which also deals with non-uniformly distributed artifacts caused by eye movements systematically affecting FRP averaging [16]. Thus, we chose to “detangle” FRPs in a regression-based approach. In the first part, we focused on the cortical activation to targets versus distractors in a time window of 200-600ms. It was hypothesized that amplitudes to targets will be larger than to distractors.

*Comparing classifiers for visual categorization:* Next, cortical activation to targets and distractors was classified with different algorithms in order to demonstrate the feasibility of EEG classification in such a motion-intensive VR task. Therefore, we compared the performance of the three most popular classifiers in BCI research as reviewed by Vărbu, Muhammad, and Muhammad [17]: linear discriminant analysis (LDA), support vector machines (SVM) and convolutional neural networks. We investigated the standard LDA as well as shrinkage LDA (sLDA) and we worked with EEGNet as neural network. LDA and sLDA both aim to find a linear combination of features that most effectively

separates two or more classes, with sLDA being less prone to overfitting [18]. SVMs find hyperplanes to maximize margins between different classes [19]. Lastly, EEGNet is a specialized compact convolutional neural network architecture tailored for the interpretation and analysis of EEG signals, designed to offer both high interpretability and robust performance in BCI tasks [20]. EEGNet performs feature extraction autonomously based on the data, whereas for LDA and SVM features must be extracted in a separate step.

For all methods, we hypothesize that the validation accuracy will be significantly above chance. Further, we expect the best classification performance for EEGNet, as it is specifically designed to classify EEG signals. Due to overfitting issues, LDA might work least accurately.

## MATERIALS AND METHODS

*Participants:* 48 participants were invited to the study and met the inclusion criteria: good health, sobriety, right-handedness, no preexisting neurological issues, normal or corrected-to-normal vision. During the experiment, 16 participants were excluded due to inaccurate eye-tracking (7), technical issues (6), motion sickness (1), pain from EEG cap (1), or below chance performance in the task (1). In total 32 participants (age 22-45,  $\bar{x} = 28.81 \pm 5.00$  years, 19 female) finished the experiment and were included into analysis.

*Study design and procedure:* After giving informed consent, participants answered demographic questions and were set up with an EEG. They then performed a workload calibration task which will be part of another analysis. The main object selection task was performed in a visually sparse VR environment created in Unity 3D. It consisted of a grey floor and sky with the controller represented by a translucent blue hand. In the experiment, spheres with a diameter of 0.5 m spawned with an angle of  $\pm 50^\circ$  left or right in front of the participants. Spheres were colored either bright blue or yellow, grayish blue or yellow, or in isoluminant gray. Spheres floated towards the participant where they disappeared either because the participant touched a sphere, or it reached the center of the virtual world. Upon destruction, three different sounds were played depending on whether the destruction was a hit, false alarm, or a miss. No sound was played for a correct rejection. Color, speed and spawn distance of the spheres depended on the condition. A 2x2x2 repeated measurements design resulted in varying difficulty levels with the factors: distinguishability (color was easy or hard to distinguish), predictability (constant or random inter-stimulus intervals), and visibility (color visible permanently or upon fixation). During 4 training blocks, participants got accustomed to the conditions. In the main part, all 8 conditions were presented in pseudo-random order, split into 2 blocks each, while the target/distractor color was counter-balanced over all participants. Each block contained 240 spheres (120 targets) and lasted 180s. After each block NASA-TLX [21] and 3D-SART [22] were collected for a separate analysis. In total, the

experiment lasted for 3.5-4.5 hours per participant, including breaks.

**Instruments:** For workload calibration a 27" HD monitor was used. The main experiment was conducted using a head-mounted display (HMD, HTC Vive) and an HTC VR controller. Positions of HMD and controller were tracked with the SteamVR Lighthouse tracking system. The HMD additionally had an inbuilt eye tracking system by SensoMotoric Instruments. Both motion and eye tracking data was streamed with a sampling rate of 90 Hz using Lab Streaming Layer (LSL, [23]). Events such as block starts and ends, sphere spawns and destructions, and eye gaze fixations were recorded in LSL. EEG was recorded using a 128-channel ANT eego sports system with passive Ag/AgCl electrodes and active cable shielding (ANT Neuro, Hengelo, Netherlands) with a backpack-worn tablet PC streaming the data wirelessly to LSL. To our knowledge, no high pass filter was applied during recording. The EEG was referenced to the vertex electrode, grounded with an electrode at the ear lobe and recorded with a 500 Hz sampling rate. Impedances were below 20 k $\Omega$ .

**EEG preprocessing:** Preprocessing of the EEG data was done in EEGLAB [24] in MATLAB using the BeMoBIL Pipeline with minor adaptations [25]. Taken together, preprocessing consisted of three steps: (1) Data import and synchronization of the different streams. (2) Data downsampling to 250 Hz, line noise removal using Zapline-plus [26], detection and interpolation of bad channels using the *clean\_rawdata* EEGLAB function, and referencing to the average. (3) Artifact removal using the adaptive mixture independent component analysis (AMICA; [27]), rejecting all non-brain components as determined by the ICLabel toolbox [28].

**Calculating fixation-related potentials with the Unfold toolbox:** To prepare for further analysis, the cleaned EEG data was filtered with a low pass filter of 35 Hz and a high pass filter of 0.2 Hz passband edges, respectively. To deconvolve overlapping EEG signals and to model the influence of artifacts the Unfold toolbox for MATLAB was used [29]. The Unfold toolbox was designed to recover isolated neuronal responses from originally overlapping cortical signals by reconstructing the deconvoluted signal mathematically. For every event of interest, a regression model is defined which is then fitted to each time point and channel relative to the onset of an event. The following events were supposed to influence the phenomenon of interest and were therefore included as a regression model: last fixation onset, fixation onset, final fixation exit, sphere spawn, sphere destruction, and sphere collision. For a more detailed description of the process including regression equations, see Rabe [30]. The Unfold toolbox returns beta weights which we then used to reconstruct the deconvoluted FRPs by summing up grand mean, regression weights of main effects, and interaction terms.

The following data was included into further analysis: (1) only trials in which participants correctly reacted to targets (hit) or distractors (correct rejection), (2) only last fixations on a sphere, i.e. before an action was performed

on it (hit a target) or not (dismiss a distractor), (3) activation from -1 to 2 ms around the fixation event (750 timepoints), (4) only Pz electrode because the most elevated P3 amplitudes can be expected over the centroparietal cortex [31]. Paired t-tests comparing target amplitudes against distractor amplitudes were conducted for each timepoint. Peak amplitudes to targets and distractors were averaged  $\pm$  10 ms around the peak in a time window of 200-600 ms. Then, a one-sided t-test (hit > distractor) was conducted. Normal distribution was assessed visually with QQ-plots and could be assumed.

**Classifying on target and distractors:** Classification was performed on the preprocessed but not unfolded dataset. All input time intervals were locked to fixation onset. In order to exclude brain responses related to motor execution we only included fixations on spheres that were more than 4 meters away (head to sphere distance), i.e. out of reach for the participant.

For classification, four algorithms were applied: LDA, sLDA, SVM, and EEGNet. Features needed to be extracted as input for LDA, sLDA, and SVM while EEGNet is designed to detect features automatically, thus it used the preprocessed data directly. Feature extraction will be described in the following.

First, epoching was done with a time window of [-1000 1500] ms and a baseline between [-400 -200] ms. Epoching steps were conducted in EEGLAB and resulted in a three-dimensional data matrix (channel x time x epoch). Second, the timeframe of [0 600] ms was used for feature extraction. This 600 ms period was split into 15 non-overlapping moving windows of 40 ms each. Third, amplitude averages were calculated for each window across all channels and epochs. Lastly, the resulting feature matrix (epoch x 1935) was fed into the classifiers (LDA, sLDA, SVM). To validate the classifiers' performance, 5-fold cross-validation was conducted. Finally, statistical significance levels were calculated with a permutation-based approach, written by Laurens Krol (based on [32]). It generates a synthetic dataset of the same size and randomly shuffles the classes 25,000 times. This produces a distribution of random correct assignments for comparison with our classifications. For  $\alpha = 0.01$ , significance is reached with 53.06% accuracy.

MATLAB R2021a and the EEGLAB 2022.1 toolbox were used for preprocessing. The SVM model used a linear kernel and a box constraint value of 0.01 to prevent overfitting. The LDA, sLDA and EEGNet models were developed using Python 3.8.8. For EEGNet, Keras (v3.0) was used. The software components were executed on a system equipped with the following hardware specifications: an AMD Ryzen 5 3600X 6-core processor running at a clock frequency of 3.80 GHz, 16 GB of RAM.

## RESULTS

**P300 to targets and distractors:** Our hypothesis was that amplitudes of the FRPs in a time window of 200-600 ms on Pz electrode would be larger to targets (i.e. hits)

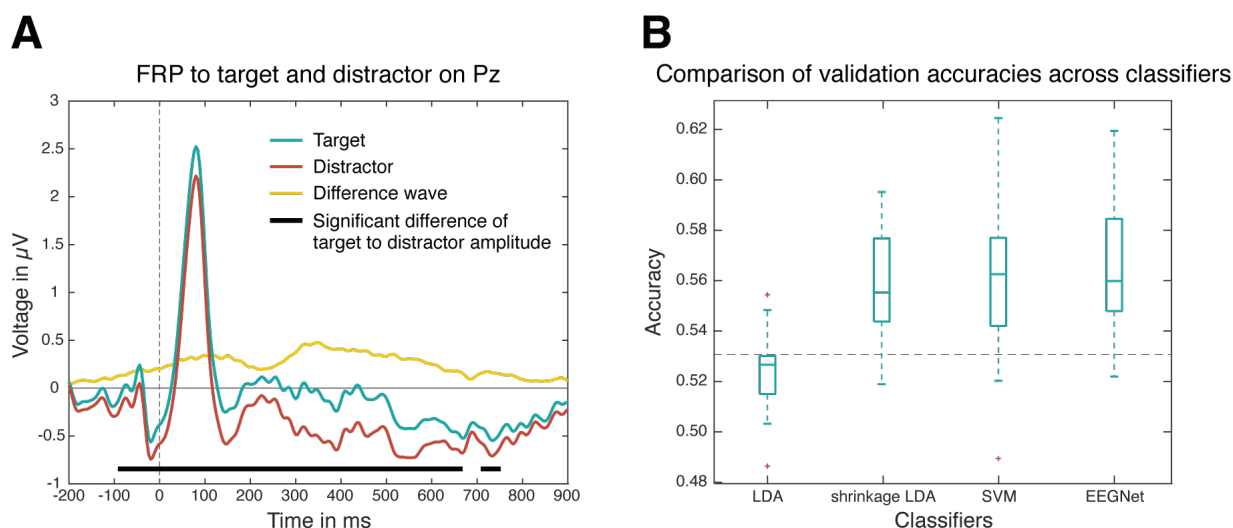


Figure 2: (A) Fixation-related potentials (FRP) to last fixation onset, 0 ms, on target or distractor. Significant differences in amplitude are indicated (pairwise t-test per timepoint,  $p < 0.05$ ) (B) Performance of different classifiers for classification of brain activation to targets or distractors. Each boxplot indicates the spread of accuracy percentages. Statistical significance level of above chance performance is indicated with a horizontal dashed line at 53.06%.

than to distractors (i.e. correct rejections). T-tests of the amplitudes on every single timepoint against each other revealed a significant difference with  $p < 0.05$  for most timepoints between around -100 ms to 750 ms (timepoints marked in Fig. 2). Comparison of the peak amplitudes indicated that during the 200-600 ms period after fixation onset, the average peak amplitude for targets was  $1.05 \mu\text{V}$  ( $SD = 0.96$ ), while for distractors it was  $0.79 \mu\text{V}$  ( $SD = 0.95$ ), showing that targets had a peak amplitude that was, on average,  $0.26 \mu\text{V}$  higher than that of distractors ( $SD = 0.47$ ). A paired one-sided t-test comparing the peak amplitudes (targets > distractors) confirmed a significant difference ( $t(31) = 3.12$ ,  $p < 0.01$ ,  $d = 0.55$ ).

**Classifier performances:** For the evaluation of EEG data classification, four algorithms—LDA, shrinkage LDA, SVM, and EEGNet—were assessed using 5-fold cross-validation. Fig. 2 demonstrates the distribution of the validation accuracies. On average, 2307 valid epochs were included for each participant, with an almost balanced ratio of targets to distractors. We hypothesized classification accuracies above chance for each method, with EEGNet performing best and LDA performing worst. Results show that all classifiers except LDA ( $\bar{x} = 52.7\%$ ) performed above random - sLDA  $\bar{x} = 55.5\%$ , SVM  $\bar{x} = 56.3\%$ , and EEGNet  $\bar{x} = 56.0\%$  -, i.e. median values exceeding the estimated threshold of 53.06 %. All classification accuracies for training and validation can be seen in Tab. 1.

## DISCUSSION

In this study, participants had to react to floating spheres of two colors in a VR environment. While the spheres were approaching towards the participants, they had to touch targets and dismiss distractors. We were interested in the cognitive reaction to targets and distractors upon

fixation and the possibility of a fixation-based BCI. Our analysis was split into two parts: (1) comparing FRP amplitudes upon last fixation onset to correctly identified targets (hits) and distractors (correct rejections), (2) assessing the performance of different classifiers (LDA, sLDA, SVM, EEGNet) to distinguish between targets and distractors. For the first part, the amplitudes in a P300 time window were larger for targets than for distractors, both during peaks and at each time point. In the second part, above chance classification accuracies were achieved for all classifiers but LDA. Performance of SVM, sLDA, and EEGNet was almost equal.

**Cortical activation around fixation:** Since the participants were in full-body motion during the experiment, we investigated whether it was generally possible to replicate a well-studied cortical response in our paradigm despite movement artifacts. The P300 was analyzed because it has been repeatedly shown to be a discriminator between targets and distractors [33–37]. Our analysis confirmed the hypotheses: during the P300 time window, FRP amplitudes to targets were higher than to distractors, both for the peak and for each timepoint. Unexpectedly, this difference emerged already 100ms before stimulus fixation which was not reported by other free-viewing studies [33,34,36]. The sustained elevation of amplitude to targets compared to distractors from -100 ms to 750 ms around fixation onset may stem from parafoveal processing before fixation onset, potentially present in all blocks where color visibility was constant

Table 1: Median training and 5-fold validation accuracies for the different classifiers

Classifier	Training	Validation
LDA	92.9%	52.7%
sLDA	71.8%	55.5%
SVM	68.8%	56.3%
EEGNet	60.1%	56.0%

(half of the trials). Such modulations might have affected the overall FRPs calculations. Indeed, early modulations of amplitudes can be seen in free-viewing tasks, and not in replay or oddball tasks [35]. Another possibility is that participants already distinguished between stimuli during earlier fixations. Unlike in the *last* fixation (which we investigated), a person is likely to decide during the *initial* fixation whether a stimulus requires further attention. This might influence the cortical activation to that same stimulus during following fixations. Supporting this idea, another study showed FRP modulations for repeated object fixation [38].

The prominent spike at approximately 40-120 ms post fixation onset most likely reflects a visually evoked lambda response, a potential unique to free-viewing studies and originating from the striate or extrastriate cortex [12].

Overall, and most importantly, we could demonstrate a substantial difference in amplitude to targets compared to distractors after stimulus fixation even in a paradigm with full-body movement and fast-paced events. This serves as a first proof of concept for the feasibility of analyzing brain responses in our novel VR interaction paradigm. Further investigations should address the potential influence of parafoveal stimulus discrimination and repeated stimulus fixation to improve understanding of cortical responses in free-viewing paradigms and to pave the way for EEG analysis in more realistic research scenarios.

*Classifying stimulus discrimination:* In an online classification of neuronal processes, it might be of interest to identify whether a participant is evaluating a stimulus as target or as non-target, for example to indicate whether the participant intends to interact with that stimulus. As a first step towards that goal, we compared the performance of four different classifiers, offline, to predict target and distractor discrimination from brain activation after fixating a stimulus. Three of the classifiers, sLDA, SVM, and EEGNet, yielded an above chance accuracy. Only LDA failed to reach that level. Low performance of LDA was predicted before because of its overfitting issues [18]. We expected the best performance of EEGNet, however all classifiers performed equally well with mean validation accuracies between 55.5-60.0%. One reason for the similar outcomes might have been the relatively noisy data, compared to other paradigms. We argue that it may create a ceiling effect in classification accuracy which could be topic of investigation in subsequent analyses.

We identified several factors that, if considered in future analyses, could improve classification accuracies. Probably, the overlap between cortical activation to subsequent stimuli hampered the classifiers' ability to effectively distinguish between targets and distractors. In the first part of our analysis, we showed that by "detangling" overlapping cortical responses with a regression-based calculation we were able to replicate results of studies without this overlap. It is advisable to come up with a similar method that can integrate well with classification to reduce the noise produced by

overlapping responses. Further, all EEG channels were used for classification without spatial filtering. Optimization could be achieved by concentrating on more relevant electrode locations. As target and distractor discrimination represented by the P300 is primarily found over the centroparietal cortex [31], we argue that respective electrodes should be elevated by a spatial filter. Finally, motion artifacts might still have obscured some of the brain signals as the paradigm was quite motion-intensive. Since all applications of BCI in motion will face similar issues we suggest the exploration and integration of more sophisticated artifact rejection techniques to ensure cleaner, more reliable data inputs.

## CONCLUSION

To the best of our knowledge, this was the first study classifying stimulus discrimination in a 3-dimensional VR environment with high stimulus frequency and full-body motion, investigating more innate interaction patterns than standard 2-dimensional monitor based experiments. The findings indicate that it is feasible to classify cortical activation patterns to stimulus discrimination despite body movements. Our results can be regarded as a promising first step to investigate and apply BCI in motion, making it more accessible for a wide range of human-computer interactions.

## REFERENCES

- [1] Zander TO, Kothe C. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J Neural Eng* 2011;8:025005.
- [2] van Erp J, Lotte F, Tangermann M. Brain-computer interfaces: Beyond medical applications. *Computer* 2012;45:26–34.
- [3] Aricò P, Borghini G, Di Flumeri G, Sciaraffa N, Babiloni F. Passive BCI beyond the lab: current trends and future directions. *Physiol Meas* 2018;39:08TR02.
- [4] Zander TO, Shetty K, Lorenz R, Leff DR, Krol LR, Darzi AW, et al. Automated Task Load Detection with Electroencephalography: Towards Passive Brain-Computer Interfacing in Robotic Surgery. *J Med Robot Res* 2017;02:1750003.
- [5] Makeig S, Gramann K, Jung T-P, Sejnowski TJ, Poizner H. Linking brain, mind and behavior. *Int J Psychophysiol* 2009;73:95–100.
- [6] Ladouce S, Donaldson DI, Dudchenko PA, Ietswaart M. Understanding Minds in Real-World Environments: Toward a Mobile Cognition Approach. *Front Hum Neurosci* 2016;10:694.
- [7] Gramann K, Ferris DP, Gwin J, Makeig S. Imaging natural cognition in action. *Int J Psychophysiol* 2014;91:22–9.
- [8] Tarr MJ, Warren WH. Virtual reality in behavioral neuroscience and beyond. *Nat Neurosci* 2002;5 Suppl:1089–92.

- [9] Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol* 1988;70:510–23.
- [10] Abiri R, Borhani S, Sellers EW, Jiang Y, Zhao X. A comprehensive review of EEG-based brain-computer interface paradigms. *J Neural Eng* 2019;16:011001.
- [11] De Vos M, Gandras K, Debener S. Towards a truly mobile auditory brain-computer interface: exploring the P300 to take away. *Int J Psychophysiol* 2014;91:46–53.
- [12] Dimigen O, Sommer W, Hohlfeld A, Jacobs AM, Kliegl R. Coregistration of eye movements and EEG in natural reading: analyses and review. *J Exp Psychol Gen* 2011;140:552–72.
- [13] Parasuraman R, Rizzo M. *Neuroergonomics: The Brain at Work*. Oxford University Press; 2008.
- [14] Smith NJ, Kutas M. Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology* 2015;52:169–81.
- [15] Smith NJ, Kutas M. Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology* 2015;52:157–68.
- [16] Nikolaev AR, Meghanathan RN, van Leeuwen C. Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain Cogn* 2016;107:55–83.
- [17] Värbu K, Muhammad N, Muhammad Y. Past, Present, and Future of EEG-Based BCI Applications. *Sensors* 2022;22.
- [18] Blankertz B, Lemm S, Treder M, Haufe S, Müller K-R. Single-trial analysis and classification of ERP components--a tutorial. *Neuroimage* 2011;56:814–25.
- [19] Samanta B, Al-Balushi KR, Al-Araimi SA. Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Eng Appl Artif Intell* 2003;16:657–65.
- [20] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J Neural Eng* 2018;15:056013.
- [21] Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock PA, Meshkati N, editors. *Advances in Psychology*, vol. 52, North-Holland; 1988, p. 139–83.
- [22] Taylor RM, Dietz AS. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In: Salas E, editor. *Situational Awareness*, Routledge; 2011.
- [23] Kothe C, Shirazi SY, Stenner T, Medine D, Boulay C, Grivich MI, et al. The Lab Streaming Layer for Synchronized Multimodal Recording. *BioRxiv* 2024:2024.02.13.580071.
- [24] Delorme A, Mullen T, Kothe C, Akalin Acar Z, Bigdely-Shamlo N, Vankov A, et al. EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Comput Intell Neurosci* 2011;2011:130714.
- [25] Klug M, Berg T, Gramann K. No need for extensive artifact rejection for ICA - A multi-study evaluation on stationary and mobile EEG datasets. *BioRxiv* 2022:2022.09.13.507772.
- [26] Klug M, Kloosterman NA. Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Hum Brain Mapp* 2022;43:2743–58.
- [27] Palmer JA, Makeig S, Kreutz-Delgado K, Rao BD. Newton method for the ICA mixture model. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE; 2008, p. 1805–8.
- [28] Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* 2019;198:181–97.
- [29] Ehinger BV, Dimigen O. Unfold: an integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ* 2019;7:e7838.
- [30] Rabe L. Investigating fixation-related brain responses underlying situation awareness in an interactive virtual reality experiment. M.Sc. Technische Universität Berlin, 2020.
- [31] Polich J. Neuropsychology of P300. In: Luck SJ, Kappenman ES, editors. *The Oxford Handbook of Event-Related Potential Components*, Oxford University Press; 2013, p. 159–88.
- [32] Mueller-Putz GR, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than Random? A closer look on BCI results. *Int J Bioelectromagn* 2008;10:52–5.
- [33] Brouwer A, Brinkhuis M, Reuderink B, Hogervorst, Erp JV. Fixation-related potentials : Foveal versus parafoveal target identification 2014.
- [34] Devillez H, Guyader N, Guérin-Dugué A. An eye fixation-related potentials analysis of the P300 potential for fixations onto a target object when exploring natural scenes. *J Vis* 2015;15:20.
- [35] Kamienkowski JE, Ison MJ, Quiroga RQ, Sigman M. Fixation-related potentials in visual search: a combined EEG and eye tracking study. *J Vis* 2012;12:4.
- [36] Kaunitz LN, Kamienkowski JE, Varatharajah A, Sigman M, Quiroga RQ, Ison MJ. Looking for a face in the crowd: fixation-related potentials in an eye-movement visual search task. *Neuroimage* 2014;89:297–305.
- [37] Brouwer A-M, Reuderink B, Vincent J, van Gerven MAJ, van Erp JBF. Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *J Vis* 2013;13:17.
- [38] Rämä P, Baccino T. Eye fixation-related potentials (EFRPs) during object identification. *Vis Neurosci* 2010;27:187–92.

# EEG MARKERS OF ACCELERATION PERCEPTION IN VIRTUAL REALITY

Gaël Van der Lee<sup>1</sup>, François Cabestaing<sup>1</sup>, Anatole Lécuyer<sup>2</sup>, Reinhold Scherer<sup>3</sup>,  
Hakim Si-Mohammed<sup>1</sup>

<sup>1</sup>Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

<sup>2</sup>Inria Rennes, IRISA, France

<sup>3</sup>University of Essex, Colchester, United Kingdom

E-mail: gael.vanderlee@univ-lille.fr francois.cabestaing@univ-lille.fr  
hakim.simohammed@univ-lille.fr

**ABSTRACT:** This study investigates neural patterns of acceleration in virtual reality (VR) using electroencephalography (EEG). Participants experienced accelerating white spheres in VR while EEG signals were recorded. Significant EEG differences were found at the fronto-central region between acceleration and slow speed, regardless of direction, and at the central region depending on the acceleration direction. Topographic responses also show differences in spacial patterns between the conditions. These findings give insights into the perception of acceleration in the brain and show potential for passive BCI applications.

## INTRODUCTION

Although there are numerous definitions of VR [1], it can be succinctly defined as “A real or simulated environment in which a perceiver experiences telepresence” [2]. The user’s psychological response to immersion in VR, termed “presence” [3], provides valuable insights into user behavior and experience. Understanding user responses to specific stimuli in VR is crucial for enhancing overall user experience and mitigating issues such as cybersickness [4].

Recently, there has been growing interest in leveraging BCIs to enhance VR experiences, aiming to create more immersive and interactive environments [5]. This paper investigates brain responses to acceleration in VR to identify potential markers of acceleration perception. Our experimental setup involved the simultaneous use of a 16-electrode EEG headset and a VR headset. Participants were presented with stimuli consisting of moving white spheres within the VR environment. These spheres initially moved at a slow constant speed before undergoing a sudden acceleration, either forward or backward, followed by a return to the initial speed. This experimental paradigm allowed us to examine the neural responses associated with the perception of acceleration in different directions.

The study of acceleration perception in VR is particularly relevant given the association between the subjective sensation of movement called vection and the user’s experi-

ence in VR. Specifically, cybersickness has been associated with vection [6] and could be better understood through that prism.

Thus, studying the perception of acceleration serves as a foundational step in investigating vection and its associated neural correlates.

We identify two potential neuromarkers:

1. A frontal marker of visual acceleration characterized by a positive potential between 300ms and 700ms after the start of the acceleration.
2. A signal differentiating the direction of the acceleration, whether it was forward or backward in the central region.

To the best of our knowledge, these findings have not been previously reported in the literature. They entail two potential meanings for the fields of neuroscience and BCIs. Firstly, they highlight fundamental cortical responses associated with acceleration perception. Secondly, they pave the way for passive BCIs that utilize this neuromarker to tailor the user experience accordingly. By identifying markers of acceleration perception, we can develop algorithms capable of discerning user attention towards acceleration events. Aligning detected accelerations with corresponding neuromarkers offers insight into user engagement with such stimuli.

In summary, this study contributes to the growing body of literature on BCIs and VR by hinting at the neural basis of acceleration perception and its potential applications in human-computer interaction and immersive technology.

## MATERIALS AND METHODS

*Objective:* This protocol aims to generate acceleration perception responses using Virtual Reality (VR). To achieve this goal, a user study was designed to trigger a potential through two types of trials: (1) sudden forward acceleration  $FA_1$  and (2) sudden backward acceleration  $BA_1$  (see Figure 1). Following this event, the participant slowed back down to their initial speed by either a forward ( $FA_2$ ) or a backward ( $BA_2$ ) acceleration.

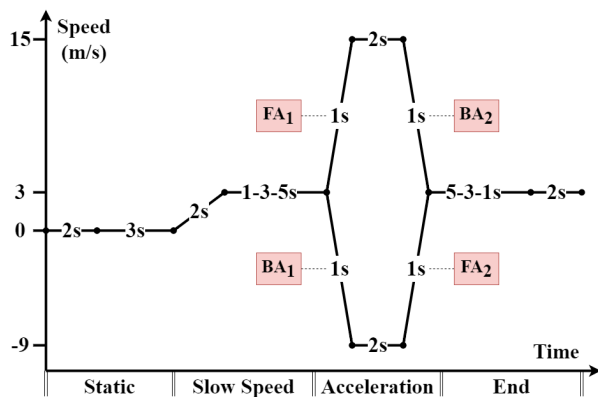


Figure 1: Illustration of a trial: Evolution of speed over time is depicted. Dashed durations represent variable delays, with one of the three delays chosen randomly. The second delay is selected to ensure that the cumulative sum of delays amounts to 6 seconds.

**Participants and Ethics:** Twenty healthy participants with normal or corrected-to-normal vision (12 men, 8 women; age  $\mu = 28.1$ ,  $\sigma = 7.99$ , min = 20, max = 56) took part in the experiment. Ethical approval for this study was obtained from the Ethics Committee of the University of Lille with approval number 2021-526-S97. The study adhered to the principles outlined in the Declaration of Helsinki. Written informed consent was obtained from all participants, who were explicitly informed of their right to withdraw from the experiment at any time without repercussion. Special attention was given to inform the participants that they can withdraw should they experience cybersickness. Data was anonymized and stored in compliance with the General Data Protection Regulation (GDPR). Participants were given their anonymized ID and could withdraw their consent at any time after the experiment, removing their recordings from the dataset.

**Experimental Setup:** The VE was displayed using a Valve Index Head-Mounted Display (HMD) connected to a DELL PRECISION 3640 personal computer with an NVIDIA RTX 3080 video card. EEG data were recorded using OpenVibe 3.1.0 software and a g.GAMMAcap2 EEG cap from g.tec medical engineering GmbH®(Austria) with 14 electrodes positioned at FPz, Fz, F1, F2, FCz, FC1, FC2, Cz, C1, C2, CPz, CP1, CP2, Pz and a reference electrode placed on the right earlobe. The VE was created using the Unity game engine software (version 2020.3.11f1).

**Trial Design:** Each trial lasted 17 seconds and consisted of four phases:

1. The *Static* phase: The environment fades in over 2 seconds and remains still.
2. The *Slow speed* phase: The environment accelerates to a speed of  $3\text{m/s}$  over 2 seconds and maintains this speed for a variable duration of 1 to 5 seconds. This phase is used to have an EEG baseline of visual stimulation without strong speed or acceleration.

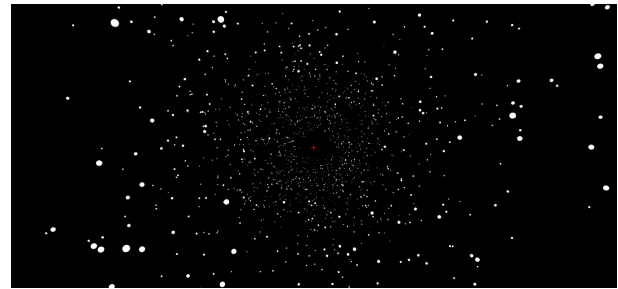


Figure 2: Depiction of the visual experience presented to participants, featuring a virtual scene composed of point clouds and a central crosshair.

3. The *Acceleration* phase: Participants experience a sudden forward acceleration  $FA_1$  or backward acceleration  $BA_1$  of  $12\text{m/s}^2$  for 1 second. The resulting speed is maintained for two seconds before returning to the initial slow speed with either a forward acceleration  $FA_2$  or a backward acceleration  $BA_2$ .
4. The *End* phase: The environment maintains a speed of  $3\text{m/s}$  for a duration matching the slow speed phase before fading out over 2 seconds.

**Environment:** The VE consisted of a minimalistic environment with stationary white spheres arranged cylindrically around the participant, following what has been done in the literature [7]. Participants were instructed to focus on a red crosshair at the center of the visual field to minimize ocular movements. Spheres gradually became visible from 150 meters away and were updated in real-time to reflect the current speed. The participant's view can be seen in Figure 2. Participants experienced 78 trials organized into four blocks. Each condition was evenly distributed amongst the trials, with 39 of each forward and backward trials, as well as 26 of each duration before acceleration (1, 3 or 5 seconds).

**Data Processing:** EEG data were processed using MNE-python for filtering and epoching. Noisy channels were identified and excluded, and data were re-referenced using common average referencing (CAR). Data were re-sampled to 128Hz and filtered from 0.3 to 10Hz using a 4th order Butterworth filter. Epochs ranged from 0.5s before stimulus onset to 1s after stimulus offset. Epochs containing voltage exceeding  $125\mu\text{V}$  were rejected. Data were stored in EEG Brain Imaging Data Structure (BIDS) standard for easy sharing. Visualization was performed using the seaborn library.

## RESULTS

In our subjects, we split the EEG data between different conditions:

- **Baseline:** This is taken during the slow speed phase, as shown in Figure 1. Where the subject is going at a constant speed of  $3\text{m/s}$ .
- $FA_1, BA_1, FA_2, BA_2$  As described in the *Trial Design* subsection.

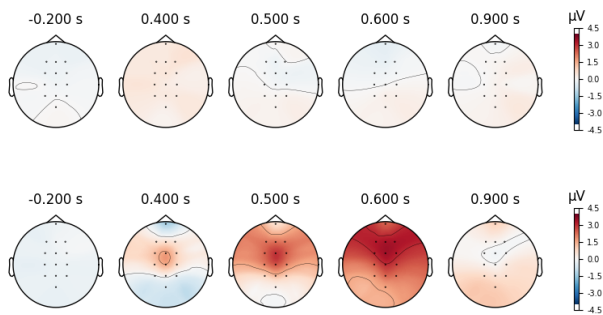


Figure 3: Topographic map comparison of the average response for all subjects to baseline (top) and the first acceleration,  $FA_1$  and  $BA_1$ , averaged (bottom)

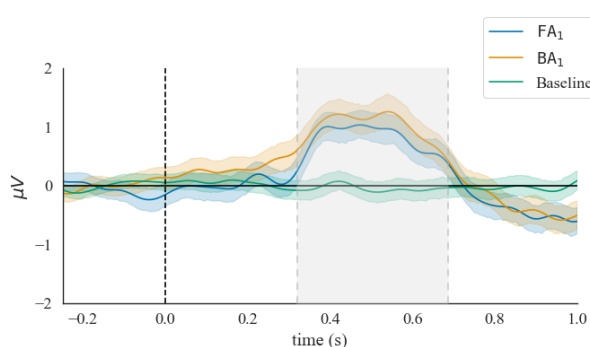


Figure 4: Mean values of the FCz electrode for  $FA_1$  (blue),  $BA_1$  (orange), and baseline (green). The 95% confidence interval is shown for each event. The shaded gray area corresponds to the time period where there was a statistically significant difference between the signals.  $FA_1$  and  $BA_1$  present similar patterns distinguishable from the baseline.

In this study, we evaluate the significance of observed differences using a non-parametric bootstrapping approach. First we perform 10,000 resamples on our data with replacement. Then we compute the 95% confidence intervals which correspond to the range between the 2.5th and 97.5th percentiles of the resampled data distribution. These confidence intervals are represented as shaded areas in the figures. Additionally, we utilize topographic maps to underscore spatial differences between conditions, with cubic interpolation applied to obtain values between electrodes.

**Marker of acceleration:** The spacial response shows differences between the baseline and the acceleration condition as shown in Figure 3 which presents a much higher positive peak along the central regions and especially the fronto-central region peaking at 600ms. A better temporal representation can be found in Figure 4, which shows a significant difference at electrode FCz between periods of strong acceleration ( $FA_1$  or  $BA_1$ ) compared to slow speed (3m/s), regardless of acceleration direction. We see the characteristic strong positive potential between 300 and 700ms after acceleration onset. Both  $FA_1$  and  $BA_1$  follow a similar pattern in this region. This particular pattern in the region could represent a marker

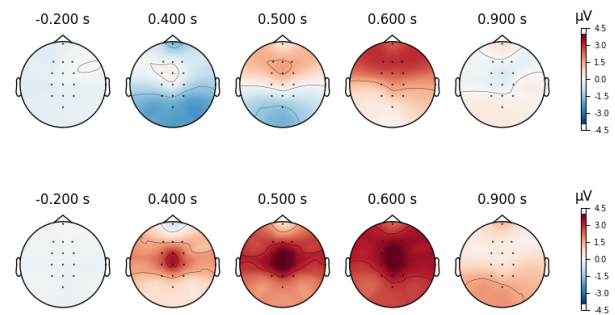


Figure 5: A comparison of the average topographic map across all subjects for  $FA_1$  (top) and  $BA_1$  (bottom)

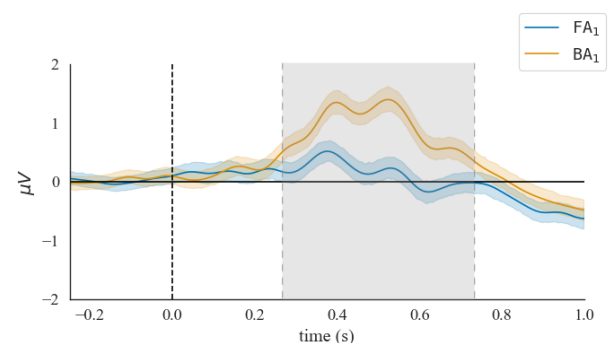


Figure 6: Mean values of the Cz electrode for  $FA_1$  (blue) and  $BA_1$  (orange). The 95% confidence interval is shown for each event. The shaded gray area corresponds to the time period where there was a statistically significant difference between the signals. The Cz presents a significant difference between  $FA_1$  and  $BA_1$ .

of acceleration, regardless of direction.

**Marker of direction:** The spacial response to the direction of acceleration also shows differences. Comparing the topographic maps in Figure 5, we find a stronger negativity in the parietal region for the  $FA_1$  condition around 400ms, that becomes a frontal and positive over the course of the next 200ms. The  $BA_1$  condition shows a much stronger negativity, especially around the central electrodes, that increases until 600ms. Looking at the Cz electrode, a significant difference was found between forward and backward acceleration, as seen in Figure 6. The  $BA_1$  condition differentiates from the  $FA_1$  condition by showing a higher peak around 400ms and keeping a stronger positive potential until 700ms after acceleration onset.

**Return to slow speed:** Responses when returning to the slower speed were not as pronounced as the responses we found in the initial acceleration condition. We did find a distinct marker for the return to slow speed section during the Acceleration phase, as seen in Figure 7. First we observe a small difference in conditions around 350–500ms, which is followed by a more pronounced late marker. This late marker showed a significant difference between the  $FA_2$  and  $BA_2$  conditions between 750–1000ms after acceleration onset.

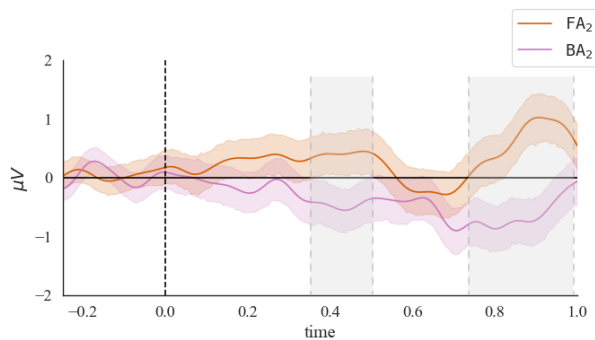


Figure 7: Mean values of the Pz electrode for  $FA_2$  (brown) and  $BA_2$  (purple). The 95% confidence interval is shown for each event. The shaded gray area corresponds to the time period where there was a statistically significant difference between the signals. The Pz presents a significant difference between  $FA_2$  and  $BA_2$ .

## DISCUSSION

**Interpretation:** The identification of distinct neural activation patterns in response to acceleration stimuli opens the door for a deeper understanding of acceleration perception circuits in the brain. Further understanding of the neural mechanisms underlying acceleration perception may reveal insights into motor control processes. Some of these responses, notably the one exhibited in Figure 6 bears resemblance to P300 responses from the literature in its timing and location [8]. As the P300 is associated with surprise and decision-making, the perception of a sudden acceleration could trigger a decision-making process in the brain.

Importantly, these findings lay the foundation for future research on passive BCIs that use acceleration perception as a neural signal. Acceleration perception could be used in a similar manner to established passive BCI signals such as mental workload [9, 10] or changes in error-related potentials [11]. Such systems could adapt a user's environment and inputs, knowing if the user perceived an acceleration and in which direction he perceived it. For example, a passive BCI could detect if a driver in a vehicle is paying attention to the road by using the markers of acceleration perception along with an accelerometer.

**Limitations and future work:** In this paper, we find patterns when going from a slow speed to a high speed ( $FA_1$  and  $BA_1$ ), but very different patterns when starting from a high speed and going back to normal ( $FA_2$  and  $BA_2$ ). All four patterns are unique, we do not explain this difference, and it warrants further study to be better understood. Moreover, while this study shows a signal specific to acceleration perception and direction, these findings are limited to specific conditions: the subject is sitting, in VR, with a stimulation consisting of white spheres.

Continued research in this direction could uncover how the response evolves as we add sensory indicators of acceleration such as sounds. One could also compare this signal to those exhibited ecological experiences of acceleration in VR as well as in real-world scenarios. Beyond

the perception of motion, this opens the door to understanding how the brain perceives self-motion induced by visual stimuli, orvection, which bears importance for the field of VR and cybersickness. Thus, another direction could be correlating this signal with the subjective perception of the participant.

## CONCLUSION

In this study, we find neural responses associated with perceptual changes induced by sudden accelerations in VR. We find different spatial responses characteristic of both acceleration and acceleration direction. We also uncover differences in EEG signals at electrodes FCz and Cz during acceleration perception, suggesting the existence of distinct neural markers for acceleration direction. These findings have implications for the development of passive BCIs and the enhancement of virtual reality experiences. By leveraging these neural markers, future research can design adaptive BCIs and create more immersive and interactive VR environments. Overall, this study contributes to advancing our understanding of the neural mechanisms underlying acceleration perception and paves the way for innovative applications in brain computer interfaces.

## ACKNOWLEDGMENTS

This research was funded by the European Union ERA-NET CHIST-ERA 2020 research and innovation program under grant number CHIST-ERA-20-BCI-003 - "GENESIS" and the French Research Agency (ANR) under grant number ANR-21-CHRA-0001-01.

## REFERENCES

- [1] Xiong J, Hsiang EL, He Z, Zhan T, Wu ST. Augmented reality and virtual reality displays: Emerging technologies and future perspectives. *Light: Science & Applications*. 2021;10(1):216.
- [2] Biocca F, Levy MR. *Communication in the Age of Virtual Reality*. Routledge (2013).
- [3] Bowman DA, McMahan RP. Virtual Reality: How Much Immersion Is Enough? *Computer*. 2007;40(7):36–43.
- [4] Nesbitt K, Davis S, Blackmore K, Nalivaiko E. Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays*. 2017;48:1–8.
- [5] Tan X, Li Y, Gao Y. Combining brain-computer interface with virtual reality: Review and prospect. In: 2017 3rd IEEE International Conference on Computer and Communications (ICCC). Dec. 2017, 514–518.
- [6] Keshavarz B, Riecke BE, Hettinger LJ, Campos JL. Vection and visually induced motion sickness: How are they related? *Frontiers in Psychology*. 2015;6.

- [7] Keshavarz B, Philipp-Muller AE, Hemmerich W, Riecke BE, Campos JL. The effect of visual motion stimulus characteristics on vection and visually induced motion sickness. *Displays*. 2019;58:71–81.
- [8] Polich J. Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*. 2007;118(10):2128–2148.
- [9] Luong T, Martin N, Raison A, Argelaguet F, Diverrez JM, Lécuyer A. Towards Real-Time Recognition of Users' Mental Workload Using Integrated Physiological Sensors Into a VR HMD. In: *ISMAR 2020 - IEEE International Symposium on Mixed and Augmented Reality*. Virtual, Brazil, Nov. 2020, 1–13.
- [10] Roy RN, Bonnet S, Charbonnier S, Campagne A. Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Jul. 2013, 6607–6610.
- [11] Dehais F *et al.* Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions. *Sensors*. 2019;19(6):1324.

# EYE-TRACKING AND SKIN CONDUCTANCE TO MONITOR TASK ENGAGEMENT DURING NEUROFEEDBACK SESSIONS

A. Fragueiro<sup>1</sup>, R.-P. Debroize<sup>1</sup>, E. Bannier<sup>1,2</sup>, C. Cury<sup>1</sup>

<sup>1</sup>Univ Rennes, Inria, CNRS, Inserm - IRISA UMR 6074, Empenn ERL U 1228, Rennes, France

<sup>2</sup>Rennes University Hospital, Rennes, France

E-mail: agustina.fragueiro@inria.fr; claire.cury@inria.fr

**ABSTRACT:** The neurofeedback (NF) inefficacy problem refers to the variability in NF success and has been associated with attentional and motivational factors. Sustaining attention on any task over an extended period is demanding and leads to attentional drops. By using eye-tracking and skin conductance, we aimed at extracting physiological features linked to cognitive work, with the further purpose of monitoring changes in task engagement during NF sessions. Here, we present preliminary results on pupil diameter (PD) and phasic skin conductance responses (ISCR) linked to cognitive task execution. We observed that changes in both features are associated with performance and time-on-task. Thus, PD and ISCR decreased along the task while the performance increased. However, this trend is affected by manipulation of the task difficulty level. We also monitored, in the same participants, PD and ISCR during one NF session. Finally, we discussed preliminary ideas for target adaptation during NF sessions based on eye-tracking and skin conductance monitoring.

## INTRODUCTION

Neurofeedback (NF) consists in feeding-back a patient with information about its neural activation to learn self-regulating its own brain activity [1]. It is therefore a powerful technique to trigger brain plasticity [2]. More importantly, NF has been postulated as a brain rehabilitation technique as it has the potential to reduce morbidity by correcting maladaptive patterns of brain function associated with a broad range of brain disorders [3]. NF is usually based on real-time electroencephalography (EEG) feature extraction, and it has been studied for several decades [4]. However, the NF inefficacy problem refers to the variability in NF success, as around 38% of participants undergoing NF training do not learn to regulate their own brain activity [5]. Among the different elements that may influence NF response, motivational [5] and attentional [6] factors have been identified as predictors of both performance and learning. Interestingly, motivation is likely to influence attention, as poorer performances can increase fear of incompetence and reduce mastering confidence which can lead to disengagement with the task and a potential label of “non-responder” [7]. To have the best success in NF training, it has been sug-

gested to monitor participant’s motivation [8], as well as, to adapt NF sessions to the participant [9]. A recent review on NF for post-stroke motor rehabilitation concluded that adaptation of NF target could lead to better meet patients’ needs [10].

Sustaining attention on task-relevant information over an extended time is crucial for successful performance in any task, however, it is demanding and leads to attentional lapses (i.e., disengagement from the task). Performance relies on the “inverted-U shape” relationship between arousal and attentional states, linked to different on-task and off-task engagement states [11]. Only intermediate arousal activity is linked to task engagement and good performance [12]. Eye-tracking (ET) [13] and skin conductance (SC) [14] tools have been extensively used to measure physiological features related to both attention and arousal levels.

In the present study, by using ET and SC, we aim to extract features linked to cognitive task execution, with the further purpose of monitoring changes during NF sessions. To synchronize our multi-modal set-up, and to extract features of interest, we first collected data while participants were engaged in cognitive tasks. As proof of concept, the same participants performed one NF training session to observe physiological changes over time. Here, we present preliminary results on pupil diameter, as it has been strongly associated with cognitive load [15], performance [16], fatigue and task engagement [17], and on the phasic component of SC activity, as it has been linked with arousal [18] and it changes faster than the tonic component [14]. Finally, we monitored the same features during the NF session.

## MATERIALS AND METHODS

*Participants:* Twenty right-handed volunteers (11 females, age range=20 – 60 yo) reporting normal vision to watch the screen participated in our study after signing informed consent. This study has been accepted by the COERLE, the Ethics Review Board at INRIA complying with the European General Data Protection Regulation.

*Procedure:* The protocol was divided into a 10-minute session of NF, followed by 30 min of randomised cognitive tasks, conducted while sitting in front of a screen with 1920 × 1080 display resolution in a dark room, while

simultaneously recording EEG, ET and SC (Fig. 1A).

**Cognitive tasks:** Participants completed two different tasks aiming to stimulate workload, (1) an adapted version of the Stroop task [19], and (2) an Addition task adapted from [20]. (1) During the Stroop task, participants were presented with two rows of words. In the row above the name of a colour (jaune, vert, rouge, bleu) is presented with the font in one of the following colours: yellow, green, red, and blue. In the row below, the name of a colour is written in black. The task of the participant was to judge if the meaning of the word below corresponds to the colour of the font of the row above or not, and press the corresponding key. Each trial was presented for 1250ms, followed by a 30ms inter-trial interval (ITI). Three task blocks (2min each), were interleaved with three rest blocks (2min each). (2) In each trial of the Addition task, a number from 1 to 9 was presented in the centre of the screen flanked by two other numbers. The task of the participant was to add the last two numbers presented in the centre of the screen and select the correct response among the flanker numbers by pressing the corresponding arrow on the keyboard. Three blocks of task (1min each) were alternated with three rest blocks (1min each). The time of appearance of the numbers decreases across the blocks to increase the difficulty level and prevent habituation (3sec in the 1st block, 2sec in the 2nd, and 1sec in the 3rd). An ITI of 30ms was set between trials. For both tasks, during rest blocks, a heart-coherence disengagement video was presented in which a blue dot increase (in 4sec) and decrease (in 4sec) repeatedly in the center of the gray screen.

**Skin conductance:** A BrainVision galvanic skin response set was used to acquire electrodermal activity from the index and middle fingers. After downsampling to 10Hz, SC responses were estimated through a Continuous Decomposition Analysis using MATLAB toolbox Ledalab [21]. Integrated phasic driver activity (ISCR), which corresponds to the area of the phasic driver within each temporal window, was extracted by setting 10sec consecutive temporal windows. Z-scores were computed along each task.

**Eye-tracking:** Eye activity was recorded by using a screen-based eye-tracker Tobii Pro X2-120 and the output was saved with Tobii Pro SDK. When pupil detection was judged as valid for both eyes according to the SDK validity codes, pupil diameter (PD) was averaged between both eyes. Z-scores were calculated for each task. To observe PD progression along time and for plotting purposes, PD was averaged within 10sec temporal windows.

**Neurofeedback:** Participants were instructed to perform a motor imagery task with both hands simultaneously. They were presented with a visual metaphor, a yellow ball moving inside a blue square rotated 90 degrees (Fig. 1B), depicting event-related desynchronization (ERD) activity of the C3 and C4 electrodes. The participant's goal was to keep the ball during all the runs in the upper corner of the metaphor, corresponding to a simultaneous motor imagery of both hands. They were

informed that if their focus was more directed to the right or the left hand, the ball would move to the respective corner of the metaphor. The NF session consisted of one calibration run, eight training runs, and one post-run. Each run was composed of one rest (30sec) followed by one task block (30sec). To record EEG activity, we used a Brain Products actiCAP set of 32 active electrodes and an actiChamp amplifier. The Cz electrode was used as the reference. The OpenVibe software [22] was used for signal processing. Signals were epoched on the last 2 sec every 0.25 sec. A discrete Laplacian spatial filter with an 8 coefficient for the C3 and C4 channels and a -1 coefficient for their respective neighbours FC5, FC1, CP5, CP1 and FC6, FC2, CP6, CP2 was applied to the resulting samples,  $Lap(C3)$  and  $Lap(C4)$ . Other channels were ignored. The power of  $Lap(C3)$  and  $Lap(C4)$  denoted  $Bp(C3)$ ,  $Bp(C4)$  is computed in the 8-30Hz frequency band. Powers  $Bp(C3)$  and  $Bp(C4)$  were continuously sent to a dedicated program via the LSL library to compute neurofeedback scores and display the ball in the right position on the metaphor. Neurofeedback scores for each electrode were computed with  $Score(t) = \frac{Bp_{ref} - Bp(t)}{Bp_{ref}}$  where  $Bp_{ref}$  is defined as  $Bp_{ref} = med(Bp_{rest}|_{[10,20]})$ , the median of the bandpower during the 10 sec central interval of the last rest block [23]. The goal target was set for each hand independently based on the 70th percentile of the scores achieved during calibration. If this score was lower than 0.15, the target was set to 0.15.

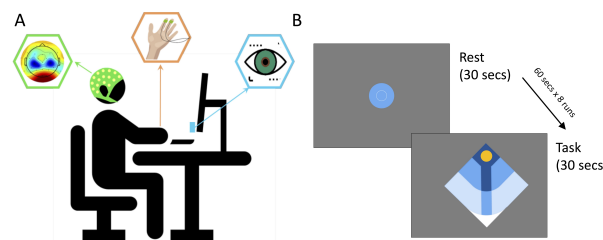


Figure 1: (A) Setup for data acquisition synchronously collecting EEG, eye-tracking and skin conductance signals, while the participant is engaged in a neurofeedback session and in resolving cognitive tasks presented on the screen. (B) A neurofeedback run composed of 30 sec rest, in which a heart-coherence video (increasing/decreasing blue dot), and 30 sec task, in which participants performed a motor imagery task of both hands simultaneously.

In addition, we preprocessed data offline with EEGLAB v2022.0 to test the difference between rest and task conditions during NF training. EEG was pass band filtered (1-40Hz) and re-referenced to average. EEG signal was corrected for ocular, muscular and noise artefacts using ICLabel 1.5 [24]. For power spectrum analyses, EEG was first selected with 1-29 sec time limits based on events corresponding to the beginning of the rest or the task conditions. For each condition, data was epoched in 2-sec no-overlapping temporal windows. Power in 8-30Hz fre-

quency band was compared between rest and task conditions. On the other hand, to observe ERS/ERD, data was epoched on 60 sec no-overlapping temporal windows, including rest (30 sec) and task (30 sec) conditions for each run (8 x subject).

**Statistical analysis:** To observe differences in performance along the task blocks, a one-way ANOVA was conducted for each task including accuracy as a dependent variable and block order (i.e. 1,2,3) as independent factor. For statistical analyses on both ISCR and PD, each task was analyzed independently and z-scores were averaged within each block. First, a one-way ANOVA was conducted including ISCR z-scores as the dependent variable and condition as the independent factor (i.e. baseline, task, rest). A two-way ANOVA was conducted including ISCR as the dependent variable and condition (i.e. task vs rest) and order (i.e. 1,2,3) as independent factors. To investigate the relationship between ISCR and accuracy, Pearson's correlations were conducted between ISCR and accuracy for each task block independently, for the total task (average among all three blocks), and for the whole task but including each block score separately. In addition, a delta score (block 3 - block 1) was calculated for both ISCR and accuracy scores. The correlation between these delta scores was also tested. The same analyses were conducted using mean PD z-scores within each block as a dependent variable on 13 subjects as the eye-tracking data for the first 7 subjects was excluded due to a change in the screen luminosity. To test learning during the NF session, paired sample t-test were conducted between calibration and post NF scores for C3, and for C4 on 18 subjects due to recording problems in the post-block for the first 2 subjects. The difference between NF scores for C3 and C4 was tested for both the calibration and the post runs. Finally, to observe changes in the NF score along the session, a one-way ANOVA was conducted one for C3 and one for C4 including NF scores as dependent variable and run as independent factor. To test changes in the synchronization between both hands, a one-way ANOVA was conducted including the difference in the NF scores between C3 and C4 as dependent variable. To compare power spectrum between the rest and task conditions during the training, paired t-tests were conducted for all channels independently (31 channels) and p-values were Bonferroni corrected. Analyses to investigate changes in ISCR along the NF session were conducted on 16 subjects due to recording problems on the first 4 subjects. A two-way ANOVA was conducted including ISCR as dependent variable and condition (i.e. rest vs task) and run (i.e. from 1 to 8) as independent factors. The same analysis was repeated for PD as dependent variable in 13 subjects. All p-values corresponding to post hoc tests included in this study were Bonferroni corrected. Statistical analyses were conducted on JASP 0.17.2.1.

## RESULTS

**Behavioral performance:** For the Stroop task, mean accuracy across participants in the task was 80% ( $sd = 0.13$ ) for the 1st block, 89% ( $sd = 0.09$ ) for the 2nd block, and 91% ( $sd = 0.07$ ) for the 3rd block. The difference in accuracy between blocks was significant ( $F(2,57) = 7.263$ ,  $p = 0.002$ ). Specifically, accuracy in the 1st block was lower than in the 2nd ( $t(38) = -2.793$ ,  $p = 0.021$ ) and 3rd ( $t(38) = -3.642$ ,  $p = 0.002$ ) blocks. Thus, performance increased along the task. For the Addition task, the mean performance across participants in the task was 89% accuracy ( $sd = 0.15$ ) for the 1st block, 80% ( $sd = 0.17$ ) for the 2nd block, and 44% for the 3rd block ( $sd = 0.18$ ). As expected, the difficulty level between blocks was different ( $F(2,57) = 40.755$ ,  $p < 0.001$ ), with the 3rd block being significantly more difficult than the others (both  $p < 0.001$ ).

**Skin Conductance:** For the Stroop task (Fig 2A), the one-way ANOVA evidenced a significant difference in ISCR between conditions (i.e. baseline, task, rest) ( $F(2,137) = 26.071$ ,  $p < 0.001$ ). Post hoc comparisons evidenced that the ISCR for the rest condition was significantly lower compared to the baseline ( $t(78) = -5.022$ ,  $p < 0.001$ ) and the task condition ( $t(118) = -6.629$ ,  $p < 0.001$ ). The two-way ANOVA, including ISCR as dependent variable and condition (i.e. task vs rest) and block order (i.e. 1,2,3) as independent factors evidenced a significant main effect of both condition ( $F(1,114) = 73.537$ ,  $p < 0.001$ ) and order ( $F(2,114) = 17.583$ ,  $p < 0.001$ ), and a significant interaction ( $F(2,114) = 10.104$ ,  $p < 0.001$ ). Post hoc comparisons showed that ISCR during rest was lower than the task condition ( $t(118) = -8.575$ ,  $p < 0.001$ ) and that ISCR in the 3rd block was lower comparing the 1st ( $t(78) = -5.898$ ,  $p < 0.001$ ) and 2nd blocks ( $t(78) = -3.480$ ,  $p = 0.002$ ). Specifically, the task 1st block was significantly different from both the 2nd ( $t(38) = 4.659$ ,  $p = 0.002$ ) and 3rd ( $t(38) = 6.673$ ,  $p < 0.001$ ) task blocks, and from all the rest blocks (all  $p < 0.001$ ). ISCR in stroop task blocks 2 and 3 was not different, but the 2nd task block was different from the 1st and 3rd rest blocks (both  $p < 0.002$ ). The rest blocks were no different. Mean ISCR during all three

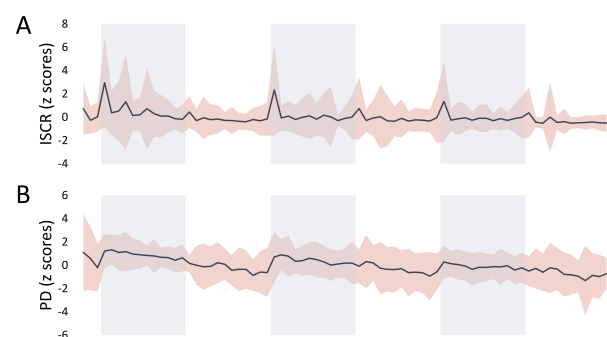


Figure 2: Stroop task. (A) Skin conductance responses (ISCR) and (B) Pupil's diameter (PD). Black line: the mean across subjects. Red area:  $\pm 2$  standard deviations. Blue shadow: task blocks. White areas: rest blocks (2 min each block). Total duration: 12 min plus 30 sec baseline at the beginning.

tasks blocks negatively correlated with mean total accuracy in the Stroop task ( $r = -0.48$ ,  $p = 0.034$ ) and, when observing the relationship for each block independently, ISCR negatively correlated with accuracy in the 2nd block ( $r = -0.44$ ,  $p = 0.05$ ) and the same tendency was observed in the 3rd block ( $r = -0.39$ ,  $p = 0.09$ ). Delta ISCR did not correlate with delta or mean accuracy. The correlation between accuracy scores and ISCR for the complete task, including all blocks independently, was significant ( $r = -0.47$ ,  $p < 0.001$ ), reflecting the decrement in ISCR and the improvement in accuracy along the task (Fig. 3A).

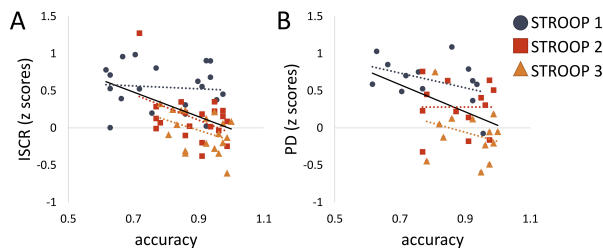


Figure 3: Stroop task. Pearson's correlations between accuracy scores and (A) Skin conductance responses (ISCR), and (B) pupil's diameter (PD). Dotted lines: linear correlations for each block. Black line: correlation over all blocks.

For the Addition task (Fig. 4A), the one-way ANOVA evidenced a significant difference in ISCR between conditions (i.e. baseline, task, rest) ( $F(2, 137) = 42.773$ ,  $p < 0.001$ ). Post hoc comparisons evidenced that the ISCR during the addition task was significantly different from both the baseline ( $t(78) = 6.210$ ,  $p < 0.001$ ) and the rest ( $t(118) = 8.607$ ,  $p < 0.001$ ). The two-way ANOVA, including ISCR as dependent variable and condition (i.e. task vs rest) and block order (i.e. 1,2,3) as independent factors evidenced, as expected, a significant effect of condition ( $F(1, 114) = 84.986$ ,  $p < 0.001$ ), a tendency for an order effect ( $F(2, 114) = 2.832$ ,  $p = 0.063$ ) but no significant interaction. Post hoc comparisons showed that ISCR was higher during the task compared to the rest ( $t(118) = 9.219$ ,  $p < 0.001$ ). Specifically, ISCR during all the task blocks (i.e. 1, 2, 3) was significantly higher compared to ISCR in all the rest blocks (all  $p < 0.001$ ). ISCR was not different among the task nor the rest blocks. The correlation was not significant between ISCR and accuracy scores for the Addition task. However, we can observe a negative tendency between ISCR and accuracy in the 1st task block ( $r = -0.40$ ,  $p = 0.081$ ), and this relationship seems to be inverted in the case of an abrupt decrement in accuracy in the 3rd task block (Fig. 5A). We observed a negative correlation between ISCR and accuracy scores along all the tasks when including each block score independently ( $r = -0.28$ ,  $p = 0.031$ ), as accuracy decreased along the task while ISCR increased with the increment on the difficulty level.

**Eye Tracking:** For the Stroop task (Fig. 2B), the one-way ANOVA including PD as dependent variable evi-

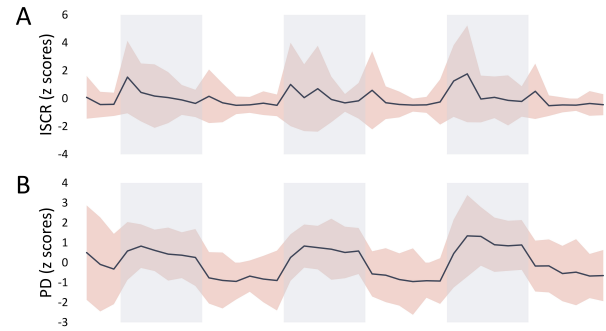


Figure 4: Addition task. (A) Skin conductance responses (ISCR) and (B) Pupil's diameter (PD). Lines and coloured areas are the same as for Fig. 2. Total duration: 6 min plus 30 sec baseline.

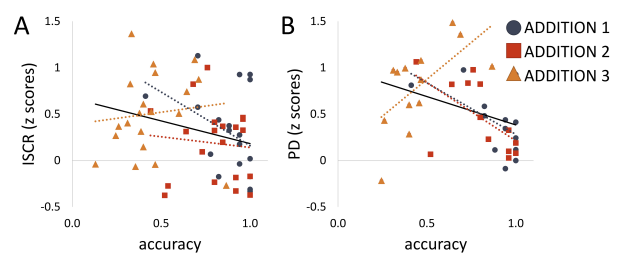


Figure 5: Addition task. Pearson's correlations between accuracy scores and (A) Skin conductance responses (ISCR), and (B) pupil diameter (PD). Dotted lines: linear correlations for each block. Black line: correlation over all blocks.

denced a significant difference between conditions (i.e. baseline, task, rest) ( $F(2, 88) = 20.407$ ,  $p < 0.001$ ). Post hoc comparisons showed that PD during the rest was significantly smaller compared to the baseline ( $t(50) = -4.734$ ,  $p < 0.001$ ) and the task ( $t(76) = -5.687$ ,  $p < 0.001$ ). The two-way ANOVA, including condition (i.e. task vs rest) and order (i.e. 1,2,3) as independent factors, evidenced a significant main effect of both condition ( $F(1, 72) = 79.651$ ,  $p < 0.001$ ) and order ( $F(2, 72) = 16.636$ ,  $p < 0.001$ ), but no significant interaction ( $F(2, 72) = 2.396$ ,  $p = 0.098$ ). Post hoc comparisons showed that PD was bigger during the task ( $t(76) = 8.925$ ,  $p < 0.001$ ) and that PD progressively decreased along the blocks, as PD was significantly different across all three blocks (all  $p < 0.05$ ). Specifically, post hoc comparisons showed that PD in Stroop 1st block was significantly different from PD in all rest blocks (all  $p < 0.001$ ) and the 3rd task block ( $t(24) = 5.556$ ,  $p < 0.001$ ). PD in Stroop 2nd block were also different compared to all the rest blocks (all  $p < 0.01$ ). Finally, PD in the 3rd task block was different from the rest 3rd block ( $t(24) = 3.893$ ,  $p = 0.003$ ), but not from rest blocks 1 and 2. When testing the correlation between accuracy and PD, no significant correlation was found. However, the correlation between PD and accuracy scores along all the tasks, including each block scores independently, was significant ( $r = -0.46$ ,  $p = 0.004$ ), as PD decreased while accuracy increased along the task (Fig. 3B).

For the Addition task (Fig. 4B), the one-way ANOVA evi-

identified a significant difference in PD between conditions (i.e. baseline, task, rest) ( $F(2, 88) = 67.208, p < 0.001$ ). Post hoc comparisons showed that PD during the addition task was significantly bigger compared to the baseline ( $t(50) = 3.711, p = 0.001$ ) and rest ( $t(76) = 11.586, p < 0.001$ ) and that PD during the rest was smaller also compared to the baseline ( $t(50) = -4.482, p < 0.001$ ). The two-way ANOVA, including PD as dependent variable and condition (i.e. task vs rest) and order (i.e. 1,2,3) as independent factors, evidenced a significant main effect of both condition ( $F(1, 72) = 237.530, p < 0.001$ ) and order ( $F(2, 72) = 7.008, p = 0.002$ ), but no significant interaction. Post hoc comparisons showed that PD was bigger during the task ( $t(76) = 15.412, p < 0.001$ ) compared to the rest and that PD increased along the task duration, as PD in the 3rd blocks were bigger compared to the 1st and 2nd blocks (both  $p < 0.01$ ). Specifically, PD did not significantly change among the addition task blocks, nor the rest blocks, and PD in all the task blocks was significantly different from all the rest blocks. Thus, in the Addition task, we did not observe a decrement in PD along the tasks. When comparing PD and accuracy for the Addition task (Fig. 5B), a negative correlation was observed in the 2nd task block ( $r = -0.59, p = 0.033$ ) and, although not significant, the same tendency was observed for the 1st task block ( $r = -0.54, p = 0.055$ ). An opposite correlation was observed for PD and accuracy in the 3rd block ( $r = 0.63, p = 0.022$ ). Averaged accuracy did not correlate with averaged PD, however, the PD delta positively correlated with accuracy in blocks 2 ( $r = 0.85, p < 0.001$ ) and 3 ( $r = 0.65, p = 0.017$ ), and with total accuracy ( $r = 0.82, p < 0.001$ ). When bigger the increment in PD is along the task, higher the performance is. There is a negative correlation between PD and accuracy scores along all the tasks when including each block score independently ( $r = -0.37, p = 0.022$ ), as accuracy decreased along the task while PD increased.

**Neurofeedback:** During the calibration, the NF score was significantly different between C3 and C4 ( $t(17) = -2.189, p = 0.021$ ). When comparing calibration and post, although not significant, there was an increment in the NF score only for C3 ( $t(17) = -1.580, p = 0.066$ ) (Fig. 6A). The one-way ANOVA conducted along the 8 runs, was not significant for NF scores in C3 nor C4, nor for the absolute difference between them (Fig. 6B). When comparing band power (8-30 Hz) between rest and task along all the NF training, the difference was significant for both C3 ( $t(19) = 4.183, p = 0.016$ ) and C4 ( $t(19) = 4.3703, p = 0.010$ ), but also for CP1 ( $t(19) = 4.2515, p = 0.013$ ) and CP2 ( $t(19) = 4.0952, p = 0.019$ ) (Fig. 6C). In the ERS/ERD conducted for training runs (i.e. 30 sec rest, 30 sec task), we observed 8-30 Hz ERD starting at the second 30, corresponding to the beginning of the task (Fig. 6D).

The two-way ANOVA including condition (rest vs task) and order (i.e. from 1 to 8) as independent factors and ISCR as dependent variable evidenced only a significant effect of order ( $F(7, 240) = 23.098, p < 0.001$ ), and a

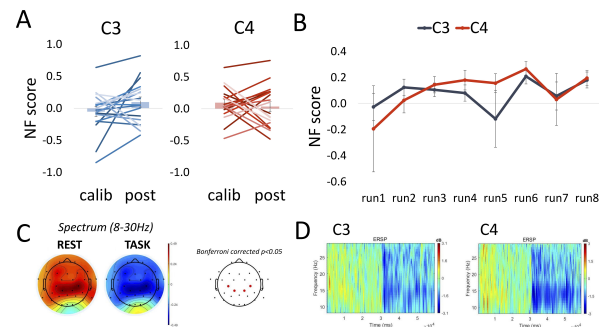


Figure 6: Neurofeedback. (A) NF scores calibration vs post blocks for C3 (blue) and C4 (red). One line per participant. (B) mean and SE for NF scores in the 8 task blocks. (C) Topoplot of task vs rest, and channels with a significant difference (in red) after Bonferroni correction. (D) ERS/ERD in 8-30Hz frequency band along NF runs (30 sec rest then 30 sec task).

significant interaction ( $F(7, 240) = 3.283, p = 0.002$ ). Post hoc comparisons confirmed that ISCR for the 1st and 2nd runs was significantly higher compared to all the other runs (all  $p < 0.01$ ). The same analysis with PD as a dependent variable evidenced a significant effect of condition ( $F(1, 192) = 18.263, p < 0.001$ ) and a tendency (not significant) for an order effect ( $F(7, 192) = 1.500, p = 0.169$ ). Specifically, PD in the rest conditions was significantly smaller compared to the task ( $t(206) = -4.160, p < 0.001$ ). Although not significant, PD in runs 1 and 2 tended to be bigger compared to run 7.

## DISCUSSION

As expected we observed differences in both ISCR and PD between the task and the rest conditions for both the Stroop and the Addition tasks. However, we observed a different trend for both physiological features (i.e. ISCR and PD) in the two different tasks. While in the Stroop task we observed a decrement in both physiological signals along the task, on the contrary, we didn't observe the same trend in the Addition task, in which the difficulty level was increasing along blocks. Thus, in the Stroop task performance increased along the task blocks while physiological signals decreased, probably linked to an expertise and/or habituation effect. In the Addition task, both ISCR and PD remained stable along the task, while the relationship between the physiological features and performance changed based on the task difficulty level. In the case of manipulation of the difficulty level, participants who showed less decrement in physiological features were able to keep a higher performance. For NF, we observed a decrement in physiological signals during the session. This may be an indicator of habituation linked to expertise or, on contrary, linked to fatigue or even to a lack of cognitive effort allocated on the task. In a future study, we aim to develop a method that progressively adapts the NF target based on individual performance and physiological features monitoring. In NF protocols, with

the repetition of sessions, participants may lose engagement on the task and this may be linked to: (1) the task has become too easy, or (2) they give up due to the difficulty in achieving the goal. For instance, if NF scores are high and physiological features are dropping during the session, we may increase the target difficulty level to make the session more challenging. However, if NF scores are high, but also physiological features are high, the difficulty level might be challenging enough. If NF scores are low and physiological signals drop, the target might be set easier to re-engage the participant's interest. Finally, we would like to briefly comment on the difference in the NF score for calibration between c3 and c4. We understand that, as all participants were right-handed, the motor imagination of a movement with the left hand might be less automatic compared to the right. Interestingly, we observed differences only for the calibration, during the NF runs, the difference tended to decrease. This study aimed to present preliminary results on the use of skin conductance and eye-tracking features to monitor task engagement during NF sessions. These first results encourage to individually adapt NF targets to keep participants engaged in the task. In future analyses, we will consider also other physiological features that might be of interest to track cognitive load and attention, such as blinks, saccades and fixations, as well as changes in the tonic component of skin conductance.

## ACKNOWLEDGEMENTS

This work is funded by the Inria Exploratory Action EyeSkin-NF.

## REFERENCES

- [1] Sitaram R *et al.* Closed-loop brain training: The science of neurofeedback. *Nature Reviews Neuroscience*. 2017;18.
- [2] Lorette C, Ziane C, Ben Hamed S. Neurofeedback for cognitive enhancement and intervention and brain plasticity. *Revue Neurologique*. 2021;177.
- [3] Stoeckel L *et al.* Optimizing real time fMRI neurofeedback for therapeutic discovery and development. *NeuroImage: Clinical*. 2014;5.
- [4] Marzbani H, Marateb HR, Mansourian M. Neurofeedback: A Comprehensive Review on System Design, Methodology and Clinical Applications. *Basic and Clinical Neuroscience*. 2016;7.
- [5] Haugg A *et al.* Predictors of real-time fMRI neurofeedback performance and improvement – A machine learning mega-analysis. *NeuroImage*. 2021;237.
- [6] Hammer EM *et al.* Psychological predictors of SMR-BCI performance. *Biological Psychology*. 2012;89.
- [7] Kadosh KC, Staunton G. A systematic review of the psychological factors that influence neurofeedback learning outcomes. *NeuroImage*. 2019;185.
- [8] Sorger B, Scharnowski F, Linden DE, Hampson M, Young KD. Control freaks: Towards optimal selection of control conditions for fMRI neurofeedback studies. *NeuroImage*. 2019;186.
- [9] Alkoby O, Abu-Rmileh A, Shriki O, Todder D. Can We Predict Who Will Respond to Neurofeedback? A Review of the Inefficacy Problem and Existing Predictors for Successful EEG Neurofeedback Learning. *Neuroscience*. 2018;378.
- [10] Le Franc S *et al.* Toward an Adapted Neurofeedback for Post-stroke Motor Rehabilitation: State of the Art and Perspectives. *Frontiers in Human Neuroscience*. 2022;16.
- [11] Unsworth N, Robison MK. Tracking arousal state and mind wandering with pupillometry. *Cognitive, Affective, & Behavioral Neuroscience*. 2018;18.
- [12] Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*. 2005;28.
- [13] Skaramagkas V *et al.* Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE reviews in biomedical engineering*. 2021;PP.
- [14] Critchley HD. Review: Electrodermal Responses: What Happens in the Brain. *The Neuroscientist*. 2002;8.
- [15] Piquado T, Isaacowitz D, Wingfield A. Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*. 2010;47.
- [16] Brink RL van den, Murphy PR, Nieuwenhuis S. Pupil Diameter Tracks Lapses of Attention. *PloS One*. 2016;11.
- [17] Hopstaken JF, Linden D van der, Bakker AB, Kompier MAJ, Leung YK. Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology: Human Perception and Performance*. 2016;42.
- [18] Lane RD, Chua PML, Dolan RJ. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*. 1999;37.
- [19] Sato D *et al.* Water immersion decreases sympathetic skin response during color-word Stroop test. *PLOS ONE*. 2017;12.
- [20] Fos LA, Greve KW, South MB, Mathias C, Benfield H. Paced Visual Serial Addition Test: An alternative measure of information processing speed. *Applied Neuropsychology*. 2000;7.
- [21] Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*. 2010;190.
- [22] Renard Y *et al.* OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain-Computer Interfaces in Real and Virtual Environments. *Presence: Teleoperators and Virtual Environments*. 2010;19.
- [23] Lioi G *et al.* Simultaneous EEG-fMRI during a neurofeedback task, a brain imaging dataset for multimodal data integration. *Scientific Data*. 2020;7.
- [24] Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*. 2019;198.

# RELIABILITY OF INDIVIDUAL TASK-RELATED FRONTAL-MIDLINE-THETA FREQUENCY FOR NEUROFEEDBACK TRAINING – EXPLORATORY INVESTIGATIONS

M. Pfeiffer<sup>1</sup>, E. Masson<sup>1</sup>, A. Kübler<sup>1</sup>, J. Rodrigues<sup>2</sup>

<sup>1</sup> Institute of Psychology, Department of Psychology I, University Würzburg, Germany

<sup>2</sup> Institute of Psychology, Department of Psychology V, University Würzburg, Germany

E-mail: maria.pfeiffer@uni-wuerzburg.de

## ABSTRACT:

Neurofeedback (NF) is a technique where participants receive real-time feedback about their brain activity to learn how to modulate it. As a non-invasive neuromodulation tool, it proves useful in both research and clinical practice. However, approximately one third of users do not respond effectively to NF, prompting efforts to improve responder rates. A promising approach involves individualizing feedback by focusing on a narrow feedback band that encompasses only the individual's peak frequency (IPF), as opposed to a fixed broadband. In some frontal-midline-theta (FMT) - NF paradigms, the IPF is determined during a single calibration session and applied over several days. In a pilot study involving five participants undergoing seven sessions of FMT-NF, we calibrated the IPF using a virtual TMaze task and conducted two follow-up sessions. Our exploratory analysis across three task sessions failed to detect a stable IPF. This, as well as the scarce literature on FMT peak frequency stability, casts first doubts on the efficacy of this calibration technique.

## INTRODUCTION

Neurofeedback (NF) is a promising technique in which individuals receive real-time feedback of their brain activity, empowering them to consciously regulate it [1]. This approach holds significant potential both in research settings and clinical applications as a non-invasive method of neuromodulation [2], [3]. However, despite its potential benefits, NF's effectiveness remains variable, with approximately one third of users not achieving tangible results [4], [5]. In response, ongoing efforts are focused on enhancing responder rates [5]. One strategy to optimize NF outcomes involves individualizing the target frequency bands to each user. In traditional electroencephalographic (EEG)-NF, electric, oscillatory brain activity was usually extracted in relatively broad, fixed frequency bands [6]. More recent approaches try to increase the signal-to-noise ratio by narrowing the target frequency band. This is done by choosing an individual peak frequency (IPF) – the frequency of the band with the most measurable activity at scalp level and providing a narrow target

frequency window around this peak. Hence, unrelated frequency responses in broad windows (i.e. noise) can be avoided.

The idea of the general IPF is rooted in the individual alpha frequency (IAF, 8-12Hz), which was shown to be a trait and hence stable over time [7]. Furthermore, it is easily detectable, as humans tend to show a peak in the alpha range (8-12Hz) of their power spectrum, when closing their eyes, being inattentive or in resting-state, with topographies depending on the respective inactivity [8].

Other frequency responses, such as theta (4-8Hz), do not show as easily detectable peaks during resting-state measurements, but during the performance of specific tasks, yet also with specific topographical distributions. For example, the task-related theta, linked to cognitive control and conflict [9], [10], [11] is localized at frontal midline electrodes (Fz, FCz), hence also named frontal-midline theta (FMT) [12], [13] or midfrontal theta (MFT) [14], [15].

Concerning peak frequencies, increases in task-related theta, when exerting cognitive control, remain in a narrower band than the entire theta band [8] supporting the idea of IPF-training if the latter is targeted. To find the respective individual theta frequency (ITF), several definition and quantification approaches exist. One classical approach bases itself on the IAF for the calibration [8], [16], resulting in an equal stability. More recent neurofeedback studies aiming at FMT modulation calibrated directly on task-related theta peaks [17], [18].

Calibrating task-related theta can be quite laborious, and often, the same calibration is applied across multiple sessions. This approach would be justifiable if individual theta frequency (ITF) measured with the task-related theta peak quantification was akin to individual alpha frequency (IAF) in terms of trait-like stability. However, to date, no studies have specifically investigated ITF stability for this new type of definition and quantification.

During a pilot study with a particular focus on FMT inhibition, involving five participants and seven NF sessions each, our design further employed a virtual TMaze task to calibrate the IPF in the initial session and

during two follow-up sessions.

In an exploratory analysis of the task sessions, we were unable to detect a stable theta peak. This unexpected finding raises critical questions regarding the reliability and efficacy of the calibration technique employed for IPF-based FMT neurofeedback.

In this paper, we present the findings of our exploratory analysis, shedding light on the difficulties of individualized NF calibration methods, particularly concerning FMT modulation.

## MATERIALS AND METHODS

**Ethical statement:** The study was carried out in accordance with the recommendations of “Ethical guidelines, The Association of German Professional Psychologists” (“Berufsethische Richtlinien, Berufsverband Deutscher Psychologinnen und Psychologen”) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki before they participated in the experiment. The protocol was approved by the local ethics committee of the department of psychology of the Julius-Maximilians-University of Würzburg (GZEK 2023-45, Ethikkommission des Institutes für Psychologie der Humanwissenschaftlichen Fakultät der Julius-Maximilians-Universität Würzburg).

**Participants:** Five participants (3 female, age:  $M = 24.4$  years,  $SD = 1.5$ ) were recruited through advertisements in an experiment online portal of the University of Würzburg. Participants were given course credits or a monetary compensation of 12,50€. All participants were at least 18 years old, righthanders, non-color blind and without a history of a psychiatric disorder. They took part in nine experimental sessions within three weeks: An initial calibration session (virtual TMaze) was followed by seven neurofeedback sessions. The virtual TMaze was recorded again directly after the last neurofeedback in session eight, as well as one week later in session nine.

**Virtual TMaze:** The virtual TMaze used in this study is an adaptation of the original design of [19], [20], [21] into a more recent games engine, the Unreal Engine 4. Participants interacted with the virtual environment using a gamepad, navigating through a TMaze in a first-person view (see Fig. 1).

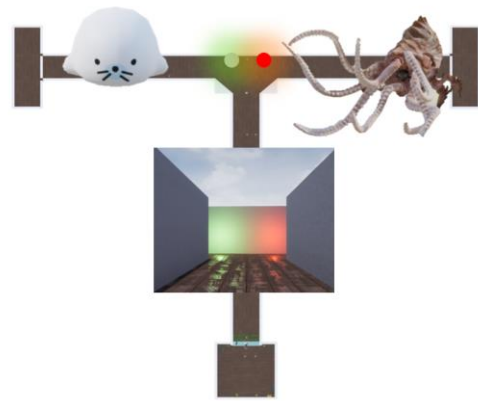
By virtually moving in the TMaze it was possible to encounter two entities: The participant could be caught by a scary kraken, leading to credit loss and an aversive sound being played, or the participant could catch a cute seal, leading to credit gain and a harmonic sound.

Each 18 second trial started with the participant positioned in a passage, facing the T-arms of the maze.

The study incorporated four distinct trial types:

**Avoidance Trials ( $n=20$ ):** A red light indicated the presence of the kraken in one of the arms. Avoidance was possible by retreating behind the starting passage into a safety-zone instead of entering the T-arms.

**Approach Trials ( $n=20$ ):** A green light indicated the



*Figure 1: Overview of an Approach-Avoidance Conflict trial of the virtual TMaze task, displaying the first-person view of the participant, the entities he may encounter and the setup of the TMaze, with the safety zone behind the participant.*

presence of the seal in one of the arms.

**Approach Avoidance Conflict Trials ( $n=30$ ):** A green and a red light indicated the presence of both entities.

**Ambiguous Events ( $n=30$ ):** A yellow light indicated the presence of an unspecified entity in one of the arms. Unbeknown to the participants, the probability of encountering one or the other entity was at 50 percent.

**Neurofeedback:** Given that the neurofeedback does not constitute the central focus of this paper, neurofeedback results are not discussed in this paper. Each session consisted of six blocks of five one-minute trials, leading to a total of 30 minutes of neurofeedback, aiming at FMT-inhibition. A three-minute resting-state was recorded before and after each session. Participants received real time feedback of their FMT activity at Fz, at the assumed ITF ( $\pm 1$ Hz). The ITF was calibrated on the event-related theta of conflict and ambiguous trials of the virtual TMaze task of the first session.

**EEG:** For the recording of the EEG ActiCap electrodes and BrainAmp EEG amplifiers (Brain Products GmbH, Gilching, Germany) were utilized. During session one, eight and nine (all three sessions in which the TMaze was recorded) 62 scalp-electrodes were placed according to the 5-10-system. Two electrodes (O1/O2) were sacrificed to be used as electrooculogram (HEOG/VEOG, right eye). The reference electrode was placed at FCz. All data was recorded using LabStreamingLayer (LSL-Connector, LabRecorder) at a sampling rate of 250Hz.

**Preprocessing:** All EEG data was processed in MATLAB using the code of and following the EPOS-Pipeline [22]. First a notch filter for line noise and resonance frequencies (50,100Hz) was applied. Next, bad channels were detected and interpolated. Detection was based on a statistical threshold of  $z > 3.29$  [23] for joint probability, kurtosis and the power spectrum. Then the data was re-referenced to common average, restoring the previously used reference electrode FCz. Epochs were cut from -1 to 5 seconds after cue (appearance of lights in the TMaze) onset. A 1Hz highpass-filter was applied before an independent-component-analysis (ICA) was performed. The components were used to select bad segments based on the same statistical criteria as before. A second ICA was

computed, this time artifactual components were automatically detected using MARA, ADJUST and SASICA. After removal of the selected components the data was finally re-referenced to current source density. To analyze time frequency responses the data was cut into shorter epochs from -1 to 2 seconds. Data was baseline corrected, using the one second before stimulus onset, and subsequently decomposed using Morlet wavelets.

*Calculation of the ITF:* As we discovered the instability of the ITF with the pipeline, we used for calibration during the study, we considered alternative ways of peak detection to calibrate the neurofeedback system. To investigate these alternative ways, we employed a “mini”-multiverse analysis, comparing combinations of multiple reasonable decisions along the pipeline.

The theta-peak was searched in a time window from 250-450ms after stimulus (light cue) onset. We varied the pipeline at five steps, with two to four alternative decisions per step, resulting in a total of 96 analyses.

- 1) Unit of time-frequency response
  - 1A: Power
  - 1B: decibel (dB) - to account for 1/f dynamics of the power spectrum.
  - 1C: dB change to baseline - to account for differences in baseline activity.
- 2) Spacing of frequency bins
  - 2A Linear – to have equally spaced bins.
  - 2B Logarithmic – to account for 1/f dynamics of the power spectrum.
- 3) Search Time (250-450ms after cue)
  - 3A: Peak Window (50ms window) – the peak is detected on a singular timepoint. Additional 25ms of data before and after this peak are included in the analysis.
  - 3B: Center of Gravity (50ms window) – to detect peaks lower in amplitude but extended in duration we utilized the average of a moving window of 50ms with 10ms steps.
- 4) Search Band (4-8Hz)
  - 4A: Broadband– detecting the peak (timewise) in the broadband, and afterwards extracting activity of each sub-band of the peak. This approach tackles an overall theta peak.
  - 4B: Sub-bands – detecting the peak (timewise) for each sub-band and correspondingly extracting the activity. This allows the investigation of frequency-interferences.
- 5) Peak Detection
  - 5A: Frequency with the highest average activity of only those trials where the specific frequency was the frequency with the most activity.
  - 5B: Frequency with the highest summed activity of only those trials where the specific frequency was the frequency with the most activity.
  - 5C: Frequency with most trials where the specific frequency was the frequency with the most activity.
  - 5D: Frequency with the highest average activity of all trials.

*Peak Timing:* To investigate whether peaks in the different frequency bins may interfere with each other we looked at the distribution of peak timings in the individual sub-bands and compared the standard deviations of peak timings for each individual trial.

*Statistical Analysis:* To assess peak stability, we calculated inter-class-correlations (ICC) across the three sessions for every pipeline. To quantify the variability between the 96 pipelines we calculated the ICC for the pipelines for each session of each participant. Due to the small number of participants (n=5) results from these statistical tests should be taken with caution.

## RESULTS

*ITF Peak Detection:* Due to the immense number of comparisons possible we display only the most important ones. Nonetheless all analyses performed are available on GitHub ([https://github.com/iamrap/FMT\\_Peak](https://github.com/iamrap/FMT_Peak)).

We observed variability in the ITF across participants, sessions as well as calibration pipelines. The employed approaches did not provide a stable peak, except for instances where it is questionable whether the stability was provided by edge artifacts of the frequency band processing [24]. The ICC for the pipelines was especially low (ICC: 0.03, 90%-CI [0.02,0.06]). The evaluation of the peak frequency stability was impossible in 55 pipelines as the strength of the edge artifacts led to zero variability in the detected peak. All three pipelines which would still be rated as fairly reliable (ICC > 0.5) [25], present a strong tendency towards edges of the frequency band. The low ICC of the other 38 pipelines challenges the assumption of a stable ITF across sessions. In the following we will refer to differences in peak detection > 1Hz as “*meaningful differences*”, since they would lead to a different setting in the neurofeedback system.

*1) Unit of time-frequency response:* The blue panels of Fig. 2 display the difference between the different choices for 1 (A-C). It is exemplary for our observation over all the performed analysis, displaying the three issues of the analysis: first, edge artifacts at the lowest frequency for power, second, edge artifacts at the uppermost frequency for the dB transformed data, and third, high fluctuation of the detected peak for dB transformed data in relation to the baseline.

*2) Spacing of frequency bins:* Decision on step 2(A-B) did not lead to such extreme effects, but nonetheless observable and *meaningful* differences, most pronounced in combination with 1C (Fig. 2, blue vs. red panels).

*3) Search Time & 4) Search Band:* Another *meaningful* difference was observed for the decision between center of gravity (3A) and peak window detection (3B), again most visible in combination with 1C. Interestingly the observed *meaningful* difference of search time (3A-B) remained relevant only in combination with 4A (broad-band) (see Fig. 2, green panels). For the sub-band peak search (4B) the

approaches 3A-B remained similar enough to not change the frequency band of the neurofeedback (see Fig., 2 green panels). Choosing the band within which to search for the peak also impacted the detection in a *meaningful* manner itself. While detecting the peaks in the sub-bands lead to a higher likelihood of the peaks being detected at edge frequencies (93% of peaks either  $>7.5\text{Hz}$  or  $<4.5\text{Hz}$ ), detection of the peak in the broad band led to the inclusion of more centered frequencies (76% of peaks either  $>7.5\text{Hz}$  or  $<4.5\text{Hz}$ ).

5) *Peak Detection*: Except for 5C, only minor differences between the different approaches for detecting the peak were notable (see Fig. 2, gold panels). An issue posed by approach 5C was the possibility of several frequencies accumulating the same number of trials, therefore not providing a single peak frequency.

*Peak Timing*: Investigations of the distribution of the peak timing for the sub-bands showed, that peaks tended to occur across the entire time for any frequency, but also a slight variation of timings between them. Nonetheless for no participant any frequency displayed a more specifically time-locked peak than the others. Comparing the divergence of peak-timing for individual trials revealed that the timing is not consistent over frequencies, as indicated by high standard deviations (on average 44ms per trial within a 100ms time window).

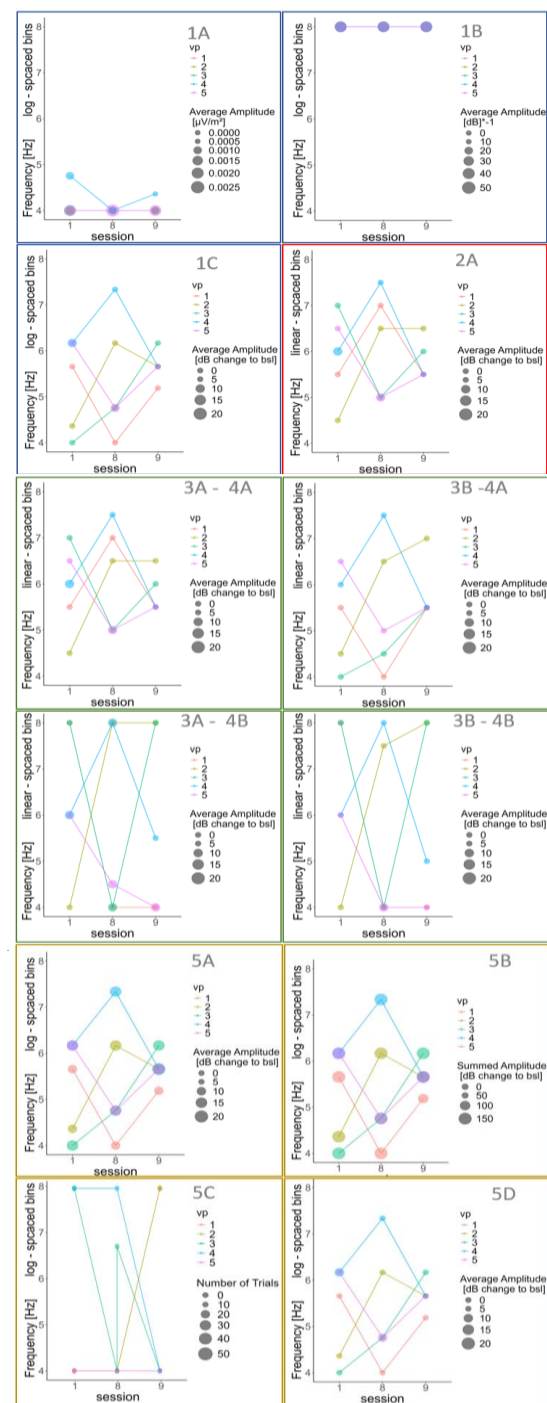
## DISCUSSION

The exploratory findings of our pilot study reveal several critical insights into the calibration of Individual Peak Frequency (IPF) for Frontal-Midline Theta (FMT) Neurofeedback (NF). The inability to establish a stable IPF across sessions raises significant questions about the reliability and effectiveness of current calibration methods, particularly in the context of FMT-NF. This discussion will critically analyze these findings, examining the implications for neurofeedback research and practice, and suggesting potential avenues for future studies.

The core challenge identified in this exploratory analysis is the stability of the ITF. Our results indicate substantial variability in ITF across participants, sessions, and calibration pipelines. This instability could be attributed to several factors:

*Trait vs. State*: Some differences are expected as human brain activity is inherently variable, influenced by factors like cognitive state, attention, and even diurnal rhythms [8]. This variability could lead to fluctuations in theta activity. Nonetheless if an IPF is supposed to be used over several sessions, it needs to be trait- and not state-dependent, hence intraindividual differences should be minimal.

*EEG-Pipelines*: The methods employed for detecting the IPF, such as the time-frequency response units and the peak detection algorithms, showed heterogenous outcomes. This suggests that the choices of methodological approach play a crucial role in the calibration process, which is supported by previous



**Figure 2: Detected frequency peaks for each participant (colors) for the three sessions (1: Initial Session (Day 0); 8: after 7 NF sessions (Day 10-14), 9: 1 week later (Day 17-21)). Each panel displays the results of a different pipeline, highlighting changes induces by design choices.**  
**Blue:** Differences for unit of time-frequency response.  
**Pipeline choice pattern:** 1A-C – 2B – 3A – 4A – 5D.  
**Red:** Linear spaced frequency bins instead of logarithmic bins. Pipeline choice Pattern: 1C – 2A – 3A – 4A – 5D.  
**Green:** Center of gravity vs. Peak Window and Broadband vs. sub-bands. Choice Pattern: 1C – 2A – 3AB – 4AB – 5D.  
**Gold:** Differences choice of peak detection. Choice pattern: 1C – 2B – 3A – 4A – 5A-D.

investigations of the effects of different design choices in EEG analysis pipelines [26]. However *meaningful* peak stability was not achieved with any of the applied approaches.

*Processing Artifacts:* The lower-edge artifacts shown for analysis with 1A (power), are explainable by the 1/f dynamics of EEG-data - lower frequencies displaying higher activity and hence are more prone to be detected as peaks. Opposingly, the decibel transforms (1B) supposedly correcting for these dynamics overcorrects it, leading to the opposite edge-artifact. The third choice 1C (dB change to baseline) introduces other artifacts which may be caused by a variance in baseline activity. The issue of edge artifacts may be tackled by emphasizing the analysis on the center of the signal by using padding or a specific window function such as Hanning windows.

Diverse padding methods could be tested in further pipelines, such as zero-padding, mirror-padding, or constant-padding, to make the edges of the signal smoother or more consistent with the central parts, whereas the Hanning window for example tapers the signal, giving more weight to the center and less to the edges and corners.

*Task Design:* The variance in baseline activity is likely caused by the task design. While we used a rather complex task to elicit conflict-related theta (the virtual TMaze), previous neurofeedback studies using task-based peak detection relied on more simplistic tasks focusing on the exertion of cognitive control (Stop-Signal-Task, Stroop) or working memory (Delayed-Match-to-Sample, n-back) [17], [18].

Even though the stability of the peak frequency was not explicitly reported in former studies, it is still plausible to assume it could have been stable. Three differences may have influenced detectable peak stability in our design. First, while the conflict-related theta elicited in the TMaze is also part of the cognitive control domain, the complexity of the task may still influence the stability of the underlying processes. It may have led to a less consistent baseline, hence influencing the baseline corrected dB transform: Second, the underlying processes of conflict-related theta and working memory-related theta may differ, which could explain a trait for one but not for the other. Third, the TMaze may also suffer from habituation, e.g. participants developing a strategy to deal with the conflict, which is not possible in the simpler designs aiming at inhibitory control (e.g. Stop-Signal Task). A comparative study, employing different tasks over several sessions would be necessary to shed further light on the suspected issue. The variability in ITF suggests a need for more personalized NF protocols. While individualizing NF based on IPFs in general is a promising approach, our study indicates that a one-time calibration may not be sufficient for FMT neurofeedback. Dependent on the intention of the modulation, e.g. if non-sleep-related processes are targeted, we believe an IPF approach to be more applicable than broadband feedback, but continuous or frequent recalibration might be necessary

to account for the dynamic nature of FMT activity.

*Peak Timing:* The investigation of peak times for the different sub-bands displayed inconsistencies across them. This may indicate a band interference, where one sub-band may cancel out the other when averaged, hence we would assume an individual investigation (4B) of each sub-band may lead to more accurate results than the peak detection on the broadband (4A).

*Outlook:* A recent meta-analysis showed, FMT-NF based in on IPFs did not outmatch broadband feedback. [27]. Chances are, the IPF did not outmatch the broadband feedback because both approaches are equally well, but due to the not yet established pipeline for ITF detection. To investigate this proposed issue of ITF peak instability and possibly for providing a stable pipeline, a larger multiverse analysis is planned, including different steps of design choices on an existing dataset of TMaze data, as well as newly recorded data from several tasks eliciting FMT.

## CONCLUSION

In conclusion, while individualizing FMT-NF by focusing on the IPF might be a promising approach, to tailor feedback to NF-users our study underlines the challenges in achieving reliable IPF calibration. Even though with such a small sample these results should be interpreted with caution, the underlying lack of literature concerning ITF stability together with the current observation necessitates a reconsideration of current calibration methods of FMT-NF and highlights the need for more sophisticated approaches. As NF continues to evolve, addressing these challenges will be crucial for maximizing its efficacy and applicability in both research and clinical settings.

## REFERENCES

- [1] R. Sitaram *et al.*, “Closed-loop brain training: the science of neurofeedback,” *Nat. Rev. Neurosci.*, vol. 18, no. 2, pp. 86–100, Feb. 2017, doi: 10.1038/nrn.2016.164.
- [2] T. L. Kvamme, T. Ros, and M. Overgaard, “Can neurofeedback provide evidence of direct brain-behavior causality?,” *NeuroImage*, vol. 258, p. 119400, Sep. 2022, doi: 10.1016/j.neuroimage.2022.119400.
- [3] J. Van Doren, M. Arns, H. Heinrich, M. A. Vollebregt, U. Strehl, and S. K. Loo, “Sustained effects of neurofeedback in ADHD: a systematic review and meta-analysis,” *Eur. Child Adolesc. Psychiatry*, vol. 28, no. 3, pp. 293–305, Mar. 2019, doi: 10.1007/s00787-018-1121-4.
- [4] B. Z. Allison and C. Neuper, “Could Anyone Use a BCI?,” in *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*, D. S. Tan and A. Nijholt, Eds., in Human-Computer Interaction Series. , London: Springer, 2010, pp. 35–54. doi: 10.1007/978-1-84996-272-8\_3.
- [5] J. H. Gruzelier, “EEG-neurofeedback for optimising performance. III: A review of

- methodological and theoretical considerations,” *Neurosci. Biobehav. Rev.*, vol. 44, pp. 159–182, Jul. 2014, doi: 10.1016/j.neubiorev.2014.03.015.
- [6] O. Alkoby, A. Abu-Rmileh, O. Shriki, and D. Todder, “Can We Predict Who Will Respond to Neurofeedback? A Review of the Inefficacy Problem and Existing Predictors for Successful EEG Neurofeedback Learning,” *Neuroscience*, vol. 378, no. 2017, pp. 155–164, 2018, doi: 10.1016/j.neuroscience.2016.12.050.
- [7] T. H. Grandy, M. Werkle-Bergner, C. Chicherio, F. Schmiedek, M. Lövdén, and U. Lindenberger, “Peak individual alpha frequency qualifies as a stable neurophysiological trait marker in healthy younger and older adults,” *Psychophysiology*, vol. 50, no. 6, pp. 570–582, Jun. 2013, doi: 10.1111/psyp.12043.
- [8] W. Klimesch, “EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis,” *Brain Res. Rev.*, vol. 29, no. 2–3, pp. 169–195, Apr. 1999, doi: 10.1016/S0165-0173(98)00056-3.
- [9] J. F. Cavanagh and M. J. Frank, “Frontal theta as a mechanism for cognitive control,” *Trends Cogn. Sci.*, vol. 18, no. 8, pp. 414–421, 2014, doi: 10.1016/j.tics.2014.04.012.
- [10] J. F. Cavanagh, L. Zambrano-Vazquez, and J. J. B. Allen, “Theta lingua franca: A common mid-frontal substrate for action monitoring processes,” *Psychophysiology*, vol. 49, no. 2, pp. 220–238, Feb. 2012, doi: 10.1111/j.1469-8986.2011.01293.x.
- [11] M. X. Cohen and T. H. Donner, “Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior,” *J. Neurophysiol.*, vol. 110, no. 12, pp. 2752–2763, 2013, doi: 10.1152/jn.00479.2013.
- [12] S. Snipes, E. Krugliakova, E. Meier, and R. Huber, “The theta paradox: 4-8 Hz EEG oscillations reflect both local sleep and cognitive control,” *Neuroscience*, Apr. 07, 2022, doi: 10.1101/2022.04.04.487061.
- [13] D. J. Mitchell, N. McNaughton, D. Flanagan, and I. J. Kirk, “Frontal-midline theta from the perspective of hippocampal ‘theta,’” *Prog. Neurobiol.*, vol. 86, no. 3, pp. 156–185, 2008, doi: 10.1016/j.pneurobio.2008.09.005.
- [14] S. J. Reznik, R. Nusslock, N. Pornpattananangkul, L. Y. Abramson, J. A. Coan, and E. Harmon-Jones, “Laboratory-induced learned helplessness attenuates approach motivation as indexed by posterior versus frontal theta activity,” *Cogn. Affect. Behav. Neurosci.*, vol. 17, no. 4, pp. 904–916, Aug. 2017, doi: 10.3758/s13415-017-0521-0.
- [15] L.-T. Hsieh and C. Ranganath, “Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval,” *NeuroImage*, vol. 85, pp. 721–729, Jan. 2014, doi: 10.1016/j.neuroimage.2013.08.003.
- [16] J. Reis *et al.*, “An alpha and theta intensive and short neurofeedback protocol for healthy aging working-memory training,” *Front. Aging Neurosci.*, vol. 8, no. JUN, pp. 1–11, 2016, doi: 10.3389/fnagi.2016.00157.
- [17] K. C. J. Eschmann and A. Mecklinger, “Improving cognitive control: Is theta neurofeedback training associated with proactive rather than reactive control enhancement?,” *Psychophysiology*, vol. 59, no. 5, May 2022, doi: 10.1111/psyp.13873.
- [18] S. Enriquez-Geppert, R. J. Huster, R. Scharfenort, Z. N. Mokom, J. Zimmermann, and C. S. Herrmann, “Modulation of frontal-midline theta by neurofeedback,” *Biol. Psychol.*, vol. 95, no. 1, pp. 59–69, 2014, doi: 10.1016/j.biopsycho.2013.02.019.
- [19] P. Ziebell, J. Rodrigues, A. Forster, J. Sanguinetti, J. J. Allen, and J. Hewig, “Inhibition of midfrontal theta with transcranial ultrasound predicts greater approach versus withdrawal behavior in humans,” *PsyArXiv*, Aug. 30, 2022, doi: 10.31234/osf.io/yvmsp.
- [20] J. Rodrigues, M. Müller, A. Mühlberger, and J. Hewig, “Mind the movement: Frontal asymmetry stands for behavioral motivation, bilateral frontal activation for behavior,” *Psychophysiology*, vol. 55, no. 1, pp. 1–19, 2018, doi: 10.1111/psyp.12908.
- [21] J. Rodrigues, P. Ziebell, M. Müller, and J. Hewig, “Standardizing continuous data classifications in a virtual T-maze using two-layer feedforward networks,” *Sci. Rep.*, vol. 12, no. 1, p. 12879, Jul. 2022, doi: 10.1038/s41598-022-17013-5.
- [22] J. Rodrigues, M. Weiß, J. Hewig, and J. J. B. Allen, “EPOS: EEG Processing Open-Source Scripts,” *Front. Neurosci.*, vol. 15, p. 660449, Jun. 2021, doi: 10.3389/fnins.2021.660449.
- [23] M. Miyakoshi, “Makoto’s preprocessing pipeline-SCCN.” 2021.
- [24] M. X. Cohen, *Analyzing neural time series data: theory and practice*. in Issues in clinical and cognitive neuropsychology. Cambridge, Massachusetts: The MIT Press, 2014.
- [25] T. K. Koo and M. Y. Li, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research,” *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016, doi: 10.1016/j.jcm.2016.02.012.
- [26] D. Trübutschek *et al.*, “EEGManyPipelines: A Large-scale, Grassroots Multi-analyst Study of Electroencephalography Analysis Practices in the Wild,” *J. Cogn. Neurosci.*, vol. 36, no. 2, pp. 217–224, Feb. 2024, doi: 10.1162/jocn\_a\_02087.
- [27] M. Pfeiffer, A. Kübler, and K. Hilger, “Modulation of Human Frontal Midline Theta by Neurofeedback: A Systematic Review and Quantitative Meta-Analysis,” *Neuroscience & Biobehavioral Reviews*, in press.

# MOTOR IMAGERY VIVIDNESS AND NATURALISTIC INNER SPEECH HABITS IN IMAGINED SPEECH CLASSIFICATION

M. Hons<sup>1</sup>, S.E. Kober<sup>1,2</sup>, S.C. Wriessnegger<sup>2,3</sup>, G. Wood<sup>1,2</sup>

<sup>1</sup> Department of Psychology, University of Graz, Graz, Austria

<sup>2</sup> BioTechMed-Graz, Graz, Austria

<sup>3</sup> Institute of Neural Engineering, Graz University of Technology, Graz, Austria

E-mail: manuel.hons@uni-graz.at

**ABSTRACT:** Research on BCI-illiteracy in the imagined speech domain has been scarce. In the current study, we therefore investigate the relationships between both motor imagery vividness as well as inner speech habits, and classification accuracy based on the neural activity evoked by speech imagination. For this purpose, we classified electroencephalography-derived brain activity with respect to four imaginatively spoken phonemes: /a/, /i/, /b/ and /k/. We found that individuals who engaged more frequently in dialogic inner speech exhibited significantly higher classification accuracies, while motor imagery vividness showed no effects. Neurophysiological findings indicate that a higher expression of dialogic inner speech is associated with a suppression of redundant or counteractive neural information. These findings extend our understanding of the substrates of classification performance, respectively, BCI-illiteracy in speech imagery-based systems.

## INTRODUCTION

*Imagery vividness and inner speech habits* Brain-computer interfaces (BCIs) rooted in electroencephalography (EEG) attempt to enable individuals with motor impairments to control a certain device by leveraging electrophysiological brain signals. Motor imagery has been used extensively for this purpose [1,2]. The classification of imagined speech emerged more recently and is considered an intuitive means for the development of speech prostheses [3]. Irrespective of whether one imagines movements or speech, some individuals fail to control BCIs via own brain signals. This phenomenon is referred to as BCI illiteracy. An estimated 15-30% of BCI-users are affected by this [4]. Regarding the roots of the illiteracy phenomenon, structural brain heterogeneity, insufficiently discriminative classification algorithms and a lack of neuroimaging diversity have been discussed in the literature [5]. Furthermore, associations between motor imagery vividness and classification performance [6-8] or corticomotor excitability [9] have been reported. Vuckovic [7], for example, showed that kinesthetic motor imagery vividness was highly correlated with the classification accuracies of a kinesthetic motor imagery task. Visual motor imagery vividness also was positively

correlated with the classification outcome, but to a lesser extent. Despite being a motor imagery process, the role of motor imagery vividness has not yet been investigated in speech imagery. Corresponding questionnaires, however, may not accurately capture the full dynamics of the underlying ability to vividly imagine speaking, as they neglect the auditory component thereof entirely. Since no speech imagery vividness questionnaire exists, previous studies [10,11] alternatively used the Varieties of Inner Speech Questionnaire (VISQ) [12] to quantify the degree to which individuals “mentally” talk to themselves on an everyday basis. However, it is important to note that conceptual and procedural discrepancies between naturalistic inner speech and experimentally manipulated imagined speech exist [13]. The former is produced naturally and spontaneously, while the latter is elicited as part of an investigation. The former often serves a function, e.g. to remind oneself of something or to monitor one’s actions, while the latter frequently consists of isolated stimuli, such as phonemes [3] and yes/no contrasts [14]. For this reason, we refer to experimentally elicited imagined speech as imagined speech and to naturally occurring inner speech as inner speech throughout the rest of this paper. Despite the differences, naturalistic inner speech habits may represent a more suitable proxy for speech imagery vividness compared to motor imagery vividness, as they capture the predominant auditory component of imagined speech.

*Neural substrates of imagined/inner speech* Neural correlates of imagined speech have been researched thoroughly over the last decades. Evidence derived from reviews [15,16] suggests the involvement of an extensive perisylvian, respectively, sylvian network including the left pars opercularis, premotor cortex (especially ventral portions), insula, supplementary motor area, inferior parietal gyrus and superior and middle temporal gyri. Similarly, findings of a recent fMRI study [11] indicate the recruitment of left-hemispheric areas, including the inferior frontal gyrus, medial frontal gyrus, insula and nucleus caudatus regarding both dialogic and monologic inner speech. Additionally, the authors found significant correlations between the activation contrast Dialogic>Monologic inner speech and self-reported

dialogic inner speech usage as assessed by the corresponding VISQ subscale in the right medial temporal gyrus as well as the right precuneus. The remaining subscales were not associated with any voxel activity cluster. With respect to identifying the drivers of BCI-illiteracy in imagined speech classification problems, discovering the neural markers of imagined speech discriminability might be of higher relevance than those of imagined speech per se. For this purpose, imagined speech classification studies frequently reported the discriminative power of features. In terms of frequency bands, higher frequency components (beta, gamma) have been shown to hold a larger amount of discriminative information [14,17,18]. Regarding the role of brain areas, respectively, EEG channels, evidence suggests that, similarly to the neural correlates of imagined speech in general, perisylvian areas, consisting of inferior frontal, inferior parietal and superior temporal channels, provide the most discriminative information [14,18]. However, it is difficult to draw general conclusions from the literature, as feature types, neuroimaging methods and classification algorithms vary strongly between studies [17].

In the current study we examine the relationships between motor imagery vividness as well as inner speech habits and classification performance regarding “mentally” spoken phonemes /a/, /i/, /b/ and /k/. In a second step we attempt to establish associations between dialogic inner speech and the discriminability of features, as quantified by mutual information scores, provided that a substantial relationship between the Dialogic inner speech subscale and classification performance emerges in step one. This choice was based on Alderson-Day and colleagues [11], who demonstrated significant associations between the Dialogic subscale and fMRI activity clusters for both dialogic and monologic inner speech. In a third step, we identify features that are significantly associated with classification accuracy. Finally, we enter them into a multiple regression model alongside dialogic inner speech in order to get insights into whether inner speech habits uniquely predict classification performance beyond the influence of neurophysiological features. Through this, we intend to provide preliminary evidence regarding the role of motor imagery vividness and - with a main focus - naturalistic inner speech habits in imagined speech classification paradigms and, consequently, imagined speech based BCI illiteracy.

## MATERIALS AND METHODS

*Participants* Twenty-seven individuals participated in this study. Due to technical errors during measurements and noisy EEG signals, five participants had to be excluded. Thus, the data of 22 individuals between the ages of 19 and 37 ( $M = 25.41$ ,  $SD = 4.02$ ) were subjected to the analyses. 10 of them were male and 12 were female. All were native German speakers, right-handed and had normal or corrected-to-normal vision. Participants reported no psychiatric or neurological

disorders, or medical diseases. Further, no use of medication that could influence the central nervous system was reported. Recruiting was performed via university-wide mailing distribution. Compensation in the form of either 28 Euros (8 Euros per hour) or course credit for psychology students was offered. To participate, individuals had to sign an informed consent document. This study was approved by the ethics committee of the University of Graz, Austria and conforms to the ethical principles of the Declaration of Helsinki.

*Procedure* Participation comprised attendance on two sessions with 1 hour and 45 minutes each (3.5 hours in sum) on two separate days. Participants were seated in a comfortable armchair approximately 100 cm in front of a 24.5-inch computer screen. A COVID-19 questionnaire, the German version of the Vividness of Movement Imagery Questionnaire 2 (VMIQ-2) [19], the Varieties of Inner Speech Questionnaire – Revised (VISQ-R) [10] and sociodemographic questions were answered prior to the beginning of the study paradigm. Instructions and test trials were presented on the computer screen. In this phase, participants were exposed to audio recordings of all four phonemes. These were carried out via near field studio monitors at a constant volume of approximately 75db at the position of the participants. Audio stimuli were self-recorded and digitally manipulated to exhibit a fundamental frequency in the gender-ambiguous range of 140-170 Hz. Two conditions were embedded into the study paradigm: Phoneme imagination and phoneme perception. However, as the phoneme perception data are not relevant for the current study and merely derived as part of a larger investigation, they were not subjected to the analyses. All participants engaged in the mental speaking of four phonemes: /a/, /i/, /b/ and /k/. The reason for this is the articulatory differences, as well as phonetic dissimilarities between these phonemes [20]. Although no specific imagination instructions were provided, the imagination condition was consistently referred to as “imagining speaking something” throughout the entire paradigm. We did not provide any specifics about the imaginative content, as we wanted to ensure that participants apply their most natural form of speech imagination. Trials consisted of 5 seconds of visual phoneme presentation. Phonemes occurred once per second for 0.5 s. Prior to and after trials a fixation cross was displayed for 1 and 2–4 s, respectively. 20 trials of the same phoneme represented one block. Eight phoneme imagination blocks (2 repetitions \* 2 conditions \* 4 phonemes) were presented in each session. An inter-block break of 15 s was incorporated. With this design 80 trials were carried out for each imaginatively spoken phoneme. The paradigm was constructed and presented in PsychoPy2 (version 1.85).

*Questionnaires* The German version of the VMIQ-2 was used in the current study. It measures the vividness of motor imagery and is comprised of three subscales: internal visual (In), external visual (Ex) and kinesthetic

(Ki) imagination. Each of the three factors has high internal consistency ( $\alpha > .7$ ). Test-retest-reliability of .69 is moderate. The VISQ-R assesses the phenomenological varieties of inner speech and consists of five factors: ‘Dialogic’ (D), ‘Evaluative/Critical’ (E), ‘Other People’ (O), ‘Condensed’ (C) and ‘Positive/Regulatory’ (P) inner speech. Dialogic inner speech denotes talking to oneself in a discursive manner, a recurring back and forth. Condensed inner speech, on the other hand, implies a rather short, fragmented inner speech rather than complete dialogue. The “Evaluative/Critical” scale comprises items that measure to which degree one tends to mentally criticize and evaluate oneself, whereas the “Positive/Regulatory” scale represents a self-praising, comforting inner speech variant. The “Other People” Scale assesses to which degree individuals experience the voices of others in their inner speech. Internal consistency is excellent, ranging from .80 to .91 across subscales.

**Data Acquisition** EEG was derived from 45 electrodes distributed across the whole scalp. Electrode positions conformed to the international 10-5 system to facilitate an even whole-head distribution. EEG was recorded by using actiCAP active wet Ag/AgCl electrodes (Brain Products GmbH), a BrainAmp EEG amplifier (Brain Products GmbH) and the accompanying recording software BrainVision Recorder (version 1.21) at a sampling rate of 500 Hz. The average of the left and right mastoid signal was used as a reference. The ground was placed at Fpz. Three ocular signals were derived from 1 cm above the nasion (vertical eye movements) and the lateral canthi (horizontal eye movements). Furthermore, fNIRS signals were concurrently acquired. However, as they are not relevant for the research questions of the current study and merely derived as part of a larger investigation, they were not subjected to the analyses.

**Preprocessing** Bandpass filtering was conducted with a 1 Hz high-pass and a 70 Hz low-pass setting. A notch filter at 50 Hz was also applied. Ocular artifacts were addressed by regressing the EOG signals out of the EEG data [21]. Subsequently, EEG data were visually inspected to mark artifact corrupted trials and channels for removal. Finally, the data were segmented into 5 s epochs (0 to 5 s) with adjacent baselines of 1 s (-1 to 0 s).

**Feature selection and classification** Power spectral densities (PSDs) were calculated for the predefined frequency bands alpha (8-12 Hz), beta (12-30 Hz) and gamma (30-70 Hz). This selection was grounded on the results of Preedapirat and Wongsawat [18] and Sereshkeh and colleagues [14], who identified higher discriminative potentials of the alpha, beta, and gamma frequency range with respect to an imagined speech classification problem. This resulted in a total of 135 features per trial (45 EEG channels \* 3 frequency bands), provided that no channels were excluded. By using 5-fold cross-validation on the training data, the number of used features was optimized. The  $k = 10, 15, 20, 25, 30$

features with the highest mutual information scores were subjected to this. Classification was performed by means of a multilayer perceptron. Individual classification results, however, are not reported; the focus of this study exclusively lies on the relationship between classification performance and motor imagery vividness, respectively, inner speech habits. However, we do want to emphasize that all classification accuracies exceeded chance level by more than 25% and exhibited an average accuracy > 70%.

**Statistical analyses** To quantify the relationships between motor imagery vividness, inner speech habits, feature discriminability and classification accuracies simple bivariate Pearson classification coefficients are reported. These underwent bootstrapping with 20000 iterations to test them for significance. In an integrative data-driven effort multiple linear regression analysis was conducted to identify more robust, unique effects. Predictors for this were selected as follows: The dialogic inner speech subscale was incorporated if it showed a substantial correlation with classification accuracy. Again, this was based on the findings of Alderson-Day and colleagues [11]. Further, features that showed significant correlations with classification accuracy were also entered. Due to large intercorrelations between features, gamma and beta features were averaged, resulting in one consolidated gamma and beta predictor each. Variance inflation factor was below 5 for all predictors.

## RESULTS

**Inner speech habits and motor imagery vividness with respect to classification performance** Correlative analyses indicated a positive relationship between the Dialogic, the Evaluative/Critical as well as the Condensed inner speech subscale and phoneme classification performance of  $r = .32$  (bootstrap 95% CI [-.18, .73]),  $r = .30$  (bootstrap 95% CI [-.13, .62]) and  $r =$

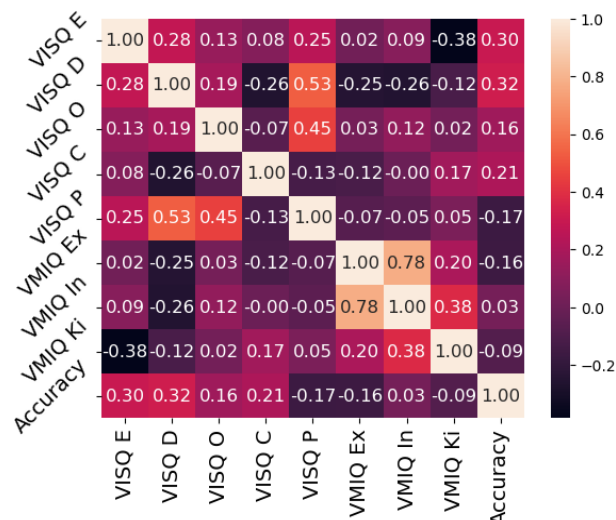


Figure 1: Correlation map containing the variables of interest.

.21 (bootstrap 95% *CI* [-.18, .61]), respectively (Fig. 1). Note that bootstrapping did not attest the correlation coefficients significance. Partial correlation analysis indicated that the correlation between the dialogic inner speech subscale (D) and classification performance (C) persists beyond the influence of Evaluative/Critical inner speech (E) ( $r_{DCE} = .26$ ). All other questionnaire scales showed absolute correlation coefficients  $< .17$ .

**Inner speech and feature discriminability** Negative correlations arose between gamma- as well as beta-based mutual information scores, and the Dialogic inner speech subscale (Fig. 2). A fronto-central beta network consisting of FC3 ( $r = -.54$ , bootstrap 95% *CI* [-.75, -.19]), FC1 ( $r = -.44$ , bootstrap 95% *CI* [-.67, -.03]), FC2 ( $r = -.57$ , bootstrap 95% *CI* [-.81, -.18]), C1 ( $r = -.51$ , bootstrap 95% *CI* [-.77, -.13]) and Cz ( $r = -.52$ , bootstrap 95% *CI* [-.78, -.18]) emerged. Further, in both the gamma and the beta frequency band CPP5h showed a significant correlation ( $r = -.45$ , bootstrap 95% *CI* [-.69, -.13];  $r = -.49$ , bootstrap 95% *CI* [-.73, -.11] for beta and gamma, respectively). Finally, mutual information values of the gamma frequency band at C4 ( $r = -.50$ , bootstrap 95% *CI* [-.74, -.18]) and the beta frequency band at TTP8h ( $r = -.35$ , bootstrap 95% *CI* [-.61, -.07]) exhibited a significant correlation with Dialogic inner speech. This constitutes a small right-hemispheric centro-temporal cluster. Lastly, O2 beta also showed a significant association ( $r = -.44$ , bootstrap 95% *CI* [-.71, -.12]) with the inner speech subscale.

**Feature discriminability and classification accuracy** Several significant correlations regarding mutual information scores of features and classification accuracy were observed (Fig. 3). A predominant left fronto-central gamma network consisting of F7 ( $r = .37$ , bootstrap 95% *CI* [.00, .66]), Fz ( $r = .50$ , bootstrap 95% *CI* [.05, .76]), FFT7h ( $r = .41$ , bootstrap 95% *CI* [.08, .66]), FC1 ( $r = .58$ , bootstrap 95% *CI* [.30, .76]), FCz ( $r = .51$ , bootstrap 95% *CI* [.23, .70]), C3 ( $r = .46$ , bootstrap 95% *CI* [.06, .71]), C1 ( $r = .50$ , bootstrap 95% *CI* [.14, .72]) and Cz ( $r = .49$ , bootstrap 95% *CI* [.12, .72]) emerged. Additionally, left parietal and right frontal contribution was observed at P3 ( $r = .39$ , bootstrap 95% *CI* [.10, .65]) and FFC6h ( $r = .43$ , bootstrap 95% *CI* [.13, .66]), respectively. A more diffuse pattern was found in the beta

band with significant correlations at FFT7h ( $r = .48$ , bootstrap 95% *CI* [.10, .73]), Cz ( $r = -.50$ , bootstrap 95% *CI* [-.76, -.11]) and CP4 ( $r = .44$ , bootstrap 95% *CI* [.10, .77]).

**Integrative model** Multiple linear regression analysis yielded the following results: gamma-based mutual information scores ( $b = .069$ ,  $p = .001$ ) as well as the Dialogic inner speech subscale ( $b = .047$ ,  $p = .009$ ) emerged as significant predictors with respect to classification accuracy (Tab. 1). Beta-based mutual information scores did not prove significant ( $b = .017$ ,  $p = .332$ ). In sum, the model explained approximately 60% of the variance of classification accuracy ( $R^2 = .595$ ,  $p < .001$ ).

Table 1. Multiple linear regression results

	<i>b</i>	<i>SE</i>	<i>t</i>
Gamma	.069**	.017	4.115
Beta	.017	.017	0.996
VISQ D	.047**	.016	2.935
$R^2 = .595$	$R^2 \text{ adj.} = .527$	$F = 8.807$	$p(F) < .001$

*Note.* Gamma: Averaged mutual information values of gamma features. Beta: Averaged mutual information values of beta features. VISQ D: Dialogic inner speech subscale of the VISQ-R

\*\* $p < .01$

## DISCUSSION

**Motor imagery vividness and inner speech habits** The aim of the current study was to investigate the relationships between motor imagery vividness as well as inner speech habits and classification performance on grounds of neural activity evoked by imaginatively spoken phonemes. Literature suggests that visual [6,8] as well as kinesthetic motor imagery vividness [8,22] have an impact on the classification of neural reaction patterns evoked by motor imagery. Dissonantly, by means of correlative methods, we were unable to establish a connection between visual or kinesthetic motor imagery vividness and classification performance in the imagined speech domain. Despite not withstanding bootstrapping-based significance tests, inner speech habits, however, indicated promising effects: dialogic and evaluative inner

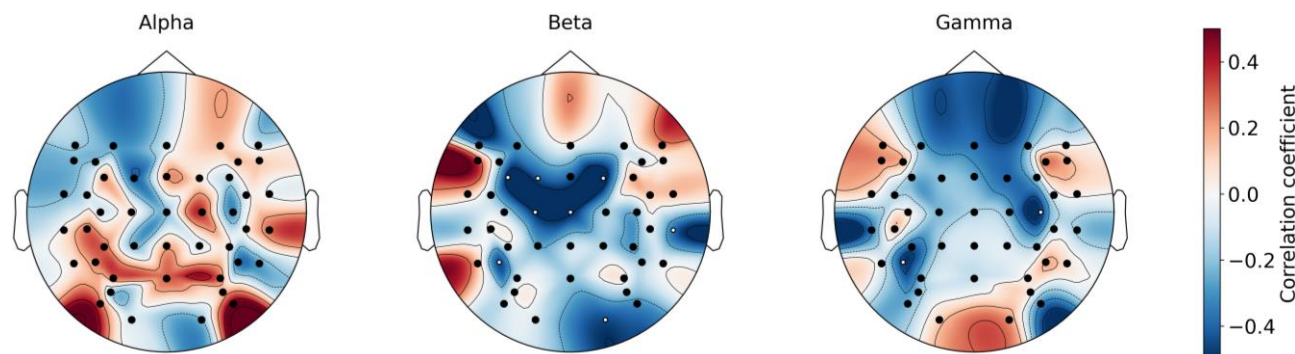


Figure 2. Topo plots of the correlation coefficients between mutual information scores of the frequency bands of interest and the VISQ-R Dialogic subscale. Significant correlation coefficients are marked in white.

speech both showed a moderate, and condensed inner speech exhibited a small to moderate correlation with classification performance. Hence our results suggest that, while individuals who can vividly imagine movements cannot produce more differentiable neural signals on grounds of imaginatively spoken phonemes, individuals engaging frequently in dialogic, evaluative inner speech can. Although, in previous studies, motor imagery vividness was shown to have reliable influence on motor imagery classification, it is not surprising that this mechanism cannot simply be translated to speech imagery. The established motor imagery vividness questionnaires do not assess speech imagery or any other form of imagery with an auditory component. Since the content of imagined speech is considered to be not only of articulatory, but also auditory nature [13], this may explain the larger influence of naturalistic inner speech on classification performance compared to motor imagery vividness.

**Neurophysiological contribution** To obtain a more comprehensive depiction of the role of naturalistic inner speech habits in an imagined speech classification paradigm, we investigated the embedding of the discriminative power of features. Negative correlations between the mutual information values, i.e., the discriminability of features and dialogic inner speech emerged. A prominent fronto-central network of beta channels emerged, along with isolated significant channels in left superior temporal, right occipital and right medial temporal regions regarding beta, and left superior temporal and right inferior central regions regarding gamma. To our knowledge, only one study exists that investigated associations between VISQ scales and neural data [11]. Our results are only partially overlapping with those reported in the referenced study, as the authors consonantly reported a significant association between the Dialogic subscale of the VISQ and an fMRI cluster in the right medial temporal gyrus, but additionally in two clusters of the right precuneus. Note that the authors did not report the direction of the correlation. Since our network is much more diffuse, caution is advised with respect to interpreting these findings. While being negatively correlated with the discriminability of EEG-channels, dialogic inner speech shows a positive relationship with the overall

classification performance. Consequently, individuals who engage more frequently in dialogic inner speech may be more successful in suppressing redundant neural information, which in turn increases accuracy. We found associations between feature discriminability and classification performance that support this notion. An increased expression of dialogic inner speech suppresses features that show a negative or no influence on classification performance. A higher expression of dialogic inner speech might therefore not only act as a noise cancelling mechanism but also directly suppress the recruitment of counteractive features. However, significant positive correlations between feature discriminability and classification performance were also found. In this regard, a predominantly left-hemispheric frontal gamma network emerged, which is in line with previous literature [11,14,18]. Despite existing findings showing more diffuse discriminability maps [14,18], comparisons are limited, as these reflect mere discriminability values and not correlations. Ultimately, we integrated these results in a data-driven approach by building a multiple linear regression model to predict classification accuracy based on dialogic inner speech and the beta as well as gamma features that showed a significant correlation with classification accuracy. In line with previous findings [14,18], gamma indicated a significant influence on classification performance that is independent from all other predictors. Similarly, dialogic inner speech yielded a significant unique effect that goes beyond the neurophysiological predictive power. As indicated in the bivariate correlative analyses, this supports the notion that a high tendency towards everyday dialogic inner speech acts as a noise cancelling mechanism, inhibiting redundant neural contributions rather than increasing feature discriminability. Lastly, beta features did not influence classification performance.

## CONCLUSION

We were able to shed light on the complex action mechanisms surrounding the role of naturalistic inner speech habits and motor imagery vividness in imagined speech classification. In opposition to previous studies, no link between motor imagery vividness and BCI performance was found. Dialogic inner speech, however,

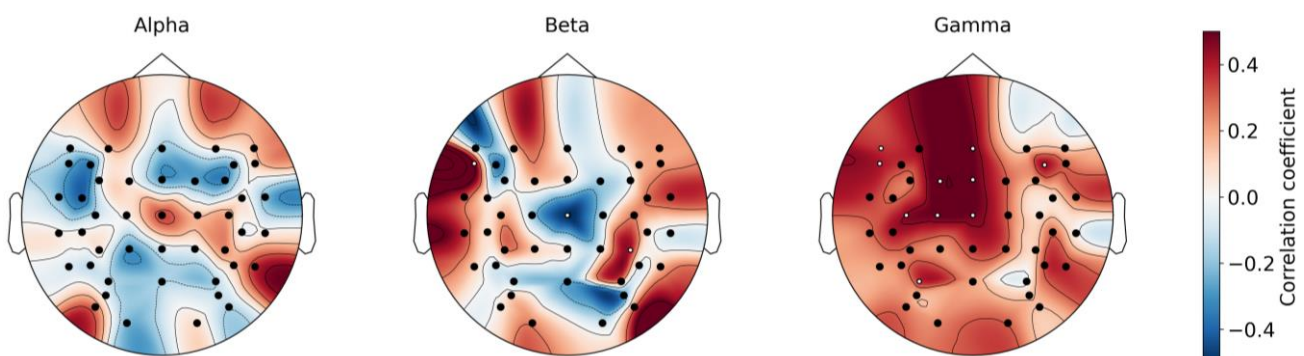


Figure 3. Topo plots of the correlation coefficients between mutual information values of the frequency bands of interest and classification accuracy. Significant correlation coefficients are marked in white.

was found to significantly predict classification accuracy. The auditory imagination component, which might be captured by the inner speech, but not by the motor imagery vividness assessment, may be the reason for this. Although promising, the presented findings are preliminary. Further studies are needed to gain an in-depth understanding about the role of naturalistic inner speech habits in speech imagery based classification performance, respectively, BCI-illiteracy.

## ACKNOWLEDGEMENTS

This study was supported by the Field of Excellence COLIBRI (Complexity of Life in Basic Research and Innovation, University of Graz).

## REFERENCES

- [1] Neuper C, Scherer R, Reiner M, Pfurtscheller G. Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. *Brain research. Cognitive brain research* 2005;25(3):668–77.
- [2] Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. *Proc. IEEE* 2001;89(7):1123–34.
- [3] DaSalla CS, Kambara H, Sato M, Koike Y. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural networks the official journal of the International Neural Network Society* 2009;22(9):1334–9.
- [4] Blankertz B, Sannelli C, Halder S, Hammer EM, Kübler A, Müller K-R et al. Neurophysiological predictor of SMR-based BCI performance. *NeuroImage* 2010;51(4):1303–9.
- [5] Allison B, Neuper C. Could anyone use a BCI? In Tan DS, Nijholt A (Ed.). *Brain-Computer Interfaces*: Springer London. London 2010, pp. 35–54
- [6] Leeuwis N, Paas A, Alimardani M. Vividness of Visual Imagery and Personality Impact Motor-Imagery Brain Computer Interfaces. *Frontiers in human neuroscience* 2021;15:634748.
- [7] Vuckovic A. Motor imagery questionnaire as a method to detect BCI illiteracy: In: 2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010), Roma, Italy, 07.11.2010 - 10.11.2010.
- [8] Vuckovic A, Osuagwu BA. Using a motor imagery questionnaire to estimate the performance of a Brain-Computer Interface based on object oriented motor imagery. *Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology* 2013;124(8):1586–95.
- [9] Lebon F, Byblow WD, Collet C, Guillot A, Stinear CM. The modulation of motor cortex excitability during motor imagery depends on imagery quality. *The European journal of neuroscience* 2012;35(2):323–31.
- [10] Alderson-Day B, Mitrenga K, Wilkinson S, McCarthy-Jones S, Fernyhough C. The varieties of inner speech questionnaire - Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and cognition* 2018;65:48–58.
- [11] Alderson-Day B, Weis S, McCarthy-Jones S, Moseley P, Smailes D, Fernyhough C. The brain's conversation with itself: neural substrates of dialogic inner speech. *Social cognitive and affective neuroscience* 2016;11(1):110–20.
- [12] McCarthy-Jones S, Fernyhough C. The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and cognition* 2011;20(4):1586–93.
- [13] Hurlburt RT, Heavey CL, Kelsey JM. Toward a phenomenology of inner speaking. *Consciousness and cognition* 2013;22(4):1477–94.
- [14] Rezazadeh Sereshkeh A, Yousefi R, Wong AT, Rudzicz F, Chau T. Development of a ternary hybrid fNIRS-EEG brain-computer interface based on imagined speech. *Brain-Computer Interfaces* 2019;6(4):128–40.
- [15] Price CJ. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* 2012;62(2):816–47.
- [16] Alderson-Day B, Fernyhough C. Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. *Psychological bulletin* 2015;141(5):931–65.
- [17] Panachakel JT, Ramakrishnan AG. Decoding Covert Speech From EEG-A Comprehensive Review. *Frontiers in neuroscience* 2021;15:642251.
- [18] Preedapirat S, Wongsawat Y. Significant Neurophysiological Features for fNIRS-EEG Brain-Computer Interfacing of Imagined Speech: In: 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Phuket, Thailand, 05.07.2022 - 08.07.2022.
- [19] Dahm SF, Bart VKE, Pithan JM, Rieger M. Deutsche Übersetzung und Validierung des VMIQ-2 zur Erfassung der Lebhaftigkeit von Handlungsvorstellungen. *Zeitschrift für sportpsychologie* 2019;26(4):151–8.
- [20] Keating PA. Phonetic and Phonological Representation of Stop Consonant Voicing. *Language* 1984;60(2):286–319.
- [21] Gratton G, Coles, M G H, Donchin, E. A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology* 1983;55(4):468–84.
- [22] Marchesotti S, Bassolino M, Serino A, Bleuler H, Blanke O. Quantifying the role of motor imagery in brain-machine interfaces. *Scientific reports* 2016;6:24076.

# REVIRE: A VIRTUAL REALITY PLATFORM FOR BCI-BASED MOTOR REHABILITATION

L. Mihić Zidar<sup>1,2</sup>, P. Raggam<sup>1,3</sup>, F. Mohammadian<sup>1</sup>, A. Barloga<sup>1</sup>, M. Grosse-Wentrup<sup>1,4,5</sup>

<sup>1</sup>Research Group Neuroinformatics, Faculty of Computer Science, University of Vienna, Vienna, Austria

<sup>2</sup>Departement for Cognitive Science, Faculty of Education, University of Ljubljana, Ljubljana, Slovenia

<sup>3</sup>Doctoral School Computer Science, Faculty of Computer Science, University of Vienna, Vienna, Austria

<sup>4</sup>Research Network Data Science, University of Vienna, Vienna, Austria

<sup>5</sup>Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria

E-mail: lucijamihiczidar@gmail.com

**ABSTRACT:** We introduce REVIRE (REhabilitation in Virtual Reality), an immersive virtual reality platform for post-stroke upper limb rehabilitation with integrated EEG recording. REVIRE immerses users in a 3D virtual environment where they can practice motor tasks that reflect everyday activities while providing comprehensive performance data with synchronized hand trajectories and EEG signals. Our proof-of-concept study tested the application on four healthy individuals across multiple training sessions. We observed significant effects of training on performance, evidenced by reduced task completion times. Changes in performance coincided with a decrease in EEG sensorimotor activity, consistent with existing motor learning research. In addition, the low incidence of cybersickness reported by participants indicates a comfortable and user-friendly experience, making our setup suitable for patient use. Our preliminary findings demonstrate the suitability of our virtual reality platform for BCI-based motor rehabilitation for clinical environments and beyond.

## INTRODUCTION

Stroke is a critical global health concern, resulting in significant motor deficits that heavily impact patients' quality of life. Upper limb (UL) impairments, in particular, severely limit patients' independence and ability to perform activities of daily living. Traditional therapy approaches that involve face-to-face, therapist-led physical exercise often fall short of fully addressing patients' needs. Constrained by limited financial and personnel resources, traditional hospital-based rehabilitation struggles to provide sufficiently engaging, intensive, and personalized therapy [1]. In response to these challenges, recent years have witnessed a surge in research on novel technology-based approaches aimed at making rehabilitation more effective, individualized, and accessible [2]. Brain-computer interfaces (BCIs) represent a significant

innovation in this field. By providing direct feedback on brain activity, BCIs encourage patients to actively engage in the self-regulation of their neural states. This approach has been shown to promote neural plasticity and improve functional outcomes [3]. When combined with immersive virtual reality (VR) as a feedback modality, BCI-VR systems can provide ecologically valid environments for task-specific and intensive practice, controlled through the patient's brain activity [4]. Despite the growing demand for rehabilitation services that extend beyond hospital settings, the widespread use of BCIs is currently limited by the cost and complexity of the required hardware. Nonetheless, with recent advancements in VR technology, the application of BCI-VR outside the traditional clinical settings is becoming more feasible.

New generations of VR head-mounted displays (HMDs), such as the Meta Quest 2<sup>1</sup>, have become lighter, more portable, and more affordable. A major advancement has been the implementation of full hand tracking, which monitors the movement of the entire hand, including all fingers, joints, and nuanced gestures. This allows users to interact with the virtual environment more naturally, without the need for controllers. It also provides clinicians with direct access to patient movement trajectories, a feature previously only available with high-end motion capture systems and robotic rehabilitation devices. In addition, newer HMDs have considerably reduced the incidence of cybersickness, the motion sickness-like adverse effects [5], which had been a concern for VR use in patient populations.

Recently, several rehabilitation games have been developed to leverage the advantages of contemporary VR-HMDs for post-stroke UL recovery. Mekbib et al. found that the addition of immersive VR UL training to occupational therapy resulted in significantly greater functional improvements compared to occupational therapy

<sup>1</sup><https://www.meta.com/at/en/quest/products/quest-2/>

alone [6]. Fregna et al. developed an immersive environment for telerehabilitation, incorporating a client app for real-time remote supervision [7]. Current BCI-VR applications predominantly employ the motor imagery paradigm to provide patients with feedback on imagined or intended movement [4]. However, emerging research suggests a potential for expanding BCI-VR paradigms. A study on BCI-VR for gait rehabilitation by Luu et al. observed an increase in cortical involvement during treadmill walking when participants controlled the virtual walking avatar via neurally decoded gait kinematics compared to when the avatar mirrored their actual steps [8]. In this paper, we introduce REVIRE, a novel VR application for post-stroke UL rehabilitation that features an immersive, easily customizable training environment. Our application supports comprehensive patient monitoring through full hand tracking and integrated EEG signal recording. Moreover, we employ commercially available low-cost VR hardware for utilization outside clinical or laboratory environments. We evaluated the feasibility of our setup with a sample of healthy participants. By collecting data across multiple recording sessions, we aimed to verify whether our system could produce viable data for assessing user performance and progress. We also sought to ensure that the setup was comfortable for users and did not induce adverse effects. In the following sections, we describe the game environment and tasks, the data collection methods, and the preliminary analysis results. We demonstrate that our application is well-suited for BCI-VR rehabilitation research, extending its potential use beyond clinical environments.

## MATERIALS AND METHODS

**REVIRE Design:** We developed an immersive VR game using the Unity 3D game engine<sup>2</sup>. The source code is freely available on GitHub<sup>3</sup>. In the 3D environment, we created motor tasks that closely mimic everyday activities. We presented the game via the Meta Quest 2 HMD.

**REVIRE Environment:** The REVIRE environment is set in a home interior. Users are seated behind a table, both in the physical and in the virtual world. Before starting the tasks, they are guided through a calibration procedure that co-locates the physical and the virtual table. This ensures that users feel a physical sensation when touching the virtual table, thereby increasing the sense of immersion and presence. Virtual hands, seen from the first-person perspective, animate the users' hand and finger movements and enable them to interact with virtual objects.

**REVIRE Tasks:** The game consists of three functional motor tasks, shown in Fig. 1. In the *Pouring* task, users pour water from the bottle into the glass, filling it to a specified line. The task is completed when the bottle and the filled glass are returned to their respective starting ar-

eas. In the *Drinking* task, users are presented with a glass filled with water. To drink, they lift the glass towards their mouth and hold it at an angle until it empties. They complete the trial by placing the empty glass in a designated area on the table. In the *Box* task, a podium with three areas at different height levels is presented on the table, with a box placed randomly in one of the areas. Users hold the box with both hands to move it to the target position.

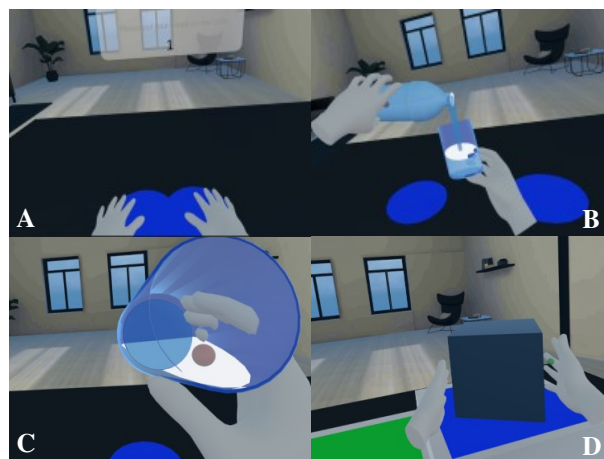


Figure 1: Gameplay: A - resting period, B - Pouring task, C - Drinking task, D - Box task.

Each task trial begins with a five-second rest period, which serves as the baseline for the EEG analysis. Users are asked to place their hands on the table and remain still while a countdown to the start of the task is displayed. Following the rest period, users are presented with the objects required to complete the task. The goal of each task is to accurately perform the intended action while minimizing the completion time. Users can perform the movements at their own pace and explore different movement strategies to complete the task. At any point in the game, they can reset the objects to their starting positions and continue the task without interrupting the trial. At the end of each task, they receive performance feedback through task completion times.

**Participants:** Participants were partially recruited using the Vienna Cognitive Science Hub Study Participant Platform, which uses the hroot software [9]. Four healthy participants, aged  $54.0 \pm 10.4$  years (one female), took part in the study. All had normal or corrected-to-normal vision and little to no previous experience with VR. All participants gave written informed consent and were compensated for their participation. The study protocol was approved by the University of Vienna Ethics Committee.

**Procedure:** Participants attended four recording sessions over  $10.0 \pm 3.6$  days. The total length of a session, including preparation time, was between 2 and 2.5 hours. Participants were seated comfortably behind a table for the duration of the recording. After EEG preparation, we placed the HMD over the EEG cap. The experimental setup is illustrated in Fig. 2. In the first session,

<sup>2</sup><https://unity.com/products/unity-engine>

<sup>3</sup><https://github.com/praggam/REVIRE>

we instructed the participants on how to navigate in the virtual environment, and they completed a practice trial of each task. Sessions consisted of four blocks of the REVIRE game, each comprising 10 trials of every task (*Pouring*, *Drinking*, and *Box*, in this fixed order). EEG was recorded continuously during every block. The average duration of the VR gameplay was  $39.4 \pm 4.8$  minutes, which is below the recommended maximum [10]. At the end of the session, participants completed the Virtual Reality Sickness Questionnaire (VRSQ) [11], administered in English language and pen-and-paper format.

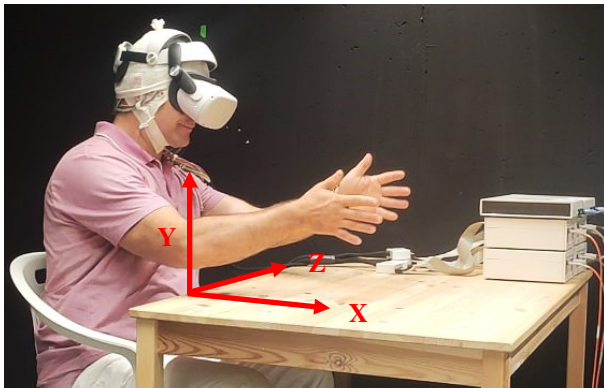


Figure 2: The experimental setup with Meta Quest 2 placed over the EEG cap. The axes represent the coordinate frame of recorded movement trajectories.

**EEG measurement:** We recorded the EEG signals with 64 gel-based passive EEG electrodes following the standard 10-20 system, with the ground and reference electrodes set at AFz and FCz, respectively. We kept the impedances below 10-15 k $\Omega$ . The EEG signals were recorded at 1 kHz using the NeurOne™ Tesla<sup>4</sup> EEG system. The amplifiers were powered by 7V batteries and connected to the main unit by fiber optic cables.

**Data recording:** For the integrated recording of the EEG and REVIRE data streams, we used the Lab Streaming Layer (LSL)<sup>5</sup> and its default recording program Lab Recorder. The REVIRE output included the position coordinates of both hands and a stream of event markers. The synchronized data streams were saved into a single XDF file. A Python<sup>6</sup> script was developed to convert the XDF files into FIF files that are compatible with the MNE-Python<sup>7</sup> library, which we used for EEG signal processing. EEG signals, event markers, and hand position data were combined into a mne.io.Raw object containing 64 EEG channels, one STIM channel with event markers, and six channels for right and left hand positions (XYZ coordinates for each hand).

**Virtual Reality Sickness Questionnaire (VRSQ):** As a measure of cybersickness, we used the VRSQ [11], which comprises nine symptoms rated on a 4-point Likert scale ranging from 0 - *Not at all* to 3 - *Severely*. The items

are divided into *Oculomotor* and *Disorientation* categories. To calculate scores for each category, corresponding items are averaged and scaled to a 100-point maximum, and the average of both categories represents the total score.

**Task performance:** As a measure of performance, we computed trial completion times based on *trial start* and *trial end* event markers. In twelve trials, completion times could not be determined due to missing markers, leaving a total of 1,908 trials ( $636 \pm 2.5$  per task).

**Movement trajectories:** To align and analyze the EEG data based on the movement events, we determined the movement onsets from the hand position data. We focused this preliminary analysis on the reaching movements in *Drinking* and *Pouring* tasks where users began trials by grabbing the bottle and/or glass. The *Box* task was excluded due to variability in the starting position of the box. We identified reach onsets based on peaks in velocity, which we confirmed manually by visually inspecting the movement trajectories. We included 1,023 trials, corresponding to 80% of *Drinking* and *Pouring* trials, where the onset was clearly identifiable and the resting period exceeded 2 seconds.

**EEG preprocessing:** We used the MNE-Python library and its built-in functions to preprocess and analyze the EEG data. We cropped the raw EEG signals to include only the *Drinking* and *Pouring* task trials. We filtered the data with a notch filter at 50 Hz and then bandpass filtered the signals at 1-49 Hz. After visually inspecting the time series of the filtered signals, we removed and interpolated channels with consistent noise ( $1.1 \pm 0.4$  per recording). We then re-referenced the signals to a common average reference. We corrected blink, eye movement, muscle, and electrode noise artifacts through manual inspection of the ICA components. Due to substantial noise contamination from muscle activity and pressure of the HMD on electrodes, we took a selective approach, retaining  $16.6 \pm 3.2$  or 26% of the components per recording. We segmented the cleaned data into epochs extending from -2 to 5 seconds around reach onset events. A total of 1,023 epochs ( $511.5 \pm 0.5$  per task) were included in the analysis.

**Event-related spectral activity:** Our preliminary EEG analysis focused on sensorimotor activity in alpha and beta frequency bands, which have been associated with motor learning, e.g., [12, 13], and are commonly used as features in EEG-BCIs [14]. We used the Morlet wavelet method to obtain time-frequency representations for the alpha (8-12 Hz) and beta (13-30 Hz) bands at a frequency resolution of 1 Hz and with the number of cycles in the wavelet at half the length of the frequency range. We summed the spectral power over the frequencies within each band and computed event-related spectral perturbation (ERSP) values by dividing time-frequency representations with the mean of the baseline period, followed by a logarithmic transformation. Focusing on motor activity, we considered the mean of ten channels at the bilateral primary motor cortex (C1, C3, C5, FC3, CP3, C2, C4,

<sup>4</sup><https://www.bittium.com/medical/bittium-neurone>

<sup>5</sup><https://github.com/scn/labstreaminglayer>

<sup>6</sup><https://www.python.org/>

<sup>7</sup><https://mne.tools/stable/index.html>

C6, FC4, and CP4). We plotted the EEG response against the movement trajectories (see Fig. 5) and estimated 0 to 3 seconds after the movement onset as the relevant time window for the reaching movement.

**Statistical analysis:** To assess the effects of practice and task type on completion times and EEG response, we opted for mixed linear models (MixedLM). The models were formulated with either 'Completion time', 'mu ERSP', or 'beta ERSP' as the dependent variable, 'Session' and 'Task' as fixed effects, and 'Participant' as a random effect. The dependent variables were aggregated across trials within a condition to obtain one value per task for every participant in every session. To evaluate whether REVIRE induced adverse symptoms, we compared the total VRSQ scores against a baseline of 25 (corresponding to the response 1 - *Slightly*, indicating mild symptoms) using a one-tailed t-test. We hypothesized that the scores would fall below this threshold.

## RESULTS

**Task performance:** Over the course of the practice sessions, participants consistently reduced the time it took them to complete the task. This trend is evident both individually and after averaging across participants, as illustrated in Fig. 3, which shows completion times aggregated across tasks. The results of the MixedLM are detailed in Table 1. The analysis revealed a significant decrease in completion times from Session 1 to 4, and a significant difference between all tasks, but no significant interaction between the two factors.

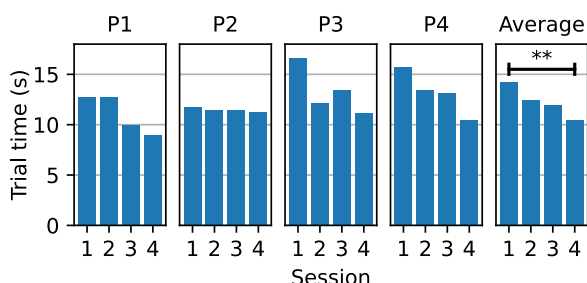


Figure 3: Mean completion times across practice sessions, averaged over tasks. Plotted for all four participants (P1 to P4), and participant average.  $**p < 0.01$

**Event-related spectral activity:** EEG analysis showed a consistent increase in mu and beta ERSP across sessions, suggesting decreased involvement of sensorimotor areas as participants increased their proficiency. Figure 4 illustrates this trend for individual participants and the participant average, with ERSP aggregated across tasks. The results of the MixedLM are detailed in Table 2. The increase in ERSP values for both mu and beta was significant in Session 4 compared to Session 1. No significant effects of task condition or task-by-game interaction were found for either beta or mu ERSP, suggesting the influence of temporal dynamics rather than task-specific factors on EEG modulation. Fig. 5 illustrates the

Table 1: Mixed Linear Model Results for Completion Time

Effect	Estimate	SE	p
Intercept (S1, T1 vs. 0)	18.54	1.19	< .001
S2 vs. S1	-2.01	1.55	.20
S3 vs. S1	-2.07	1.55	.18
S4 vs. S1	-4.30	1.55	.01
T2 vs. T1	-3.12	1.55	.04
T3 vs T1	-9.83	1.55	< .001
S2×T2	0.45	2.19	.84
S3×T2	-0.73	2.19	.74
S4×T2	-0.67	2.19	.76
S2×T3	0.21	2.19	.93
S3×T3	0.18	2.19	.94
S4×T3	2.11	2.19	.34

Note: The participant variance is 0.87. S is session, T1-3 are Pouring, Drinking, and Box tasks. SE is the standard error of the estimate. Statistical significance was set at  $p < .05$ .

EEG response to the bimanual reaching movements in the Pouring task. A characteristic desynchronization of the sensorimotor rhythms was observed at movement onset. The magnitude of the response consistently decreased as the training sessions progressed. Diversification of ERSP levels across sessions was particularly pronounced during the first two seconds of the reaching movement. Alongside changes in motor activity, topographical plots showed a decrease in frontal activity across the sessions, suggesting reduced involvement of attentional and executive processes [13].

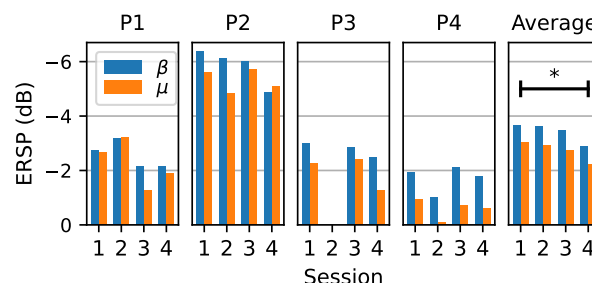


Figure 4: Motor beta and mu ERSP across practice sessions, averaged over tasks. Plotted for all participants (P1 to P4), and participant average.  $*p < 0.05$ . Note: Missing ERSP values for P3 in session 2 due to a technical issue.

**Cybersickness:** The descriptive statistics for VRSQ are detailed in Table 3. We found REVIRE to be well tolerated, with minimal cybersickness symptoms reported by all participants. Specifically, total VRSQ scores ( $8.91 \pm 7.64$ ) were significantly below the mild symptom threshold of 25,  $t(3) = -18.26$ ,  $p < .001$ . Fig. 6 depicts the cybersickness ratings of all participants, which consistently fell in the range of no to negligible symptoms.

## DISCUSSION

This paper introduced REVIRE, an immersive VR platform for BCI-based motor rehabilitation. We presented the results of a proof-of-concept study in which we tested

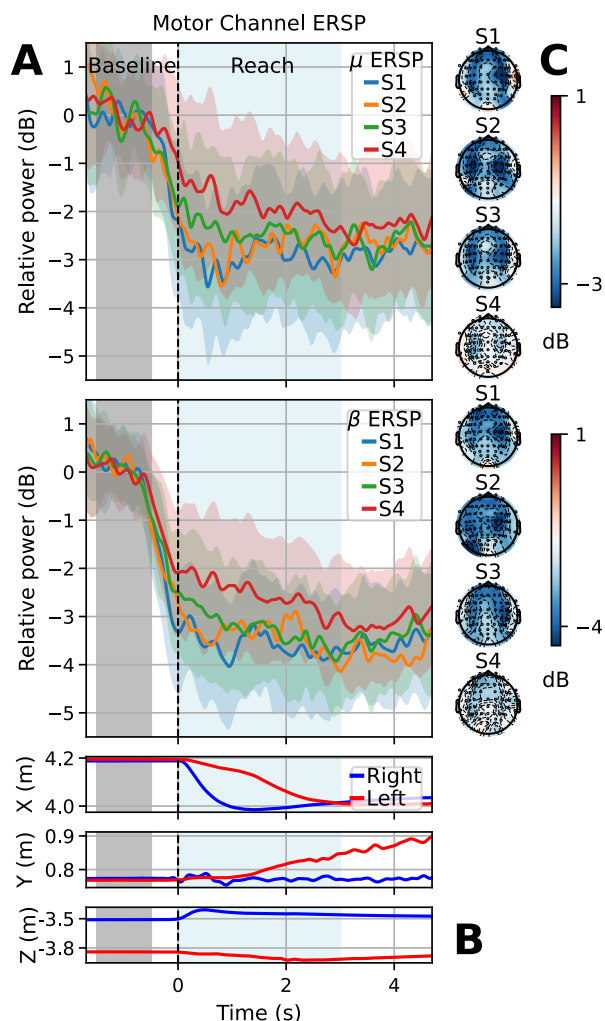


Figure 5: Grand-average of EEG responses and corresponding hand trajectories in the Pouring task. A: Within-epoch progression of mu and beta ERSP at motor channels. B: Left- and right-hand trajectories of the bimanual reaching motion on XYZ axes. C: Alpha (top) and beta (bottom) topographies of mean ERSP of the reaching movement duration (0-3s). S is session.

our VR-EEG setup on four healthy participants.

**Task performance:** Our results showed significant effects of training on participants' task performance across sessions, indicated by reduced trial completion times. We observed this trend not only in aggregated data but also at the participant level, demonstrating that completion times may provide an indicator for monitoring the progress of individual patients. Completion times in VR games have been found to correlate with clinical assessments [7, 15], supporting the potential of VR performance metrics as a meaningful assessment of rehabilitation progress.

**Hand trajectories:** REVIRE supports full hand tracking and stores hand movement trajectories. We have shown that the trajectories can be used to determine the precise movement onset or to identify specific actions, such as reaching movements. This can be a useful tool for analyzing EEG data recorded during natural self-paced movements. In addition, hand movement trajectories

Table 2: Mixed Linear Model Results for Mu and Beta ERSP

Effect	Estimate	SE	p
<b>Mu</b>			
Intercept (S1, T1 vs. 0)	-3.17	1.05	< .001
S2 vs. S1	0.44	0.59	.46
S3 vs. S1	0.64	0.54	.23
S4 vs. S1	1.23	0.54	<b>.02</b>
T2 vs. T1	0.02	0.54	.97
S2×T2	-0.20	0.83	.81
S3×T2	-0.48	0.76	.53
S4×T2	-1.04	0.76	.18
<b>Beta</b>			
Intercept (S1, T1 vs. 0)	-3.65	0.95	< .001
S2 vs. S1	0.28	0.59	.64
S3 vs. S1	0.30	0.54	.58
S4 vs. S1	1.09	0.54	<b>.04</b>
T2 vs. T1	-0.11	0.54	.84
S2×T2	-0.12	0.82	.89
S3×T2	-0.20	0.76	.79
S4×T2	-0.67	0.76	.38

*Note:* The participant variance is 3.84 for mu and 3.05 for beta. S is session, T1-2 are Pouring and Drinking tasks. SE is the standard error of the estimate. Statistical significance was set at  $p < .05$ .

Table 3: Descriptive Statistics of VRSQ Scales Across Sessions

Variable	Mean	Mdn	St. Dev.	IQR
Oculomotor	11.98	8.33	12.16	16.67
Disorientation	5.83	6.67	5.90	13.33
Total	8.91	9.58	7.64	11.04

have applications beyond what has been demonstrated in this preliminary analysis. By providing a direct quantitative measure of movement quality, clinicians can use trajectories to assess patient progress and highlight specific areas that may need improvement, such as range of motion, speed, or accuracy of movements.

**Event-related spectral activity:** Our study demonstrated the feasibility of capturing meaningful neural signals amidst considerable movement and the placement of a VR-HMD over EEG electrodes. Specifically, we observed a significant practice-related decrease in mu and beta ERSP magnitude, which is consistent with previous research on motor learning-related spectral changes, e.g., [12]. Our findings support the potential of EEG features as markers of rehabilitation progress.

**Cybersickness:** Participants' feedback indicated minimal cybersickness symptoms, confirming that our setup is well tolerated by users and further underscoring the suitability of REVIRE for patient use.

**Limitations and future work:** The main limitation of our study is the small sample size, consisting of only healthy participants, which limits the statistical validity and generalizability of our findings. Nevertheless, we observed promising trends and demonstrated the capability of our setup to provide meaningful data.

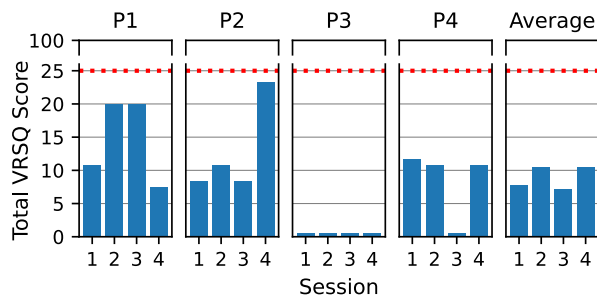


Figure 6: VR Sickness Questionnaire scores across sessions. Plotted separately for all four participants (P1 to P4), and for participant average. Scores below the red line indicate negligible to no cybersickness symptoms.

Our results demonstrated that REVIRE is well-suited for VR-BCI rehabilitation studies. Therefore, a future direction is to integrate the application into a closed-loop BCI system. By implementing real-time EEG processing and modifying the VR software to respond to neural signals, our application can provide neurofeedback or real-time customization of the training paradigm to the patient's neural state. Our results motivate further research to validate the application for clinical use and beyond, particularly with stroke patients and larger participant samples.

## CONCLUSION

In this paper, we demonstrated the potential of REVIRE to provide an immersive, task-specific rehabilitation environment that yields comprehensive behavioral, hand motion, and EEG data for patient monitoring. We established the feasibility of our VR-EEG setup and its utility for further clinical research, offering a promising foundation for integrating VR and EEG in accessible BCI-based motor rehabilitation.

## ACKNOWLEDGEMENTS

We thank our colleagues Tryggvi Edwald, Peter Fugger, Akshey Kumar, and Christoph Luther for their technical and creative support.

## REFERENCES

- [1] Levin MF, Weiss PL, Keshner EA. Emergence of Virtual Reality as a Tool for Upper Limb Rehabilitation: Incorporation of Motor Control and Motor Learning Principles. *Phys Ther*. 2015;95(3):415–425.
- [2] Everard G *et al*. New technologies promoting active upper limb rehabilitation after stroke: an overview and network meta-analysis. *European Journal of Physical and Rehabilitation Medicine*. 2022;58(4).
- [3] Mansour S, Ang KK, Nair KP, Phua KS, Arvaneh M. Efficacy of Brain–Computer Interface and the Impact of Its Design Characteristics on Poststroke Upper-limb Rehabilitation: A Systematic Review and Meta-analysis of

Randomized Controlled Trials. *Clinical EEG and Neuroscience*. 2022;53(1):79–90.

[4] Wen D *et al*. Combining brain–computer interface and virtual reality for rehabilitation in neurological diseases: A narrative review. *Annals of Physical and Rehabilitation Medicine*. 2021;64(1).

[5] Kourtesis P, Collina S, Doumas LAA, MacPherson SE. Technological Competence Is a Pre-condition for Effective Implementation of Virtual Reality Head Mounted Displays in Human Neuroscience: A Technological Review and Meta-Analysis. *Frontiers in Human Neuroscience*. 2019;13.

[6] Mekbib DB *et al*. A novel fully immersive virtual reality environment for upper extremity rehabilitation in patients with stroke. *Annals of the New York Academy of Sciences*. 2021;1493(1):75–89.

[7] Fregna G, Schincaglia N, Baroni A, Straudi S, Casile A. A novel immersive virtual reality environment for the motor rehabilitation of stroke patients: A feasibility study. *Frontiers in Robotics and AI*. 2022;9.

[8] Luu TP, Nakagome S, He Y, Contreras-Vidal JL. Real-time EEG-based brain-computer interface to a virtual avatar enhances cortical involvement in human treadmill walking OPEN. *Scientific Report*. 2017;7:8895.

[9] Bock O, Baetge I, Nicklisch A. hroot: Hamburg Registration and Organization Online Tool. *European Economic Review*. 2014;71:117–120.

[10] Kourtesis P, Collina S, Doumas LAA, MacPherson SE. Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology. *Frontiers in Human Neuroscience*. 2019;13.

[11] Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*. 2018;69:66–73.

[12] Zabielska-Mendyk E, Francuz P, Jaśkiewicz M, Augustynowicz P. The Effects of Motor Expertise on Sensorimotor Rhythm Desynchronization during Execution and Imagery of Sequential Movements. *Neuroscience*. 2018;384:101–110.

[13] Jeunet C, Glize B, McGonigal A, Batail JM, Micoulaud-Franchi JA. Using EEG-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: Theoretical background, applications and prospects. *Neurophysiologie Clinique*. 2019;49(2):125–136.

[14] Wang W *et al*. Neural Interface Technology for Rehabilitation: Exploiting and Promoting Neuroplasticity.

[15] Adamovich S *et al*. A virtual reality based exercise system for hand rehabilitation post-stroke: transfer to function. In: *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 4936–4939.

## WHICH IMAGINED SENSATIONS MOSTLY IMPACT ELECTROPHYSIOLOGICAL ACTIVITY?

E. Savalle<sup>1,\*</sup>, F. Le Jeune<sup>1,\*</sup>, L. Driessens<sup>2</sup>, M. J-M. Macé<sup>1</sup>, L. Pillette<sup>1</sup>

<sup>1</sup>Univ. Rennes, Inria, CNRS, IRISA, Rennes, France

<sup>2</sup>Inria, Univ. Rennes, IRISA, CNRS, Rennes, France

\*Co-first authorship. Both authors contributed equally.

E-mail: emile.savalle@inria.fr

**ABSTRACT:** Motor imagery brain-computer interfaces (MI-BCI) user training aims at teaching people to control their sensorimotor cortex activity using feedback on the latter, often acquired using electroencephalography (EEG). During training, people are mostly asked to focus their imagery on the sensations associated with a movement, though very little is known on the sensations that mostly favor sensorimotor cortex activity. Our goal was to assess the influence of imagining different sensations on EEG data. Thirty participants performed MI tasks involving the following sensations: (i) interoceptive, arising from the muscles, tendons, and joints, (ii) exteroceptive, arising from the skin, such as thermal sensations, or (iii) both interoceptive and exteroceptive. The results indicate that imagining exteroceptive sensations generates a greater neurophysiological response than imagining interoceptive sensations or both. Imagining external sensations should thus not be neglected in the instructions provided during MI-BCI user training. Our results also confirm the negative influence of mental workload and use of visual imagery on the resulting neurophysiological activity.

### INTRODUCTION

Controlling one's own brain activity when receiving direct information regarding the former is a skill that can be acquired using neurofeedback training. During such training, people's brain activity is acquired, often using electroencephalography (EEG), and converted into a feedback that people have to learn to control [1]. The ability to modulate one's own brain activity can be used for two main types of applications. First, to use brain-computer interfaces (BCIs), that enable the control of external digital systems by producing discriminatory and stable brain patterns each associated with a specific command for the system [1, 2]. For instance, BCIs can be used to control the direction of a character in a video game or the direction of a wheelchair by imagining right or left-hand movements [3, 4]. Second, for neurofeedback (NF) applications for which the end goal is that the modifications occurring in the brain activity lead to cognitive improvements, often in clinical applications [5].

For instance, neurofeedback can be used for motor rehabilitation after a stroke [6].

Many of these applications are based on the users' ability to control their sensorimotor brain activity. To do so, people are often asked to perform motor imagery (MI) tasks, such as imagining hand movements. Indeed, observing, executing, or imagining sensorimotor tasks induces a similar desynchronization over the sensorimotor cortex [7–9]. Two main non-exclusive MI methods are discussed in the literature [8, 9]. The first one is visual motor imagery (VMI) when people imagine the visual characteristics of the movement, which notably involves the visual cortical network. The second one is kinesthetic motor imagery (KMI), when people imagine the somatosensations associated with the movement. Those somatosensations include both (i) exteroceptive sensations, i.e., all the sensations arising from the skin, such as thermal, touch, or vibration sensations, and (ii) interoceptive sensations, i.e., all the information arising from the muscles, tendons, and joints, such as muscle contraction but also higher-level information such as knowing where our limbs are located in space. We recommend the review from Hillier et al. on the history of the terms related to proprioception and the assessment of proprioception [10].

When training to perform motor imagery, the learners are most frequently instructed to perform KMI [1]. As stated in the first paragraph, most researchers use EEG to acquire brain activity, most likely because it is a portable and relatively cheap method of acquisition. In the rest of the article, we will mostly focus on the results obtained using EEG and specify if the results were obtained using another acquisition method. The use of KMI instructions is mostly justified by the results obtained by Neuper et al. in 2005 [8]. Among others, they investigated the neurophysiological activation resulting from KMI and VMI. They found that classification performances of the data acquired when people were doing KMI were significantly higher than the ones obtained based on the data acquired when they were performing VMI. The highest classification accuracy was reached over the left central electrode site (i.e., electrode C3, which is coherent with the task performed by the participants to imagine right-

hand movements) with 67% of good classification with kinesthetic imagery and 56% with visual imagery. These results are in line with the results found by [11]. Conversely, recent results using EEG measures of connectivity found a better classification accuracy of VMI compared to KMI [9]. The type of imagery to perform could also depend on the task that needs to be learned and the stage of learning [12]. For instance, visual imagery seems more appropriate to learn the technical motor skill of drawing complex forms, while kinesthetic imagery enabled better temporal representation of the task [13].

The results on the influence of the modalities of feedback on BCI/NF efficiency could also provide insights regarding the sensations that should be associated with the motor imagery tasks. As such, the advantage provided by KMI compared with VMI is consistent with the results indicating that tactile and proprioceptive feedback (e.g., provided with vibrotactile actuators and orthosis) is more efficient than visual feedback in terms of classification performance, neurophysiological modifications, and user preferences [14].

As presented above, KMI involves many different sensations, among which the participants are left to choose from. For instance, the participants can decide to focus their imagination on the sensations arising from their muscles, and/or from their skin. Imagining exteroceptive sensations (i.e., sensation of pressure arising from squeezing a ball) in addition to interoceptive ones could significantly improve the classification performances based on EEG [15] or fNIRS [16] data, in particular, the ones of participants with poor performances (participants with performances below 70% for a BCI with 2 classes) [15]. Imagining the exteroceptive sensations associated with a movement could activate sensorimotor cortical structures and thereby improve BCI/NF user training [16]. Imagining exteroceptive sensations, i.e., vibrations on the back of the hand, does elicit desynchronization in alpha and beta bands (8-26 Hz), particularly in the upper alpha band and lower beta band (10-16Hz), over the sensorimotor cortex, i.e., C3 and C4 electrodes [17, 18].

Very little is currently known about the MI instructions that should be provided during BCI/NF user training, most of all regarding the potential influence of different external sensations. Our experiment therefore aims to study the influence of different somatosensory imagery tasks on neurophysiological activity with the aim of better advising our participants on the tasks they must imagine during MI user training.

## MATERIALS AND METHODS

The neurotypical participants included in this experiment took part in a 2-hour long session where they had to imagine five types of somatosensations, corresponding to the different conditions. A within-participant comparison of the mental tasks was chosen. The order of presentation of the conditions was randomized across participants.

*Participants:* Thirty right-handed participants with good or corrected vision took part in this experiment (7 women and 23 men; age 21-60,  $M = 29.4$ ,  $SD = 9.4$ ). None of them had any history of neurological or psychiatric disorder. The study was conducted following the relevant guidelines for ethical research according to the Declaration of Helsinki. Participants gave written informed consent before participating in the study. The study has been reviewed and approved by Inria's ethical committee, the COERLE (approval number: 2023-30).

*Experimental protocol:* The experiment lasted about two hours during which the participants were seated in a comfortable armchair, in front of a monitor. The participants first answered two questionnaires notably assessing demographic information, e.g., age and handedness. The EEG headset was then placed on their heads and a video presenting the experimental instructions was presented to them. EEG data was then acquired for 2 minutes while the participants were asked to focus on the visual scenery of their choice. The maximum force that participants were capable of exerting was then measured using a dynamometer placed inside a foam ball that the participants had to squeeze as strongly as they could for 30 seconds. This measure was used to provide instructions calibrated to the maximal force of the participants during the experiment. Following that, the main phase of the experiment began. It was composed of 5 different conditions during which the participants imagined sensorimotor imagery tasks varying according to the type of imagined somatosensation, i.e., interoceptive, exteroceptive, or both, and the number of exteroceptive sensations, i.e., pressure only or pressure and vibration. For each condition, participants watched a video presenting specific instructions for the movements and sensations to imagine. They also performed and experienced the movement and sensations associated with the conditions (see Figure 1):

- **Interoceptive sensation (I)** – Hand grasping with force on an invisible object, i.e., without fully closing the hand to avoid exteroceptive stimulation.
- **Exteroceptive sensation of pressure (E1)** – Ball pressed on the inside of the hand without voluntary movements of the hand.
- **Exteroceptive sensation of pressure and vibration (E2)** – Vibrating ball pressed on the inside of the hand without voluntary movements of the hand.
- **Interoceptive sensation and exteroceptive sensation of pressure (IE1)** – Hand movement to squeeze a ball.
- **Interoceptive sensation and exteroceptive sensation of pressure and vibration (IE2)** – Hand movement to squeeze a vibrating ball.

The pressure exerted voluntarily, or involuntarily by the experimenter, on the participants' hand was controlled at 20% of the maximum force produced by the participant. Indeed, previous results found that imagining movement

Type of sensations			
Internal	External	Internal & external	
			Pressure
			Pressure & Vibration

Figure 1: Experimental conditions: Interoceptive (I), Exteroceptive with pressure (E1), Exteroceptive with pressure & vibration (E2), Interoceptive & Exteroceptive with pressure (IE1), Interoceptive & Exteroceptive with pressure & vibration (IE2).

with different amount of force impacted the resulting brain activity [15]. Participants then had to imagine these different tasks during 20 trials each lasting 10 seconds. Runs lasted 4 min 30 seconds each. A break was offered to the participants between runs. At the end of these runs and for each condition, participants completed a questionnaire assessing their user experience. The questionnaire was composed of (i) the questions from the NASA-TLX [19] to assess mental workload and (ii) two questions based on the kinesthetic and visual imagery questionnaire (KVIQ) [20] to assess how clear the motor imagery task was in terms of visual and somatosensory representation with scales ranging from "No image" to "Image as clear as a movie" and from "No sensations" to "Sensations as intense as when performing the movement/feeling the sensations". All the questions were answered using an analogical scale ranging from 0 to 20. Finally, the EEG headset was removed and the participants completed the final questionnaire that evaluated which imagery task they preferred and which seemed the most effective and simple to imagine. This experimental protocol was presented and discussed at the French national BCI conference in 2023 [21].

**EEG Recordings & Signal Processing:** The electroencephalographic (EEG) data was recorded using 20 active electrodes, using a g.USBamp EEG amplifier (g.tec, Austria). The electrodes were placed on the scalp of the participant over the sensorimotor area (at locations FC5, FC3, FC1, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CP2, CP4, CP6 and Pz in the 10-20 system). They were referenced to the left earlobe and grounded to AFz. The data was sampled at 512 Hz, and processed online using OpenViBE 3.4.0 [22].

EEG data was preprocessed with MNE-Python [23]. The signal was filtered using a zero-phase notch filter with a 50 Hz cut-off and a finite impulse response band-pass filter with cut-off frequencies of 1 and 49 Hz and then average-referenced. We extracted epochs from 1 second before MI instruction cue to 10 seconds after. Epochs with peak-to-peak amplitude greater than 100  $\mu$ V were rejected. Participants with more than 50% of total epochs rejected were removed from the analysis. As a result, 3 participants were rejected from the analysis.

**Variables:** Among the neurophysiological characteristics, we investigated the event-related potentials (ERPs), corresponding to either a desynchronisation (i.e., ERD)

or a synchronisation (i.e., ERS) in the brain activity of our participants while they performed the different sensorimotor imagery tasks. To assess the ERPs, we first computed time-frequency representation using Morlet wavelets. We resampled the data at 256 Hz, and used a Morlet wavelet transform to calculate the EEG signal power between 12 and 20 Hz.

To have an idea of the evolution of the ERPs throughout the different trials, the resulting data was averaged across participants and electrodes (CP3, C3, C4, and CP4) for each conditions. Previous experiments mostly focused on C3 and C4. However, somatosensory data is primarily processed in posterior areas leading to the inclusion of CP3 and CP4 in our analyses [24]. The data was then normalized relative to baseline (the first second before cue) using a log-ratio of power at each time point relative to the mean power of the baseline, that we call *Power Evolution over Trial* in this analysis :

$$Power\ Evolution\ over\ Trial = \log(Task/\overline{Baseline})) \quad (1)$$

Then, *Power Evolution over Trial* was averaged across time, excluding the first two seconds and last second of the trial providing the *Average ERD/S value*.

Finally, to investigate the potential reasons for the difference between our conditions, we used the answers to post-conditions questionnaires, NASA-TLX and the adapted KVIQ, to observe their correlation with *Average ERD/S values*. There are 7 different variables calculated from the questionnaires: "Mental demand", "Temporal demand", "Performance", "Workload" and "Frustration" for the NASA-TLX; "Visual imagery" and "Kinesthetic imagery" for the KVIQ. Each variable was evaluated on an analogical scale ranging from 0 to 20, with lower values indicating lower workload for the NASA-TLX variables, and less vivid or clear imagery for the KVIQ variables.

## RESULTS

In the first step of our analysis, we assessed the potential influence of our different experimental conditions on the cortical activity over the sensorimotor cortex throughout our trials.

To assess this, we first plotted *Power Evolution over Trial*, i.e., the evolution of *ERD/S* over time in the beta band (12.5-20 Hz), for our 3 major conditions (I, E and IE), averaged across participants and across "*Nb ext. sensations*" for condition E and IE (see Figure 2). A 1-second sliding average window was used for readability. On average, it seems that all the participants managed to desynchronize their brain activity over the sensorimotor cortex for the E and IE conditions, albeit seeming stronger for the E conditions. For the I condition, the brain activity over the sensorimotor cortex seems to have desynchronized until the third second and then steadily synchronized until the end of the trial.

Then, we were interested in knowing if these observed differences were significant, to assess this we use *ERD/S*

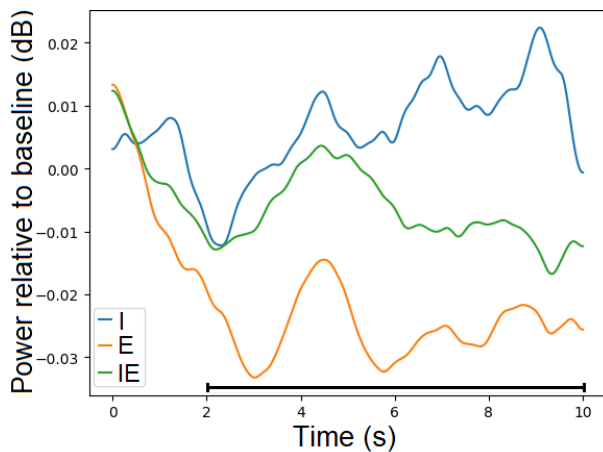


Figure 2: Average Power Evolution over Trial computed for each condition, with data smoothed using a 1-second window moving average. The black bar indicates the time window used for the analysis.

value. First, we used a Shapiro-Wilk test to verify the normality of *ERD/S values*. We found no significant deviation from normality ( $W = 0.98$ ,  $p = 0.66$ ) in our data. To avoid any distortion in the statistical results, we also checked the data for outliers. Any participant with an *ERD/S value* plus or minus two standard deviations relative to the median *ERD/S value* was considered an outlier and removed from our analyses. In total, 4 participants were removed from the following analyses.

We first studied the impact of the type and number of somatosensations imagined on the *Average ERD/S values*. As the interoceptive condition (I) was not associated with a number of exteroceptive sensations, it was first removed from the analysis. Thus, we assessed if the number of exteroceptive sensations, i.e., “*Nb ext. sensations*” (1 or 2) and the type of somatosensations, i.e., “*Type of sensations*”, (E or IE) had an impact on *Average ERD/S values* using a 2-way ANOVA with “*Type of sensations*” (E and IE) and “*Nb ext. sensations*” (Pressure and Pressure & Vibration) as independent variables. Our results indicate that “*Type of sensations*” influences *Average ERD/S values* [ $F(1, 22) = 7.52$ ;  $p = 0.01$ ;  $\eta = 0.047$ ], but “*Nb ext. sensations*” [ $F(1, 22) = 1.2$ ;  $p = 0.208$ ;  $\eta < 10^{-2}$ ] does not. A small trend could be present for “*MI type \* Nb sensation*” [ $F(1, 23) = 3.02$ ;  $p = 0.1$ ;  $\eta < 10^{-2}$ ].

To compare condition I to the others and since no significant influence of “*Nb ext. sensations*” was found on *Average ERD/S values*, we averaged E1 with E2 into E, and IE1 with IE2 into IE. We performed a one-way repeated measures ANOVA with “*Type of sensations*” (I, E and IE) as independent variable and *Average ERD/S values* as dependent variable. The results show a **significant influence of “Type of sensations”** [ $F(2, 44) = 7.64$ ;  $p < 10^{-2}$ ;  $\eta = 0.13$ ]. To gain insight on this significant result, post-hoc analyses were performed and revealed a **significant difference between I and E** ( $p < 10^{-2}$ ), **E and IE** ( $p = 0.02$ ) but not between I and IE ( $p = 0.27$ ).

Figure 3 reports the distribution of these values, showing

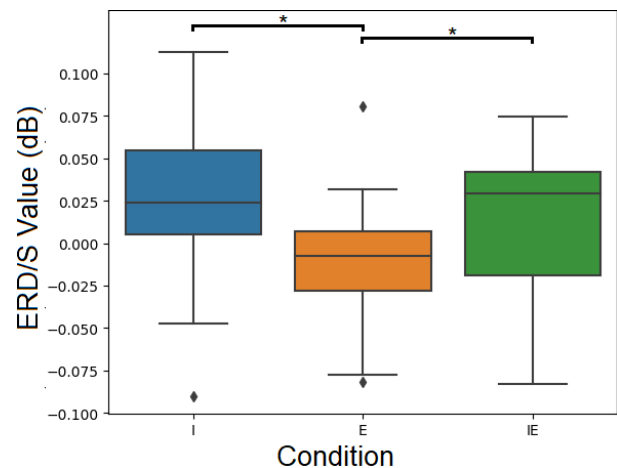


Figure 3: ERS/D recordings for conditions I, E and IE. Asterisks meanings:  $p < 0.05$ : \*;  $p < 0.01$ : \*\*;  $p < 0.001$ : \*\*\*.

a greater desynchronization for the condition E compared to both I and IE.

Finally, to assess the potential causes of these significant differences, we studied the correlations between *Average ERD/S values* and the data acquired after each condition regarding the user experience. The correlations were assessed through Pearson correlation with False Discovery Rate (FDR) to correct for multiple comparisons. **A small significant positive correlations was found between Average ERD/S values and Visual Imagery** ( $r = 0.16$ ,  $p < 10^{-3}$ ,  $DoF = 478$ ) **and between Average ERD/S values and Mental workload** ( $r = 0.12$ ,  $p = 0.02$ ,  $DoF = 478$ ). These correlations indicate that the more participants felt mentally overloaded, or felt that they performed realistic visual imagery, the less they produced strong ERDs.

## DISCUSSION

In this study, the electrophysiological analysis focused on *Event-Related Desynchronization/Synchronization (ERD/S)* occurring over the sensorimotor cortex (i.e., C3/C4 and CP3/CP4) resulting from different sensorimotor imagery tasks performed by the participants. Visually, it seems that the participants managed to desynchronize their brain activity over their sensorimotor cortex in most conditions, except when imaging purely interoceptive sensation. This result might be related to the unfamiliarity of performing a hand-grasping gesture without any object in the center, as familiarity with a task was found to positively influence single-trial detectability of imagined movements [25]. There is a trend ( $p = 0.1$ ) of interaction between the number of exteroceptive sensations and the type of somatosensations although further experiments should verify this interaction.

Our statistical analyses mostly revealed significantly stronger desynchronization when our participants imagined exteroceptive sensations compared to when they imagined (i) only interoceptive and (ii) both interoceptive and exteroceptive sensations. A distinct pattern of activation when imagining exteroceptive sensations is in line

with the work of Yao et al. who worked on somatosensory imagery-based BCI and found relatively high average classification accuracy ranging from 77% to 85% depending on their studies [17, 18]. The classification performances obtained by Yao et al. using somatosensory imagery-based BCIs are similar to the ones obtained in motor imagery-based BCI with maybe a smaller number of persons who could not control the system compared to other studies. It should be noted that these results remain difficult to compare in terms of type of somatosensations imagined by the participants as they are left free to choose among different strategies [17, 18].

Based on previous work, we were expecting stronger ERDs when participants imagined both interoceptive and exteroceptive sensations compared to when they imagined interoceptive sensations alone [15, 16, 26]. This discrepancy could be the result of choices made in the design of our experimental paradigm. Indeed, when pressing the ball against participants' hands, the contact with the table added exteroceptive sensations at the back of the hand and could add a passive movement with small amplitude of the participant's hand producing interoceptive sensations. Thus, our exteroceptive sensations conditions (E1 and E2) were not purely composed of exteroceptive sensations and the amount of exteroceptive sensation was sensibly more important than in the two other conditions. This difference in exteroceptive sensations between E and IE did influence our results. We hypothesize that this increase in the amount of exteroceptive sensations could explain the discrepancy between our E and IE conditions but could also stress the importance of spatially distributed exteroceptive stimulation. Future experiment should consider adding symmetric external force to the back of the hand while the participants grasp the ball in both interoceptive and exteroceptive sensation conditions to counteract this bias. Additionally, we used the same ball for all participants regardless of their hand size leading to possible differences in the movement executed and imagined by participants. Following studies could use balls with different sizes to take into account the diversity of participants and keep the movement consistent across participants. Although, our analyses were comparing intra-participant data, so this factor should not have influenced our results.

Finally, we also discovered significant positive correlations between the ERDs and the workload experienced by participants. A negative influence of workload on BCI performance was already suspected as workload is one of the main factors influencing learning in general [27] and poor BCI performances were associated with high theta waves, which is an indicator of high workload [28]. Our results are in contradiction with the recent results from Gu et al. who found a positive influence of high workload on MI-BCI accuracy and desynchronization over the sensorimotor cortex [29]. However, the positive correlation between the ERDs and how realistically participants visually imagined the task is consistent with previous results from the literature [8, 11].

## CONCLUSION

In this study, we designed and conducted an experiment studying the influence of sensorimotor imagery tasks to determine which sensations would elicit the largest desynchronization of the sensorimotor cortex. Our 30 participants performed five sensorimotor imagery tasks, each one with a different imagined somatosensation.

Results are consistent with the literature stating that we should encourage the participants to imagine exteroceptive sensations, such as pressure or vibrations, during BCI user training. We found significantly stronger desynchronization in the low beta band (12-20 Hz) over the sensorimotor cortex when our participants imagined exteroceptive sensations (either only pressure or both pressure and vibrations) compared to when they imagined only interoceptive sensations (from their muscles, tendons, and joints) and compared to when both interoceptive and exteroceptive sensations were imagined. Even though these results tend to indicate a limited role of interoceptive sensations imagery on resulting sensorimotor activity, our participants might still have imagined some interoceptive sensations while performing passive small movements. Higher level interoceptive sensations, such as the spatial perception of limbs in space, were also still present in the exteroceptive condition even though the exteroceptive tasks solicited them much less than the interoceptive condition. The differences in our conditions could also be explained by a greater surface of exteroceptive stimulation in the exteroceptive-only conditions. Future experiments with more complex experimental setup should further investigate this hypothesis.

In this article, we presented preliminary analyses of our results, the processing of our data is still ongoing. Future analyses will study more specifically the influence of our conditions on the temporal, spatial (notably between sensorimotor and motor cortex and lateralization), and frequency bands characteristics of our EEG data. Classification accuracy performances should also be computed to better compare our results to the ones reported in previous articles. Additionally, potential differences in user experience among our conditions will be investigated to gain insights on potential underlying factors explaining the differences observed in this article, especially the cognitive workload or the use of visual imagery.

Also, our results were obtained offline as our participants performed the tasks without any feedback on their brain activity. Thus, our results are not entirely comparable with previous ones that provided feedback to the participants regarding their brain activity. Future experiments should investigate how providing different sensorimotor imagery instructions influences participants' ability to learn to control their brain activity in BCI user training.

To build on these first results, future analysis will focus on more frequency bands such as alpha/mu rhythms, and attempt single trial classification to determine which sensation are easier to discriminate from each other.

## REFERENCES

- [1] Roc A *et al.* A review of user training methods in brain computer interfaces based on mental tasks. *Journal of Neural Engineering*. 2021;18(1):011002.
- [2] Wolpaw J, Wolpaw EW. *Brain-computer interfaces: Principles and practice*. OUP USA (2012).
- [3] Tonin L *et al.* Learning to control a bmi-driven wheelchair for people with severe tetraplegia. *Iscience*. 2022;25(12).
- [4] Benaroch C *et al.* Long-term BCI training of a tetraplegic user: Adaptive riemannian classifiers and user training. *Frontiers in Human Neuroscience*. 2021;15:118.
- [5] Batail JM *et al.* EEG neurofeedback research: A fertile ground for psychiatry? *L'Encéphale*. 2019;45(3):245–255.
- [6] Le Franc S *et al.* Toward an adapted neurofeedback for post-stroke motor rehabilitation: State of the art and perspectives. *Frontiers in Human Neuroscience*. 2022;16:917909.
- [7] Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans. *Neuroscience letters*. 1997;239(2-3):65–68.
- [8] Neuper C, Scherer R, Reiner M, Pfurtscheller G. Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial eeg. *Cognitive brain research*. 2005;25(3):668–677.
- [9] Yang YJ, Jeon EJ, Kim JS, Chung CK. Characterization of kinesthetic motor imagery compared with visual motor imageries. *Scientific Reports*. 2021;11(1):3751.
- [10] Hillier S, Immink M, Thewlis D. Assessing proprioception: A systematic review of possibilities. *Neurorehabilitation and neural repair*. 2015;29(10):933–949.
- [11] Davidson RJ, Schwartz GE. Brain mechanisms subserving self-generated imagery: Electrophysiological specificity and patterning. *Psychophysiology*. 1977;14(6):598–602.
- [12] Dickstein R, Deutsch JE. Motor imagery in physical therapist practice. *Physical therapy*. 2007;87(7):942–953.
- [13] Féry YA. Differentiating visual and kinesthetic imagery in mental practice. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*. 2003;57(1):1.
- [14] Pillette L. Redefining and adapting feedback for mental-imagery based brain-computer interface user training to the learners' traits and states. Ph.D. dissertation. Bordeaux. 2019.
- [15] Park S, Ha J, Kim DH, Kim L. Improving motor imagery-based brain-computer interface performance based on sensory stimulation training: An approach focused on poorly performing users. *Frontiers in Neuroscience*. 2021;15:732545.
- [16] Wriessneger SC, Bauernfeind G, Kurz EM, Raggam P, Müller-Putz GR. Imagine squeezing a cactus: Cortical activation during affective motor imagery measured by functional near-infrared spectroscopy. *Brain and cognition*. 2018;126:13–22.
- [17] Yao L, Mrachacz-Kersting N, Sheng X, Zhu X, Farina D, Jiang N. A multi-class bci based on somatosensory imagery. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2018;26(8):1508–1515.
- [18] Yao L, Jiang N, Mrachacz-Kersting N, Zhu X, Farina D, Wang Y. Performance variation of a somatosensory bci based on imagined sensation: A large population study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2022;30:2486–2493.
- [19] Cegarra J, Morgado N. Étude des propriétés de la version francophone du nasatlx. In: *Communication présentée à la cinquième édition du colloque de psychologie ergonomique (Epique)*. 2009, 28–30.
- [20] Malouin F, Richards CL, Jackson PL, Lafleur MF, Durand A, Doyon J. The kinesthetic and visual imagery questionnaire (kviq) for assessing motor imagery in persons with physical disabilities: A reliability and construct validity study. *Journal of Neurologic Physical Therapy*. 2007;31(1):20–29.
- [21] Le Jeune F, Driessens L, Savalle E, Pillette L. Which sensations should you imagine? In: *Collectif pour la Recherche Transdisciplinaire sur les Interfaces Cerveau-Ordinateur CORTICO 2023*. 2023.
- [22] Renard Y *et al.* Openvibe: An open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. *Presence*. 2010;19(1):35–53.
- [23] Gramfort A *et al.* Meg and eeg data analysis with mne-python. *Frontiers in neuroscience*. 2013:267.
- [24] Dykes RW. The anatomy and physiology of the somatic sensory cortical regions. *Progress in neurobiology*. 1978;10(1):33–88.
- [25] Gibson RM, Chennu S, Owen AM, Cruse D. Complexity and familiarity enhance single-trial detectability of imagined movements with electroencephalography. *Clinical Neurophysiology*. 2014;125(8):1556–1567.
- [26] Yao L *et al.* A stimulus-independent hybrid bci based on motor imagery and somatosensory attentional orientation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2017;25(9):1674–1682.
- [27] Baddeley A. Working memory. *Current biology*. 2010;20(4):R136–R140.
- [28] Ahn M, Cho H, Ahn S, Jun SC. High theta and low alpha powers may be indicative of bci-illiteracy in motor imagery. *PloS one*. 2013;8(11):e80886.
- [29] Gu B *et al.* The effects of varying levels of mental workload on motor imagery based brain-computer interface. *International Journal of Embedded Systems*. 2020;12(3):315–323.

# ONLINE DETECTION OF EPILEPTIC SPIKES FOR USE IN EPILEPSY MONITORING

M. Mohammadpour<sup>1,2</sup>, C. Kapeller<sup>2</sup>, K. Kamada<sup>3</sup>, J. Scharinger<sup>1</sup>, M. Schwarzgruber<sup>2</sup>,  
M. Korostenskaja<sup>4</sup>, C. Guger<sup>2</sup>

<sup>1</sup> Department of Computer Science, Johannes Kepler University, Linz, Austria

<sup>2</sup> g.tec medical engineering GmbH, Linz, Austria

<sup>3</sup> Department of Neurosurgery, Megumino Hospital, Eniwa, Japan

<sup>4</sup> g.tec neurotechnology, NY, USA

E-mail: mohammadpour@gtec.at

**ABSTRACT:** Epileptic spikes, indicative of the seizure onset zone (SOZ), provide meaningful insight for neurosurgeons looking to find seizure locations, particularly during intraoperative procedures. Many algorithms have been proposed to detect epileptic spikes, primarily based on offline data analysis. However, none of these algorithms have been successfully adapted for online applications. In this study, we introduce a novel method for online detecting epileptic spike patterns in electrocorticography (ECoG) data. This algorithm dynamically models statistical distributions of signal envelopes, which could discriminate between signals containing epileptic spikes and those showing background activity. The effectiveness of the proposed algorithm is evaluated using resting-state data from two patients. The results reveal a sensitivity of 73% and a specificity of 95% for detecting epileptic spikes online, with an overall accuracy of 93% and an f1 score of 52%. Overall, these results validate the potential of online detection as a valuable method for epilepsy monitoring and diagnosis.

## INTRODUCTION

Epileptic seizures result from the excessive and synchronized activity of large neuronal groups, making epilepsy one of the most common neurological disorders globally, impacting around 50 million individuals [1]. While many epilepsy patients effectively manage seizures with medication, approximately one-third continue to experience seizures despite treatment [2]. For these cases, surgical resection of the brain tissue responsible for seizures becomes a feasible treatment, which needs to identify SOZ accurately.

Clinical localization of the seizure onset zone (SOZ) requires implanting intracranial EEG (iEEG) electrodes, recorded over several days, to capture spontaneous seizures [3]. Electrodes within the SOZ are identified through visual inspection of iEEG recordings taken during seizures, guiding the removal of surrounding tissue during surgery. Despite serving as the current

gold standard for mapping the epileptic brain clinically, this manual process is time-consuming, costly, and carries potential risks of morbidity [4]. Consequently, there is growing interest in automating SOZ localization to simplify epilepsy monitoring and facilitate the identification of the SOZ [5].

Interictal epileptiform discharges (IEDs) are transient electrographic events observed in patients with epilepsy. They serve various diagnostic and monitoring purposes, aiding in the identification of epileptic activity and the localization of epileptogenic tissue and SOZ. During presurgical evaluations [6], neurosurgeons often use information from interictal discharges to understand where the seizures start in the brain and to plan where and how much tissue to remove [7, 8]. Studies have indicated that resecting regions exhibiting frequent spikes correlates with improved surgical outcomes [9, 10]. Research has demonstrated that IEDs can effectively localize the seizure onset [11], with the most common types of IEDs identified through visual and semi-automated detection in long-term monitoring (LTM) and visual detection in high-density EEG (hdEEG) significantly aligning with the SOZ [12].

This study aims to address the growing need and to create an automated online epileptic spike detection method for SOZ localization. It leverages signal envelopes to model the statistical distributions of ECoG signals. This approach aims to enable real-time epilepsy, including intraoperative application. To achieve this goal, electrocorticography (ECoG) data from patients with epilepsy were used, in which half-second segments of data were analyzed to detect epileptic spikes. Later, a sequence detection algorithm was applied to the identified spikes to capture spatial information. This algorithm identifies spikes occurring across multiple electrodes in close temporal proximity, improving the understanding of epileptic activity distribution.

## MATERIALS AND METHODS

The study utilized ECoG data from two patients

undergoing diagnostic subdural grid implantation at Megumino Hospital in Japan. This data collection occurred within the Epilepsy Monitoring Unit (EMU) during the patients' resting-state sleep at night. The recordings were conducted under the influence of antiepileptic drugs (AEDs). As a result, it was expected that the recorded data would show a decrease in interictal epileptiform activity due to the administration of AEDs.

The patients underwent implantation of subdural grid electrodes to localize the SOZ and perform real-time functional mapping to identify critical brain function areas, aiming to minimize resection before surgery. The implanted grids, sourced from Ad-Tech in Racine, WI, USA, comprised platinum discs with a diameter of 4.0 mm, spaced apart at 5-10 mm intervals.

The ECoG signals were acquired in the EMU using a 256-channel g.HIamp biosignal amplifier (g.tec medical engineering GmbH, Austria). The signals were digitized with a high resolution of 24 bits at a sampling rate of 4800 Hz. Ground and reference electrodes were in the dorsal parietal cortex to ensure signal stability and consistency. 144 channels were recorded for patient 1 and 136 for patient 2 (280 channels total).

The online detection system for epileptic spikes was developed using MATLAB Simulink (MathWorks, Inc.) and comprises several key components. Initially, the signal was down-sampled to 200 Hz and subjected to a high-pass filter with a cutoff frequency of 2 Hz to eliminate DC offset. Later, each channel underwent an 8th-order Butterworth filtering within the 10-60 Hz band. Following this preprocessing step, the signals were processed using a common average reference (CAR) technique to mitigate noise and non-cerebral artifacts, a critical step for ensuring the accuracy of spike detection algorithms by reducing false positive events. The input signal was then segmented into half-second intervals to facilitate spike detection. The output of the spike detection block, depicted in scope with markers denoting detected events (Figure 1), relies on a real-time adaptation of the method outlined in [13].

To achieve real-time spike detection, the instantaneous envelope of each filtered channel was computed using the absolute value of the Hilbert transform. Spikes typically produce an energy increase, resulting in peaks in the envelope within the 10-60 Hz frequency band. Statistical distribution of the envelope was computed for each segment, and a model was fitted using a maximum likelihood algorithm (MLE). Later, the mode and median of the normalized (log-normal distribution) data were used to establish a threshold for detecting segments containing spikes from those displaying background activity. Statistical parameters such as mean and standard deviation were calculated for each window and accumulated for ongoing threshold adjustment to ensure adaptability. This adaptive thresholding mechanism optimizes spike detection performance across varying signal conditions.

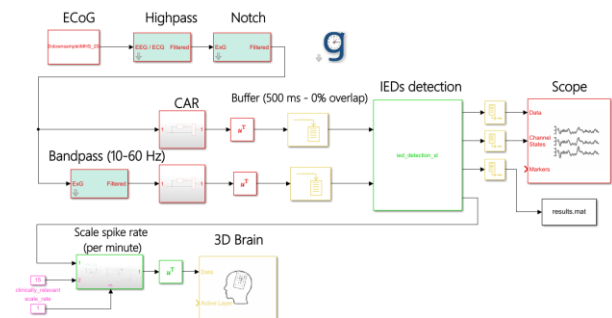


Figure 1: The Simulink model of real-time epileptic spike detection.

ECoG data were examined for interictal discharges to assess the model's performance. For this purpose, a 10-minute segment of ECoG data was selected. The signals were analyzed using g.BSanalyze software (g.tec medical engineering GmbH) to identify spikes. IEDs typically start with a sharp wave or spike, indicating a brief, high-amplitude deviation from baseline. Subsequently, a slow wave component may follow, characterized by a slower and more prolonged deflection than the sharp wave [14]. Figure 2 illustrates the labeling of epileptiform discharges.

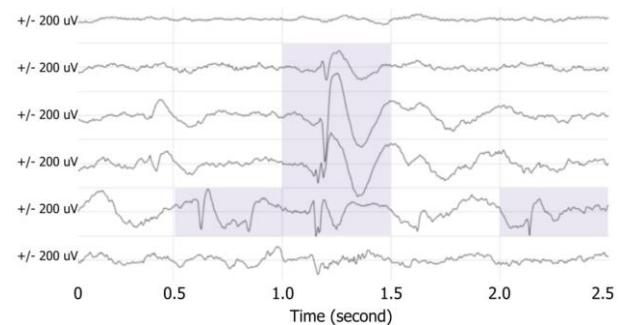


Figure 2: Annotation of epileptiform discharges for use compared to the detected event.

After the completion of the simulation, the detected events are saved in a file, which are indicators of spike occurrence in the signal. Later, these events are organized into vectors for each channel. Each vector represents the spike times detected for the corresponding channel. For the evaluation, true positive (TP) is defined as an event where the predicted event is located on the marked event in ground truth, i.e., at least some samples overlap. A false positive (FP) is defined as an event where the predicted event does not overlap with the marked event. A false negative (FN) is defined as an event where the predicted event does not overlap with the event in ground truth. A true negative (TN) is defined as an event where no event is predicted, and no event is in the ground truth. Several performance metrics based on these definitions are calculated. Accuracy measures the proportion of correctly identified events, computed as  $\text{accuracy} = (\text{TP} + \text{TN}) / N$ . Sensitivity reflects the model's ability to correctly identify true events, calculated as  $\text{sensitivity} = \text{TP} / (\text{TP} + \text{FN})$ . Specificity gauges the model's proficiency in identifying true negative events, determined by

specificity =  $TN / (TN + FP)$ . The F-score, the harmonic mean of precision and recall, is computed as  $F\text{-score} = (2 \times FP) / (2 \times TP + FP + FN)$ .

## RESULTS

The proposed system was applied to two patients, including data collected under the influence of AEDs. Figure 3 displays an example output from the scope, illustrating the detected epileptic spikes for patient 1. This visual representation clearly depicts the identified epileptic activity within the recorded ECoG data, aiding in assessing and analyzing epileptic spike detection performance.

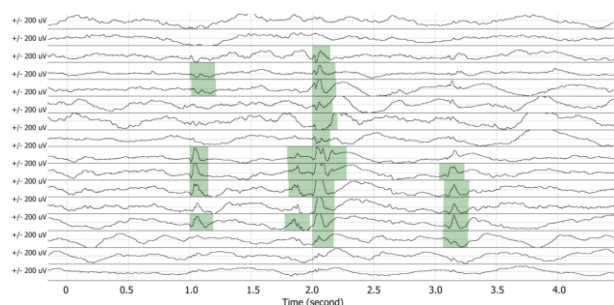


Figure 3: Visualization of online detection of epileptic spike.

Upon running the model, sensitivity, specificity, accuracy, and F-score are calculated for model evaluation, as depicted in Table 1. This table presents the average performance metrics for patients, where the predicted values of the proposed system are compared with the ground truth labels of epileptic spikes. The data used for analysis were extracted from minutes 3 to 10 of the recording. This duration includes 7 minutes of data, during which the model's performance was considered stable. This specific timeframe was focused on ensuring that the analysis is conducted on data where the spike detection algorithm has reached a consistent and reliable performance level. First, thresholds were established to detect epileptic spikes with 95% specificity for the entire patient. Then, sensitivities for detecting epileptic spike events were calculated on the patients with those thresholds. A comparison from Table 1 shows that the performance of the online system closely resembles that of offline reference methods. This observation highlights the effectiveness of the online system in accurately detecting epileptic spikes, demonstrating its potential as a feasible alternative to traditional offline methods. The low F1 score in the comparison is primarily due to the high number of false positive (FP) events, which arise from the complexity of the comparison. In this scenario, the time samples of detected spikes with those of marked spikes (ground truth) across all channels were compared. This presents a significant challenge because the detected spike should be aligned with the time sample of a spike in the ground truth. In contrast, some other research groups solely compare the spike detection algorithm's spike rate

with the SOZ or spike rate in the ground truth across all channels, resulting in higher scores. However, this approach overlooks the temporal validation aspect.

Table 1: Performance results of the proposed system for datasets

Methods	TP	TN	FP	FN	Sens.	Spec.	Acc.	F1
ONLINE	44	737	43	17	0.73	0.95	0.93	0.52
REF [13]	47	743	39	11	0.70	0.95	0.94	0.45
REF [15]	18	768	34	20	0.39	0.95	0.94	0.23
REF [16]	46	742	39	13	0.69	0.95	0.94	0.48

Figure 4 demonstrates the calibration time of the proposed method for detecting spike events in the ECoG signals. Here, it can be seen that after minute 3, the method could reach a stable detection period.

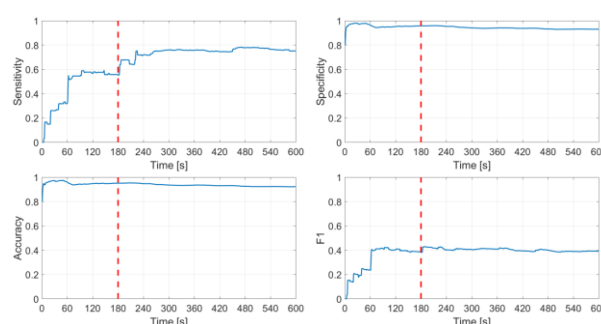
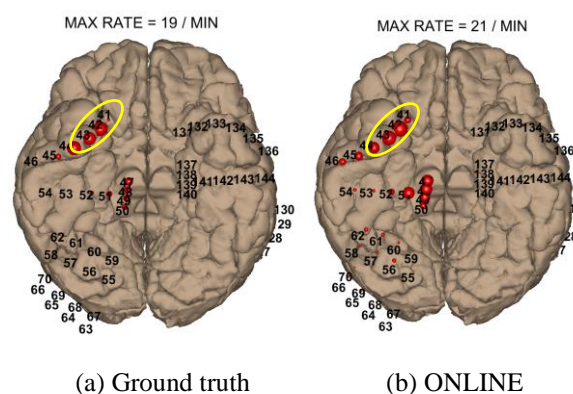


Figure 4: Calibration time of epileptic spike detection for use in real-time BCI application.

Figure 5 illustrates the spatial distributions of epileptic spikes for different methods in a 3D brain schema for patient 1. To represent the spatial distribution, the spike rate for each channel is calculated in terms of spikes per minute. Notably, the seizure onset location for patient 1 is in channels 41-43, shown in yellow electrodes. By reviewing Figure 5, it becomes evident that channels in the models exhibiting high spike rates are either located within the SOZ or close to it, which is why the proposed method aligns with the offline method for pointing the SOZ location. This observation strengthens the correlation between epileptic activity and the SOZ, emphasizing the importance of accurately localizing the SOZ for effective diagnosis and treatment of epilepsy.



(a) Ground truth

(b) ONLINE

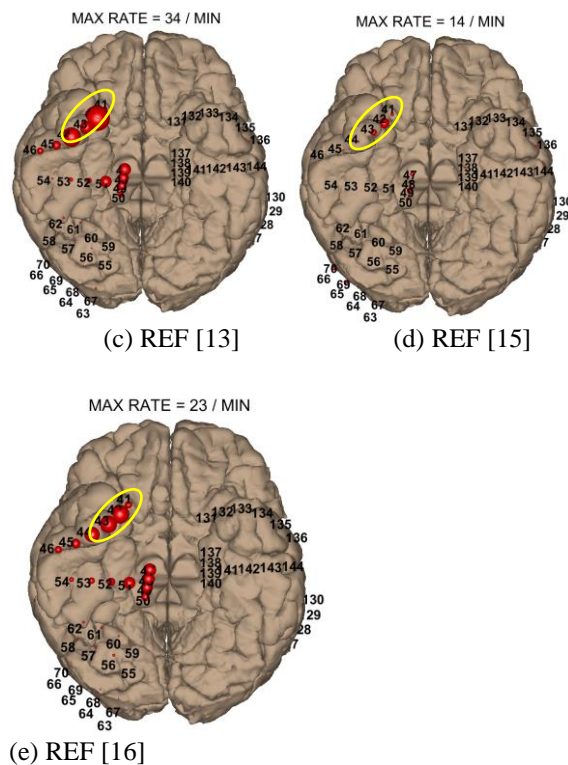


Figure 5: Epileptic spike rate in 3D brain schema for the models. Electrodes 41-43 indicate SOZ. Subfigures a-e indicate the epileptic spike rate for ground truth, online method, method ref [13], and method ref [15].

## DISCUSSION

In contrast to the approach described in [16], the current study presents a novel online epileptic spike detection method. Unlike the existing offline methods [13, 15, 16], which necessitate the use of the full length of the signal for detection, the approach in this work utilizes only a half-second segment to model the statistical distribution of ECoG signals for detecting epileptic spikes. Changes to the method described in reference [13] have enhanced its online performance. This distinction highlights the efficiency and effectiveness of the online method, offering potential advantages in terms of computational resources and speed of analysis compared to traditional offline methods.

In this research, the proposed approach involves relying solely on a reviewer for signal labeling, which introduces potential limitations. This could be a reason why comparing spike occurrences solely in the time domain can be challenging due to the inherent variability in EEG signals and the potential for small temporal deviations between ground truth and detected spikes. Furthermore, inherent limitations, such as mislabeling or biases towards specific markers, may exist. These factors could contribute to the observed low sensitivity, specificity, and accuracy values when comparing spikes in the time domain to ground truth.

However, analyzing spike activity in the spatial domain, such as plotting the rate of spikes across different channels, offers a more comprehensive understanding of the underlying neural activity. Spatial information provides insights into the specific regions or electrodes where spikes frequently occur, allowing for a more robust comparison between ground truth and predicted data. Therefore, integrating spatial analysis can reduce some of these limitations and provide a more reliable assessment of spike detection performance.

One of the primary challenges hindering the practical implementation of a brain-computer interface (BCI) is the long calibration period required. However, this paper proposes a novel approach utilizing adaptive thresholding, which accumulates statistical characteristics of the ECoG signal. This method demonstrates stable epileptic spike detection in terms of specificity after approximately 3 minutes, thereby significantly reducing the calibration time required for spike detection in epilepsy monitoring applications. The experimental results depicted in Figure 4 illustrate that the proposed algorithm rapidly achieves a predefined performance level. This capability suggests that the algorithm can facilitate real-world applications of spike detection without the need for extensive data to train a model. Instead, it leverages only a small amount of initially available data, making it highly practical for deployment in clinical settings.

One of the limitations of this study lies in the limited data population of epileptic patients and the reliance on a single signal reviewer for marking epileptic spikes. It would be beneficial to involve multiple signal raters to enhance the robustness and reliability of the statistical distribution used to establish thresholds for spike detection. By incorporating input from multiple reviewers and considering the intersection of their marked events across time samples, biases toward specific spike patterns can be reduced. This approach can improve sensitivity and specificity in spike detection, as it captures a more comprehensive range of epileptic activity patterns. Therefore, future studies should consider involving a larger pool of signal reviewers, a longer length of ECoG data, and more patient data to address this limitation and enhance the accuracy of spike detection algorithms.

## CONCLUSION

The study introduces a novel automated method for real-time detection of epileptic discharges to support evaluation for epilepsy surgery. The method employs adaptive thresholding based on the statistical characteristics of signal envelopes. This approach also incorporates spikes' spatial information to capture spike propagation patterns. The automated real-time epileptic discharge detection system can potentially reduce the duration of long-term ECoG monitoring in the EMU.

Furthermore, it could be utilized in intraoperative monitoring to assist neurosurgeons in localizing the SOZ, thus enhancing surgical precision and patient outcomes.

## REFERENCES

- [1] Vidyaratne, L. S., & Iftekharuddin, K. M. (2017). Real-time epileptic seizure detection using EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11), 2146-2156.
- [2] Kwan, P., & Brodie, M. J. (2000). Early identification of refractory epilepsy. *New England Journal of Medicine*, 342(5), 314-319.
- [3] Lüders, H. O., Najm, I., Nair, D., Widdess-Walsh, P., & Bingman, W. (2006). The epileptogenic zone: general principles. *Epileptic disorders*, 8, S1-S9.
- [4] Van Gompel, J. J., Rubio, J., Cascino, G. D., Worrell, G. A., & Meyer, F. B. (2009). Electrocorticography-guided resection of temporal cavernoma: is electrocorticography warranted and does it alter the surgical approach?. *Journal of neurosurgery*, 110(6), 1179-1185.
- [5] Demuru, M., Kalitzin, S., Zweiphenning, W., van Blooij, D., van't Klooster, M., Van Eijnden, P., ... & RESPECT Group Braun Kees 2 Ferrier Cyrille 2 Gebbink Tineke 2 Gosselaar Peter 2 Huiskamp Geertjan 2 van Klink Nicole 2 Ophorst Janine 2 van Rijen Peter 2 van der Salm Sandra 2 Winter Anouk Velders-de 2. (2020). The value of intra-operative electrographic biomarkers for tailoring during epilepsy surgery: from group-level to patient-level analysis. *Scientific Reports*, 10(1), 14654.
- [6] Rosenow, F., & Lüders, H. (2001). Presurgical evaluation of epilepsy. *Brain*, 124(9), 1683-1700.
- [7] Asano, E., Muzik, O., Shah, A., Juhász, C., Chugani, D. C., Sood, S., ... & Chugani, H. T. (2003). Quantitative interictal subdural EEG analyses in children with neocortical epilepsy. *Epilepsia*, 44(3), 425-434.
- [8] Marsh, E. D., Peltzer, B., Brown III, M. W., Wusthoff, C., Storm Jr, P. B., Litt, B., & Porter, B. E. (2010). Interictal EEG spikes identify the region of electrographic seizure onset in some, but not all, pediatric epilepsy patients. *Epilepsia*, 51(4), 592-601.
- [9] Kim, D. W., Kim, H. K., Lee, S. K., Chu, K., & Chung, C. K. (2010). Extent of neocortical resection and surgical outcome of epilepsy: intracranial EEG analysis. *Epilepsia*, 51(6), 1010-1017.
- [10] Dworetzky, B. A., & Reinsberger, C. (2011). The role of the interictal EEG in selecting candidates for resective epilepsy surgery. *Epilepsy & Behavior*, 20(2), 167-171.
- [11] Jacobs, J., LeVan, P., Chander, R., Hall, J., Dubeau, F., & Gotman, J. (2008). Interictal high-frequency oscillations (80–500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain. *Epilepsia*, 49(11), 1893-1907.
- [12] Heers, M., Böttcher, S., Kalina, A., Katletz, S., Altenmüller, D. M., Baroumand, A. G., ... & Dümpelmann, M. (2022). Detection of interictal epileptiform discharges in an extended scalp EEG array and high-density EEG—a prospective multicenter study. *Epilepsia*, 63(7), 1619-1629.
- [13] Janca, R., Jezdik, P., Cmejla, R., Tomasek, M., Worrell, G. A., Stead, M., ... & Marusic, P. (2015). Detection of interictal epileptiform discharges using signal envelope distribution modelling: application to epileptic and non-epileptic intracranial recordings. *Brain topography*, 28, 172-183.
- [14] Hirsch, L. J., Fong, M. W., Leitingner, M., LaRoche, S. M., Beniczky, S., Abend, N. S., ... & Gaspard, N. (2021). American clinical neurophysiology society's standardized critical care EEG terminology: 2021 version. *Journal of Clinical Neurophysiology*, 38(1), 1-29.
- [15] Brown III, M. W., Porter, B. E., Dlugos, D. J., Keating, J., Gardner, A. B., Storm Jr, P. B., & Marsh, E. D. (2007). Comparison of novel computer detectors and human performance for spike detection in intracranial EEG. *Clinical neurophysiology*, 118(8), 1744-1752.
- [16] Conrad, E. C., Tomlinson, S. B., Wong, J. N., Oechel, K. F., Shinohara, R. T., Litt, B., ... & Marsh, E. D. (2020). Spatial distribution of interictal spikes fluctuates over time and localizes seizure onset. *Brain*, 143(2), 554-569.

# COMPARISON OF CNN-BASED EEG CLASSIFICATION IN SENSOR AND SOURCE SPACE

M. Maurer<sup>1</sup>, D. Baumgarten<sup>1,2</sup>, J. Vorwerk<sup>1</sup>

<sup>1</sup> Institute of Electrical and Biomedical Engineering, UMIT TIROL, Hall in Tirol, Austria

<sup>2</sup> Institute of Mechatronics, University of Innsbruck, Innsbruck, Austria

E-mail: magdalena.maurer@umit-tirol.at

**ABSTRACT:** Electroencephalography (EEG) is a popular tool in brain-computer interfacing (BCI), due to its unique time resolution and simplicity of application. For the design of BCIs, rapid and accurate classification algorithms are needed to classify the brain state correctly in real-time. Recent technological advancements facilitate the use of novel methods for signal processing and analysis such as real-time source estimation and classification via deep learning approaches. In this work a previously established convolutional neural network (CNN) architecture, the EEGNet, was applied to a publicly available motor imagery EEG dataset for classification of sensor measurements and source estimates that were computed with three different inverse approaches. Both for sensor signals and source estimates similar classification accuracies as in the literature could be achieved. However, no significant difference in performance between sensor and source space analysis was observed.

## INTRODUCTION

In the field of brain-computer interfaces (BCI), the electroencephalogram (EEG) remains one of the most popular measurement tools for acquiring brain signals, due to its many positive characteristics for the BCI use case. BCIs offer the possibility to interact with an external machine such as a computer or a rehabilitation robot based primarily on modulations of brain activity. With EEG, electrical brain activity can be measured non-invasively on the surface of the scalp. EEG is comparatively cheap and easy to use, and it is portable. In addition, due to its high temporal resolution in the millisecond range, the EEG allows for measuring ongoing brain activity basically in real-time. EEG-based BCI systems are therefore used in a great variety of settings such as in rehabilitation, health and attention monitoring, entertainment and skill improvement [1, 2]. To allow for the control of an external device it is necessary to analyze data and derive a control command in real time. Various measurement paradigms for EEG-based BCIs exist to elicit specific EEG patterns from which control signals can be generated. Common paradigms include motor imagery (MI), visually evoked potentials or error-related potentials [2].

A control signal is extracted by detecting suitable features to distinguish the respective signal patterns and then classifying them. To train the classifier in the best possible way, a time-consuming recording of training data and an offline training phase are usually required. Since the goal is to provide BCI users with good, precise, and intuitive control of the external device, achieving high classification accuracies in real-time is essential [1, 3]. One of the greatest difficulties in the real-time derivation of control signals lies in the low signal-to-noise ratio (SNR) of single-trial evaluations. Further challenges are the large variability of signals between subjects, but also within subjects from session to session, as well as the large amount of training data required to train a classifier [3].

One frequently used BCI paradigm, especially in the field of rehabilitation, is based on Motor Imagery (MI), where the user has to imagine a movement without actually performing it. MI-based BCIs have been used for example to control a robotic rehabilitation device in motor training of the upper extremity after stroke [4], to trigger functional electrical stimulation (FES) in muscles when using an orthosis after spinal cord injury [5], or to elicit electrical stimulation with transcranial magnetic stimulation (TMS) during neurorehabilitation to enhance cortex excitability [6]. Typically, control commands are generated from these signal types using classification methods such as linear discriminant analysis (LDA) or support vector machines (SVM) [3, 7].

Recently, the use of convolutional neural networks (CNN) has been increasingly investigated for the classification of brain activity. It was stated that CNNs work reasonably well even in the presence of artifacts and noisy data. Their application is therefore interesting to tackle the challenge of poor SNR in real-time analyses. In addition, CNNs offer the possibility of being used in transfer learning, which can help to reduce time-consuming data collection from individuals and lengthy training cycles. Here, the influence of possible differences in the measurement setup and electrode configurations between recordings must be considered, when data from multiple measurements and various individuals are combined [3, 8, 9].

The required classification steps for BCIs are usually carried out directly with sensor space data, i.e., with measurements obtained directly from the scalp surface.

However, EEG data can be analyzed not only in sensor space but also in source space. In general, source modeling is widely used across many fields of EEG analysis. This involves transforming the superficially measured signals into the source space using inverse operations. The signals derived at the scalp are influenced and distorted on their way from the brain to the head surface by the different tissue layers and their diverse electrical properties. In order to counteract this so-called volume conduction effect, an attempt is made to reconstruct the underlying activity of electrical sources in the brain using the signals measured externally on the scalp. Popular inverse methods for the transformation into the source space are minimum norm estimation (MNE), weighted MNE, and beamformers [10–13]. While source estimation has already been used occasionally in the context of BCIs, the combination with CNN offers a new perspective.

The aim of this preliminary work is to combine CNN-based classification of EEG data with three different methods of source analysis to compare the classification performance between sensor space and source space. For this purpose, a preestablished CNN architecture, the EEGNet [9] is applied to a publicly available MI-based EEG dataset from the Graz BCI competition IV [14]. CNN-based classification is conducted both directly on sensor measurements and source estimates generated with three different inverse approaches: Minimum Norm Estimation (MNE), weighted MNE (wMNE), and Beamformer.

## METHODS

During this work, all analyses were performed using the Graz BCI competition IV dataset 2A [14]. This dataset contains EEG measurements from nine healthy subjects, who performed four different MI tasks. The four classes include cue-based imagination of left-hand, right-hand, feet, and tongue movement. There were two sessions recorded on different days each containing 288 trials in total, yielding 72 trials per class. EEG measurements were derived from 22 electrodes arranged according to the international 10-20 system with a sampling frequency of 250Hz. A more detailed description of this dataset can be found in [14].

In the context of this work classification of only two classes was desired, therefore only data from left-hand and right-hand MI were used. For each trial, ten overlapping 2s MI-epochs were extracted from the data spanning the MI period from 0.5s to 3.5s after cue onset, as indicated in [14].

To perform classification in sensor space, EEG measurements from the BCI dataset were provided as input to the EEGNet-based CNN, and the performance was evaluated using a 10-fold cross-validation. Before classification in source space, three different representations of source activity were computed using MNE, wMNE and beamformer. Then, classification was performed on each source estimation using CNN again in combination with a 10-fold cross-validation.

The three different inverse approaches for source transformation were computed using MATLAB version R2023a (The MathWorks Inc., <https://www.mathworks.com>). As individual head geometries were not available, the required lead field matrix for inverse computation was taken from the New York Head [15], which is a standardized finite element human head model that provides lead fields for 231 electrode locations and approximately 75,000 source locations. The scalar lead field matrix for fixed dipole orientation orthogonal to the cortex surface was used. First, the 22 channels that were also used in the BCI dataset were selected and the entire head model was down-sampled using cortex2K provided by the New York Head to reduce the number of sources. Next, regions of interest (ROI) were selected including post- and precentral gyrus, central sulcus and paracentral lobule according to the Destrieux atlas [16]. In the end, a total of 122 source locations remained for estimation. For ease of comparison, the same number of sources was used for all three inverse operations. Prior to classification and source estimation EEG data were preprocessed with common average referencing (CAR) and the lead field matrix was restricted to the 22 channels from the BCI dataset and rereferenced. No further preprocessing or artifact removal was performed.

*Source estimation:* In general, the relationship between the cortical sources and the resulting signals at sensors on the head surface can be described by the formulation in equation (1), where  $y$  denotes sensor signals,  $x$  denotes the underlying source activity,  $L$  denotes the lead field matrix, and  $n$  represents noise. The underlying sources  $\hat{x}$  can be reconstructed from measurements  $y$  following equation (2), where  $M$  describes the inverse operator that maps the sensor signals to the source space.

$$y = Lx + n \quad (1)$$

$$\hat{x} = My \quad (2)$$

The first and simplest approach for transforming the sensor measurements to the source space is MNE which uses Tikhonov regularization and minimizes the expression described in equation (3).

$$\min_x \|Lx - y\|^2 + \lambda \|Wx\|^2 \quad (3)$$

For the simple case of MNE, the weight matrix  $W$  corresponds to the identity matrix  $I$ . By solving this minimization problem, the inverse mapping operator  $M_{\text{MNE}}$  can be computed as described in equation (4). Here,  $\lambda$  denotes the regularization parameter, which was estimated from the SNR, and  $I$  corresponds to the identity matrix.

$$M_{\text{MNE}} = L^T(LL^T + \lambda I)^{-1} \quad (4)$$

A more refined version of MNE is weighted MNE where additional weights  $W$  are introduced to compensate for the preference of weak superficial sources. The resulting

mapping operator  $M_{wMNE}$  is presented in equation (5). The weights  $W$  were computed based on the norm of the columns of the lead field matrix  $L$ .

$$M_{wMNE} = W^{-1}L^T(LW^{-1}L^T + \lambda I)^{-1} \quad (5)$$

Finally, another approach for source estimation is provided by the linearly constrained minimum variance (LCMV) beamformer. In beamforming, a set of weights is computed for each predefined source location constructing a spatial filter that relates each sensor's contribution to each source estimate. These weights serve as the inverse operator  $M_{LCMV}$  and the computation of its components  $m_{LCMV}(r)$  for the scalar case is described by equation (6). Here,  $R$  denotes the data covariance matrix and  $l(r)$  the column of the lead field matrix for a dipole of fixed orientation at location  $r$ .

$$m_{LCMV}(r) = (l^T(r)R^{-1}l(r))^{-1}l^T(r)R^{-1} \quad (6)$$

The resulting three mapping operators,  $M_{MNE}$ ,  $M_{wMNE}$ , and  $M_{LCMV}$  were applied to EEG measurements from the BCI dataset according to equation (2), to compute three different time course representations in source space.

*Classification with CNN:* Subsequently, feature extraction and classification were performed with an EEGNet-based [9] CNN on four different signal modalities: sensor signals, source estimates generated with MNE, source estimates generated with wMNE, and source estimates from LCMV beamformer. The EEGNet is a compact CNN architecture consisting of two blocks, where at first a 2D temporal convolution is performed to learn frequency-specific features followed by depthwise convolution to learn spatial filters. In the second block 2D separable convolution is performed to summarize the individual features. In the end, classification is performed using a softmax algorithm, however, in the context of this work classification based on sigmoid function was used. As optimization algorithm the Adam optimizer was used. A more detailed explanation of the network can be found in [9]. Prior to training the CNN, a hyperparameter search was performed to find the most suitable model parameters. Based on that eight temporal and two spatial filters were implemented, the learning rate was set to 0.001, batch size was set to 32, and the number of epochs was set to 12 both for sensor and source space data. The EEGNet-based CNN was implemented in Python using TensorFlow (<https://www.tensorflow.org/>), Keras (<https://keras.io/>), and SciKit-Learn (<https://scikit-learn.org/stable/>) libraries. In the end, classification accuracies of all four signal modalities were compared via the means of Kruskal-Wallis-Test, to determine whether there is a statistically significant difference between the methods.

## RESULTS

For the classification in sensor space, an average classification accuracy of  $79.25 \pm 13.90$  % for all nine subjects was found. Individual classification results for

each subject for sensor space data are presented in Tab. 1, where the second column shows testing accuracies highlighted in bold, and column three shows training accuracies. The lowest accuracy in sensor space was found for subject 02 with  $56.87 \pm 2.92$  %, and the highest accuracy was obtained for subject 08 with  $96.04 \pm 1.35$  %.

Table 1: Classification accuracies in sensor space

Subject #	Test accuracies in %	Training accuracies in %
01	<b>85.49 <math>\pm</math> 2.71</b>	88.26 $\pm$ 7.95
02	<b>56.87 <math>\pm</math> 2.92</b>	79.86 $\pm$ 7.39
03	<b>90.21 <math>\pm</math> 3.10</b>	98.19 $\pm$ 1.54
04	<b>65.42 <math>\pm</math> 5.65</b>	85.76 $\pm$ 4.32
05	<b>73.68 <math>\pm</math> 5.06</b>	88.26 $\pm$ 1.72
06	<b>67.29 <math>\pm</math> 4.09</b>	87.92 $\pm$ 2.97
07	<b>84.44 <math>\pm</math> 2.41</b>	93.06 $\pm$ 2.51
08	<b>96.04 <math>\pm</math> 1.35</b>	99.03 $\pm$ 0.82
09	<b>93.82 <math>\pm</math> 0.89</b>	93.06 $\pm$ 3.29
AVG	<b>79.25 <math>\pm</math> 13.90</b>	90.38 $\pm$ 6.09

The classification accuracies for source estimates generated with MNE and wMNE can be seen in Tab. 2 and Tab. 3, respectively. Again, obtained accuracies on the test data are presented in column two and training accuracies are presented in column three.

Table 2: Classification accuracies in source space generated with MNE

Subject #	Test accuracies in %	Training accuracies in %
01	<b>71.39 <math>\pm</math> 8.87</b>	73.47 $\pm$ 12.56
02	<b>50.21 <math>\pm</math> 2.62</b>	78.06 $\pm$ 8.33
03	<b>83.68 <math>\pm</math> 13.13</b>	91.46 $\pm$ 12.36
04	<b>51.11 <math>\pm</math> 2.08</b>	56.67 $\pm$ 9.33
05	<b>74.44 <math>\pm</math> 6.24</b>	83.61 $\pm$ 6.60
06	<b>52.71 <math>\pm</math> 2.33</b>	62.36 $\pm$ 15.05
07	<b>69.31 <math>\pm</math> 10.25</b>	86.94 $\pm$ 12.30
08	<b>83.33 <math>\pm</math> 11.95</b>	94.72 $\pm$ 5.44
09	<b>88.19 <math>\pm</math> 3.91</b>	82.85 $\pm$ 11.03
AVG	<b>69.37 <math>\pm</math> 14.84</b>	78.90 $\pm$ 12.79

Table 3: Classification accuracies in source space generated with wMNE

Subject #	Test accuracies in %	Training accuracies in %
01	<b>65.07 <math>\pm</math> 9.47</b>	86.46 $\pm$ 4.30
02	<b>51.67 <math>\pm</math> 2.95</b>	77.99 $\pm$ 9.62
03	<b>93.40 <math>\pm</math> 0.94</b>	97.64 $\pm$ 2.48
04	<b>50.00 <math>\pm</math> 2.07</b>	81.53 $\pm$ 4.53
05	<b>58.12 <math>\pm</math> 5.37</b>	81.46 $\pm$ 7.74
06	<b>60.42 <math>\pm</math> 6.15</b>	83.68 $\pm$ 5.65
07	<b>67.50 <math>\pm</math> 6.40</b>	83.68 $\pm$ 10.64
08	<b>93.06 <math>\pm</math> 1.64</b>	99.10 $\pm$ 0.98
09	<b>90.21 <math>\pm</math> 3.30</b>	97.99 $\pm$ 1.48
AVG	<b>69.94 <math>\pm</math> 17.63</b>	87.72 $\pm$ 8.21

The average classification accuracy over all subjects for the MNE-based source space was  $69.37 \pm 14.84$  % and  $69.94 \pm 17.63$  % for wMNE. Here, the minimal accuracy

for MNE was found for subject 02 with  $50.21 \pm 2.62$  %, and in wMNE for subject 04 with  $50.00 \pm 2.07$  %. Similarly, the maximal accuracy in MNE source space was obtained for subject 09 with  $88.19 \pm 3.91$  % and in wMNE source space for subject 03 with  $93.40 \pm 0.94$  %. In Tab. 4 the classification accuracies for sources reconstructed using LCMV beamformer are provided. The average accuracy achieved with this inverse method was  $78.14 \pm 13.12$  %. The lowest accuracy was achieved for subject 02 with  $55.83 \pm 2.60$  % and maximum accuracy was achieved for subject 08 with  $95.56 \pm 1.10$  %.

Table 4: Classification accuracies in source space generated with LCMV beamformer

Subject #	Test accuracies in %	Training accuracies in %
01	<b><math>85.83 \pm 3.63</math></b>	$95.35 \pm 3.09$
02	<b><math>55.83 \pm 2.60</math></b>	$89.17 \pm 5.00$
03	<b><math>85.07 \pm 6.87</math></b>	$98.47 \pm 2.29$
04	<b><math>64.51 \pm 6.28</math></b>	$87.71 \pm 7.34$
05	<b><math>81.46 \pm 4.31</math></b>	$91.11 \pm 2.61$
06	<b><math>65.00 \pm 3.50</math></b>	$90.69 \pm 2.89$
07	<b><math>83.26 \pm 5.86</math></b>	$94.58 \pm 4.62$
08	<b><math>95.56 \pm 1.10</math></b>	$99.51 \pm 0.74$
09	<b><math>86.74 \pm 3.72</math></b>	$97.99 \pm 1.62$
AVG	<b><math>78.14 \pm 13.12</math></b>	$93.84 \pm 4.34$

The distribution of classification accuracies across all four modalities for all nine subjects is depicted in Fig. 1. The comparison of all four methods with the Kruskal-

Wallis-Test showed that there is no significant difference (Chi square = 2.93,  $p = 0.40$ ,  $df = 3$ ) in classification performance between the different approaches.

## DISCUSSION

In this study, the classification performance of a previously introduced CNN architecture, the EEGNet, was assessed for four different input datasets generated from a publicly available MI BCI dataset. The first input provided were unprocessed sensor measurements and the other three modalities were corresponding source estimates computed with three distinct inverse methods. Considering the classification accuracies for both sensor and source space data, the proposed EEGNet-based CNN was able to achieve a comparable performance to other standard classification methods from previous studies that were also evaluated on the BCI competition IV dataset [9, 17–19], with a relatively small number of epochs. The presented network yielded an average classification accuracy of about 79% on sensor space data and thereby shows slightly higher classification accuracies than e.g. Schirrmeister et al. who reported an average accuracy of about 74% with their ShallowConvNet [18] or Kar et al. who achieved an average accuracy of about 70% with their CNN-model [19]. It has to be considered, however, that the present work discriminated only two classes whereas most other studies performed classification of all four tasks contained in the dataset.

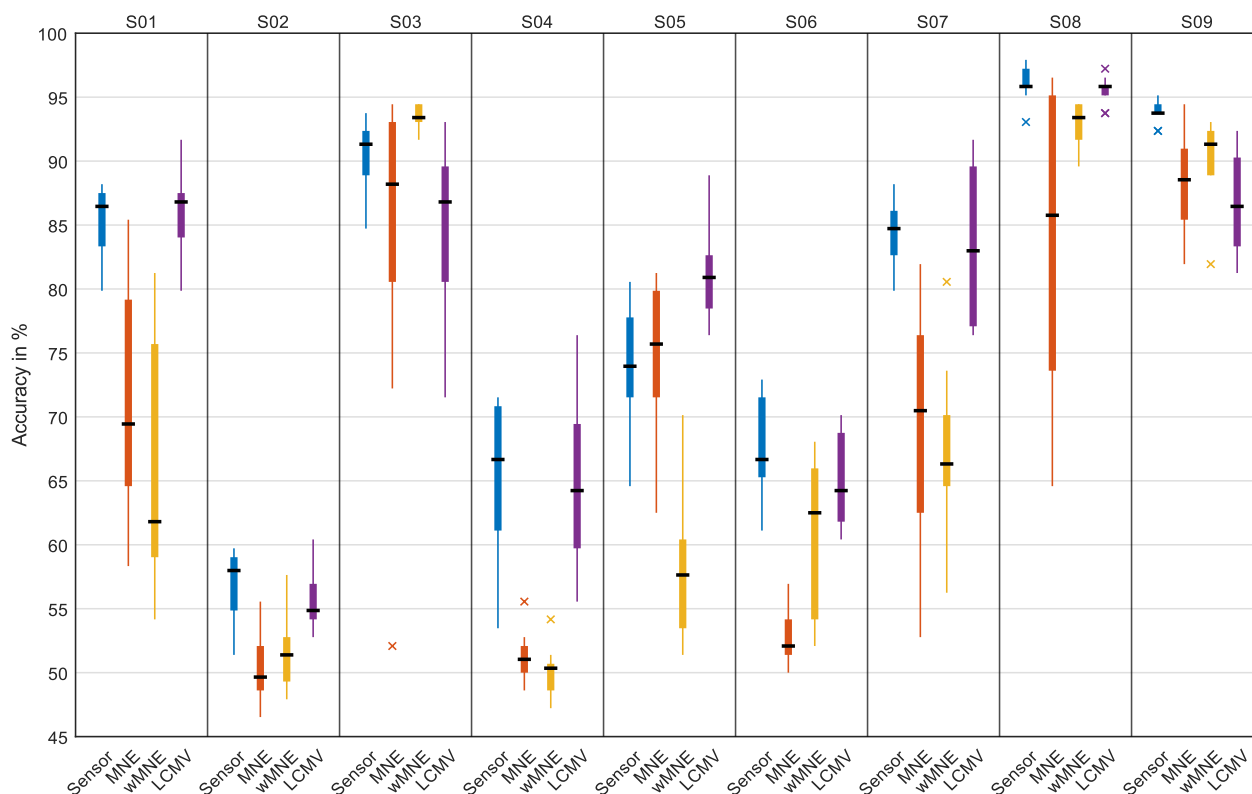


Figure 1: Distribution of classification accuracies across all four modalities (sensor measurements in blue, MNE estimates in orange, wMNE estimates in yellow, beamformer estimates in purple) for all nine subjects. For better visualization the y-axis was cut to contain only accuracy values between 45 and 100 %.

To keep conditions across models as similar as possible the same hyperparameters, including number of epochs, were used for all four modalities. Overall, the training accuracies in all four cases were quite high while still matching the general trend, meaning that subjects with lower testing accuracies also had lower training accuracies. In some cases, like for example with subject 02, the training accuracies were considerably higher than the accuracies achieved in testing, however, this is likely due to a higher intra-subject variability between the first and second recording session rather than due to overfitting. When comparing the classification results from sensor space to those achieved with source estimates in Fig. 1, it is evident that the accuracy values do not differ noticeably between the different modalities, even though it seems that source estimation with MNE and wMNE yielded slightly lower accuracies for most of the subjects. The Kruskal-Wallis-Test confirms that there is no significant difference between sensor space and any of the source space data. These findings are in contrast to other studies that found an improvement in BCI performance when applying source analysis prior to classification, although these studies did not use CNNs for classification [11, 12]. The results of the present work support the idea that CNNs are good at finding optimal separation criteria despite noisy data, as the compensation for the volume conduction effect via source estimation did not seem to significantly affect the classification performance. One reason for the comparatively weaker classification performance of source estimates could be the rather small number of measurement channels that are provided with the BCI dataset. In this work, 122 source estimates were generated using a standardized head model and only 22-channel EEG measurements with poor head coverage. No individual head geometries and exact electrode positions were available. It was previously shown, however, that an increased electrode density and good head coverage allow for improved source reconstruction [20, 21]. Other studies that found an improvement in BCI performance in combination with source estimation used a larger number of sensors distributed over the entire head [11, 12]. Results of source estimation could further be improved by additional preprocessing such as data whitening with noise covariances, regional clustering for optimally reducing the source space, or introducing spatial normalization strategies to beamformers [12, 13, 22, 23]. So, with measurements providing better electrode configurations spanning the entire head surface, computation of more accurate source estimates and thereby an improvement of classification performance might be feasible.

Nonetheless, as comparable classification accuracies could be achieved with source estimates, future implementations using CNN and source reconstruction for transfer learning in BCIs are promising. The transformation of sensor measurements to the source space offers the possibility to jointly process data of several individuals and different recording sessions in a

common signal space, as the effect of changes in electrode configuration can be circumvented.

## CONCLUSION

Accurate real-time classification of EEG data is essential to provide users with precise and intuitive BCI control. In this study, it was shown that CNNs can achieve comparable classification performances based on sensor and source space data respectively. Future investigations will focus on the real-time applicability of CNNs and transfer learning for BCI systems will be explored. The reliable classification of source estimates as shown by this work is promising for the application in transfer learning. Furthermore, CNN-based classification of source estimates will also be performed on own recorded data during MI and arithmetic learning tasks, to explore whether alternative tasks can be discriminated from EEG measurements and could be used for providing neurofeedback.

## ACKNOWLEDGEMENTS

This research was funded in whole or in part by the Austrian Science Fund (FWF), Grant DOI 10.55776/P35949, and the Tyrolian Science Fund (TWF), project F.45065. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

## REFERENCES

- [1] K. Värbu, N. Muhammad, and Y. Muhammad, "Past, present, and future of EEG-based BCI applications," *Sensors*, vol. 22, no. 9, p. 3331, 2022.
- [2] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain-computer interface paradigms," *Journal of Neural Engineering*, vol. 16, no. 1, p. 11001, 2019.
- [3] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 31005, 2018.
- [4] K. K. Ang *et al.*, "A clinical study of motor imagery-based brain-computer interface for upper limb robotic rehabilitation," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 5981–5984.
- [5] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "'Thought'-control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neuroscience letters*, vol. 351, no. 1, pp. 33–36, 2003.
- [6] A. Gharabaghi *et al.*, "Coupling brain-machine interfaces with cortical stimulation for brain-state dependent stimulation: enhancing motor cortex excitability for neurorehabilitation," *Frontiers in Human Neuroscience*, vol. 8, p. 122, 2014.

- [7] P. Wierzgała, D. Zapala, G. M. Wojcik, and J. Masiak, "Most popular signal processing methods in motor-imagery BCI: a review and meta-analysis," *Frontiers in Neuroinformatics*, vol. 12, p. 78, 2018.
- [8] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, vol. 16, no. 3, p. 31001, 2019.
- [9] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 56013, 2018.
- [10] R. Grech *et al.*, "Review on solving the inverse problem in EEG source analysis," *Journal of Neuroengineering and Rehabilitation*, vol. 5, no. 1, pp. 1–33, 2008.
- [11] B. J. Edelman, B. Baxter, and B. He, "EEG source imaging enhances the decoding of complex right-hand motor imagery tasks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 4–14, 2015.
- [12] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss, "Beamforming in noninvasive brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1209–1219, 2009.
- [13] B. U. Westner *et al.*, "A unified view on beamformers for M/EEG source reconstruction," *NeuroImage*, vol. 246, p. 118789, 2022, doi: 10.1016/j.neuroimage.2021.118789.
- [14] C. Brunner, R. Leeb, and G. Müller-Putz IV, "BCI Competition IV Dataset 2a: 4-Class Motor Imagery," *Graz University of Technology: Graz, Austria*, 2008.
- [15] Y. Huang, L. C. Parra, and S. Haufe, "The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting," *NeuroImage*, vol. 140, pp. 150–162, 2016.
- [16] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *NeuroImage*, vol. 53, no. 1, pp. 1–15, 2010.
- [17] I. Dolzhikova, B. Abibullaev, R. Sameni, and A. Zollanvari, "An Ensemble CNN for Subject-Independent Classification of Motor Imagery-based EEG," in *IEEE 2021 – 2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 319–324.
- [18] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017, doi: 10.1002/hbm.23730.
- [19] A. Kar, S. Bera, S. K. P. Karri, S. Ghosh, M. Mahadevappa, and D. Sheet, "A deep convolutional neural network based classification of multi-class motor imagery with improved generalization," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 5085–5088.
- [20] J. Song *et al.*, "EEG source localization: Sensor density and head surface coverage," *Journal of Neuroscience Methods*, vol. 256, pp. 9–21, 2015, doi: 10.1016/j.jneumeth.2015.08.015.
- [21] G. Lantz, R. G. de Peralta, L. Spinelli, M. Seeck, and C. M. Michel, "Epileptic source localization with high density EEG: how many electrodes are needed?," *Clinical Neurophysiology*, vol. 114, no. 1, pp. 63–69, 2003.
- [22] F.-H. Lin, T. Witzel, S. P. Ahlfors, S. M. Stufflebeam, J. W. Belliveau, and M. S. Hämäläinen, "Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates," *NeuroImage*, vol. 31, no. 1, pp. 160–171, 2006, doi: 10.1016/j.neuroimage.2005.11.054.
- [23] C. Dinh *et al.*, "Real-Time MEG Source Localization Using Regional Clustering," *Brain Topography*, vol. 28, no. 6, pp. 771–784, 2015, doi: 10.1007/s10548-015-0431-9.

# IMPACT OF MENTAL FATIGUE ON REGAINING MOTOR FUNCTIONALITY: A PRELIMINARY EEG STUDY ON STROKE SURVIVORS

Kaniska Samanta<sup>1\*</sup>, KongFatt Wong-Lin<sup>1</sup>, Girijesh Prasad<sup>1</sup>, Saugat Bhattacharyya<sup>1</sup>

<sup>1</sup>Intelligent Systems Research Centre, Ulster University, Derry Londonderry, Northern Ireland, United Kingdom

E-mail: Samanta-k@ulster.ac.uk

**ABSTRACT:** In the past few decades, research has demonstrated that brain-computer interface (BCI) based neurorehabilitation for stroke survivors can enhance the re-learning of lost motor functionality better than traditional physiotherapy involving professional physiotherapists. Though BCI-aided systems have several advantages over traditional rehabilitation methods, one of the major shortcomings of such intervention is its inability to recognize the relevant motor activity of the brain when the user gets mentally fatigued, which eventually causes the deterioration of the BCI performance. In this paper, a preliminary EEG study on stroke survivors has been reported on how mental fatigue can potentially hinder the enhancement of motor re-learning and elongate the rehab process. From the study, it has been inferred that objective measurement of mental fatigue is essential to prevent any subjective bias, and the rehabilitation paradigm should be adaptive to the participants' mental status to optimize the rehab outcomes.

## INTRODUCTION

Stroke is a severe and potentially life-threatening medical condition that occurs when the blood flow to a specific brain area is interrupted. The majority of the survivors face minor to major motor disability depending upon the severity of stroke [1]. Hence, it has become necessary to develop efficient assistive neurotechnology for stroke survivors to regain the motor functionality of their impaired limb. A previous study reported that traditional physiotherapy involving professional physiotherapists has less impact on regaining motor functionality [2]. One possible reason for this observation could be that the patient is not actively engaged in relearning the lost motor functions during this type of motor rehabilitation program [3]. The participant's active engagement is crucial in BCI-aided intervention for stroke rehabilitation to restore motor function effectively and efficiently.

Brain-computer interface (BCI) is an emerging technology that is increasingly becoming important to establish effective communication pathways between the brain and computer-controlled devices (e.g., robotic exoskeleton) by using invasive (e.g., electrocorticography) or

non-invasive (e.g., electroencephalography, magnetoencephalography) neuroimaging modalities [4, 5]. Combining motor imagery (MI) or the imagined movement of a limb with BCI for neurorehabilitation enables users to control an exoskeleton attached to their impaired limb [6, 7]. Existing studies show BCI-aided neurorehabilitation exhibits promising improvements in restoring motor functionality compared to traditional physiotherapy and other assistive technologies [8].

Though the BCI-aided neurorehabilitation framework shows promising results for the rehabilitation of stroke survivors, there is a need to develop a reliable and efficient assistive system for neurorehabilitation to account for the variability in brain activity over time because brain activity changes its characteristics due to changes in different mental states (e.g., mental fatigue, boredom, etc.) while carrying out the same activity. In previous studies [9, 10], it has been observed that a shift in mental state, especially the induction of mental fatigue, negatively impacts the performance of BCI-based rehabilitation. Slower motor movement during a mentally fatigued state has been reported in [11]. While the impact of mental fatigue on motor performance has been explored in the existing literature, the impact of fatigue on the enhancement of motor learning has been less studied. Branscheidt et al., in [12], reported a long-lasting detrimental impact of muscle fatigue on long-term motor skill learning. Still, the research did not include the possible effect of mental fatigue caused by the prolonged cognitive load imposed during the experiment. Persistence of mental fatigue on motor control has been reported for healthy individuals in [13]. Moghani et al. have reported that mental fatigue causes a loss of self-controlled feedback in motor learning for healthy people [14].

Though the existing research has explored the potential impact and persistence of mental fatigue on long-term motor learning, there is a significant knowledge gap that correlates the impact of mental fatigue with regaining lost motor functionality for actual patients with motor impairment.

The study of mental fatigue and its effect on motor learning is mainly limited to healthy individuals, and to the best of the authors' knowledge, no research has been

done so far to study the impact of mental fatigue on the re-learning of motor activities in stroke survivors. In this present contribution, a novel EEG-based preliminary study on stroke survivors has been reported, aiming to establish a relationship between motor learning and mental fatigue.

## MATERIALS AND METHODS

### *Dataset Description:*

This contribution used an EEG dataset acquired in [8] during a stroke rehabilitation program. The original data contains EEG recordings from 5 chronic stroke (ischemic) patients suffering from hemiparesis. The mean time since the first occurrence of stroke was  $21.8 \pm 4.49$ , within the range of 17 to 28 months. As revealed by the testimonials of the patients, their motor functionality stopped improving for the last one year [8]. The dataset contains 12 channel EEG recordings from F3, F4, FC3, FCz, FC4, C3, C4, CP3, CPz, CP4, P3, and P4 brain regions. The data were recorded with a g.USBamp (g.tec, Graz, Austria) biosignal amplifier, along with active ring electrodes (g.LADYbird having sintered Ag/AgCl crown) attached to the EEG cap (g.GAMMAcap). The signal was sampled at 512 Hz, and initially, a band-pass filter with cut-off frequencies of 0.1 Hz to 100 Hz was applied with a notch filter at 50 Hz to avoid the power-line noise. The participants were instructed to perform grasping attempts with their left and right hands, as shown on a computer screen. The rehabilitation program consists of up to 12 BCI-controlled hand-exoskeleton therapy sessions for each participant, spanning over 5 weeks with 2-3 sessions per week. Each session consists of 5 runs (the first two for calibration of the BCI exoskeleton and the last three with real-time feedback). Each run consists of 40 trials (20 trials for the left hand and 20 trials for the right hand), and each trial lasts for 8 seconds with a random 2s to 3s interval as the inter-trial interval between the two consecutive trials, which makes one run roughly 7.5 minutes. Short inter-run breaks of around 5–6 minutes were provided for the participants to rest. From the 5 volunteers, only 3 patients who completed the entire rehabilitation program (with 12 sessions) were selected for the present analysis.

The rehabilitation outcomes were measured every week in terms of standard motor recovery measures: Action Research Arm Test (ARAT) and Grip Strength (GS) (in kg). The ARAT measures grasp (score: 0–18), grip (score: 0–12), pinch (score: 0–18), and gross movements (score: 0–9). Thus, the total range of ARAT is 0–57. In this paper, only the ARAT score has been used for analysis, as it already includes measurements related to grip.

Participants were instructed to report their own evaluation of fatigue and motivation level before and after each session using the visual analogue scale from 0 to 10.

### *Data Pre-processing:*

At first, the continuous data were epoched based on the trigger information. Then, the ICA-based automated arte-

fact rejection technique was applied to filter ocular and other common muscle artefacts (using the 'EEGLAB' extension 'ICLabel' in MATLAB R2022b), and cleaned data were visually inspected to check the quality and manually rejected any remaining artefacts. A band-pass filter with cut-off frequencies of 0.1 and 50 Hz was applied, as high-frequency noise had been observed after artefact correction.

The epoched and pre-processed data had been decomposed into different EEG rhythms, namely, Delta (0.1–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–50 Hz).

### *Methodology:*

In the existing literature, it has been well established that event-related desynchronization/synchronization (ERD/S) [15], a relative power decrease/increase of EEG in a specific frequency band, is highly associated with physical motor execution and mental motor imagery [16]. The same study also reported that the ability to generate ERD/S is highly subjective and involves sufficient neurofeedback training [16].

In this present contribution, ERD/S have been calculated for each EEG rhythm separately for all trials from all sessions. As effective ERD/S generation involves rigorous neurofeedback training, the present hypothesis assumes the enhancement of ERD/S as a potential indication of the enhancement of lost motor functionality.

As the pre- and post-task subjective fatigue scores varied in every session, the amount of fatigue exclusively induced by the rehabilitation session has been calculated by subtracting the pre-task fatigue score from the post-task fatigue score.

Rhythm-wise band-power estimation was also calculated for inter-trial intervals to study brain activities during the short resting states between two consecutive trials. As the original data were recorded during a rehabilitation session, which was not primarily focused on studying the impact of mental fatigue on BCI performance, no EEG recordings are available during the inter-run resting periods.

## RESULTS AND DISCUSSIONS

In this present contribution, we studied how behavioural change occurs in patients with gradual motor functionality improvement. In the BCI community, it has been established that learning motor skills enhances the ERD/S at the motor cortex region of the brain [17]. Hence, the present analysis has been primarily focused on the C3 and C4 EEG channels based on the impaired limb of the participant.

### *Inter-session behavioural data analysis:*

The variation of fatigue levels across different sessions has been shown in Fig.1. It is to be noted that the stroke severity of the participants differed for each individual, and the final regain of motor functionality was also not at the same level after completing all 12 sessions. Hence, to identify the motor ability improvement trend, ARAT

scores, which were initially rated between 0-57, have been normalized between 0 and 1 using the "min-max normalization" technique. Similarly, the induced fatigue level has also been normalized between 0 to 1.

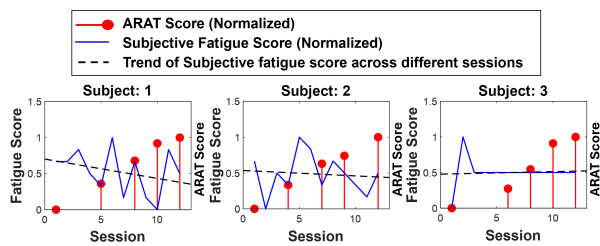


Figure 1: Trend of fatigue over 12 sessions. Trend line (dashed) fitted using linear regression

From Fig.1, it can be observed that subjective fatigue level is decreasing for two participants out of three over different sessions (trend based on linear regression has been shown by the dashed line in Fig.1) when ARAT score clearly demonstrates the improvement of motor functionality. The observation has been found to be statistically significant using Wilcoxon's two-tailed signed rank test with a 5% threshold for two participants (Subject 1 and 2). In contrast, no significant change has been observed for Subject 3.

#### Inter-session ERD/S analysis:

As reported earlier, ERD/S has been computed for all aforesaid EEG rhythms, and it has been found that the beta (13-30 Hz) band reflects the impact of mental fatigue on motor performance better than all other EEG rhythms. Such observation may be due to its rich content of information related to motor activity. Average ERD/S over different sessions for the motor cortex region is reported in Fig.2.

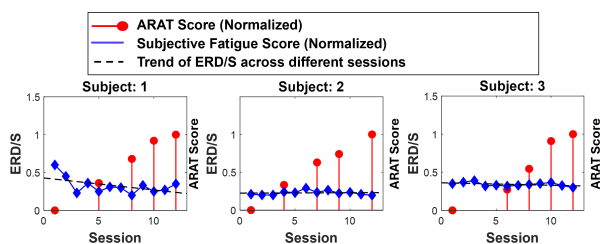


Figure 2: Trend of ERD/S across 12 sessions. Trend line (dashed) fitted using linear regression

From Fig.2, it can be observed that average ERD/S is decreased over different sessions for all three participants, indicating enhancement of ERD/S, which is a potential indication of motor function development for the impaired hand (statistical significance has been observed only for subject 1 while subject 2 and 3 did not exhibit any statistical significance). Hence, from Fig.1 and Fig.2, it can be inferred that subjective fatigue score and ERD/S have a positive correlation while it shows a negative correlation with ARAT score.

#### BCI classification performance across different sessions:

BCI classification accuracy averaged across 3 real-time feedback runs for every session, has been computed to understand how BCI performance changes across the 12 sessions with gradual improvement of motor functionality. The BCI performance for 3 participants has been reported in Fig3.

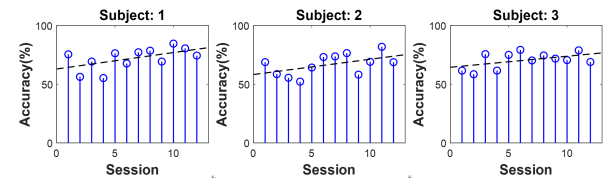


Figure 3: Average BCI classification accuracy during 3 feedback runs across 12 sessions. Trend line (dashed) fitted using linear regression

From Fig.3, it can clearly be observed that the performance of the BCI system was improving along with the improvement of motor functions of the impaired hand and exhibits a negative correlation with subjective fatigue level. It is interesting to observe that the BCI performance of Subject 1 was highest in the session in which the reported subjective fatigue score was observed to be the lowest (session 10). But for participants 2 and 3, no such observation can be made. One of the possible reasons for this may be the understanding of mental fatigue is highly subjective, and that can bias self-rating of mental fatigue. Previous literature indicates that volunteers participating in BCI or related experiments often confuse mental fatigue with other mental states, such as effortful attention (where the self-perceived effort of the participant is very high) or boredom [18]. Hence, it is necessary to evaluate mental fatigue objectively to quantify it without any subjective bias.

#### Intra-session analysis of ERD/S and beta bandpower:

Along with inter-session analysis, an intra-session study has also been done to obtain better insight into the underlying neural dynamics. For intra-session analysis, the first 10 trials and the last 10 trials from every run have been taken into consideration to calculate average ERD/S during the commencement and ending of each run. In Fig.4, intra-session analysis for three subjects for the beta band has been reported.

Based on the findings depicted in Fig.4, it is intriguing to note that there is a noticeable decline in the ERD/S enhancement following the completion of a run (group of trials without a significant resting period) on multiple occasions. This observation can be attributed to the potential induction of fatigue. Additionally, it is worth mentioning that in the majority of cases, the ERD/S improves at the beginning of the subsequent run compared to the ERD/S at the end of the previous run (statistically significant changes in ERD/S after short inter-run breaks have been marked in the figure with '\*' when two-tailed Wilcoxon's signed-rank test has been done with 5% threshold). One plausible explanation for this trend is the restoration of the brain from a fatigued state, potentially facilitated by a short break between runs. In a few cases,

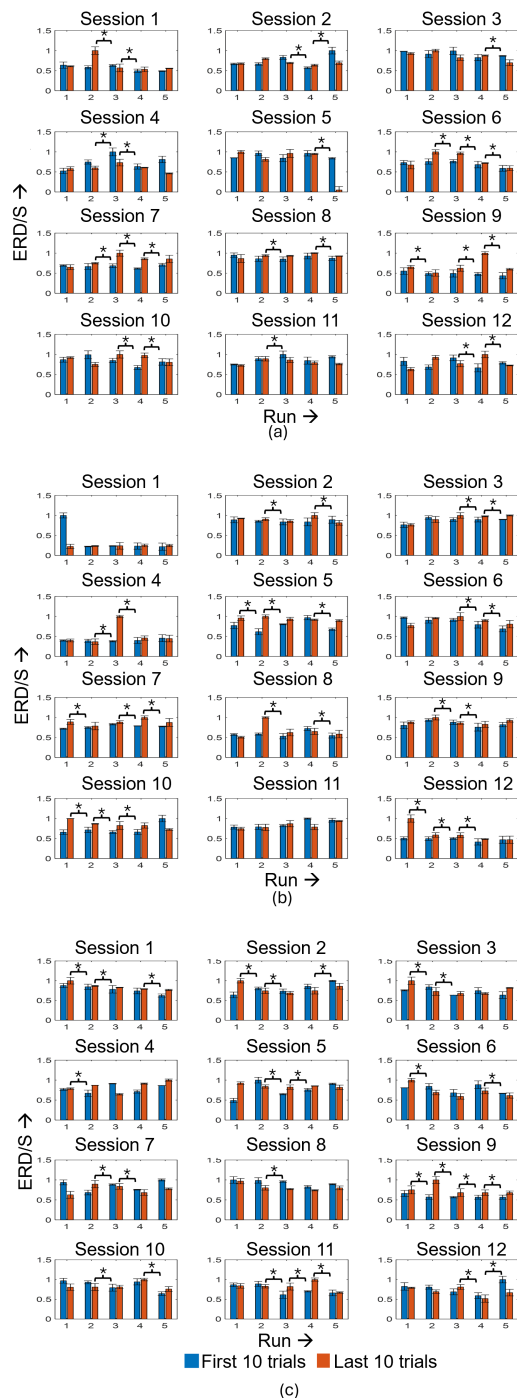


Figure 4: Intra-session ERD/S activities for different runs for (a) Subject 1, (b) Subject 2, and (c) Subject 3

for example, in Subject 3 session 2, it can be observed that there is a significant decline in ERD/S enhancement between run 4 and 5, which might indicate that the participant got too fatigued to recover during the small break between run 4 and 5.

Hence, the above analysis shows a possible relationship between fatigue and re-learning of motor functions by evaluating ERD/S, subjective fatigue score, and ARAT score.

Along with ERD/S, beta band power was also investigated for inter-trial intervals where participants were expected to have low or no motor activity. The analysis results are reported in Fig.5.

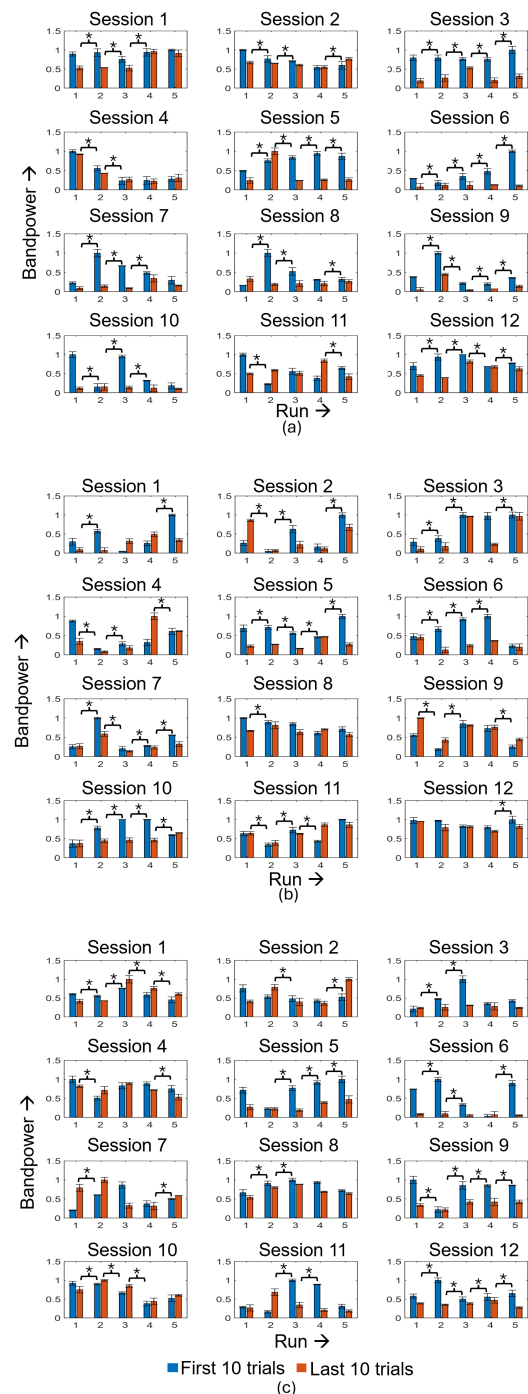


Figure 5: Intra-session beta band power activities for different runs for (a) Subject 1, (b) Subject 2, and (c) Subject 3

From Fig.5, it has been observed that, for the majority of the runs across different sessions, beta band power (which is also associated with focus and arousal [19]) has been reduced at the end of runs compared to the com-

mencement of that run. Furthermore, this observation is found to be consistent across different subjects. The plausible explanation of this observation may be linked to the potential inclination of mental fatigue. Similar to Fig.4, improvement of brain arousal state (increase in beta band power) has also been identified after short inter-run breaks.

#### *Intra-session BCI classification performance:*

The average classification accuracy of the BCI system for the 3 feedback runs across all sessions has also been analyzed to obtain more insight into the relationship between motor re-learning and subjective mental fatigue. It is to be noted that the first two runs correspond to the calibration phase, and the classifier did not generate any output. The intra-session classification performance for all 3 participants has been reported in Fig.6.

From Fig.6 and Fig.4, it can be observed that classification accuracy exhibits an improvement compared to its previous run only in the cases where enhancement of ERD/S occurred after a short mid-task break, indicating the negative impact of mental fatigue on BCI performance which may hinder improvement of motor abilities. A noteworthy observation is that Subject 3 consistently reported no subjective increase in fatigue levels during the rehabilitation sessions. However, the neural activity exhibited indications of possible fatigue induction. This reinforces the importance of establishing reliable, objective markers for fatigue.

From the psychological standpoint, mental fatigue manifests itself in terms of loss of attention and cognitive decline. Shift/loss of attention due to mental fatigue can disrupt participants' active engagement on task-related stimuli and can compromise the production of ERD/S. The intra-session analysis clearly indicates that BCI performance and regaining of motor functionalities are subject to the participant's active engagement in the rehabilitation exercise. This observation unfolds another vital aspect of BCI research. In existing BCI research, the majority of experimental paradigms related to motor imagery or motor execution tasks are designed as an open loop system where session length, number of trials per session, and mid-session breaks are fixed by the researchers and kept the same for all participants while the induction of mental fatigue is highly subjective and depends on the mood, emotional condition, and cognitive capacity of the participant at that moment. From this present contribution, it can be observed that adaptation and modulation of the experimental paradigm, based on participants' mental and cognitive states, are essential to ensuring participants' active engagement to optimize the enhancement of neurorehabilitation more quickly and effectively.

While the analysis mentioned above suggests that band power and ERD/S activity have the potential to serve as neuromarkers for monitoring fatigue objectively, further comprehensive investigations are required. A more detailed study is necessary to delve deeper into these measures and establish their effectiveness and reliability in fatigue monitoring. Moreover, previous studies on bore-

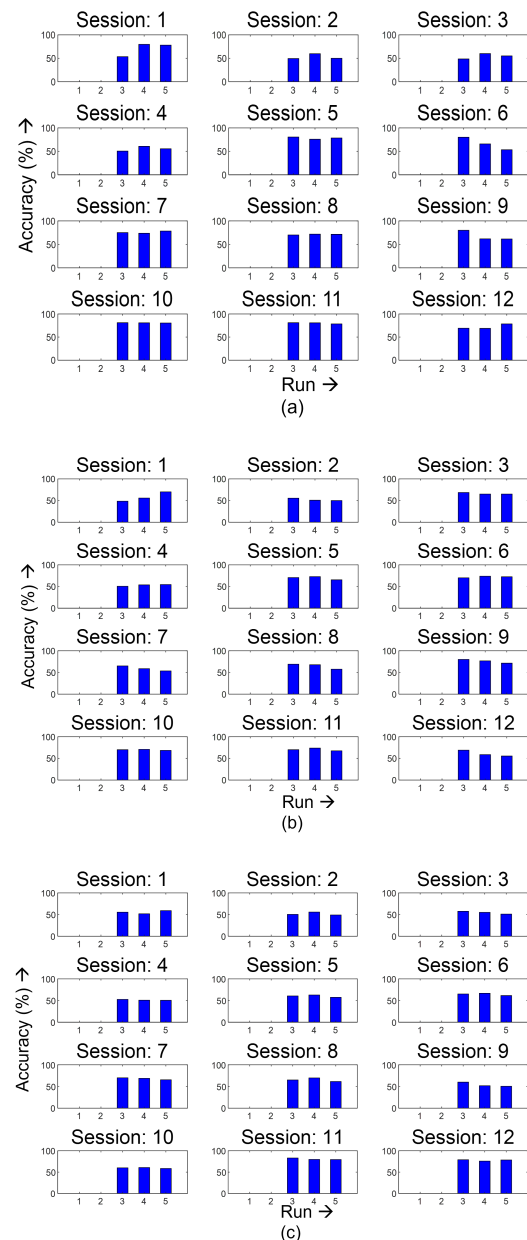


Figure 6: Intra-session classification accuracy of the BCI model for different runs for (a) Subject 1, (b) Subject 2, and (c) Subject 3

dom [20, 21] and mental fatigue [22] report similar types of decrement in vigilance and increased reaction time. This indicates a possible overlap between mental fatigue and boredom in terms of behavioural outcomes. Hence, it is necessary to identify relevant neuromarkers exclusively associated with mental fatigue while developing a passive BCI system to monitor mental fatigue.

## CONCLUSION

In this present contribution, a preliminary study on the impact of mental fatigue on the improvement of motor functionality and BCI performance has been done on 3

stroke survivors. The study exhibits the potential negative impact of mental fatigue on enhancing motor functions. This paper also highlights the necessity of objectively monitoring mental fatigue and designing an adaptive BCI rehabilitation paradigm for optimal enhancement outcomes. The future scope of this study includes consolidating current observations on a larger patient population and identifying neuromarkers explicitly associated with fatigue for developing a passive BCI-based status monitoring system that will assist an active BCI-based exoskeleton in adapting the rehab sessions based on patients' mental and cognitive conditions.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support in part provided by the *Ulster University-Vice Chancellor Research Scholarship* and the *DfE ISPF Project 610124: A UK-LMIC Research Network For Autism Spectrum Disorder (ULMiRN-ASD)*.

## REFERENCES

- [1] Asmar R, O'Brien E, Mieke S, Murray A, Parati G, Steenvoorden G, *et al.* Affordable technology: Blood pressure measuring devices for low resource settings. 2005.
- [2] Sousa Fernandes MS de *et al.* Effects of physical exercise on neuroplasticity and brain function: A systematic review in human and animal studies. *Neural plasticity*. 2020;2020.
- [3] Putrino D, Zanders H, Hamilton T, Rykman A, Lee P, Edwards DJ. Patient engagement is related to impairment reduction during digital game-based therapy in stroke. *Games for health journal*. 2017;6(5):295–302.
- [4] Ang KK, Guan C. Brain-computer interface for neurorehabilitation of upper limb after stroke. *Proceedings of the IEEE*. 2015;103(6):944–953.
- [5] Colucci A *et al.* Brain-computer interface-controlled exoskeletons in clinical neurorehabilitation: Ready or not? *Neurorehabilitation and Neural Repair*. 2022;36(12):747–756.
- [6] Robinson N, Mane R, Chouhan T, Guan C. Emerging trends in bci-robotics for motor control and rehabilitation. *Current Opinion in Biomedical Engineering*. 2021;20:100354.
- [7] Gordleeva SY *et al.* Exoskeleton control system based on motor-imaginary brain-computer interface.. 2017;9(3 (eng)):31–36.
- [8] Chowdhury A, Dutta A, Prasad G. Corticomuscular co-activation based hybrid brain-computer interface for motor recovery monitoring. *IEEE Access*. 2020;8:174542–174557.
- [9] Talukdar U, Hazarika SM, Gan JQ. Motor imagery and mental fatigue: Inter-relationship and eeg based estimation. *Journal of computational neuroscience*. 2019;46:55–76.
- [10] Talukdar U, Hazarika SM, Gan JQ. Adaptation of common spatial patterns based on mental fatigue for motor-imagery bci. *Biomedical Signal Processing and Control*. 2020;58:101829.
- [11] Gentili R, Han CE, Schweighofer N, Papaxanthis C. Motor learning without doing: Trial-by-trial improvement in motor performance during mental training. *Journal of neurophysiology*. 2010;104(2):774–783.
- [12] Branscheidt M *et al.* Fatigue induces long-lasting detrimental changes in motor-skill learning. *Elife*. 2019;8:e40578.
- [13] Jacquet T, Poulin-Charronnat B, Bard P, Lepers R. Persistence of mental fatigue on motor control. *Frontiers in psychology*. 2021;11:588253.
- [14] Khojasteh Moghani M, Zeidabadi R, Shahabi Kaseb MR, Bahreini Borujeni I. Mental fatigue reduces the benefits of self-controlled feedback on learning a force production task. *Perceptual and Motor Skills*. 2021;128(5):2398–2414.
- [15] Pfurtscheller G. Functional brain imaging based on erd/ers. *Vision research*. 2001;41(10-11):1257–1260.
- [16] Kitahara K, Kondo T. Modulation of erd/s by having a conscious target during lower-extremity motor imagery. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015, 6630–6633.
- [17] Rimbart S, Bougrain L, Fleck S. Learning how to generate kinesthetic motor imagery using a bci-based learning environment: A comparative study based on guided or trial-and-error approaches. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2020, 2483–2498.
- [18] Myrden A, Chau T. Effects of user mental state on eeg-bci performance. *Frontiers in human neuroscience*. 2015;9:308.
- [19] Lim S, Yeo M, Yoon G. Comparison between concentration and immersion based on eeg analysis. *Sensors*. 2019;19(7):1669.
- [20] Yakobi O, Boylan J, Danckert J. Behavioral and electroencephalographic evidence for reduced attentional control and performance monitoring in boredom. *Psychophysiology*. 2021;58(6):e13816.
- [21] Pattyn N, Neyt X, Henderickx D, Soetens E. Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & behavior*. 2008;93(1-2):369–378.
- [22] Thompson C, Fransen J, Beavan A, Skorski S, Coutts A, Meyer T. Understanding the influence of a cognitively demanding task on motor response times and subjective mental fatigue/boredom. *Brazilian Journal of Motor Behavior*. 2020;14(1):33–45.

# MAPPING NEUROMUSCULAR REPRESENTATION OF GRASPING MOVEMENTS USING ULTRA-HIGH-DENSITY EEG AND EMG

Leonhard Schreiner<sup>1,2\*</sup>, Pauline Schomaker<sup>4\*</sup>, Sebastian Sieghartsleitner<sup>1,3</sup>, Michael Schwarzgruber<sup>1</sup>, Harald Pretl<sup>2</sup>, Andreea I. Sburlea<sup>4</sup>, Christoph Guger<sup>1</sup>

<sup>1</sup> g.tec medical engineering GmbH, Schiedlberg, Austria

<sup>2</sup> Institute for Integrated Circuits, Johannes Kepler University, Linz, Austria

<sup>3</sup> Institute of Computational Perception, Johannes Kepler University, Linz, Austria

<sup>4</sup> Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, the Netherlands

E-mail: schreiner@gtec.at

\*These authors contributed equally to this work.

**ABSTRACT:** Understanding the intricate coordination between the brain and muscles during movement tasks is crucial for advancing our knowledge of motor control and enhancing Brain-Computer Interface (BCI) devices. This study investigates the mechanisms underlying grasping movements using diverse objects and grasping techniques. Employing a novel ultra-high-density (uHD) EEG/EMG system, the study examines neural and muscular activity with high spatial resolution. Results of three healthy subjects highlight event-related desynchronization/synchronization (ERD/S) patterns and classification accuracies for EEG and EMG signals during grasping tasks. Temporal analysis reveals a strong relationship between EMG/EEG activation and classification outcomes, supported by kinematic data as evidence of motion. S02 achieved the highest average EEG and EMG classification accuracies at 69.4% and 97.8%, respectively, while S01 had the lowest at 64% and 85.4%. The observed dependencies between accuracies imply an interconnected and synergistic relationship between EEG and EMG modalities, which holds promise for enhancing overall performance in future BCIs.

## INTRODUCTION

Examining our brain's and muscles' coordination during movement tasks elucidates the intricate mechanisms underlying motor control. The primary aim of this study is to investigate the grasping of objects, including a multitude of grasp types and objects, to enhance future brain-computer interface (BCI) devices. Researchers have explored the mechanics of grasping in both animals and humans using various methodologies [1], [2], [3], [4], [5], [6], [7]. Nonetheless, a substantial gap in knowledge remains regarding the precise mechanisms through which our brains govern these movements, particularly as they evolve. Non-invasive techniques such as EEG offer high temporal resolution, allowing for studying neural dynamics during grasping. However, their spatial resolution is often limited, constraining the

capacity to attain a more nuanced comprehension of neural control [8], [9], [10], [11]. Similarly, EMG devices, often characterized by low resolution [12], are used to investigate muscular activities. Our research utilizes a novel ultra-high-density (uHD) EEG/EMG system to explore the intricate interplay between neural and muscular activity in greater spatial detail. The system demonstrates improvements due to its increased sensor density, outperforming other high-density EEG systems. It has been effectively utilized in research studies focused on decoding finger movements [11], hand gestures [14], and mapping the central sulcus using somatosensory evoked potentials [13].

We aim to expand the scope of discrimination by employing various objects and grasping types, thus delving into multiple dimensions for precise analysis. Sburlea et al. [3] investigated the slow-frequency EEG components with a similar paradigm. They found that the grasp types are encoded in motor cortex areas, while object properties activate the frontoparietal regions. Additionally, they discovered that the grasp types are significantly better decoded during the execution and release stages than the observation stage. Building upon their findings, we focus our investigations on the motor cortex area contralateral to the moved hand in the execution and release stages. However, we focused on extracting EEG frequencies in the 8-30 Hz range, which are substantial in movement decoding in EEG research [14], [15].

## MATERIALS AND METHODS

Our system incorporates flexible surface electrode grids that were applied on the scalp as well as arm, and hand muscles. The uHD EEG/EMG system (g.Pangolin, g.tec medical engineering GmbH, Austria) has an inter-electrode distance of 8.6 mm and an exposed sensor diameter of 5.9 mm. We used the system to measure data from three healthy subjects (two right-handed and one left-handed).

Two biosignal amplifiers (see Fig. 1A(a)) allow the acquisition of an average amount (across subjects) of 330 channels of biosignal data, with an average of 235 channels dedicated to scalp recordings (176 (S01), 256 (S02), and 272 (S03)). We recorded EMG data from 96 channels across all subjects, focusing on intrinsic (hand) and extrinsic (forearm) arm muscles. (see Fig. 1A). Data was acquired with a sample rate of 512 Hz. Additionally, the kinematics of the hand and arm movements are acquired and digitized using the leap motion camera [16] at a variable sample rate between 80 and 120 Hz. The grasping task and six different grasps are visible in Fig. 1 B. We distinguished between four objects (big sphere, small sphere, big cylinder, small cylinder) and three grasping types (power, precision, pinch), resulting in a total of 12 grasping conditions, as demonstrated by Sburlea et al. [3]. Each grasping type was executed 27 times per object per run. Two runs per object were performed, with randomized order, resulting in 54 trials for further analysis. Additionally, within each run, the grasping types were shuffled.

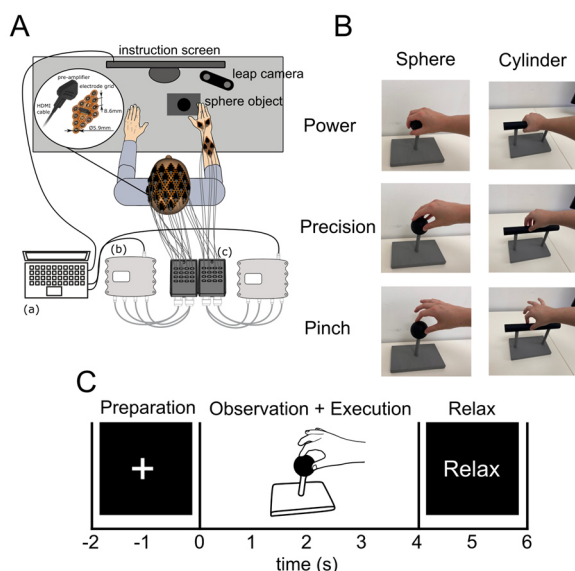


Figure 1: uHD EEG/EMG system for decoding object grasping task. (A) schematic system setup: two g.HIamp (b) (g.tec medical engineering GmbH) biosignal amplifiers, each with 256 channels, connected to the acquisition computer (a) for synchronous EEG and EMG recording. The uHD EEG system comprises electrode grids with 16 channels and a pre-amplifier attached to each grid, connected to the connector box (c). For EEG acquisition, an average of 235 channels covers sensory and motor areas on the contralateral hemispheres. Six grids (96 channels) are placed on the extrinsic and intrinsic hand muscles for EMG data acquisition. (B) Two objects (small sphere, small cylinder) and three grasp types (power, precision, and pinch) are depicted; see [3] for all 12 classes of the object grasps. (C) Paradigm Procedure: The procedure commences with the display of a fixation cross for 2 seconds, followed by randomized instructions for the grasping task (4 seconds), succeeded by a 2-second relaxation period.

Subjects are then directed to prepare for the next 8-second cycle. In total, 54 trials per class were recorded.

Instructions were given on a computer screen placed approximately 1.5 meters in front of the subjects. First, a fixation cross was displayed for 2 seconds, followed by a 4-second observation and execution phase, during which pictures depicting the object and grasp to be performed were presented, as illustrated in Fig. 1B. Subsequently, upon completing the grasping task, participants were instructed to return to the starting position, relax their arm/hand, and prepare for the upcoming trial (Fig. 1C).

### EEG Preprocessing

The raw EEG recordings were first notch-filtered at 50 Hz and its harmonics using a 2nd-order Butterworth filter. After notch-filtering, bad channels were identified and removed using the approach described in [11], except for the band-pass filter ranging from 0.5 to 40 Hz instead. Finally, EEG data were common average referenced.

### Feature Extraction and Epoching

For the classification of grasp types, band power features based on 8 to 30 Hz (4th-order Butterworth band-pass filter) were extracted as this frequency range encompasses both mu (8-12 Hz) and beta (13-30 Hz) rhythms which are associated with motor functions [17], [18]. Note that mu and beta rhythms may be analyzed separately. However, this was not done for the current classification analysis to keep dimensionality (i.e., number of features) low. On the other hand, beta band power features were extracted for topography plots. The band power was estimated by squaring EEG time samples and applying a centered moving average using a window length of 0.75 seconds and a step size of 0.1 seconds. Furthermore, band power estimates were log-transformed to improve Gaussianity (i.e., normality) [19]. Finally, the log-transformed band power features were epoched using 1-second pre- and 8 second post-cue.

### EEG Classification

Classification models were employed to investigate if the extracted band power features can differentiate between the grasp types. Specifically, pairwise classification of grasp types was performed for each object, respectively, leading to 12 two-class classification problems per subject. Pairwise classification was employed instead of a 3-class problem to allow for easier interpretation of results. Note that a classification analysis between objects was not carried out as objects were not shuffled on a trial-by-trial basis, which would lead to inflated accuracies due to the non-stationarity observed in EEG. A regularized linear discriminant analysis (rLDA) was utilized as a classification model, with the regularization parameter  $\alpha$  set to 0.1 [20]. The classification framework used was a 10-times 10-fold cross-validation, in which the random seed was set to the respective iteration (i.e., 1 to 10) to allow for reproducible results.

### Topography Plots

Topography plots were created according to [11] and [21] using a custom montage creator software (g.tec medical engineering GmbH, Austria). Event-related desynchronization/synchronization (ERD/S) was

calculated using the log-transformed band power features of the beta band, with the baseline reflecting 0.5 seconds pre-cue. The brain models depicted in the topography plots were created using anatomical MRI scans from all participants. The brain and skull were reconstructed using FreeSurfer software (developed at the Martinos Center for Biomedical Imaging in Cambridge, MA, United States) based on the T1-weighted MRI data [22].

### EMG

The raw EMG recordings were notch-filtered at 50 Hz and its harmonics, using a 2nd-order Butterworth filter. EMG features reflected simple root-mean-square values (RMS) of the band-pass filtered EMG (20 to 200 Hz, 4th-order Butterworth filter) data. The window length and step size were set to 0.2 and 0.05 seconds, respectively. Finally, RMS features were epoched using 1 second pre- and 8 second post-cue.

### MVC

Maximum Voluntary Contraction (MVC) was recorded using a commercially available dynamometer. Participants performed a maximum contraction with their dominant hand for five seconds, followed by a one-minute break. The contraction was repeated three times and was performed by using a power grip on the dynamometer. The EMG signals obtained during MVC were then used to normalize EMG signals obtained during the grasping paradigm [23].

### EMG Classification

EMG classification of grasps per object was analogous to EEG classification, except that 3-class classification for the grasp types was performed and that  $\alpha$  was set to 0.25 for the rLDA. Thus, four classification problems, one for each object, were carried out per subject. The three-class classification was carried out as EMG is expected to result in much greater accuracy and, thus, easier interpretable results.

## RESULTS

Fig. 2 illustrates the ERD/S topographies from all subjects, with the time points set to 1 second (for S01 and S02) and 1.5 seconds (for S03) after task instruction, as these time points reveal the most pronounced ERD for all subjects, respectively. The large and small object conditions were averaged for each grasp type, resulting in 6 conditions. During the sphere power grasp, S01 and S02 exhibited the most significant ERD at the contralateral hemisphere, around the C3 and C4 electrode positions, for S02 and S01 respectively. S03 showed the greatest consistency across all objects and grasps and demonstrated a clear focal spot around the C1 and C3 electrode positions. In other words, S01 and S02 exhibit more lateralized ERD, whereas S03's ERD is slightly more central. A weaker ERD was observed for S01, and the sphere object led to a stronger ERD than the cylinder.

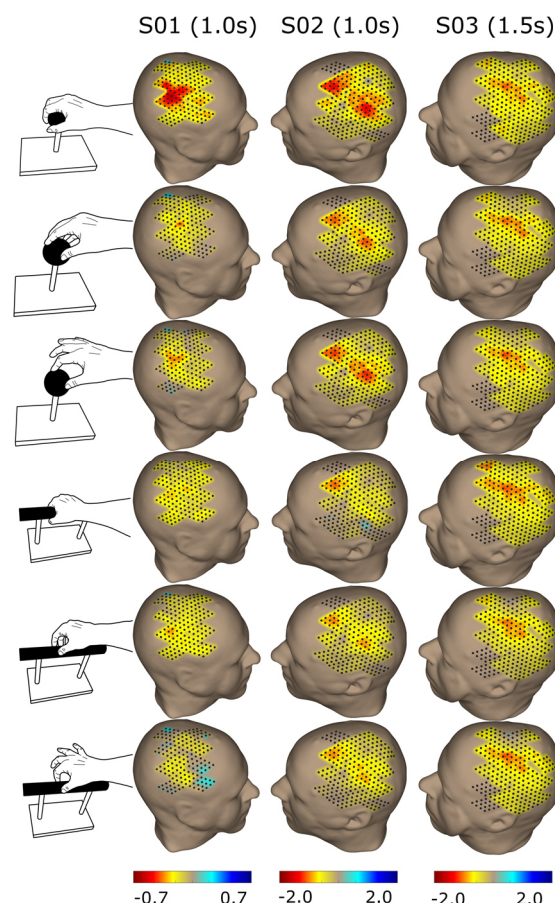


Figure 2: ERD/S Topographical maps were created for each participant and the corresponding grasping type. The hemispheres are shown contralateral to the hand-side of the grasp, with S01 being left-handed and S02 and S03 right-handed. The maps were constructed utilizing the beta frequency range (13-30 Hz) at 1 second for S01 and S02 and 1.5 seconds for S03—values in dB.

Fig. 3 shows the topographical ERD/S time course obtained for the beta frequency band features, averaged over both sphere objects (big and small), and power grasps for subject S03. The time window was set to -1 s pre-task instruction and 8 s post-task instruction (1 s steps) for a detailed temporal representation. The representation of the EMG RMS power for the same tasks was separated into intrinsic (blue) and extrinsic (green) muscles and depicted as mean (SD). The EMG RMS filtered from 20-200 Hz was calculated as a percentage according to the MVC measurement. The data from the leap motion camera was used for kinematic analysis. The retrieved velocity of the vertical movement from the hand/arm is depicted as mean (SD) in m/s (magenta). The EEG Accuracy graph in the last row (orange) is calculated in 0.1 s steps as mean (SD) for the ten runs from the 10 times 10-fold cross-validation. The EMG accuracy graph (cyan) reflects the outcome of a 3-class problem. Similar to the EEG classification, the mean (SD) from the ten classification runs was plotted. Empirical chance levels were obtained by generating null models in which the class labels were shuffled and marked as dashed lines in Fig. 3.

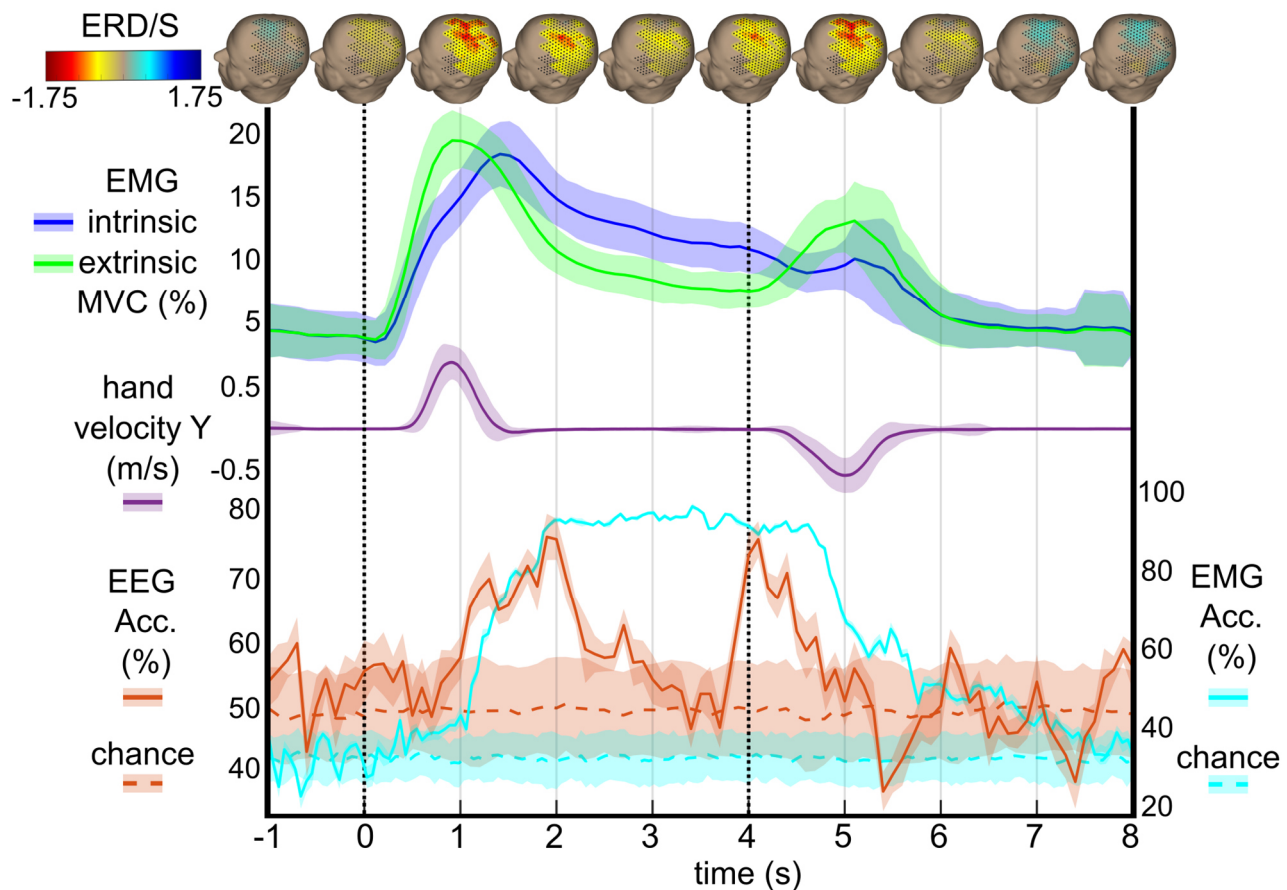


Figure 3: Subject S03 detailed analysis of the sphere power condition (small / big sphere merged). The time axis ranges from -1 to 8 seconds, with 0 as the task presentation on screen and the relax instruction at 4 seconds. The first row shows the ERD/S topographical time course. The maps were constructed utilizing the beta frequency band (13-30 Hz) with power shown in dB (color bar encodes ERD in red and ERS in blue). Row 2 represents the EMG MVC (%) time course from the 20-200 Hz filtered RMS signal as mean (SD) in blue for the intrinsic muscles and green for the extrinsic hand muscles. Row 3 depicts the kinematics of the arm movement as Velocity (m/s) in vertical (Y) direction (lifting/lowering) as mean (SD) in magenta. The last row represents the classification accuracy for EEG big sphere power vs. precision (orange) and the EMG accuracies for the 3-class problem classification outcome for the big sphere object (cyan). The EEG axis scaling was set to 40-80% (left side), whereas the EMG Accuracies are drawn from 20-100% (right side). Real chance levels drawn as dashed lines: EEG (orange) and EMG (cyan).

Table 1: EEG Accuracies for all classification pairs

Grasps / Subjects		S01	S02	S03
<b>Big Sph.</b>	Power vs. Precision	47.9%	69.4%	<b>76.9%</b>
	Power vs. Pinch	63.5%	61.1%	70.4%
	Precision vs. Pinch	57.3%	57.4%	66.7%
<b>Big Cyl.</b>	Power vs. Precision	65.6%	<b>78.7%</b>	63.9%
	Power vs. Pinch	55.2%	74.1%	70.4%
	Precision vs. Pinch	<b>77.1%</b>	71.3%	64.8%
<b>Sm. Sph.</b>	Power vs. Precision	58.3%	76.9%	69.4%
	Power vs. Pinch	69.8%	74.1%	66.7%
	Precision vs. Pinch	67.7%	64.8%	57.4%
<b>Sm. Cyl.</b>	Power vs. Precision	66.7%	65.7%	71.3%
	Power vs. Pinch	72.9%	75.9%	75.0%
	Precision vs. Pinch	65.6%	63.9%	67.6%

Tab. 1 shows the pairwise EEG classification accuracies of all the grasping techniques across objects. For each subject, 12 accuracies are depicted. The classification pair with the highest averaged accuracy was found for

Small Cylinder Power vs. Pinch with 74.6% across subjects. Subject S02 reached the highest average accuracy of 69.4% for all pairs, S03 reached 68.4%, and S01 had the lowest accuracy of 64%.

Table 2: EMG accuracies for 3-class problems

Grasps / Subjects	S01	S02	S03
<b>Big Sph.</b>	90.4%	97.7%	96.4%
<b>Big Cyl.</b>	98.1%	<b>98.8%</b>	95.9%
<b>Sm. Sph.</b>	<b>100%</b>	98.6%	98.1%
<b>Sm. Cyl.</b>	53.1%	96.2%	<b>99.4%</b>

Tab. 2 shows the EMG classification accuracies computed as 3-class problems (all grasping types included). The object with the best accuracy was the small sphere, with an average accuracy of 98.9% across subjects. The subject with the highest accuracy is similar to EEG S02, with 97.8% across objects. S02 reached 97.5%, and S01 has the lowest average accuracy of 85.4%.

## DISCUSSION

As illustrated in Fig. 2 and Fig. 3, the uHD EEG system's high spatial resolution enables a thorough examination of ERD/S topographies. The spatiotemporal dynamics of ERD/S offer valuable insights into the brain patterns triggered by grasping movements and their neural representation. In Fig. 3 (top row), a desynchronization occurs at the time point (1s) after task instruction, declining from seconds 2-3, followed by another increase towards the offset of the grasping task. Following the ERD, an ERS becomes apparent 2 to 3 seconds after the movement termination. The topographies show a similar focal spot as found in prior studies using the uHD EEG system [11], [15] and other studies examining temporal ERD/S dynamics [24].

The EMG results from the intrinsic and extrinsic muscles have a similar temporal pattern (see Fig. 3) with an activation start around 200 ms after task instruction. However, the extrinsic group shows a steeper ascent of the curve and a slightly higher MVC percentage value at the beginning of the movement, which can be explained by a temporal difference in the activation of the muscles, meaning lifting off the arm incorporates more extrinsic muscles [25]. This is also reflected in the peak values, which reach around 1 second for extrinsic muscles and 1.4 seconds for intrinsic muscles. The intrinsic muscles show higher activation throughout the grasping task due to active grasping control and the possibility of slightly resting the arm on the object. When moving the arm back to the resting position, the extrinsic muscles show higher activation than the intrinsic. The results from the EMG analysis correspond closely with the temporal behavior observed in the data from the Leap camera, indicating a parallel trend in their patterns over time, with a slightly longer delay of approx. 0.5 seconds from task instruction to velocity onset can be explained by prior muscle activation and real movement detected by the camera. The EEG and EMG accuracy traces (Fig. 2 last row) started to incline from the chance level at around 1 second after task instruction, which is closely associated with the temporal behavior of the EMG and kinematics, however showing an additional 0.5 seconds delay for the movement onset [26]. The maximum EMG accuracy is stable at around the maximum of 95%; however, in comparison, the EEG accuracy declines again after the initial peak. This phenomenon may be attributed to diminished attention towards movement execution, as automatic patterns governed by lower-level brain structures entail reduced involvement of higher-order control mechanisms [27]. Interestingly, the EEG accuracy shows a second increase at the movement offset but peaks earlier than the movement onset. This temporal phenomenon could stem from anticipating returning the arm to the starting position [28].

S02 reached the highest accuracy for EEG and EMG classification with 69.4% and 97.8%, respectively (see Tab.1 and Tab. 2). S01 achieved the worst accuracy with 64% for EEG and 85.4% for EMG classification.

Considering that S01 has the lowest amount of scalp channels used for classification, it also indicates that the high spatial density of EEG electrodes is beneficial for decoding motor tasks. A similar analysis was done by [11] using the uHD EEG, where they showed that when subsampling the electrode count, there was a decrease in accuracy for decoding individual finger movements. A lower ERD power was also observed for S01, which adds to the lower performance. For EMG, only the small sphere object grasp for S01 showed lower accuracies, which could be attributed to the removal of roughly half of the trials due to technical issues in the acquisition. The observed trend of superior performance in both EEG and EMG across subjects suggests a nuanced interplay between the examined modalities. This pattern highlights how the biosignals are interconnected and complement each other, suggesting they work synergistically to produce a combined effect greater than the sum of their individual contributions. Consequently, future research endeavors should prioritize utilizing parameters such as corticomuscular coherence alongside advanced classification methodologies to optimize the performance and functionality of brain-computer interfaces (BCIs) and neuroprosthetic systems.

## Acknowledgements

The authors would like to thank the Institute of Neural Engineering, Graz University of Technology, for providing the objects for this study.

## REFERENCES

- [1] J. A. Michaels and H. Scherberger, 'Population coding of grasp and laterality-related information in the macaque fronto-parietal network', *Sci Rep*, vol. 8, no. 1, p. 1710, Jan. 2018, doi: 10.1038/s41598-018-20051-7.
- [2] S. Schaffelhofer and H. Scherberger, 'Object vision to hand action in macaque parietal, premotor, and motor cortices', *eLife*, vol. 5, p. e15278, Jul. 2016, doi: 10.7554/eLife.15278.
- [3] A. I. Sburlea, M. Wilding, and G. R. Müller-Putz, 'Disentangling human grasping type from the object's intrinsic properties using low-frequency EEG signals', *Neuroimage: Reports*, vol. 1, no. 2, p. 100012, Jun. 2021, doi: 10.1016/j.ynirp.2021.100012.
- [4] A. I. Sburlea and G. R. Müller-Putz, 'Exploring representations of human grasping in neural, muscle and kinematic signals', *Sci Rep*, vol. 8, no. 1, Art. no. 1, Nov. 2018, doi: 10.1038/s41598-018-35018-x.
- [5] R. D. Flint, J. M. Rosenow, M. C. Tate, and M. W. Slutzky, 'Continuous decoding of human grasp kinematics using epidural and subdural signals', *J Neural Eng*, vol. 14, no. 1, p. 016005, Feb. 2017, doi: 10.1088/1741-2560/14/1/016005.
- [6] G. Lange, C. Y. Low, K. Johar, F. A. Hanapiah, and F. Kamaruzaman, 'Classification of

- Electroencephalogram Data from Hand Grasp and Release Movements for BCI Controlled Prosthesis', *Procedia Technology*, vol. 26, pp. 374–381, Jan. 2016, doi: 10.1016/j.protcy.2016.08.048.
- [7] A. Schwarz, P. Ofner, J. Pereira, A. I. Sburlea, and G. R. Müller-Putz, 'Decoding natural reach-and-grasp actions from human EEG', *J. Neural Eng.*, vol. 15, no. 1, p. 016005, Dec. 2017, doi: 10.1088/1741-2552/aa8911.
- [8] A. K. Robinson, P. Venkatesh, M. J. Boring, M. J. Tarr, P. Grover, and M. Behrmann, 'Very high density EEG elucidates spatiotemporal aspects of early visual processing', *Sci Rep*, vol. 7, no. 1, p. 16248, Dec. 2017, doi: 10.1038/s41598-017-16377-3.
- [9] C. M. Michel and D. Brunet, 'EEG Source Imaging: A Practical Review of the Analysis Steps', *Front Neurol*, vol. 10, p. 325, 2019, doi: 10.3389/fneur.2019.00325.
- [10] V. Brodbeck *et al.*, 'Electroencephalographic source imaging: a prospective study of 152 operated epileptic patients', *Brain*, vol. 134, no. 10, pp. 2887–2897, Oct. 2011, doi: 10.1093/brain/awr243.
- [11] H. S. Lee *et al.*, 'Individual finger movement decoding using a novel ultra-high-density electroencephalography-based brain-computer interface system', *Front Neurosci*, vol. 16, p. 1009878, Oct. 2022, doi: 10.3389/fnins.2022.1009878.
- [12] M. Zia ur Rehman *et al.*, 'Multiday EMG-Based Classification of Hand Motions with Deep Learning Techniques', *Sensors*, vol. 18, no. 8, Art. no. 8, Aug. 2018, doi: 10.3390/s18082497.
- [13] L. Schreiner *et al.*, 'Mapping of the central sulcus using non-invasive ultra-high-density brain recordings', *Sci Rep*, vol. 14, no. 1, p. 6527, Mar. 2024, doi: 10.1038/s41598-024-57167-y.
- [14] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, 'EEG-based discrimination between imagination of right and left hand movement', *Electroencephalography and clinical Neurophysiology*, vol. 103, no. 6, pp. 642–651, 1997, doi: 10.1016/S0013-4694(97)00080-1.
- [15] L. Schreiner, S. Sieghartsleitner, K. Mayr, H. Pretl, and C. Guger, 'Hand gesture decoding using ultra-high-density EEG', in *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, Apr. 2023, pp. 01–04. doi: 10.1109/NER52421.2023.10123901.
- [16] G. Marin, F. Dominio, and P. Zanuttigh, 'Hand gesture recognition with leap motion and kinect devices', in *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, France: IEEE, Oct. 2014, pp. 1565–1569. doi: 10.1109/ICIP.2014.7025313.
- [17] K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, and R. P. N. Rao, 'Cortical activity during motor execution, motor imagery, and imagery-based online feedback', *Proceedings of the National Academy of Sciences*, vol. 107, no. 9, pp. 4430–4435, Mar. 2010, doi: 10.1073/pnas.0913697107.
- [18] D. O. Cheyne, 'MEG studies of sensorimotor rhythms: a review', *Exp Neurol*, vol. 245, pp. 27–39, Jul. 2013, doi: 10.1016/j.expneurol.2012.08.030.
- [19] J. Gruenwald, A. Znobishchev, C. Kapeller, K. Kamada, J. Scharinger, and C. Guger, 'Time-Variant Linear Discriminant Analysis Improves Hand Gesture and Finger Movement Decoding for Invasive Brain-Computer Interfaces', *Front. Neurosci.*, vol. 13, p. 901, Sep. 2019, doi: 10.3389/fnins.2019.00901.
- [20] Y. Guo, T. Hastie, and R. Tibshirani, 'Regularized linear discriminant analysis and its application in microarrays', *Biostatistics*, vol. 8, no. 1, pp. 86–100, Jan. 2007, doi: 10.1093/biostatistics/kxj035.
- [21] J. Kubanek and G. Schalk, 'NeuralAct: A Tool to Visualize Electrocortical (ECoG) Activity on a Three-Dimensional Model of the Cortex', *Neuroinformatics*, vol. 13, no. 2, pp. 167–174, Apr. 2015, doi: 10.1007/s12021-014-9252-3.
- [22] A. M. Dale, B. Fischl, and M. I. Sereno, 'Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction', *NeuroImage*, vol. 9, no. 2, pp. 179–194, Feb. 1999, doi: 10.1006/nimg.1998.0395.
- [23] B. P. T. Ngo and R. P. Wells, 'Evaluating protocols for normalizing forearm electromyograms during power grip', *Journal of Electromyography and Kinesiology*, vol. 26, pp. 66–72, Feb. 2016, doi: 10.1016/j.jelekin.2015.10.014.
- [24] G. Pfurtscheller, 'Functional brain imaging based on ERD/ERS', *Vision Research*, vol. 41, no. 10, pp. 1257–1260, May 2001, doi: 10.1016/S0042-6989(00)00235-2.
- [25] C. I. Long, P. W. Conrad, E. A. Hall, and S. L. Furler, 'Intrinsic-Extrinsic Muscle Control of the Hand in Power Grip and Precision Handling: AN ELECTROMYOGRAPHIC STUDY', *JBJS*, vol. 52, no. 5, p. 853, Jul. 1970.
- [26] E. Trigili *et al.*, 'Detection of movement onset using EMG signals for upper-limb exoskeletons in reaching tasks', *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, p. 45, Mar. 2019, doi: 10.1186/s12984-019-0512-1.
- [27] R. M. Hardwick, S. Caspers, S. B. Eickhoff, and S. P. Swinnen, 'Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution', *Neuroscience & Biobehavioral Reviews*, vol. 94, pp. 31–44, Nov. 2018, doi: 10.1016/j.neubiorev.2018.08.003.
- [28] P. Cisek and J. F. Kalaska, 'Neural Correlates of Reaching Decisions in Dorsal Premotor Cortex: Specification of Multiple Direction Choices and Final Selection of Action', *Neuron*, vol. 45, no. 5, pp. 801–814, Mar. 2005, doi: 10.1016/j.neuron.2005.01.027.

# DECODING MORAL JUDGEMENT FROM TEXT: A PILOT STUDY

Diana E. Gherman<sup>1</sup>, Thorsten O. Zander<sup>1</sup>

<sup>1</sup> Brandenburg University of Technology Cottbus–Senftenberg, Germany

E-mail: diana.gherman@b-tu.de

**ABSTRACT:** Moral judgement is a complex human reaction that engages cognitive and emotional dimensions. While some of the morality neural correlates are known, it is currently unclear if we can detect moral violation at a single-trial level. In a pilot study, here we explore the feasibility of moral judgement decoding from text stimuli with passive brain-computer interfaces. For effective moral judgement elicitation, we use video-audio affective priming prior to text stimuli presentation and attribute the text to moral agents. Our results show that further efforts are necessary to achieve reliable classification between moral congruency vs. incongruency states. We obtain good accuracy results for neutral vs. morally-charged trials. With this research, we try to pave the way towards neuroadaptive human-computer interaction and more human-compatible large language models (LLMs).

## INTRODUCTION

*Passive BCIs.* Passive brain-computer interfaces (pBCIs) can seamlessly decode mental states from a user's brain activity [1]. Active BCIs require the conscious and intentional modulation of one's brain activity, while reactive BCIs make use of external stimuli such as flickering lights to evoke a desired reaction [2]. Meanwhile, pBCIs operate in the background, capturing the spontaneous reactions to specific stimuli in the environment. Most commonly, electroencephalography (EEG) signals are collected and used for mental state classification. Once decoded, pBCIs can provide this real-time information to a computer that can then adapt its outputs to cater to individual needs and preferences. This new form of interaction has previously been described as neuroadaptive [3]. Thus, pBCIs could upgrade human-computer interaction (HCI) to a more natural, fluid type of communication that can be employed in various fields. The potential for safer and more efficient occupational environments through neuroadaptivity has been shown for driving [4], aviation [5] and medicine [6], but also for leisure activities such as gaming [7]. Among others, cognitive states like workload [8], error-perception [9] and surprise [10] have been successfully decoded with pBCI. While extensive research has been done to explore average EEG correlates of emotions, there are relatively few studies that demonstrate robust

capabilities for emotional state detection at a single trial level [11,12]. The most common types of features used for emotion classification are event-related potentials (ERPs), frontal EEG asymmetry and event-related desynchronization / synchronization [13]. To investigate single-trial emotion detection from ERPs, a recent study combined workload and stress detection in a social evaluation context [14]. Using a cross-subject classification technique with transfer learning, stress vs. relaxation levels were detected with an average accuracy of over 80%. Single-trial classification of emotion based on ERPs was also achieved for different levels of valence and arousal with a definite advantage for arousal discrimination in [15] and [16]. Another study using EEG recorded while participants were watching music videos managed high classification accuracies for stress levels by using entropy-based features [17]. Our study proposes exploring how well pBCI systems can perform in classifying a specific type of emotion, moral emotion [18]. According to the well-known arousal-valence dimension model of emotions [19], moral violations could evoke high arousal and negative valence emotions such as anger or disgust [20,21]. In contrast, congruent moral stimuli could be associated with low arousal and positive valence. In this investigation, we try to decode moral emotions with pBCI through moral judgements.

*Moral judgement.* We operationalize here moral judgement as the degree of agreement or disagreement to morally-charged contexts. Moral judgement is a complex human reaction that can include both a cognitive and emotional dimension [22,23]. As an automatic and emotional response, moral judgement can be triggered at an unconscious, intuition-based level, determined by a combination of factors such as personality, culture or motivation [24,25] and is associated with deeper structures of the brain [26]. On the other hand, especially when explicit moral reasoning is required, cognitive functions such as inhibition, cognitive conflict, memory and theory of mind processes are engaged and different prefrontal cortical areas become more active [27,28]. A morally-charged stimulus can either resonate with or challenge an individual's moral perspective, thereby evoking a meaningful moral reaction. This depends on the congruency moral stance with one's personal values and experience with a particular topic. This reaction can be recorded with brain imaging methods such as EEG and potentially decoded with pBCI. While some EEG

studies looked at the signal patterns associated with neutral, positive, and negative moral judgements, there has not been much work investigating the feasibility of single-trial moral judgement detection for text stimuli [29]. In [30], 90 morally consistent and inconsistent statements were presented to pre-selected groups consisting of Christian and non-Christian male participants while recording their electroencephalography (EEG) data. The statements were displayed one word at a time, with the final word of each determining the overall moral meaning. In reaction to these key words, a small N400 event related-potential (ERP) was found for morally-incongruent words. Also, a late positive potential (LPP) was found around 500-600 ms. The congruency of the moral words was determined based on participants' religiosity for relevant topics (e.g. "I think euthanasia is acceptable/unacceptable"). Another similar study [31] used morally acceptable or unacceptable statements (aligned or misaligned with social norms) presented word by word to elicit moral agreement or disagreement. They also found an LPP around the fronto-parietal region in the case of unacceptable statements. A more recent study that used a multivariate pattern classification (MVPA) showed that agreement or disagreement to morally-charged statements (e.g. "Wars are acceptable / unacceptable") could be predicted from 180ms following the critical ending words, based on the approval or disapproval with these statements indicated via button presses ("yes" and "no") [32]. Moral attitudes regarding particular topics are acquired throughout one's life and are strongly correlated with views and values assimilated within family, society, and personal experiences. The context in which statements appear is also important in eliciting corresponding moral reactions. Previous studies have shown that negative emotion can that trigger a signalling mechanism, making moral situations more salient [22]. Thus, a realistic emotional context used as an affective priming for the textual stimulus could significantly help in this elicitation, as compared to passive statements devoid of context [33,34]. This might be especially relevant for single trial detection. Also, existing theories on effective emotion elicitation attest to the importance of constructing agents for moral assessments to be attributed to, which also improve the elicitation of moral reactions, making the experience more relatable and impactful [35,36]. In this paper, we investigate the feasibility of moral judgement decoding with pBCI for morally-charged statements presented following affective priming represented by emotional videos on specific topics. Previous work has identified video-based stimuli with audios to be considerably more efficient in emotion elicitation, as they are more realistic [37] and produce the highest number of statistically significant features [38]. While most studies that used affective priming in the context of moral judgement assessment so far have used text-based priming, we explore the use of videos with audio here. In light of an increasingly digitized world and advanced artificial

intelligence systems (AI) such as large language models (LLMs) [39], successful real-time decoding of moral judgement could open a new realm of possibilities for better and more human-compatible HCI through neuroadaptivity.

## MATERIALS AND METHODS

*Participants* This pilot study included 3 participants (2 males, and 1 female) with a mean age of 31 years. The experimental procedure was approved by the Research Ethics Committee of the Brandenburg University of Technology Cottbus-Senftenberg (ID: EK2024-03).

*EEG recording.* Their EEG data was recorded using an ActiCHamp amplifier with 64 active actiCAP slim gel electrodes (Brain Products GmbH, Gilching, Germany). The system provides an electrode montage along the extended international 10-20 system (see <https://www.brainproducts.com/downloads/cap-montages/> for detailed positions). On the used hardware platform the recorded data is natively reference-free and was common-average referenced after recording. The signal was sampled at 500Hz.

*Experiment overview.* The task involved watching videos and reading statements related to 4 social justice issues: immigration, racial discrimination, sexism, and homosexuality. Sixteen videos were presented in a random order, followed by 10 randomized statements (5 morally agreeable/congruent and 5 morally disagreeable/incongruent). The utilized videos were collected directly from YouTube or compiled together using sequences from a longer Youtube video, such that each video lasted approximately 1 minute. They represented a segment from real TV or media news and they were generally found on channels of multimedia news organizations. Each video included audio as well. After each visualisation, the participants would read an instruction informing them the upcoming statements would be comments left under the respective video by people on the internet. Thus, we are framing strangers on the internet as moral agents responsible for their actions, here agreeable or disagreeable statements. In reality, statements were created by experimenters with the help of the publicly available LLM, ChatGPT. The written statements are presented word by word, in a Rapid Serial Visualization Presentation (RSVP) manner [41] with an Optimal Recognition Point (ORP) alignment of the words [42]. Corresponding with the ORP position, the aligned letters in each word were presented in red, while the rest of the letters remained black. While pre-defined labels were set for these key words, the participants also indicated their agreement, disagreement, or uncertainty regarding the read statement by using keyboard buttons (left arrow for *agree*, downward arrow for *uncertain* and right arrow for *disagree*). The position of the agreement level buttons on the screen corresponded with the position of the response keyboard buttons. The ending, key word of each sentence determined the entire moral stance of the

sentence and represented the events used for classification. In total, there were 160 statements and therefore, 160 key words in the task: 80 morally-congruent and 80 morally-incongruent. Following, we will illustrate one video-statements example. One of the videos included in the study was a short news piece on the persecution and abuse gay people experience in Uganda due to strict anti-homosexuality laws. In this, there are depictions of people expressing their disappointment and fear regarding these laws and sequences of politicians communicating morally controversial statements such as: “We are going to reinforce the law enforcement officers to make sure that homosexuals have to space in Uganda.” After this video the sentences in Tab. 1 were presented in a randomized, word by word manner. The speed of the word presentation differed based on the character length of each non-key word, with 700ms base time and 20ms added for each character besides the first one. For instance, the word *must* was presented for 760ms. The ending, morally-charged words were all presented for 1500ms. The task lasted one hour.

Table 1. Example of morally congruent and incongruent statements

Congruent
1. Uganda's laws for gays should be <b>humane</b> .
2. Gay people in Uganda deserve <b>freedom</b> .
3. Gay people in Uganda should be treated with <b>dignity</b> .
4. Gay Ugandan citizens merit more <b>respect</b> .
5. Equality or gay individuals in Uganda is <b>essential</b> .
Incongruent
6. In Uganda, laws regarding gays must be <b>harsh</b> .
7. Uganda's gay people deserve <b>prison</b> .
8. Gay Ugandans should be shown <b>disrespect</b> .
9. Ugandan gays merit more <b>punishment</b> .
10. For Uganda's gays, equality is <b>unacceptable</b> .

**Classification method.** The classification was performed offline, using MATLAB R2022a (The Mathworks, Inc., Natick, MA, USA) and BCILAB 1.4-devel [43]. Responses that did not align with the predefined classes (*congruent* vs. *incongruent*) were excluded from the classification. Thus, in the sentence “Gay people in Uganda deserve freedom.” the predefined label for the word freedom was congruent. If the participants pressed on the “disagree” or uncertain buttons instead, this trial was excluded from the classification. We also explored the classification of moral (congruent and incongruent moral combined trials) vs. neutral trials. The neutral trials were categorized based on list of 86 words that appeared within sentences. Examples of neutral words include: “eventually, ultimately, casual, concept, idea, fact”. A windowed means approach [44] was used for the feature extraction. The data was bandpass-filtered between 0.1 and 15 Hz. Regularized linear discriminant analysis (LDA) with a (5x5)-fold cross-validation was used for

the classification of congruent vs. incongruent trials and moral vs. neutral trials. Epochs of 1 second were extracted with a start time at stimulus onset (key word presentation). We explored two sets of 50 ms time windows in which amplitude is averaged. One set of time windows we used were between 300 and 600 ms after the stimulus, with 6 consecutive time windows. The second set of time windows were set between 400ms and 1000 ms, with 12 consecutive time windows. These windows align with the assumed occurrence of N400 and LPP effects as discussed in [30].

## RESULTS

The average classification results on congruent vs. incongruent classes (*CvsI*) and neutral vs. moral (*NvsM*) for both sets of time can be seen in Tab. 2. Only one participant reached classifier significance for the 400-1000 set, with an accuracy of 65%. The chance level in this case is at 57%, which coincides with the associated average accuracy. In contrast, all classifiers for both time window sets reached significance for the neutral vs. moral trials. Averaged ERP potentials for channels Fz and Cz were obtained for both types of classes after independent component analysis (ICA) and non-brain component removal. ERPs for morally congruent vs. incongruent trials are illustrated in Fig. 1 and ERPs for morally-charged vs. neutral trials are illustrated in Fig. 2.

Table 2. Classification results for congruent vs. incongruent (*CvsI*) and neutral vs. moral (*NvsM*) trials

Time windows	TP (%)	TN (%)	Accuracy (%)
300 - 600 <i>CvsI</i> / <i>NvsM</i>	49 / 83	52 / 69	50 / 78
400 - 1000 <i>CvsI</i> / <i>NvsM</i>	59 / 80	54 / 72	57 / 77

*TP* = True positives (*incongruent*); *TN* = True negatives (*incongruent*)

## DISCUSSION

While decoding accuracy for morally congruent and incongruent trials was not successful with this simple approach, we could observe good decoding accuracies for neutral vs. morally-charged words. This was also reflected in the grand-average ERP. Our results are not entirely surprising, given the difficulty of emotion detection from EEG at a single-trial level [12] and the complexity of moral emotions. A recent pBCI investigation [29] also found chance-level results when looking at the potential of single-trial detection for morally acceptable and objectionable trials on data collected in [30] and [31]. However, we found good performance classification for neutral vs. morally-charged trials. We postulate that while the chosen moral words are relevant enough to produce genuine reactions

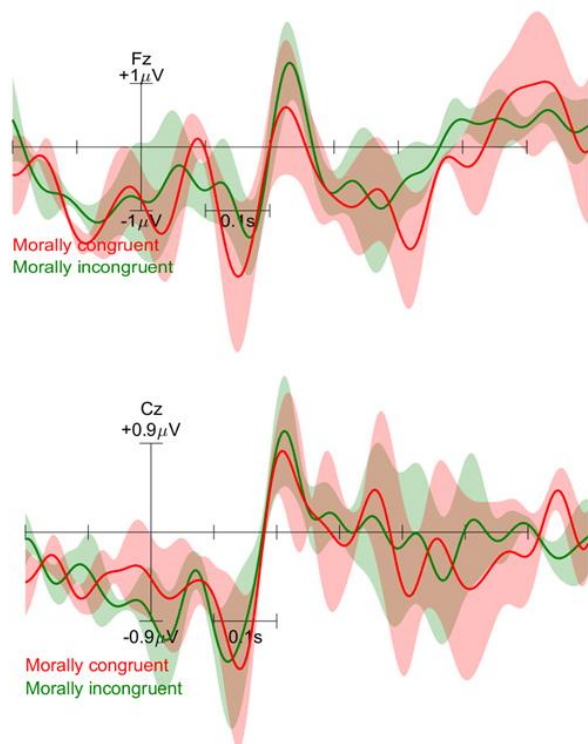


Figure 1. Subject-average ERP potentials for morally congruent and incongruent words.

in comparison to neutral stimuli, the current feature extraction and classification approach might need improvements to better capture potential signal differences between morally congruent and incongruent trials. Encouraging results come from recent studies that explored more sophisticated algorithms and feature extraction methods for emotion detection [17,45]. Another way we plan to improve our results in a larger study is to only include participants that align with a certain profile, such that we can ensure they hold clear moral stances towards the topics. Previous studies have identified the importance of moral attitude strength for effective moral emotion elicitation [46] and the corresponding impact on neural signals [23]. In this study we assumed that participants will have the expected, coherent moral value system. We excluded trials, where the manual answers were incompatible with our assumptions. As we only discarded a few trials, we think the participants shown here share our assumed morality. In the recruitment for the main study following up this pilot, we will pre-assess the moral value system of each participant. More specifically, we will include questionnaires meant to assess the participants' attitudes towards sexism [47], immigration [48], racism [49] and homosexuality [50]. Hence, only participants who highly agree with immigration and homosexuality and highly disagree with sexism and racism will be invited to the study. Successful real-time decoding of mental states in reaction to written stimuli

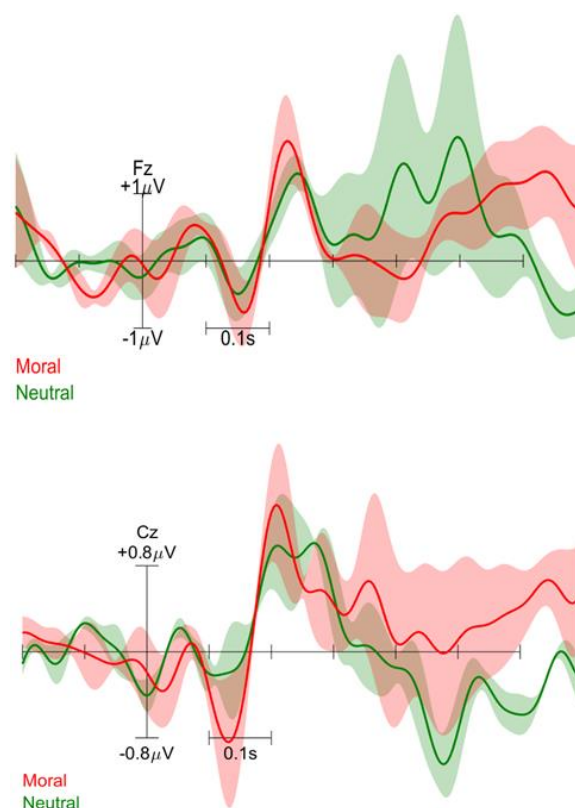


Figure 2 Subject-average ERP potentials for moral and neutral words.

could transform human-computer communication in the context of LLMs. For instance, training of LLM could benefit from replacing or augmenting human explicit feedback in Reinforcement Learning with Human Feedback (RLHF) [51,52] with neural-based implicit feedback [53], potentially offering new solutions for a better synergy between humans and machines.

## CONCLUSION

In this pilot investigation, we looked at the feasibility of single-trial detection of moral judgement from text after video-based affective priming. Our work offers insights into the neural correlates of moral judgement, as well as ideas for classification improvement for a study that includes more participants and better-suited participant profiles.

## REFERENCES

- [1] Zander TO, Kothe C. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J Neural Eng.* 2011 Mar 24;8(2):025005.
- [2] Wolpaw JR. Chapter 6 - Brain-computer interfaces. In: Barnes MP, Good DC, editors. *Handbook of Clinical Neurology.* Elsevier; 2013. p. 67–74.

- [3] Krol LR, Zander TO. Chapter 2 - Defining neuroadaptive technology: the trouble with implicit human-computer interaction. In: Fairclough SH, Zander TO, editors. *Current Research in Neuroadaptive Technology*. Academic Press; 2022. p. 17–42.
- [4] Lin CT, Wu RC, Liang SF, Chao WH, Chen YJ, Jung TP. EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Trans Circuits Syst I Regul Pap*. 2005 Dec;52(12):2726–38.
- [5] Dehais F, Dupres A, Di Flumeri G, Verdiere K, Borghini G, Babiloni F, et al. Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE; 2018. p. 544–9.
- [6] Zander TO, Shetty K, Lorenz R, Leff DR, Krol LR, Darzi AW, et al. Automated Task Load Detection with Electroencephalography: Towards Passive Brain-Computer Interfacing in Robotic Surgery. *J Med Robot Res*. 2017 Mar 1;02(01):1750003.
- [7] Krol LR, Freytag SC, Zander TO. Meyendtris: a hands-free, multimodal tetris clone using eye tracking and passive BCI for intuitive neuroadaptive gaming. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. New York, NY, USA: Association for Computing Machinery; 2017. p. 433–7. (ICMI '17).
- [8] Gerjets P, Walter C, Rosenstiel W, Bogdan M, Zander TO. Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Front Neurosci* [Internet]. 2014;8. Available from: <https://www.frontiersin.org/articles/10.3389/fnins.2014.00385>
- [9] Parra LC, Spence CD, Gerson AD, Sajda P. Response error correction--a demonstration of improved human-machine performance using real-time EEG monitoring. *IEEE Trans Neural Syst Rehabil Eng*. 2003 Jun;11(2):173–7.
- [10] Pawlitzki J, Klaproth O, Krol LR, Zander TO. 2021. Automation surprise in the neuroadaptive cockpit. In *Neuroergonomics Conference*.
- [11] Alarcao SM, Fonseca MJ. Emotions recognition using EEG signals: A survey. *IEEE Trans Affect Comput*. 2019 Jul 1;10(3):374–93.
- [12] Alimardani M, Hiraki K. Passive brain-computer interfaces for enhanced human-robot interaction. *Front Robot AI*. 2020 Oct 2;7:125.
- [13] Al-Nafjan A, Hosny M, Al-Ohali Y, Al-Wabil A. Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review. *NATO Adv Sci Inst Ser E Appl Sci*. 2017 Dec 1;7(12):1239.
- [14] Bagheri M, Power SD. Simultaneous Classification of Both Mental Workload and Stress Level Suitable for an Online Passive Brain-Computer Interface. *Sensors* [Internet]. 2022 Jan 11;22(2). Available from: <http://dx.doi.org/10.3390/s22020535>
- [15] Mathieu NG, Bonnet S, Harquel S, Gentaz E, Campagne A. Single-trial ERP classification of emotional processing. In: 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE; 2013. p. 101–4.
- [16] Liu YH, Wu CT, Kao YH, Chen YT. Single-trial EEG-based emotion recognition using kernel Eigen-emotion pattern and adaptive support vector machine. *Annu Int Conf IEEE Eng Med Biol Soc*. 2013;2013:4306–9.
- [17] Gao Y, Wang X, Potter T, Zhang J, Zhang Y. Single-trial EEG emotion recognition using Granger Causality/Transfer Entropy analysis. *J Neurosci Methods*. 2020 Dec 1;346(108904):108904.
- [18] Haidt J. The Moral Emotions. In: *Handbook of Affective Sciences*. Oxford University Press New York, NY; 2002. p. 852–70.
- [19] Russell JA. A circumplex model of affect. *J Pers Soc Psychol*. 1980;39(6):1161.
- [20] Hutcherson CA, Gross JJ. The moral emotions: a social-functionalist account of anger, disgust, and contempt. *J Pers Soc Psychol*. 2011 Apr;100(4):719–37.
- [21] Chapman HA, Kim DA, Susskind JM, Anderson AK. In bad taste: evidence for the oral origins of moral disgust. *Science*. 2009 Feb 27;323(5918):1222–6.
- [22] Decety J, Michalska KJ, Kinzler KD. The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. *Cereb Cortex*. 2012 Jan;22(1):209–20.
- [23] Hundrieser M, Stahl J. How attitude strength and information influence moral decision making: Evidence from event-related potentials: ERPs in moral decision. *Psychophysiology*. 2016 May 1;53(5):678–88.
- [24] Haidt J. The emotional dog and its rational tail: a social intuitionist approach to moral judgement. *Psychol Rev*. 2001 Oct;108(4):814–34.
- [25] Gaertner SL, McLaughlin JP. Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics. *Soc Psychol Q*. 1983;46(1):23–30.
- [26] Cunningham WA, Raye CL, Johnson MK. Implicit and explicit evaluation: FMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *J Cogn Neurosci*. 2004 Dec;16(10):1717–29.
- [27] Fede SJ, Kiehl KA. Meta-analysis of the moral brain: patterns of neural engagement assessed

- using multilevel kernel density analysis. *Brain Imaging Behav.* 2020 Apr;14(2):534–47.
- [28] Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. The neural bases of cognitive conflict and control in moral judgement. *Neuron.* 2004 Oct 14;44(2):389–400.
- [29] Andreeßen LM. Towards real-world applicability of neuroadaptive technologies: investigating subject-independence, task-independence and versatility of passive brain-computer interfaces [Internet]. BTU Cottbus-Senftenberg; 2023. Available from: [https://opus4.kobv.de/opus4-btu/files/6652/Andreessen\\_Lena.pdf](https://opus4.kobv.de/opus4-btu/files/6652/Andreessen_Lena.pdf)
- [30] Van Berkum JJA, Holleman B, Nieuwland M, Otten M, Murre J. Right or wrong? The brain's fast response to morally objectionable statements: The brain's fast response to morally objectionable statements. *Psychol Sci.* 2009 Sep;20(9):1092–9.
- [31] Leuthold H, Kunkel A, Mackenzie IG, Filik R. Online processing of moral transgressions: ERP evidence for spontaneous evaluation. *Soc Cogn Affect Neurosci.* 2015 Aug;10(8):1021–9.
- [32] Hundrieser M, Mattes A, Stahl J. Predicting participants' attitudes from patterns of event-related potentials during the reading of morally relevant statements - An MVPA investigation. *Neuropsychologia.* 2021 Mar 12;153:107768.
- [33] Demel R, Waldmann M, Schacht A. The Role of Emotions in Moral Judgements: Time-resolved evidence from event-related brain potentials [Internet]. *bioRxiv.* bioRxiv; 2019. Available from: <https://www.biorxiv.org/content/10.1101/541342v1.full.pdf>
- [34] Greenaway KH, Kalokerinos EK, Williams LA. Context is Everything (in Emotion Research). *Soc Personal Psychol Compass.* 2018 Jun;12(6):e12393.
- [35] Gray K, Wegner DM. Dimensions of moral emotions. *Emot Rev.* 2011 Jul 28;3(3):258–60.
- [36] Pantazi M, Struiksma M, Van Berkum J. An EEG study on the role of perspective-taking in the assessment of value-loaded statements [Internet]. Available from: [https://studenttheses.uu.nl/bitstream/handle/20.500.12932/11511/Thesis\\_Pantazi.pdf?sequence=1&isAllowed=y](https://studenttheses.uu.nl/bitstream/handle/20.500.12932/11511/Thesis_Pantazi.pdf?sequence=1&isAllowed=y)
- [37] Rahman MM, Sarkar AK, Hossain MA, Hossain MS, Islam MR, Hossain MB, et al. Recognition of human emotions using EEG signals: A review. *Comput Biol Med.* 2021 Sep;136(104696):104696.
- [38] Masood N, Farooq H. Comparing neural correlates of human emotions across multiple stimulus presentation paradigms. *Brain Sci.* 2021 May 25;11(6):696.
- [39] Levy S. What OpenAI Really Wants. *Wired* [Internet]. 2023 Sep 5 [cited 2023 Sep 13]; Available from: <https://www.wired.com/story/what-openai-really-wants/>
- [40] Kothe C, Medine D, Grivich M. Lab streaming layer (2014). URL: <https://github.com/sccn/labstreaminglayer>.
- [41] Potter M. Rapid serial visual presentation (rsvp): a method for studying language processing. 2018 Apr 17; Available from: <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429505379-5/rapid-serial-visual-presentation-rsvp-mary-potter>
- [42] Brysbaert M, Nazir T. Visual constraints in written word recognition: evidence from the optimal viewing-position effect. *J Res Read.* 2005 Aug;28(3):216–28.
- [43] Kothe CA, Makeig S. BCILAB: a platform for brain-computer interface development. *J Neural Eng.* 2013 Aug 28;10(5):056014.
- [44] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components--a tutorial. *Neuroimage.* 2011 May 15;56(2):814–25.
- [45] Arjun, Rajpoot AS, Panicker MR. Subject independent emotion recognition using EEG signals employing attention driven neural networks. *Biomed Signal Process Control.* 2022 May 1;75:103547.
- [46] Ugazio G, Lamm C, Singer T. The role of emotions for moral judgements depends on the type of emotion and moral scenario. *Emotion.* 2012 Jun;12(3):579–90.
- [47] Glick P, Fiske ST. The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *J Pers Soc Psychol.* 1996;70(3):491–512.
- [48] Pratto F, Sidanius J, Stallworth LM, Malle BF. Social dominance orientation: A personality variable predicting social and political attitudes. *J Pers Soc Psychol.* 1994 Oct;67(4):741–63.
- [49] McConahay JB. Modern racism, ambivalence, and the Modern Racism Scale. 1986; Available from: <https://psycnet.apa.org/record/1986-98698-004>
- [50] Fisher TD, Davis CM, Yarber WL. Handbook of Sexuality-Related Measures. Routledge; 2013. 680 p.
- [51] Stiennon N, Ouyang L, Wu J, Ziegler DM, Lowe R, Voss C, et al. Learning to summarize from human feedback [Internet]. *arXiv [cs.CL]*. 2020. Available from: <https://proceedings.neurips.cc/paper/2020/file/1f89885d556929e98d3ef9b86448f951-Paper.pdf>
- [52] Casper S, Davies X, Shi C, Gilbert TK, Scheurer J, Rando J, et al. Open problems and fundamental limitations of reinforcement learning from human feedback [Internet]. *arXiv [cs.AI]*. 2023. Available from: <http://arxiv.org/abs/2307.15217>
- [53] Xu D, Agarwal M, Fekri F, Sivakumar R. Playing Games with Implicit Human Feedback.

# REAL-TIME NEUROFEEDBACK ON INTER-BRAIN SYNCHRONY: CURRENT STATES AND PERSPECTIVES

K. Won<sup>1</sup>, L. Pillette<sup>2</sup>, M. J.-M. Macé<sup>2</sup>, A. Lécuyer<sup>1</sup>

<sup>1</sup>Inria, Univ. Rennes, IRISA, CNRS, Rennes, France

<sup>2</sup>Univ. Rennes, Inria, CNRS, IRISA, Rennes, France

E-mail: kyung-ho.won@inria.fr

**ABSTRACT:** During neurofeedback (NFB) user training, participants learn to control the feedback associated with specific components of their brain activity, also called neuromarkers, to improve the cognitive abilities related to these neuromarkers, such as attention and mental workload. The recent development of methods to record the activity of several people's brains simultaneously opens up the study of neuromarkers related to social interactions, computed from inter-brain synchrony (IBS). Here, we review the previous articles that trained participants to control electroencephalographic neuromarkers computed from inter-brain metrics. The topic remains relatively unexplored as we only identified seven articles in the literature. We specifically studied the characteristics of the user's training, i.e., instruction, task and feedback, and the neuromarkers used to provide feedback. The reported results are promising as four studies including subjective measures of interaction report higher interaction and relationship scores with higher IBS during NFB training. Finally, we draw guidelines, identify open challenges, and suggests recommendations for future studies on this topic.

## INTRODUCTION

Neurofeedback (NFB) refers to a paradigm that trains participants to voluntarily modulate specific components of their brain activity, also called neuromarkers. The desired modulation of these neuromarkers is rewarded by sensory stimuli based on the acquired neurophysiological data [1]. The aim is to improve cognitive abilities associated with these neuromarkers as a results of the NFB training. NFB has been widely used for clinical and non-clinical purposes, such as a treatment of social anxiety disorder (SAD) [2] and improving brain-computer interface (BCI) performance [3]. The development of simultaneous multi-brain recording enabled the study of neuromarkers that are specific to social interaction with many new promising applications. In this regard, Saul et al. [2] suggested developing new NFB-based treatments of SAD based on neuromarkers acquired from multi-brain recordings. This suggestion is based on the previous finding of a relationship between inter-brain synchronization and SAD [4]. In addition to self-centered individual NFB, the extended use of NFB aims to integrate individual par-

ticipants' brain activities in a common framework to detect synchronous brain activities during social interaction paradigms.

However, there is a lack of literature specifically focused on brain-to-brain real-time interactions, as the majority of NFB studies concern individual brain activities [1, 5] rather than inter-brain activities. Also, the majority of hyperscanning studies are investigating offline neuromarkers rather than online synchronization-related neuromarkers [6]. In this regard, the current review aims to provide an overview of the current state of research on inter-brain synchrony-based NFB to identify related challenges, provide recommendation and suggest ideas for future research. We targeted online inter-brain NFB experiments, focusing on feedback scenarios involving more than one participant and online inter-brain neuromarkers. We report on the validation of these neuromarkers, in addition to common essential information in these NFB studies: goals, instructions, sensory stimuli and outcomes. Moreover, we focused this review on electroencephalography (EEG) based on the analysis from Saul et al. that EEG would be the neuroimaging technique of choice for inter-brain NFB studies as EEG provides brain activity measures with high temporal resolution, a requirement for duration-bound experimental stimuli [2]. EEG provides also great portability, which is an advantage for studies in naturalistic environments, as the EEG is not limited to specialized spaces, such as shielded rooms.

## MATERIALS AND METHODS

*Study selection:* This review mainly focuses on both the user training and the neuromarkers characteristics of these NFB protocols. To retrieve all relevant papers, a systematic literature search has been conducted in the Scopus and Web of Science databases as described in Figure 1. The following keywords were used: (*EEG OR Electroencephalography*) AND (*Brain-Computer Interface OR Brain-Machine Interface OR Neural Interface OR Neurotherapy OR Neurofeedback*) AND (*IBS OR Inter-Brain OR Hyperscanning OR Brain Synchrony OR Social interaction*). Papers published until the end of December 2023 were included. Two of the authors (KW and LP) reviewed the titles, abstracts, and KW read

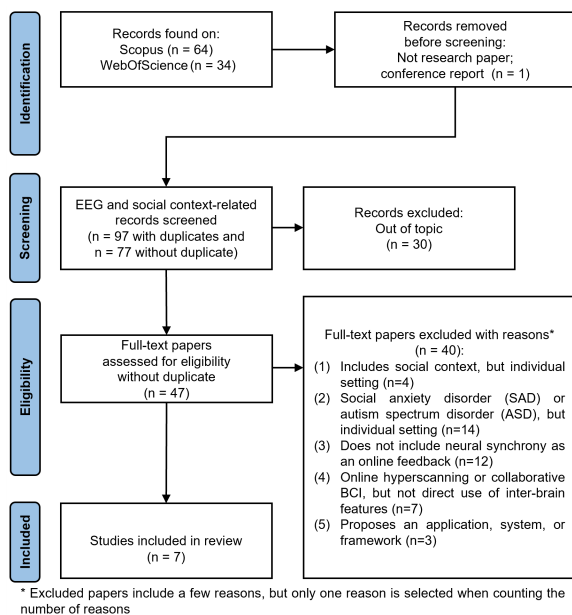


Figure 1: PRISMA flowchart for the inclusion of studies.

the whole papers if there was a doubt. Both reviewers agreed on the selection of the studies. The inclusion criteria used were the following ones: (1) presentation of experimental results on NFB training using inter-brain synchrony/similarity for the online feedback (2) electroencephalographic neuromarker used for the NFB training. Studies that did not satisfy both of the inclusion criteria were excluded. The exclusion criteria were the following ones: (1) non-clinical studies including social context, but individual recording and neurofeedback, (2) clinical studies including social context, but individual recordings and neurofeedback, (3) studies not incorporating neural synchrony as online feedback but as offline measures (4) hyperscanning studies not incorporating inter-brain features to limit our focus on feedback training based on direct measure of inter-brain synchrony (voting for decision-making and competition are also interactions, but the performances are often calculated in online scenarios by combining individual features or classification results rather than inter-brain synchrony) (5) studies proposing the development of an algorithm, an application, or a framework instead of a neurophysiological analysis.

**Data extraction:** In this review, we extracted information related to the: (1) objective and results of the NFB study, (2) online neuromarkers used for the NFB, (3) interaction task and feedback, (4) instructions provided to the participants. Additional validation or analyses are described, such as re-calculation of synchrony with statistical validation and correlation analysis with subjective measurements.

## RESULTS

Our search led to the inclusion of seven papers in this review. A summary of experimental scenarios, feedback, online features, and overall characteristics for each se-

lected paper is provided in Tab. 1. Among the selected studies, the two studies [7, 8] from the same group include different numbers of participants and offline data analysis, but they had the same characteristics that we focused on in this review. Eventually, we summarized these two studies in the same row (the fourth row in Tab. 1).

**Neurofeedback outcomes:** To investigate neurofeedback outcomes, five studies out of the selected studies conducted subsequent offline analyses to validate synchrony and investigate correlations to subjective measurements through more sophisticated artifact reduction and complex synchrony metrics with statistical validation [7–11]. For example, Susnoschi Luca et al. [9] re-calculated synchrony using phase-locking value (PLV) instead of their online measure (relative alpha) and validated it through a permutation test. As a result, they found that the PLV obtained during the baseline period did not pass the permutation test, meaning that participants did not exhibit synchrony during this phase. In contrast, the PLV during the task showed significant synchrony in theta, alpha, and beta bands, showing that the synchrony-based NFB could also affect different bands other than the target bands. Müller et al. found a significant positive correlation between the synchrony in theta band and the self-reported partner’s likeability [10], and Salminen et al. found higher self-reported empathy toward partners with synchrony-based feedback blocks compared to no-feedback blocks [7, 8]. Lastly, Dikker et al. re-calculated synchrony with two different measures, imagery part of coherence (iCOH) and projected power correlation (PPC), and found significant correlations between PPC (7-8 Hz) and relationship duration, iCOH (21-22 Hz) and social closeness, and PPC (14-15 Hz) and personal distress.

**Neurofeedback scenario and stimuli:** Among the selected papers, five studies provided visual feedback [7–11], and two studies involved auditory feedback [12, 13]. As visual feedback, Dikker et al. [11] used a mutual wave machine in a dome-like environment during public exhibitions where two participants were seated face-to-face. The participants were engaged in a 10-minute face-to-face interaction, and the light pattern of the mutual wave machine was rendered so that higher synchrony between partners corresponded to brighter lighting projected onto each surface. Salminen et al. [7, 8] designed a shared virtual environment of meditation called DYNECOM, where multiple avatars representing participants were sitting in a ring on a small shrine-like platform with natural wind sounds. A bridge connected the two facing avatars, and the participants were instructed to concentrate on empathetic feelings toward their partner. The synchrony between the two participants modulated the intensity and color of the light shining on both sides of the bridge between the two avatars. Susnoschi Luca et al. [9] designed a GUI that consists of two gauge bars on the left and right-sides representing the relative alpha for each participant and a seesaw in the middle indicating the balance between their relative alpha. During the collaborative task,

the participants were instructed to maintain their relative alpha levels within 5% to keep the balance of the seesaw without speaking and body movements. Likewise, Müller et al. [10] designed an NFB task that consists of two balls. During the task, the balls got closer or further away from each other depending on the neural synchrony between the two participants. The participants were instructed to move the balls towards each other by controlling their brain activity accordingly, relying on various mental strategies such as relaxation, mental calculation, thoughts generation, etc.

Regarding auditory feedback, two studies controlled music, such as volume and beats, using neural synchrony. Winters and Koziej [12] mapped neural synchrony to control the volume of an ambient music stream (higher synchrony was mapped to increased volume) played through speakers during natural face-to-face interactions in public exhibitions. Ceccato et al. [13] developed a Brain-Computer Musical Interface (BCMI) that changed music through the neural synchrony estimation by increasing or decreasing the interval between musical notes, yielding more pleasant music in case of high synchrony between two participants.

*Task instructions:* We investigated whether the task instructions provided in the studies were explicit, i.e., participants were told that the synchrony in their brain activity modulated the perceived stimuli. Salminen et al. [7, 8] implicitly instructed participants to utilize the information provided by the environment and concentrate on empathetic feelings toward their pair, whereas Winters and Koziej explicitly informed participants that they would hear the music when their brain activities were synchronized [12]. Dikker et al. compared implicit vs explicit instructions, and found that explicit instructions increased synchrony significantly over time, whereas implicit instructions induced no significant changes [11].

Additionally, we divided the studies into goal-oriented and not goal-oriented instructions depending on whether the participants had a goal to achieve. For example, Susnoschi Luca et al. and Müller et al. [9, 10] gave their participants the goal to keep the balance of the seesaw and to move balls towards each other by modulating their brain activity. The participants focused on the balls and seesaw controlled by their synchrony during the NFB. On the other hand, the rest of the studies places participants in a natural face-to-face interaction and did not specify any goal. Even though synchrony between participants modulated the environment (light patterns, music, etc.), there was no direct instruction on whether they should modulate their environment to desired states, such as maintaining specific light patterns or volume or beat of the music. In these face-to-face experiments, the participants just focused on their partner.

*Online neuromarkers:* To calculate inter-brain synchrony as a neurofeedback feature during online scenarios, amplitude or phase coupling measures between two participants were assessed. Amplitude coupling was obtained by comparing band power and filtered EEG ampli-

tudes. For example, Susnoschi Luca et al. [9] compared the relative alpha power (RA) at Pz electrode for each participant and instructed them to maintain their RA within 5% of each other to maintain a collaborative state within pairs. Likewise, Winters and Koziej [12] compared the average alpha band power of two participants at AF7, AF8, TP9, and TP10 electrodes. Dikker et al. [11] assessed the similarity by calculating average and highest Pearson correlation coefficients from all electrode pairs between two participants (AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, and O2) in delta (1-4 Hz), theta (4-7 Hz), alpha (7-12 Hz), and beta (12-30 Hz) frequency bands. Salminen et al. [7, 8] compared frontal alpha asymmetry between F3 and F4 electrodes between two participants.

Phase coupling was obtained by extracting the instantaneous phase from EEG data and calculating the phase difference or consistency of the difference. Müller et al. [10] calculated the coupling strength as a sum of all possible electrode pairs within and between the two participants through frontocentral (F3, Fz, F4, C3, Cz, and C4) Absolute Coupling Index (ACI) in four frequency bins (2.5, 5, 10, and 20 Hz). The ACI counts the number of samples achieving phase differences ranging between  $-\pi/4$  and  $\pi/4$ . Ceccato et al. [13] calculated phase locking value (PLV) between two participants (Fz, F3, F4, C3, Cz, C4, Pz, and Oz electrodes). The PLV calculates instantaneous phase difference between two signals for each trial and measures variability of the difference across trials, assuming the difference between phase-coupled signals varies little, which is called phase-locking.

## DISCUSSION

*Neurofeedback outcomes:* The selected studies reported neural synchrony during the feedback blocks and correlations between the neural synchrony and self-reported questionnaires, such as social relationship duration and likeability. These results show the presence of neural synchrony during the NFB and its relationship with subjective measures of interaction. However, it is difficult to investigate the progress of these outcomes over time since all the selected studies designed single-session experiments, unlike classical NFB studies, which often includes multi-session [5]. This is mostly likely related to the significant additional time and efforts for installing EEG headsets on multiple people. The selected studies observed significant neural synchrony between participants during the feedback period. However, without comparing before and after the feedback, it is unclear how the NFB modulates neural synchrony over time compared to the pre-feedback period. Furthermore, self-reported questionnaires were collected once either at the beginning or at the end of the experiment with a single condition, so it is difficult to investigate whether the synchrony-based NFB changed the subjective experience and feeling of partners along the experiment, even though the correlation between the synchrony and self-reported

questionnaires post NFB were obtained. A few selected studies designed control condition, such as a no-feedback condition in addition to a feedback condition [7, 8, 10]. Salminen et al. designed the NFB with various feedback conditions and observed higher self-reported empathy after EEG-based feedback compared to no-feedback condition [7, 8]. Interestingly, Müller et al. designed conditions with normal, fake, and negative feedback (rewarded on weaker synchrony). They found that negative feedback achieved the highest neurofeedback performance, and normal neurofeedback performance achieved the lowest performance in the ball task, meaning that the participants could achieve their best performance when their synchrony was negatively mapped to the balls [10]. Taken together, more evidence is still needed to test the feasibility of the inter-brain NFB, such as how long the feedback effects could last, how many sessions are required to train the synchrony, and how much inter-brain NFB improves social interaction. Therefore, it seems necessary to design a multi-session experiment including pre/post-baseline analysis and various control conditions to confirm whether the neural synchrony could be trained by an inter-brain NFB and investigate the progresses over time.

*NFB scenarios and stimuli:* The selected studies designed NFB scenarios, including 2D object control [9, 10], natural face-to-face interaction [11–13], and meditation [7, 8]. During the NFB condition, feedback was delivered via visual and auditory stimuli. Visual stimuli are effective sensory stimuli for a single task as they are intuitive, and the two studies that used it used objects visually moving according to the neural synchrony [9, 10]. On the other hand, for multitasking situations, using only visual stimuli may not be the ideal stimuli because participants should separate their visual attention on each task. During the NFB tasks designed by Dikker et al. and Salminen et al., participants had to look at the same time at their partner and at the visual feedback (light) controlled by the neural synchrony. The switch of their gaze back and forth between their partner and the visual feedback could have decreased their attention. Attention to their partner and to the feedback stimuli was probably more easily shared in the two studies delivering auditory feedback (music control) during face-to-face interactions [12, 13]. Participants were able to look at their partner while listening to the music controlled by the neural synchrony. Among unused sensory stimuli in the selected studies, tactile stimulation could be used as feedback for the multitasking paradigm. Jeunet et al. incorporated tactile feedback in a multitasking environment consisting of motor imagery BCI and counting visual distracters task [14], showing a better performance than in a condition with visual feedback only. In this regard, the tactile stimulation could be applied within natural face-to-face interactions while maintaining visual attention to partners. However, the effects of those multi-sensory stimuli on neural synchrony and workload should be investigated.

*Task instructions:* Among the selected studies, we in-

vestigated task instructions in terms of explicitness and goal-orientation. Regarding goal-orientation instructions, Susnoschi Luca et al. and Müller et al. [9, 10] provided specific goals to achieve to their participants, such as moving balls towards each other and keeping the balance of a seesaw. On the other hand, the rest of the selected studies used indirect instructions in the feedback scenarios. Even though the participants were informed that their synchrony would change their environments, their main task was to interact face-to-face with their partner or to concentrate on empathetic feelings. Those instructions can relate to active and passive NFB, as Saul et al. discussed that their difference is whether the NFB platform responds to a participant who is trying to modulate brain activity to reach or maintain a certain pattern, or to a participant who does not consciously attempt to modulate their brain activity but rather interacts naturally with the setup [2]. Even though it is unclear whether the participants who received non goal-oriented instructions tried to modulate their brain activity, the fact that they did not have a specific goal may have led to differences from the participants who received goal-oriented instructions.

Comparing directly explicit and implicit instructions, Dikker et al. found that a group who was explicitly told that their neural synchrony would be reflected in light patterns showed significantly increased neural synchrony over time, whereas a group with implicit instructions did not show significant changes [11]. They hypothesized that the explicit instruction would function as an incentive for participants to remain focused on the interaction. It shows that instructions can increase or decrease the effectiveness of the NFB on neural synchrony by influencing participants' comprehension of the task and environment. It remains unclear which instructions and scenarios effectively enhance neural synchrony through feedback, so future research should further investigate the effect of different instructions.

*Online neuromarkers:* In general, inter-brain synchronization measures, such as phase synchronization, could be obtained after dedicated pre-processing, excluding bad channels and trials, and reducing noise components, such as motion artifacts and eye blinking. Afterwards, instantaneous phase is obtained for each channel and compared to different channels (intra and inter-brain), yielding adjacency matrices. Each connectivity is then compared to surrogate data to validate that the connectivity is statistically significant. However, feedback must be delivered nearly in real time, which limits the processing time to calculate common synchrony metrics with statistical validation. As a result, we found that the selected studies frequently used band power comparisons and correlation of EEG amplitudes to calculate synchrony as online features instead of complex synchrony metrics used in previous hyperscanning studies, as summarized in the following reviews [15, 16]. A few selected studies conducted the subsequent analysis offline to validate the effects of the NFB [7–11]. Besides band power-based features, one selected study utilized phase-locking value (PLV) from

eight electrode channels as neural synchrony, but it needs to be further validated, as they only included one pair, and the PLV was not investigated in-depth with statistical validation [13]. Another selected study calculated phase coupling measures online from six electrode channels [10], showing that some measures could be obtained online. In this regard, it would be interesting to investigate the relationship between the online and offline neuromarkers, as the selected studies that conducted further offline analysis and obtained different synchrony measures did not compare those online and offline measures [9, 11].

*Future direction:* We discussed the characteristics of the current inter-brain NFB studies and suggested a few recommendations for future research regarding training features, instructions, and feedback scenarios to address the current concerns. In summary, inter-brain NFB should consider comparing online features to complex synchrony features, use explicit instructions, and investigate the training effects over time. With proper validation of neural synchrony enhancement between two individuals, it seems to have the potential to address difficulties with daily social interactions, as the selected studies found correlations in the neural synchrony between partners during NFB tasks with pairs' likability [10], empathy [7, 8], and social closeness [11].

## CONCLUSION

This review provided a comprehensive overview of the current state of inter-brain synchrony-based NFB. Online synchrony features in amplitude and phase coupling and subsequent offline analyses were identified. Regarding the feedback scenarios and outcomes, we observed the importance of instruction and the necessity of a multi-session experimental design. We hope this review contributes to the groundwork of future investigations into inter-brain NFB based on inter-brain synchrony.

## ACKNOWLEDGEMENT

This work was supported in part by a grant from the GENESIS project (CHIST-ERA ANR-21-CHRA-0001-01, January 2022).

## REFERENCES

- [1] Vernon DJ. Can Neurofeedback Training Enhance Performance? An Evaluation of the Evidence with Implications for Future Research. *Applied Psychophysiology and Biofeedback*. 2005;30(4):347–364.
- [2] Saul MA, He X, Black S, Charles F. A Two-Person Neuroscience Approach for Social Anxiety: A Paradigm With Interbrain Synchrony and Neurofeedback. *Frontiers in Psychology*. 2022;12:568921.
- [3] Wan F, Da Cruz JN, Nan W, Wong CM, Vai MI, Rosa A. Alpha neurofeedback training improves SSVEP-based BCI performance. *Journal of Neural Engineering*. 2016;13(3):036019.
- [4] Wang MY, Luan P, Zhang J, Xiang YT, Niu H, Yuan Z. Concurrent mapping of brain activation from multiple subjects during social interaction by hyperscanning: A mini-review. *Quantitative Imaging in Medicine and Surgery*. 2018;8(8):819–837.
- [5] Kuznetsova E, Veilahti AVP, Akhundzadeh R, Radev S, Konicar L, Cowley BU. Evaluation of Neurofeedback Learning in Patients with ADHD: A Systematic Review. *Applied Psychophysiology and Biofeedback*. 2023;48(1):11–25.
- [6] Barde A, Gumilar I, Hayati AF, Dey A, Lee G, Billingham M. A Review of Hyperscanning and Its Use in Virtual Environments. *Informatics*. 2020;7(4).
- [7] Salminen M, Ruonala A, Timonen J. DYNECOM: Augmenting Empathy in VR with Dyadic Synchrony Neurofeedback DYNECOM: Augmenting Empathy in VR with Dyadic Synchrony Neurofeedback. no. January. 2019.
- [8] Salminen M *et al.* Evoking Physiological Synchrony and Empathy Using Social VR With Biofeedback. *IEEE Transactions on Affective Computing*. 2022;13(2):746–755.
- [9] Susnoschi Luca I, Putri FD, Ding H, Vucković A. Brain Synchrony in Competition and Collaboration During Multiuser Neurofeedback-Based Gaming. *Frontiers in Neuroergonomics*. 2021;2.
- [10] Muller V, Perdakis D, Mende MA, Lindenberger U. Interacting brains coming in sync through their minds: An interbrain neurofeedback study. *Ann N Y Acad Sci*. 2021;1500(1):48–68.
- [11] Dikker S *et al.* Crowdsourcing neuroscience: Inter-brain coupling during face-to-face interactions outside the laboratory. *Neuroimage*. 2021;227:117436.
- [12] Winters RM, Koziej S. An auditory interface for real-time brainwave similarity in dyads. In: *Proceedings of the 15th International Audio Mostly Conference*. 2020, 261–264.
- [13] Ceccato C, Pruss E, Vrins A, Prinsen J, Alimardani M. BrainiBeats: A dual brain-computer interface for musical composition using inter-brain synchrony and emotional valence. In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, 1–7.
- [14] Jeunet C, Vi C, Spelmezan D, N’Kaoua B, Lotte F, Subramanian S. Continuous Tactile Feedback for Motor-Imagery Based Brain-Computer Interaction in a Multitasking Context. In: *Human-Computer Interaction – INTERACT 2015*. Springer International Publishing: Cham, 2015, 488–505.
- [15] Dumas G, Nadel J, Soussignan R, Martinerie J, Garnero L. Inter-Brain Synchronization during Social Interaction. *PLoS ONE*. 2010;5(8):e12166.
- [16] Dumas G, Fairhurst MT. Reciprocity and alignment: Quantifying coupling in dynamic interactions. *Royal Society Open Science*. 2021;8(5):rsos.210138, 210138.

Ref.	Dyad	EEG (nb elec.)	Objective	Online feature	Feedback Scenario	Instructions	Outcomes
[9]	10	g.USBamp (16)	To investigate how indirect, non-muscular social interaction during neurofeedback affects players' oscillatory brain activity	Difference of relative alpha	Visual (Screen, 2D seesaw)	Keep the balance of the seesaw	Inter-brain synchrony existed only during the task.
[10]	25	Brain Products (6)	To explore whether inter-brain synchronization (IBS) is amenable to neurofeedback (NFB)	Absolute coupling index (ACI)	Visual (Screen, 2D ball and pendulum)	Move balls towards each other	NFB showed stronger IBS in delta, theta, and beta, and weaker IBS in alpha than rest.
[11]	360	Emotiv (14)	To test the hypothesis that real-time reflection of synchrony would motivate participants to remain socially engaged with each other for the duration of the interaction	Correlation between EEG	Visual (FtF, light pattern)	Explicit group were told that the light patterns reflected the IBS. Non-explicit group: no instruction	Low alpha and beta IBS were correlated to relationship duration and social closeness. Explicit groups showed stronger IBS at beta.
[7], [8]	21, 36	Brain Products (6)	To examine how the different neurofeedback types affected physiological synchrony between the users of the shared space and their self-reported empathy	Difference of alpha asymmetry	Visual (VR-FtF meditation)	Concentrate on empathetic feelings and direct them at partners	NFB induced higher empathy and synchrony of alpha asymmetry compared to non-feedback.
[12]	NI*	Muse (4)	To design the neurofeedback instrument to be easy to understand and related and pleasant for members of the general public in an exhibition context	Similarity of alpha power	Auditory (FtF, music control)	"When your brainwaves are similar, you will hear the music."	-
[13]	1	Modified g.tec Unicorn (8)	To apply inter-brain synchronization metrics for the development of a multi-user passive brain-computer musical interfaces (BCMI)	Phase locking value (PLV)	Auditory (FtF, music generation)	-	-

Table 1: Summary of the information extracted from our selected papers. \* 4 exhibitions without any number of participants specified. Screen means that participants looked at the screen, and FtF means that participants were seated Face-to-Face during the task.

## TO REPEAT OR NOT TO REPEAT? ERP-BASED ASSESSMENT OF THE LEVEL OF CONSCIOUSNESS - A CASE STUDY

S. Halder<sup>1</sup>, A. Matran-Fernandez<sup>1</sup>, R. Nawaz<sup>1</sup>, M. Lopes da Silva<sup>2</sup>, T. Bertoni<sup>3</sup>, J.-P. Noel<sup>3</sup>, J. Jöhr<sup>2</sup>, A. Serino<sup>3</sup>, K. Diserens<sup>2</sup>, R. Scherer<sup>1</sup>, S. Perdikis<sup>1</sup>

<sup>1</sup>School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom

<sup>2</sup>Department of Clinical Neurosciences, Acute Neurorehabilitation Unit, Division of Neurology, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

<sup>3</sup>MySpace Lab, Department of Clinical Neurosciences, University Hospital of Lausanne, University of Lausanne, Lausanne, Switzerland

E-mail: [serafeim.perdikis@essex.ac.uk](mailto:serafeim.perdikis@essex.ac.uk)

**ABSTRACT:** Determination of the wakefulness and consciousness state in patients with disorders of consciousness (DOC) is vital for clinical decision-making. Typically, behavioral indicators and motor responses are employed. Recent advancements in neuroimaging have enabled motor independent assessment of DOC patients.

We present a single-case analysis of a 24-year-old female, selected from a sample of  $n=77$  patients, diagnosed with a DOC. We investigated the single-trial classification of stimuli within the peri-personal space (PPS) using event-related potential (ERP) features. Data from two sessions, conducted ten days apart, were analysed.

We observed significant differences in classification accuracies between sessions (high in session one, low in session two), which did not correspond to the patient's recovery from UWS to MCS. ERP analyses confirmed the difference between sessions, supporting the observed changes in classification accuracies.

Our study underscores the importance of longitudinal assessments to accurately diagnose DOC patients. In future research we aim to expand our analyses to the full dataset.

### INTRODUCTION

Reliably determining the state of wakefulness and consciousness of patients with disorders of consciousness (DOC) is crucial for clinical decision-making, providing appropriate care and ensuring patient rights. Usually, behavioral indicators and motor performance in response to specific instructions are used to determine this. The Glasgow Coma Scale [1], for example, assesses a person's level of consciousness based on their ability to open their eyes and perform verbal and motor responses. The Coma Recovery Scale-Revised (CRS-R) [2] is a more comprehensive assessment tool that covers multiple domains and allows for a more detailed assessment and differentiation between states of consciousness such as coma, vegetative state (VS), minimally conscious state (MCS), and locked-in syndrome.

In recent times, researchers have been exploring functional neuroimaging technologies and brain-computer interface-based approaches with the aim to detect unique cognitive patterns when assessing the consciousness state of patients who cannot exhibit motor behavior due to brain injuries [3–5]. One of the several brain networks that have been targeted for this purpose is the cortical network that encodes the Peri-Personal Space (PPS). The PPS is the space surrounding the body that defines the immediate physical domain and is relevant to the interaction between self and others or self and the environment [6]. It is assumed that the related cortical network is linked to bodily self-consciousness and therefore hypothesised to be altered in patients with DOC. Indeed, a physiological index of PPS was identified in evoked electroencephalogram (EEG) responses to tactile, auditory, or audio-tactile stimulation at distances within and outside the PPS [7]. Seventeen patients with DOC participated in the study. The results suggest that the extracted multi-sensory evoked responses degrade in patients with DOC and correlate with the Lempel-Ziv complexity, a metric used to predict global states of consciousness in continuous EEG signals, but not with CRS-R scores [7]. Although these results seem to be in line with neuroscientific findings, they are not yet conclusive and more data is needed to make more precise statements. Among other things, because the Lempel-Ziv complexity as a measure of conscious experience has been called into question [8].

To minimise errors in diagnosis, it is recommended to repeat the CRS-R at least five times within a time period of a few weeks [9]. Given the non-stationarity and inherent variability of EEG signals, the question arises of how many repetitions are required before reliable clinical decision about the state of consciousness of a patient with DOC can be made when using EEG. To answer this question, the experiment in [7] was repeated in a larger cohort of patients with DOC and the EEG and CRS-R assessments were repeated several times per patient. The study is still ongoing, but in this paper we present initial

results and a case study that highlights and emphasizes the need for repeated measurements for making informed decisions.

## MATERIALS AND METHODS

**Patients:** A dataset of 84 patients (23 female, median age 53 years, range=18–84) with a disorder of consciousness (CRS-R at assessment median=15, range=0–23) was recorded at the University Hospital of Lausanne (CHUV), Switzerland. Seven patients were excluded from the analysis in this paper due to incomplete data. For the remaining  $n=77$  patients a total of 202 sessions (median=2 sessions, range=1–7 sessions) were recorded.

The patient (Patient A) selected for detailed analysis in this paper was 24 years old at the time of the experiment, female and admitted with a traumatic brain injury. Session 1 was performed two days and Session 2 twelve days post-admittance to the acute care unit. The patient was diagnosed with unresponsive wakefulness syndrome (UWS; CRS-R 7) in Session 1 and minimally conscious state minus (MCS-; CRS-R 11) in Session 2. Approximately three months post-injury the patient emerged from the MCS.

**Experiment:** Three different stimuli were administered: (1) auditory close (AC; distance 5 cm from extended arm; 65.2 dB SPL; 50 ms of white noise via speaker), (2) auditory far (AF; distance 75 cm; 64.1 dB SPL; 50 ms of white noise via speaker) and (3) tactile (T; two FES electrodes attached to dorsal part of arm near elbow; 50 ms of continuous, sub-threshold stimulation at 35 Hz). Furthermore, auditory and tactile were combined (tactile + auditory close (TAC); tactile + auditory far (TAF)). One experimental block consisted of 50 presentations for each of the five stimuli (250 total). Sessions were planned as a set of three blocks. However, this was not always possible due to constraints of the clinical environment. Further details can be found in the description of the original study [7].

**EEG recording:** EEG data were recorded with 16 channels positioned at Fz, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, and CP4 (all referenced to the right earlobe) at a sampling rate of either 500 Hz or 512 Hz (g.USBamp or g.Nautilus respectively, both by g.tec medical engineering GmbH, Graz, Austria.).

**Preprocessing:** Recordings were read in GDF format using Python MNE [10]. Continuous data were high-pass filtered at 1 Hz. On epoched data (-1 to 2 s) bad channels were marked using RANSAC [11] and bad trials were marked using AutoReject [12, 13]. Next a 20 Hz low-pass filter was applied to the continuous data. The continuous data was then epoched (-1 to 2 s referred to stimulus onset), bad channels were interpolated, and trials rejected. All the 512 Hz epochs were resampled to 500 Hz. No baseline was applied. This resulted in a median number of 1358 trials per patient (range=319–4810 trials).

**Classification:** For classification, epochs from 0 to 1 s

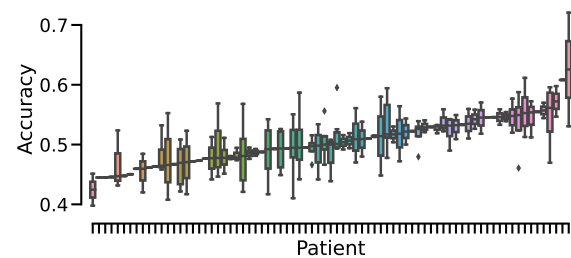


Figure 1: Single-trial accuracy at distinguishing close and far in auditory only mode (i.e., AC vs. AF). Each boxplot represents the data of one patient across all sessions of that patient. Patients on the x-axis were sorted by median accuracy. A high variance of accuracy was expected as we assumed that some patients in the sample would not be conscious. For this paper we decided to investigate the patient with the highest median accuracy (on the far right; referred to as Patient A in this paper).

after stimulus presentation were used. These epochs were resampled to 10 Hz. We trained a shrinkage linear discriminant analysis (LDA) classifier on two classes (AC vs. AF and TAC vs. TAF) and five classes (All vs. All) using scikit-learn [14]. Performance was assessed via mean accuracy using stratified 10-fold cross-validation independently on each session. We used the median accuracy across sessions to select the patient for discussion in this paper.

**ERP Analysis:** We compared the ERPs between Session 1 and Session 2 by computing the median response across channels CP1, CPz and CP2 in the time window from -0.25 s to 1.0 s around stimulus presentation using the pre-processed data. Furthermore we performed a time-frequency decomposition using eight Morlet Wavelets in a range from 2–18 Hz using 1–9 cycles per frequency band and used this to compute the inter-trial coherence (ITC) for Session 1 and Session 2 separately using all trials. A baseline from -0.25 s to 0.0 s was applied both for ERP and ITC visualisation.

## RESULTS

**Patient selection:** Binary classification accuracies (AC vs. AF) of single-trial ERPs ranged from a median of 42% to 62% (median of whole dataset 50%; see Figure 1). For this paper we chose to investigate the patient with the highest median accuracy (referred to as Patient A in this paper).

**Patient A accuracies:** Accuracies dropped for all classification approaches from Session 1 to Session 2 (see confusion matrices for two-class AC vs. AF in Figure 2, two-class TAC vs. TAF in Figure 3, and five-class All vs. All in Figure 4). Binary single-trial classification accuracies for Session 1 are close to 70% for both the auditory (Session 1 median=0.69, SD=0.11, range=0.53–0.87; Session 2 median=0.54, SD=0.11, range=0.38–0.76) and tactile stimulus modalities (Session 1 median=0.66, SD=0.07; range=0.62–0.83; Session 2 median=0.58, SD=0.09, range=0.38–0.7). Five-class classification accuracy drops in Session 2 in par-

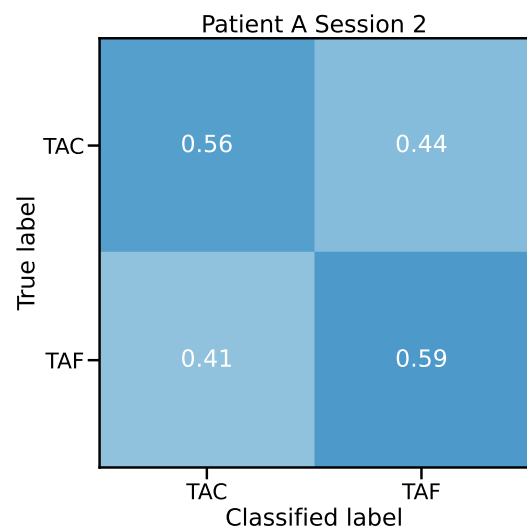
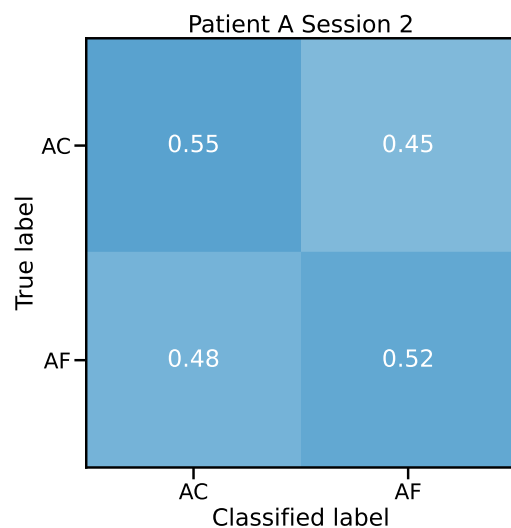
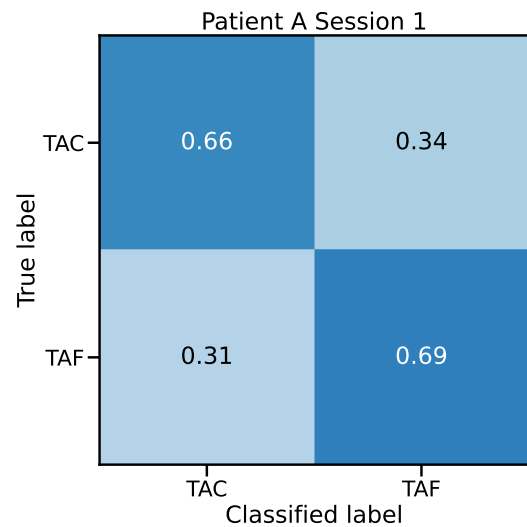
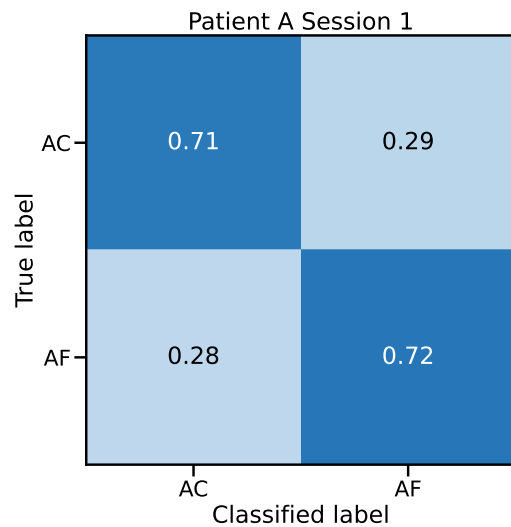


Figure 2: Two-class (AC vs. AF) normalised confusion matrices for Patient A. The accuracy drops considerably from session 1 (top) to session 2 (bottom).

Figure 3: Two-class (TAC vs. TAF) normalised confusion matrices for Patient A. The accuracy drops considerably from session 1 (top) to session 2 (bottom).

ticular due to incorrect classification of pure tactile (T) and tactile-auditory-close (TAC) trials (Session 1 median=0.54, SD=0.06; range=0.46–0.66; Session 2 median=0.35, SD=0.06, range=0.27–0.47). The difference in accuracy between Session 1 and 2 is statistically significant according to t-tests for independent samples with Bonferroni correction (see Figure 5; AC vs. AF Session 1 vs. Session 2  $t_{18} = 3.73, p = .004$ ; TAC vs. TAF Session 1 vs. Session 2  $t_{18} = 2.72, p = .04$ ; All vs. All Session 1 vs. Session 2  $t_{18} = 6.99, p < .0001$ ).

**Patient A ERPs:** Investigation of the event-related potentials aligns with the classification results. See Figure 6 for responses to tactile-auditory close and far stimuli shown separately for Session 1 and Session 2. Phase-locked responses to the tactile stimulus are visible in Session 1 between 100 and 300 ms post stimulus. Differences in the response to close and far stimuli were clearly visible in Session 1 between 300 and 600 ms.

No phase-locked responses were observed in Session 2. The time-domain responses align with inter-trial coherence (see Figure 7). In Session 1 strong phase-locking was observed between 0 and 500 ms post-stimulus with the strongest response around 7 Hz. The visualisation of the ITC has no coherent pattern in Session 2.

## DISCUSSION

The case study presented in this paper offers insights into the delineation of PPS in patients with disorders of consciousness using ERP features and single-trial classification. To limit the analysis, as a starting point we selected the patient with the highest median classification accuracy in our database (Patient A), who also exhibited high variability in classification performance across two sessions which took place 10 days apart.

The results demonstrate our approach is capable to dif-

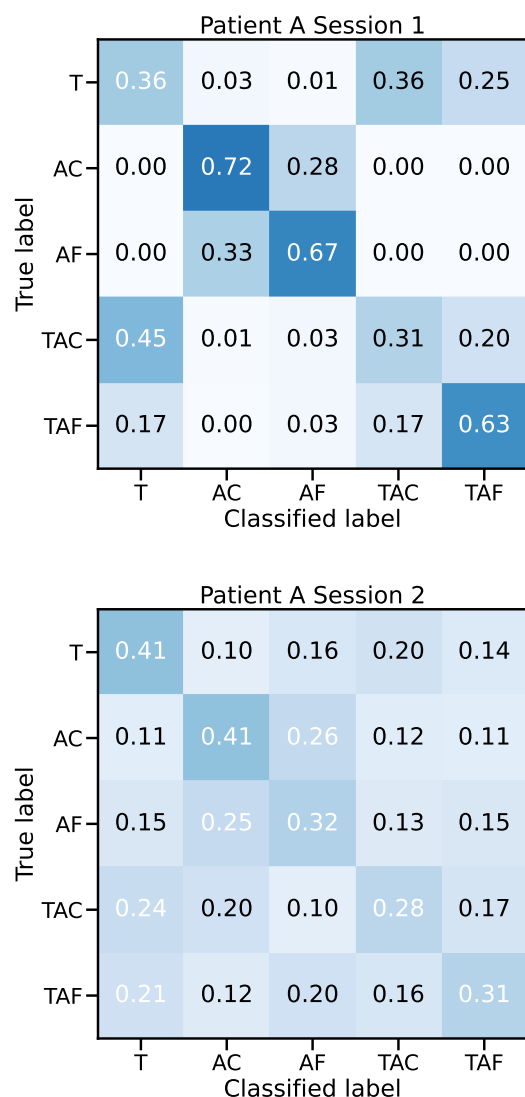


Figure 4: Five-class confusion matrices for Patient A. The accuracy drops considerably from session 1 (top) to session 2 (bottom).

ferentiate stimuli presented within the PPS across various conditions (T, AC, AF, TAC, TAF), with notably high classification accuracies in the first session (particularly as these results are obtained on single-trial ERPs). These findings underscore the sensitivity of PPS delineation as a potential tool for assessing consciousness in DOC patients independently of motor output [7, 15].

After suffering a car accident, Participant A's recordings were performed 2 and 12 days post-injury (Session 1 and Session 2, resp.), with the patient being deemed unresponsive (CRS-R=7, UWS) in the first session and minimally conscious (CRS-R=11, MCS-) in the second one. She was discharged from acute care 21 days post-injury and emerged from minimally conscious state (CRS-R=21) 2.5 months post-discharge.

The variation in classification accuracy between the first and second sessions, particularly with the observed recovery from DOC as indicated by the final available CRS-

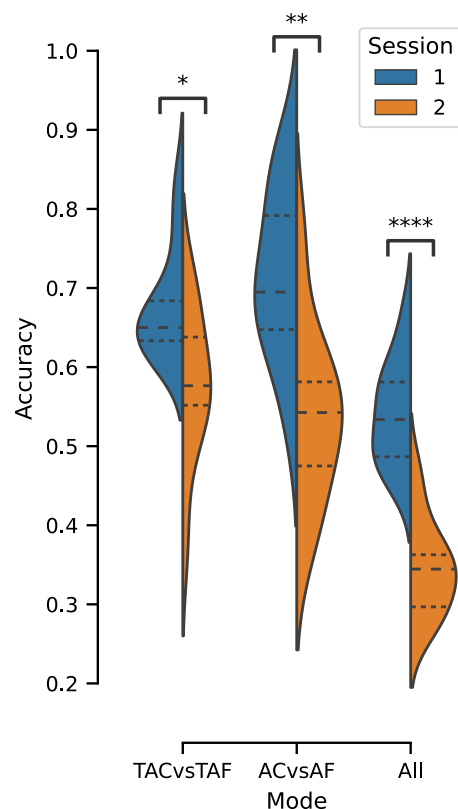


Figure 5: Distribution of accuracy in the 10-fold cross-validation performed on the data of each session. Colours indicate session: blue Session 1 and orange Session 2. The first and second violins represent the binary classification results (TAC vs. TAF; AC vs. AF, resp.) and the third five-class classification results (All vs. All). Stars indicate significance according to t-test for independent samples with Bonferroni correction: \*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*\*:  $p < 0.0001$ . The central dashed line was placed at the median of each violin, the finer dashed lines at the first quartiles.

R score, may suggest an alteration in sensory processing or awareness levels as the patient regained consciousness. This is further supported by the presence of clear ERP peaks for combined tactile and auditory conditions in the first session, which seem to diminish alongside improved consciousness levels. However, the reduction in classification accuracy during the second session poses questions about the dynamics of PPS and its neural correlates as patients recover, such as the representation of PPS and how it might be centered around the body, offering insights into the neural mechanisms that could be involved in PPS processing [16]. Thus, the lower accuracy might occur as a consequence of a reorganization of sensory processing networks or changes in the salience of peripersonal stimuli as the patient's cognitive state evolves. These observations are critical for developing a nuanced understanding of consciousness and its manifestations in DOC patients, providing a foundation for future investigations into the mechanisms underlying consciousness recovery.

On the other hand, there are other plausible explanations

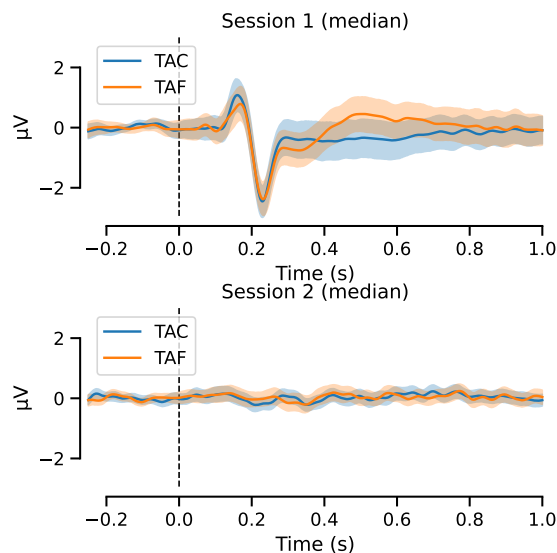


Figure 6: Median ERPs (across channels CP1, CPz and CP2) from Session 1 (top) following combined tactile-auditory stimulation and Session 2 (bottom) of Patient A. Epochs were baselined to the average of the 250 ms before stimulus presentation. The blue lines indicate the response to the close stimulus (tactile-auditory close; TAC), the orange the response to the far stimulus (tactile-auditory far; TAF). The vertical dashed line indicates the timepoint of stimulus presentation. The shaded area indicates the parametric confidence interval (95%).

for the observed low performance in the second session, which is particularly surprising as the patient was discharged shortly after and was associated with a higher CRS-R than Session 1. Firstly, it is conceivable that the patient may have been in a state of sleep during the second session, akin to the absence of responsiveness we would expect to observe if the patient was unconscious [17, 18]. Alternatively, technical issues with the recording equipment or environmental factors, the likelihood of which increases due to the harsh experimental conditions at bedside in an acute unit, could have influenced the quality of the recordings and therefore the final performance of our classifiers. Finally, the discrepancy between the results and the CRS-R scores for each of the sessions could be due to the inherent limitations of the CRS-R themselves [19, 20], particularly as clinical underestimation of conscious awareness may occur (which might have happened in Session 1 of Patient A). However, it is important to note that we cannot definitively conclude which of the four scenarios is applicable in this particular case and further exploration of the larger database is needed. The seeming contradiction of CRS-R and ERP classification results may be resolved by including further measures of consciousness in the analysis, such as measures of EEG signal diversity, such as the Lempel-Ziv Complexity [21, 22] as suggested in [7]. This measure can be computed from the spontaneous EEG enabling us to include this in a future analysis.

A key takeaway from our study is the recognition that relying solely on data from the second session would have led to an erroneous diagnosis for Patient A. This is par-

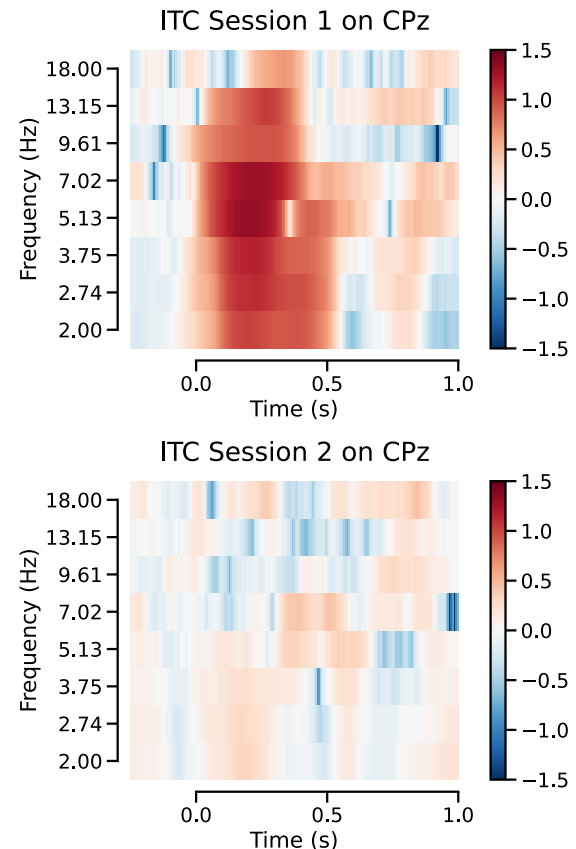


Figure 7: Inter-trial coherence computed for all trials of Session 1 (top) and 2 (bottom) of Patient A. Stimulus presentation occurred at 0.0 s. ITC was calculated for 8 frequency bands in the range of 2-18 Hz. Epochs were baselined to the average of the 250 ms before stimulus presentation. Colour indicates the strength (red stronger, blue weaker) of the ITC.

ticularly evident in light of the patient's reemergence to a state of minimum consciousness shortly after the recording of this session. Consequently, our findings align with the perspective advocated by Wannez and colleagues [9] regarding the necessity of conducting multiple recording sessions over time, encompassing various times of the day to account for circadian rhythms. This approach is essential for accurately assessing DOC and avoiding potentially misleading interpretations based on discrete observations. In essence, our study underscores the importance of adopting a longitudinal approach to clinical assessments in this domain.

## CONCLUSIONS

In conclusion, our study adds to the body of literature advocating in favour of conducting multiple recording sessions over time when assessing DOC to minimise erroneous diagnoses of patients. However, it is important to acknowledge that we are basing this stance on results from a single patient selected from a larger dataset. In the future, we plan to expand our analyses to the other patients, with the aim of providing a more comprehensive understanding of the factors influencing variability

in performance and the implications for clinical practice. By broadening our scope and methods (e.g., by including the Lempel-Ziv Complexity measure), we can contribute to enhancing the accuracy and reliability of diagnostic assessments in this challenging area of healthcare.

## REFERENCES

- [1] Teasdale G, Jennett B. Assessment of coma and impaired consciousness. a practical scale. *The Lancet*. 1974;304.
- [2] Giacino JT, Kalmar K, Whyte J. The JFK coma recovery scale-revised: Measurement characteristics and diagnostic utility. *Archives of Physical Medicine and Rehabilitation*. 2004;85.
- [3] Owen AM, Coleman MR, Boly M, Davis MH, Laureys S, Pickard JD. Detecting awareness in the vegetative state. *Science*. 2006;313.
- [4] Horki P *et al.* Detection of mental imagery and attempted movements in patients with disorders of consciousness using EEG. *Frontiers in Human Neuroscience*. 2014;8.
- [5] Real RG *et al.* Information processing in patients in vegetative and minimally conscious states. *Clinical Neurophysiology*. 2016;127.
- [6] Rizzolatti G, Fadiga L, Fogassi L, Gallese V. The space around us. *Science*. 1997;277(5323):190–191.
- [7] Noel JP *et al.* Peri-personal space encoding in patients with disorders of consciousness and cognitive-motor dissociation. *NeuroImage: Clinical*. 2019;24:101940.
- [8] Orłowski P, Bola M. Sensory modality defines the relation between EEG Lempel–Ziv diversity and meaningfulness of a stimulus. *Scientific Reports*. 2023;13.
- [9] Wannez S, Heine L, Thonnard M, Gosseries O, Laureys S, Collaborators CSG. The repetition of behavioral assessments in diagnosis of disorders of consciousness. *Annals of Neurology*. 2017;81(6):883–889.
- [10] Gramfort A *et al.* MNE software for processing MEG and EEG data. *NeuroImage*. 2014;86:446–460.
- [11] Fischler MA, Bolles RC. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*. 1981;24(6):381–395.
- [12] Jas M, Engemann D, Raimondo F, Bekhti Y, Gramfort A. Automated rejection and repair of bad trials in MEG/EEG. In: 2016 International Workshop on Pattern Recognition in NeuroImaging (PRNI). 2016, 1–4.
- [13] Jas M, Engemann DA, Bekhti Y, Raimondo F, Gramfort A. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*. 2017;159:417–429.
- [14] Pedregosa F *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825–2830.
- [15] Patané I, Cardinali L, Salemme R, Pavani F, Farnè A, Brozzoli C. Action planning modulates peripersonal space. *Journal of Cognitive Neuroscience*. 2019;31(8):1141–1154.
- [16] Serino A *et al.* Body part-centered and full body-centered peripersonal space representations. *Scientific reports*. 2015;5(1):18603.
- [17] Chennu S, Bekinschtein TA. Arousal modulates auditory attention and awareness: Insights from sleep, sedation, and disorders of consciousness. *Frontiers in psychology*. 2012;3:17334.
- [18] Cote KA. Probing awareness during sleep with the auditory odd-ball paradigm. *International Journal of Psychophysiology*. 2002;46(3):227–241.
- [19] Cruse D *et al.* Bedside detection of awareness in the vegetative state: A cohort study. *The Lancet*. 2011;378(9809):2088–2094.
- [20] Jöhr J, Halimi F, Pasquier J, Pincherle A, Schiff N, Diserens K. Recovery in cognitive motor dissociation after severe brain injury: A cohort study. *PLoS One*. 2020;15(2):e0228474.
- [21] Lempel A, Ziv J. On the complexity of finite sequences. *IEEE Transactions on information theory*. 1976;22(1):75–81.
- [22] Schartner M *et al.* Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. *PloS one*. 2015;10(8):e0133532.

## EXPLORING NEW TERRITORY: CALIBRATION-FREE DECODING FOR C-VEP BCI

J. Thielen<sup>1</sup>, J. Sosulski<sup>2</sup>, M. Tangermann<sup>1</sup>

<sup>1</sup>Donders Institute, Radboud University, Nijmegen, the Netherlands

<sup>2</sup>Department of Computer Science, University of Freiburg, Freiburg, Germany

E-mail: jordy.thielen@donders.ru.nl

**ABSTRACT:** This study explores two zero-training methods aimed at enhancing the usability of brain-computer interfaces (BCIs) by eliminating the need for a calibration session. We introduce a novel method rooted in the event-related potential (ERP) domain, unsupervised mean maximization (UMM), to the fast code-modulated visual evoked potential (c-VEP) stimulus protocol. We compare UMM to the state-of-the-art c-VEP zero-training method that uses canonical correlation analysis (CCA). The comparison includes instantaneous classification and classification with cumulative learning from previously classified trials for both CCA and UMM. Our study shows the effectiveness of both methods in navigating the complexities of a c-VEP dataset, highlighting their differences and distinct strengths. This research not only provides insights into the practical implementation of calibration-free BCI methods but also paves the way for further exploration and refinement. Ultimately, the fusion of CCA and UMM holds promise for enhancing the accessibility and usability of BCI systems across various application domains and a multitude of stimulus protocols.

### INTRODUCTION

A brain-computer interface (BCI) records the user's brain activity and converts these into computer commands, offering an alternative output channel that does not rely on muscular activity. Electroencephalography (EEG) is commonly used to record brain activity due to its affordability, practicality, and non-invasiveness. The primary application of BCIs lies in restoring lost control, particularly in communication. One notable example is the visual BCI speller, where users can select symbols by focusing their gaze on them when displayed on a screen. This technology proves invaluable for individuals facing challenges such as amyotrophic lateral sclerosis, where progressive loss of voluntary motor control makes speaking and typing difficult [1].

Prior to BCI usage, a machine learning model capable of classifying unseen brain signals needs to be calibrated on labelled EEG data from the same user, as individuals display different patterns of brain activity. Additionally, the same user might show different patterns over multiple days of use (session-to-session variability) and even

within-session non-stationarity. To mitigate any negative effects of these confounders, the user is guided through an initial stage to record brain activity while being instructed which symbol to attend to.

While a trained classification model is necessary for using the intended BCI application, the calibration recording delays a deployment and may be prohibitive specifically for users with a limited attention span. In general, the necessity of calibration may impede the acceptance and widespread adoption of BCIs by patients and healthy users. Encouragingly, recent advancements in BCI technology have surfaced which offer potential solutions to alleviate this challenge.

The first advancement involves selecting an informative brain signal feature to minimize the duration of the calibration phase. BCIs can be driven by various brain signals, often evoked by advanced stimulus protocols. Many popular stimulus protocols induce one or more event-related potentials (ERPs). Among these, the visual evoked potential (VEP), triggered by a flash, stands out. VEP-based BCIs are widely embraced due to their effectiveness across diverse user populations [2]. The VEP can be effectively used in three different ways [3].

Firstly, in a BCI based on time-modulated VEP (t-VEP), stimuli are sequentially presented to reduce temporal overlap, resulting in a relatively slow paradigm. Secondly, in a frequency-modulated (f-VEP)-based BCI, each stimulus simultaneously and rapidly flashes at a unique frequency and phase [4]. Despite its speed, f-VEP faces limitations due to the restricted range of narrow-band options and potential artefacts that may obscure signals. Thirdly, in a code-modulated (c-VEP)-based BCI, each stimulus rapidly flashes with a pseudo-random noise-code [5]. In this protocol, stimulus sequences are optimized to be dissimilar, ensuring that their corresponding brain activity is dissimilar as well. Remarkably, c-VEP BCI has recently demonstrated unprecedented performance [6].

The second advancement involves the choice of the decoding approach. In the case of c-VEP, a method was developed, termed 'reconvolution', which relies on a forward model embedded in a canonical correlation analysis (CCA) [7]. This model characterizes the response to a sequence of flashes as the linear summation of responses to individual flashes. This reconvolution approach substan-

tially reduced the number of trainable parameters while it simultaneously increased the number of samples that were available as training data points. This not only decreased the required training data but also empowered the model to predict responses to unseen stimulus sequences. Recently, this reconvolution CCA method was shown to achieve remarkable performances on c-VEP data event without the need for a calibration session, by finding the stimulus sequence that best fits the data in a trial [8].

In a recent development, a novel classification approach was introduced for ERP-based BCI. The approach, termed ‘unsupervised mean-difference maximization’ (UMM) does not require labeled calibration data [9]. UMM does not act on every single epoch, i.e., the evoked response of a single stimulus, but on a set of epochs belonging to one control command, e.g., one trial that leads to the selection of a symbol in a spelling application. Therefore, UMM has similarities to the aforementioned CCA method. While the objective of CCA is to maximize the explained variance, UMM identifies an attended target symbol by maximizing the distance between target and non-target ERPs belonging to one trial. Diverging from the CCA method, UMM incorporates several regularization approaches, including the use of block-Toeplitz covariance matrices [10] for determining domain-specific distances, and it can take advantage of a built-in confidence metric. Collectively, UMM has demonstrated impressive performance across various ERP datasets, without the need for a training session [9]. In this study, we aim to combine the efficiency of the c-VEP stimulus protocol and the carefully regularized UMM approach for zero-training. This will be the first instance that UMM is applied to c-VEP data, arguably a much faster stimulus protocol than the conventional ERP stimulus protocol. In the analysis, we draw a comparison with the CCA zero-training pipeline that was already evaluated on c-VEP data. This study not only sheds light on the efficacy of CCA and UMM, but also deepens our understanding of their underlying distinct mechanisms, unraveling insights into constructing effective BCIs for communication and control. By eliminating the need for a calibration session, this research paves the way for plug-and-play BCIs, marking a significant stride towards user-friendly and accessible BCI technology.

## MATERIALS AND METHODS

**Dataset:** We assessed the efficacy of the CCA and UMM zero-training approaches using an open-access c-VEP dataset [11]. Comprehensive details about this dataset can be found in the original study [8]. For the current study, we only used the part of this dataset labeled as ‘offline experiment’, in which 30 participants engaged in a copy-spelling task. EEG data were recorded from 8 electrodes placed following the 10-10 system (Fz, T7, O1, POz, Oz, Iz, O2, T8) amplified using a Biosemi Active2 amplifier and sampled at a frequency of 512 Hz. Throughout the experiment, participants interacted with

a  $4 \times 5$  matrix speller displayed on a 12.9in iPad Pro with a 60 Hz refresh rate and a  $1920 \times 1080$  px resolution. The  $N = 20$  cells within this matrix each measured  $3.1 \text{ cm} \times 2.8 \text{ cm}$ , with a  $0.4 \text{ cm}$  separation both horizontally and vertically between cells. The cells were presented against a mean-luminance gray background. Each cell  $i \in \{1, \dots, N\}$  was luminance modulated using a unique binary stimulus sequence at full contrast, with a 1 encoding a white cell and a 0 a black cell. The stimulus sequences were carefully chosen from an optimized subset of Gold codes [12]. These sequences were modulated such that they contained flashes of only two durations: a short flash of 16.67 ms and a long flash of 33.33 ms. The sequences had a length of  $126 = 2 * (2^6 - 1)$  bits and at 60 Hz cycling through a stimulation code once took 2.1 s. Participants completed 5 identical runs, with each run comprising 20 trials, one for each of the 20 cells presented in a random order. Each trial started with a 1-second cue highlighting the target cell in green. Subsequently, all cells started flashing with their respective stimulus sequences for a duration of 31.5 s (equivalent to 15 code cycles), during which participants maintained fixation on the target cell. Post-trial, no feedback was provided and the subsequent trial commenced without delay. To sum up, each participant contributed 100 trials of 31.5 s, including 5 repetitions for each of the 20 stimuli. The EEG data underwent preprocessing using Python version 3.10.9 and MNE version 1.6.0. Initially, a notch filter at 50 Hz was applied to eliminate line noise. This was followed by a band-pass filter with a lower cut-off at 6 Hz and an upper cut-off at 50 Hz, which was optimized in an initial analysis. Subsequently, the data were segmented into single-trials, spanning from 500 ms before stimulus onset to 31.5 s after stimulus onset. The dataset was then downsampled to 180 Hz, which is a multiple of the monitor refresh rate at 60 Hz. Finally, the initial 500 ms of data per trial, which may have caught artefacts resulting from the initial slicing and subsequent filtering processes, were removed.

**Canonical correlation analysis (CCA):** Using CCA, let’s assume that the current trial  $\mathbf{X} \in \mathbb{R}^{C \times T}$  contains  $T$ -many temporal features extracted from each of the  $C$ -many channels. Here,  $C = 8$ . To decode the attended target symbol  $\hat{y}$  of a new trial via CCA, each of the  $i \in \{1, \dots, N\}$  possible hypotheses about which cell, i.e., which stimulus sequence, may have represented the target, are considered. Here,  $N = 20$ .

As CCA operates at the trial level, its  $i$ th stimulus sequence is described by the event time-series  $\mathbf{E}_i \in \mathbb{R}^{E \times T}$  for  $E$ -many events and  $T$ -many temporal features. Here, we modeled  $E = 3$  events including the two flash durations and an onset event for the sudden start of the stimulation.

Subsequently, the event time-series are transformed into a structure matrix  $\mathbf{M}_i \in \mathbb{R}^{M \times T}$  with  $M$ -many event time-points and  $T$ -many temporal features. Let’s assume equally long responses to each of the  $E$  events, then  $M = E * L$ . Here,  $L = 54$ , which corresponds to 300 ms

at 180 Hz. This structure matrix is a Toeplitz matrix describing the onset, duration and overlap of the responses to each of the events in the  $i$ th stimulation sequence. We fit a CCA model for each of the  $N$  candidate stimulus sequences  $i \in \{1, \dots, N\}$  by learning sequence-specific spatial filters  $\mathbf{w}_i \in \mathbb{R}^C$  and temporal filters  $\mathbf{r}_i \in \mathbb{R}^M$ :

$$\arg \max_{\mathbf{w}_i, \mathbf{r}_i} \frac{\mathbf{w}_i^\top \mathbf{X} \mathbf{M}_i^\top \mathbf{r}_i}{\mathbf{w}_i^\top \mathbf{X} \mathbf{X}^\top \mathbf{w}_i \mathbf{r}_i^\top \mathbf{M}_i \mathbf{M}_i^\top \mathbf{r}_i} \quad (1)$$

Instantaneous classification of the current trial, i.e., determining the one attended target symbol  $\hat{y}$  from the  $N = 20$  symbols, is then performed by maximizing the correlation, which is equivalent to the square root of the explained variance:

$$\hat{y} = \arg \max_i \frac{\mathbf{w}_i^\top \mathbf{X} \mathbf{M}_i^\top \mathbf{r}_i}{\mathbf{w}_i^\top \mathbf{X} \mathbf{X}^\top \mathbf{w}_i \mathbf{r}_i^\top \mathbf{M}_i \mathbf{M}_i^\top \mathbf{r}_i} \quad (2)$$

Alternatively to this instantaneous CCA, CCA can learn across trials. Specifically, previous trials can be included to improve the estimates for the current trial, as described in [8]. Equation 1 can be formulated using the spatio-temporal cross-covariance  $\mathbf{\Sigma}_{\mathbf{X}\mathbf{M}_i} \in \mathbb{R}^{C \times M}$ , the spatial covariance  $\mathbf{\Sigma}_{\mathbf{X}} \in \mathbb{R}^{C \times C}$  and the temporal covariance  $\mathbf{\Sigma}_{\mathbf{M}_i}$ :

$$\arg \max_{\mathbf{w}_i, \mathbf{r}_i} \frac{\mathbf{w}_i^\top \mathbf{\Sigma}_{\mathbf{X}\mathbf{M}_i} \mathbf{r}_i}{\mathbf{w}_i^\top \mathbf{\Sigma}_{\mathbf{X}} \mathbf{w}_i \mathbf{r}_i^\top \mathbf{\Sigma}_{\mathbf{M}_i} \mathbf{r}_i} \quad (3)$$

The estimation of these covariance matrices can be improved by accumulating the data  $\mathbf{X}$  and predicted structure matrix  $\mathbf{M}_j$  of the previous trial(s). This cumulative CCA is an optimistic one, as it assumes that previous trials were classified correctly (i.e., naive labeling).

In summary, in this work, we applied two versions of CCA. For both we used the empirical covariance matrix identical to the original work [8]. The first version, denoted CCA\_e1, was instantaneous and estimated the covariance from the current trial only. The second version, denoted CCA\_ec, was cumulative and used previous trials for covariance estimation to facilitate decoding of the current trial. Code for the CCA approach is available at <https://github.com/thijor/pyntbci>.

*Unsupervised mean-difference maximization (UMM):*

Using UMM, we first slice the current trial into the contained  $K$ -many epochs, which are synchronized to each bit in the stimulus sequences, i.e., the monitor refresh rate at 60 Hz.

Let's assume that an epoch  $\mathbf{x} \in \mathbb{R}^D$  is described by a  $D$ -dimensional feature space, which contains  $T$ -many temporal features extracted from each of the  $C$ -many channels, i.e.,  $D = C * T$ . Here,  $C = 8$  and  $T = 54$  for epochs of 300 ms long at 180 Hz. To decode the attended target symbol  $\hat{y}$  via UMM for the current trial, each of the  $i \in \{1, \dots, N\}$  possible hypotheses about which cell may have represented the target, are considered. Here,  $N = 20$ . For every possible hypothesis  $i \in \{1, \dots, N\}$  we then estimate the mean-difference vector  $\Delta \mu_i \in \mathbb{R}^D$ , which is the difference between the flash ERP and non-flash ERP:

$$\Delta \mu_i = \frac{1}{|A_i^+|} \sum_{j \in A_i^+} \mathbf{x}_j - \frac{1}{|A_i^-|} \sum_{j \in A_i^-} \mathbf{x}_j \quad (4)$$

where  $\mathbf{x}_j \in \mathbb{R}^D$  is the  $D$ -dimensional EEG feature vector of the  $j$ -th epoch, and  $A_i^+$  and  $A_i^-$  denote the sets of epochs for which a flash was either presented (bit is 1) or not (bit is 0) under the current hypothesis of  $i$  being the target stimulus sequence.

Instead of determining the attended symbol of the current trial by maximizing the Euclidean distance between flash and non-flash ERPs across all  $N$ -many hypotheses, the metric is first normalized using the inverse of the global feature covariance matrix  $\mathbf{\Sigma} \in \mathbb{R}^{D \times D}$  to better cope with non-spherical feature distributions in the feature space  $\mathbb{R}^D$ , known as the Mahalanobis distance:

$$\hat{y} = \arg \max_i (\Delta \mu_i)^\top \mathbf{\Sigma}^{-1} (\Delta \mu_i) \quad (5)$$

Please note that the covariance matrix can be estimated based on the epochs of the current trial only, which makes UMM an instantaneous decoding approach that does not require calibration data. Due to the challenging ratio of the feature dimensionality and the number of epochs contained in a single trial, we used a block-Toeplitz regularization with tapering to obtain a more robust estimate of the covariance matrix [10].

Alternatively to this instantaneous use of UMM, knowledge from previous trials about both, the estimated class means and the covariance matrix can be included for obtaining improved estimates for the current trial, as described in [9]. Specifically, as the covariance matrix can be calculated without label information the covariance matrix can more robustly be estimated by also using epochs from previous trials [9]. A similar approach can be used to more robustly estimate the flash and non-flash ERPs, by using information from previous trials. However, for this, label information is required. UMM simply uses its own predictions from previous trials as pseudo labels (sometimes also referred to as naive labeling). This approach is made more robust, by weighting the mean estimates from previous trials by UMM's confidence in these previous trials [9]. Specifically, if UMM is very certain of its own prediction, these ERP means will have more weight in later trials and vice versa.

In summary, for this work, we used two versions of UMM. For both versions we used the block-Toeplitz regularized covariance matrix. The first version, denoted UMM\_t11, is instantaneous by estimating the covariance and means only from the current trial. The second version, denoted UMM\_tcw, is cumulative by using previous trials to facilitate decoding of the current trial. For an overview of all method's abbreviations used, please refer to the legend of Fig. 1. Code for the UMM approach is available at: [https://github.com/jsosulski/umm\\_demo](https://github.com/jsosulski/umm_demo).

*Analysis:* This study assessed the effectiveness of CCA and UMM in classifying c-VEP data without calibration, including both an instantaneous approach, where trials are classified without any prior calibration (CCA\_e1, UMM\_t11), and a cumulative approach, where trials are classified while leveraging information from previously analyzed trials (CCA\_ec, UMM\_tcw).

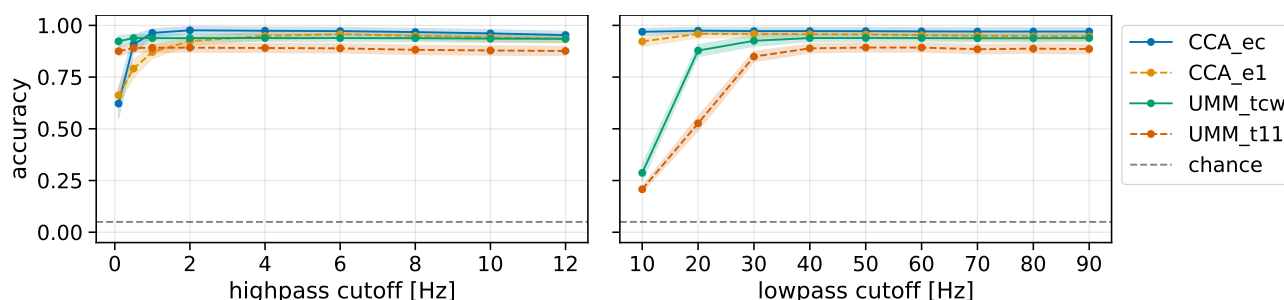


Figure 1: **Bandpass hyper-parameters for CCA and UMM.** Depicted are the grand average classification accuracy for CCA and UMM across varying highpass (left) and lowpass (right) cutoff values. Here, a single-trial duration of 31.5 s is used. When varying the highpass, the lowpass remained at 40 Hz, and when varying the lowpass, the highpass remained at 6 Hz. In order, the symbols behind a method refer to: the type of covariance matrix being empirical (e) or block-Toeplitz (t); covariance matrices computed instantaneously (i) or cumulative (c); and the mean vectors (of UMM) computed either instantaneously (i) or using a weighted cumulative average (w). The dashed gray line denotes the theoretical chance level (5%).

As previously mentioned, we first varied the cutoff frequency of the highpass and lowpass filter in the bandpass spectral filter used for preprocessing, to explore the influence of these hyper-parameters on the performance of the analyzed methods. For the highpass we tested 0.1, 0.5, 1.0, 2.0, 4.0, 6.0, 8.0, 10.0 and 12.0 Hz, for a total of 9 evaluations, all with a lowpass at 40 Hz. For the lowpass we tested 10 to 90 Hz in 10 Hz increments, for a total of 9 evaluations, all with a highpass at 6 Hz. These evaluations were always carried out using the full trial durations of 31.5 s. Recognizing that each method may respond uniquely to these variations, our final comparative analysis focused on the bandpass cutoff frequencies that resulted in the highest classification accuracy.

To assess the methods, we generated decoding curves by varying trial durations from 1.05 s (half a code cycle) to 10.5 s (5 code cycles) in 1.05 s increments, and from 10.5 s to 31.5 s in 2.1 s increments, for a total of 20 decoding time steps. Across these time steps, the number of bits ranged from 63 to 630 in 63-bit increments, and 630 to 1.890 in 126-bit increments, directly corresponding to the available epochs for UMM at each decoding time step. Because modulated Gold codes have an equal number of ones and zeros, the number of flash and non-flash epochs were always equal or deviated at most by 1 when using half a code cycle, for any of the stimulus sequences. In this analysis, a bandpass filter of 6–50 Hz was used, as it turned out optimal for all methods.

In this study, all statistics were carried out using a one-sided paired Wilcoxon signed-rank test to test for a larger classification accuracy of one versus another method. The significance level was set to  $\alpha = 0.025$ . Reported *p*-values were not corrected for multiple comparisons.

## RESULTS

In this study, we assessed the performance of two calibration-free methods when applied to c-VEP data. These methods, CCA and UMM, were evaluated as an instantaneous version that classified each new trial without prior information, treating it as the first, and a cumula-

tive version that learned from the insights gained through previously classified trials.

Acknowledging the potential for each method to exhibit distinct responses to varying bandpass filter hyper-parameters, our initial focus involved determining the optimal hyper-parameters for each method. The classification accuracy of 31.5-second trials across different cutoff frequencies and methods is illustrated in Fig. 1.

In the highpass analysis (Fig. 1, left side), it was evident that CCA is more sensitive to a low highpass value than UMM. For both CCA's instantaneous and cumulative versions, a decline in accuracy was observed for highpass values below 2 Hz. In contrast, UMM appeared to be less affected. In the pursuit of the highest accuracy, CCA\_ec achieved a classification accuracy of 0.97 at 2 Hz, CCA\_e1 0.96 at 6 Hz, UMM\_tcw 0.94 at 4 Hz, and UMM\_t11 0.89 at 2 Hz. When adopting a common highpass at 6 Hz for all methods, the methods still attained these peak performances.

In the lowpass analysis (Fig. 1, right side), an inverse trend emerged, revealing that UMM is more sensitive to the lowpass value than CCA. Specifically, UMM achieved a peak performance only when the lowpass value was set no lower than 40 Hz, whereas for CCA, this peak was already attained at 20 Hz. In the pursuit of optimal performance, we identified a peak classification accuracy of 0.97 at 20 Hz for CCA\_ec, 0.96 at 30 Hz for CCA\_e1, 0.94 at 50 Hz for UMM\_tcw, and 0.89 at 50 Hz for UMM\_t11. When applying a common lowpass at 50 Hz for all methods, they continued to operate at these peak performance levels.

The above mentioned results led to the selection of a common passband set to 6 to 50 Hz. We then continued analyzing the behavior of the different methods by estimating so-called decoding curves that show the classification accuracy across different amounts of data available in each single-trial, see Fig. 2.

From these decoding curves, two general trends could be identified. Firstly, the cumulative versions of both methods (CCA\_ec and UMM\_tcw) outperformed their instantaneous counterparts (CCA\_e1 and UMM\_t11).

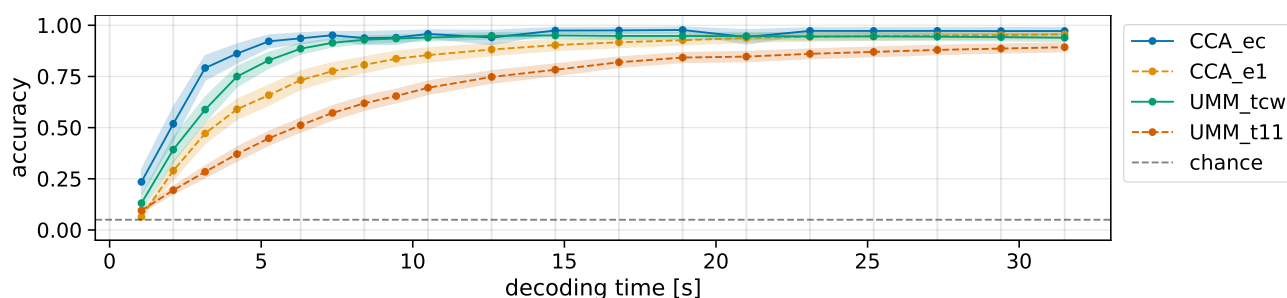


Figure 2: **Decoding curve for CCA and UMM.** Depicted are the grand average classification accuracy for CCA and UMM across varying single-trial durations. Here, a bandpass of 6-50 Hz is used. For a definition of method names, see 2. The dashed gray line denotes the theoretical chance level (5%).

Secondly, overall, the CCA methods (CCA\_ec and CCA\_e1) outperformed the UMM methods (UMM\_tcw and UMM\_t11). For an overview of the classification accuracy at some time points, please see Tab. 1.

Table 1: **Classification accuracy.** Listed are the grand average accuracy reached by the four methods at distinct trial durations.

	1.05 s	2.1 s	4.2 s	10.5 s	31.50 s
CCA_ec	0.24	0.52	0.86	0.96	0.97
CCA_e1	0.06	0.29	0.59	0.85	0.96
UMM_tcw	0.13	0.39	0.75	0.94	0.94
UMM_t11	0.09	0.19	0.37	0.69	0.89

The cumulative CCA method (CCA\_ec) achieved an accuracy of 0.24 at the smallest trial duration of 1.05 s (half a code cycle) and 0.52 at 2.1 s (one code cycle). At these two early time points, CCA\_ec did not significantly outperform UMM\_tcw, which reached 0.13 and 0.39, respectively. At all further time points, CCA\_ec did significantly surpass UMM\_tcw. At the maximum trial length of 31.5 s, CCA\_ec achieved a classification accuracy of 0.97, significantly outperforming UMM\_tcw, which achieved 0.94.

The instantaneous CCA method (CCA\_e1) achieved a performance of 0.06 at 1.05 s, while the instantaneous UMM (UMM\_t11) reached 0.09. At this time point, UMM\_t11 significantly outperformed CCA\_e1. Instead, at any further time point CCA\_e1 significantly surpassed the accuracy of UMM\_t11. At the maximum trial length of 31.5 s, CCA\_e1 reached a classification of 0.96 which was significantly higher than UMM\_t11 with 0.89.

The cumulative versions always outperformed the instantaneous version for both CCA and UMM. Notably, the instantaneous CCA (CCA\_e1) performed almost on par with its cumulative version (CCA\_ec) for trials longer than 25.2 s. Specifically, at 27.3 s, CCA\_e1 reached an accuracy of 0.95 while CCA\_ec reached 0.97, which was not significantly higher. Also at the 29.4 and 31.5 s trials, CCA\_ec did not significantly surpass CCA\_e1.

## DISCUSSION

We introduced UMM to c-VEP BCI, a calibration-free method originating from the ERP domain. We conducted

a systematic offline comparison with the conventional c-VEP method employing CCA. Both methods underwent evaluation in an instantaneous manner, classifying each single trial without prior knowledge, as well as in a cumulative way, utilizing previously classified trials as training data. The ultimate goal was to establish a c-VEP BCI with enhanced usability and broader adoption potential by eliminating the initial calibration session. We gave both methods access to EEG data only that would be available in an online experiment such that we don't expect a major problem for online use. For CCA approaches, online performance was reported in [8] where the non-instantaneous version matched a supervised model after a warm-up period.

Both the conventional method, CCA, and the novel approach, UMM, hinge on the principle that selecting a single symbol from a set of candidate symbols is considerably simpler than the reconstructing the entire stimulus sequence. Using 20 symbols in this study, one has to evaluate 20 candidate stimulus sequences only for a symbol selection. In contrast, the exhaustive reconstruction of the stimulus would involve an exponential growth, reaching  $2^{60}$  potential sequences for a 1-second stimulus at a presentation rate of 60 Hz.

Without involving a calibration session, the cumulative CCA already reached higher than 90 % performance for trials of 5.25 s, with the cumulative UMM just behind reaching a similar performance at 7.35-second trials. Despite CCA regularly outperforming UMM, astonishingly, these results demonstrate that UMM performs rather well, despite not being optimized for c-VEP data.

Furthermore, we demonstrated the capability of instantaneous classification, where CCA reached higher than 90 % accuracy at 14.70-second trials and UMM 89 % at 29.40 s. Instantaneous decoding does not learn from, or consider in any way, previous trials. Under stationary conditions, this scarcity of data may not allow the model to reach peak performance. On the other hand, non-stationary feature distributions like latency- or amplitude changes of ERP components over time, will likely not affect instantaneous decoding, while models mainly trained on data collected *before* the appearance of such feature drifts may suffer from performance degradation. Both CCA and UMM make different assumptions and ex-

exploit them. These assumptions may be met by different datasets to various degrees. For instance, CCA strongly leverages the sequential structure and large overlap between responses to adjacent stimuli in c-VEP datasets. Additionally, the event definition used focusses on the target ERP while discerning between ERPs associated with short and long flashes, but more and different event types can easily be implemented. Furthermore, UMM maintains its operations within the original EEG feature space, whereas CCA operates in the component space.

In contrast, UMM searches for the stimulus sequence with the largest target to non-target ERP distance, potentially rendering it less susceptible to slow drifts in the data, as evidenced by the highpass analysis. Besides, CCA uses the empirical covariance matrix, which can be challenging to estimate with limited data, while UMM employs domain-specific regularization techniques such as shrinkage and a block-Toeplitz covariance matrix [10]. Lastly, in the non-instantaneous UMM formulation, ERP mean estimates are improved using previously classified trials, by carefully weighting mean updates based on the confidence of each previous trial. Gaining a comprehensive understanding of the strengths and limitations of both methods bears the potential to develop refined versions tailored to specific characteristics of novel datasets through thoughtful hyper-parameterization.

Essentially, CCA and UMM necessitate a specific stimulus protocol involving repetitions, and both require knowledge about the precise timing and sequence of stimuli within a single trial (i.e., the selection of one symbol). While such information is typically available in BCI protocols using evoked responses, it may not seamlessly extend to other protocols like those based on sensorimotor rhythms. Moreover, these decoding methods are only applicable for benchmarks if the sequence information is provided, as demonstrated in MOABB [13]. This characteristic classifies the studied CCA and UMM methods as semi-supervised, given that stimulus information is requisite, while label information is not required.

## CONCLUSION

We showed that both CCA and UMM offer the potential to eliminate the necessity for a calibration session, thereby enhancing the usability for BCI applications, especially when integrated with the c-VEP protocol. These findings mark an initial stride toward combining the robust capabilities of machine learning methods across diverse domains. They inspire the exploration of their cross-application and cross-pollination, unlocking new possibilities for advancing BCI technologies.

## ACKNOWLEDGEMENTS

This work was part of the project ‘Dutch Brain Interface Initiative’ (DBI2) with project number 024.005.022 of the research programme ‘Gravitation’ which is (partly) financed by the Dutch Research Council (NWO).

## REFERENCES

- [1] Verbaarschot C *et al.* A visual brain-computer interface as communication aid for patients with amyotrophic lateral sclerosis. *Clinical Neurophysiology*. 2021;132(10):2404–2415.
- [2] Volosyak I, Rezeika A, Benda M, Gembler F, Stawicki P. Towards solving of the illiteracy phenomenon for VEP-based brain-computer interfaces. *Biomedical Physics & Engineering Express*. 2020;6(3):035034.
- [3] Bin G, Gao X, Wang Y, Hong B, Gao S. VEP-based brain-computer interfaces: Time, frequency, and code modulations. *IEEE Computational Intelligence Magazine*. 2009;4(4):22–26.
- [4] Nakanishi M, Wang Y, Chen X, Wang YT, Gao X, Jung TP. Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Transactions on Biomedical Engineering*. 2017;65(1):104–112.
- [5] Martínez-Cagigal V, Thielen J, Santamaría-Vázquez E, Pérez-Velasco S, Desain P, Hornero R. Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): A literature review. *Journal of Neural Engineering*. 2021.
- [6] Shi N *et al.* Estimating and approaching the maximum information rate of noninvasive visual brain-computer interface. *NeuroImage*. 2024;120548.
- [7] Thielen J, Broek P van den, Farquhar J, Desain P. Broad-band visually evoked potentials: Re(con)volution in brain-computer interfacing. *PLOS ONE*. 2015;10(7):e0133797.
- [8] Thielen J, Marsman P, Farquhar J, Desain P. From full calibration to zero training for a code-modulated visual evoked potentials for brain-computer interface. *Journal of Neural Engineering*. 2021;18(5):056007.
- [9] Sosulski J, Tangermann M. UMM: Unsupervised mean-difference maximization. *arXiv preprint arXiv:2306.11830*. 2023.
- [10] Sosulski J, Tangermann M. Introducing block-Toeplitz covariance matrices to remaster linear discriminant analysis for event-related potential brain-computer interfaces. *Journal of Neural Engineering*. 2022;19(6):066001.
- [11] Thielen J, Marsman P, Farquhar J, Desain P. From full calibration to zero training for a code-modulated visual evoked potentials brain computer interface. Version 3. 2023. [Online]. Available: <https://doi.org/10.34973/9txv-z787>.
- [12] Gold R. Optimal binary sequences for spread spectrum multiplexing. *IEEE Transactions on Information Theory*. 1967;13(4):619–621.
- [13] Aristimunha B *et al.* *Mother of all BCI benchmarks*. Version 1.0.0. 2023. [Online]. Available: <https://github.com/NeuroTechX/moabb>.

# MACHINE LEARNING-BASED IDENTIFICATION OF TES-TREATMENT NEUROCORRELATES

Pasquale Arpaia<sup>1</sup>, Lidia Ammendola<sup>2</sup>, Maria Cropano<sup>2</sup>, Matteo De Luca<sup>1</sup>, Anna Della Calce<sup>1</sup>,  
Ludovica Gargiulo<sup>1</sup>, Giacomo Lus<sup>3</sup>, Luigi Maffei<sup>1</sup>, Daniela Malangone<sup>2</sup>, Nicola Moccaldi<sup>1</sup>,  
Simona Raimo<sup>4</sup>, Elisabetta Signoriello<sup>3</sup>, Paolo De Blasiis<sup>5,6</sup>

<sup>1</sup>Dep. of Electrical Engineering and Information Technology (DIETI), University of Naples  
Federico II, Naples, Italy

<sup>2</sup>Multiple Sclerosis Center, II Neurological Clinic, University of Campania "Luigi Vanvitelli",  
Naples, Italy

<sup>3</sup>Second Division of Neurology, Dep. of Advanced Medical and Surgery Science, University of  
Campania "Luigi Vanvitelli", Naples, Italy

<sup>4</sup>Dep. of Medical and Surgical Sciences, "Magna Graecia" University of Catanzaro, Catanzaro, Italy

<sup>5</sup>University of Basilicata, Potenza, Italy

<sup>6</sup>University of Campania "Luigi Vanvitelli", Naples, Italy

E-mail: pasquale.arpaia@unina.it

**ABSTRACT:** This study presents a Machine Learning-based identification of electroencephalographic (EEG) features related to transcranial Electrical Stimulation (tES) in Multiple Sclerosis (MS) patients. The contribution is a first step toward an automated system capable of adjusting electrical stimulation according to the EEG feedback (EEG-based adaptive tES). Five MS patients underwent both tES or sham treatments and a Theory of Mind (ToM) training, and the EEG signal before and after treatments was acquired both in Eyes-Open (EO) and in Eyes-Closed (EC) condition. tES was administered by fixed cathode electrodes on the right deltoid muscle. Power differences between post and pre tES treatment in six bands of interest were explored. Support Vector Machine classifier achieved 92.5 % and 100.0 % accuracy in classifying a subject treated with tES, by exploiting power differences within high beta in T3 and gamma in T3 and P3 in EO condition and power differences within gamma in T3, Pz, Cz in EC condition, respectively. In particular, absolute power in gamma band was reduced after the treatment. The result is clinically significant due to the tendency of MS patients to have high values in this band, caused by the compensation determined by the neurons as a result of the demyelination process.

## INTRODUCTION

Multiple Sclerosis (MS) is a neurological disease with recurrent episodes of focal disorders influenced by the location and extent of demyelinating lesions within the Central Nervous System (CNS) [1]. MS can cause deficits of Theory of Mind (ToM), namely the capacity to infer mental states causing action and to reason on the contents of one's own and others' minds [2, 3]. How-

ever, how ToM subcomponents (cognitive and affective) are affected in MS it's not yet well understood. Typically, MS symptoms are pharmacologically treated. Currently, multidisciplinary approaches based on symptom type are emerging, including disease-modifying therapies, lifestyle modifications, psychological support, and rehabilitation interventions. [4–6].

In recent years, also transcranial Electrical Stimulation (tES) has been considered among the treatments used to relieve MS symptoms.

tES consists in the administration of small currents applied to the scalp [7]. The applied current can induce acute or long-lasting effects depending on the signal specifications [8]. Three tES modalities, distinguished by current administration methods, are commonly used: low-intensity Direct Current (tDCS), Alternating Current (tACS) and Random Noise Current (tRNS) [9, 10].

tDCS is the most widely used technique for the treatment of various medical conditions [7]. tDCS microscopically produces a series of effects including resting threshold modification, changes in synaptic processes, enhancement of synaptic plasticity and effects on glial cells [11–14]. tDCS treatment effects on MS-related impairments have rarely been addressed, despite stimulating neuronal activity is an important promoter of the remyelination process [15]. Some studies suggest positive effects, including attention enhancement, executive function and motor improvement, as well as reduction of associated symptoms [16, 17][18].

tDCS technique has been demonstrated to be effective but without following precise guidelines on stimulation parameters and brain areas to be stimulated. In most cases, the effectiveness of the treatment is assessed only by the improvement of pathology symptoms [19].

Studies on healthy patients demonstrated the EEG-based effects of tDCS treatment by describing changes in the five frequency bands: delta ([0.5-4] Hz), theta ([4-8] Hz), alpha ([8-12] Hz), beta ([12-30] Hz) and gamma ([30-45] Hz). According to Boonstra et al. [20], stimulation causes an increase in power especially below 5 Hz (delta band). Theta band is also involved in this study and the literature observes a specific increase in spectral power during tDCS stimulation specifically in the cingulate cortex and the dorsolateral prefrontal cortex (DLPFC) [21, 22]. Some results showed significantly changes in the alpha band power after anodal tDCS over the left DLPFC [23, 24], while Mangia et Al.[21] reported variations in the beta band during and after stimulation. Song et al. [25] observed an increase in beta frequency power after tDCS treatment, resulting in a change in the state of efficient cognitive functioning of the brain. Also an increase in gamma power associated with the engagement of proactive control in DLPFC was demonstrated [26]. Changes in EEG frequency bands indicating neuron deterioration are also connected with Multiple Sclerosis. MS patients have shown abnormally low Posterior Dominant Rythm (PDR), a reliable predictor of baseline neural activity, with a significantly lower mean value [27]. Literature demonstrates a significant increase in power in the delta band in the fronto-temporo-central regions and significant increase in delta and theta waves has been observed in MS patients with a high load of subcortical lesions [28, 29]. Increased amplitude in gamma band was also observed in Relapsing-Remitting Multiple Sclerosis (RRMS) patients. Moreover, decreased alpha frequency during rest indicates pathological desynchronization of widespread neural networks regulating cortical arousal fluctuation and tonic attention [29]. Although the use of EEG signal has been widely used for the treatment and diagnosis of Multiple Sclerosis, there are few studies using tES treatments referring to changes in EEG features to improve health condition of MS patients [30, 31]. This study aims to identify EEG features on tES-treated MS patients and the analysis of start and finish EEG condition correlated with clinical condition changes in ToM framework.

## MATERIALS AND METHODS

*Clinical protocol:* Five MS patients (4 women, 1 man) ranging from 18 to 75 years with the Expanded Disability Status scale (EDSS) score between 1 to 7 were enrolled in this study [32]. The exclusion criteria were the following:

- History of psychiatric illness, head injury or other neuro-degenerative diseases (dementia or global cognitive impairment);
- Surgery;
- Intracranial metal implantation and pacemaker;
- Severe disability (score > 7 on the EDSS scale);
- Pregnancy or lactation;

- Illiteracy.

The experimental protocol involved the following phases:

- Collection of demographic and clinical history data, a neuropsychological assessment and self-administration of questionnaires;
- 3 minutes EEG recording before treatment sessions in Eyes-Open (EO) condition;
- 3 minutes EEG recording before treatment sessions in Eyes-Closed (EC) condition;
- Non-pharmacological treatment conducted twice a week for 16 weeks, for a total of 32 sessions.
- 3 minutes EEG recording at the end of all treatment sessions in EO condition;
- 3 minutes EEG recording at the end of all treatment sessions in EC condition;
- Monitoring of the neuropsychological profile by administering the same battery of tests used during the first phase.

The order of EO and EC acquisitions was randomized both pre and post tES treatment. The non-pharmacological treatment was characterized by:

- Structured training on Theory of Mind conducted by a psychologist [33]. It consists of the viewing of short videos selected from films, or specially made, portraying various human social interactions requiring the recognition of emotions (happiness, sadness, anger, surprise, fear, and disgust) and ToM skills (decoding beliefs, irony, misunderstandings, and intentions) to be understood correctly. The training was conducted twice a week for 30 minutes, simultaneously with the tDCS intervention. In total, 64 short videos were prepared, of which 32 focus on the recognition of basic emotions (happiness, sadness, anger, surprise, fear, and disgust), and 32 represent social situations that require cognitive ToM skills to be understood (such as understanding irony and gaffes, attributing beliefs and intentions). In each session, two or three videos were presented which can be viewed several times, according to the patients' requests. Patients be asked to become 'social investigators', making interpretations on the emotions, mental states, and intentions of the protagonists of the social scenes presented.
- An electrical brain stimulation intervention using tDCS. During the active stimulation sessions, a 2.5x2.5 cm anode was applied on the left DLPFC area, while the cathode was placed on the right deltoid muscle. The tDCS was applied for 20 minutes at an intensity of 2 mA. The same procedure was used for the sham condition, but in this case, the electric current was applied only in the first 20 seconds of tDCS. The sham treatment, as well as the

active ones, was applied twice a week for 16 weeks for a total of 32 sessions.

A flow chart of the experimental protocol is showed in Fig. 1



Figure 1: Experimental protocol flow chart.

**Instrumentation:** EEG data were acquired by Mitsar EEG 201 system. It is a QEEG 21 channel device arranged according to the international 10/20 positioning system. The sampling rate is 2000 Sa/s. For tDCS treatment, 1x1 tDCS mini-CT (Soterix Medical, New York, USA) was used. It is a low intensity transcranial stimulator with current intensity ranging between 0.1 and 5 mA, and current duration between 5 and 40 minutes.

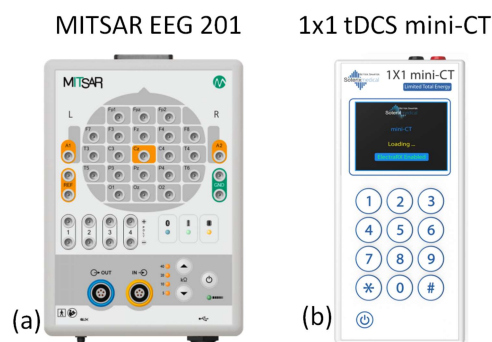


Figure 2: EEG acquisition (a) and tES (b) systems.

**EEG data processing:** EEG data were filtered by a fourth-order bandpass Butterworth filter ([0.5 - 45] Hz), then the *Artifact Subspace Reconstruction* (ASR) [34] procedure was used. ASR splits the EEG signal into components and once a threshold based on the signal variance distribution is identified, rejects noisy components above the threshold, reconstructing the signal by considering the remaining components. ASR was used with a cutoff equal to 15 to remove artifacts. Then EEG tracks were divided into 1-s epochs, organized in the form of [Epochs x Channels x Features].

In the features extraction phase, PDR amplitude and frequency and absolute and relative powers for all channels in delta, theta, alpha, low beta, high beta and gamma bands were computed in EC and EO conditions. Later, differences between pre and post both tES and sham treatment were calculated. The *Sequential Feature Selection* (SFS) [35] was applied in the feature selection phase. SFS is used to identify the most significant features to discriminate among different conditions. In this phase, the Support Vector Machine was the classifier embedded within the SFS. After creating the label vector with the 0 and 1 values associated with tES and sham treatments respectively, the training phase was carried out. Four patients were employed for training the classifiers:

two from the stimulation group and two from the sham group. Subsequently, the epochs of a fifth patient were allocated for the test set. The training set comprised 480 epochs, while the test set 120 epochs. The most informative features were selected during the training phase. The number of the features selected by the SFS algorithm was defined as the minimum number maximizing the classification accuracy. Only the features selected during the training phase were considered for the test phase on the epochs from the fifth patient.

**Results:** Three EEG features maximizing the discriminability between treated and no-treated patients both in EO and in EC conditions were selected by the SFS algorithm (Figs 3, 4). SFS algorithm on the train subjects reported a mean accuracy of 88.75% with a standard deviation of 8.45% in EC condition and a mean accuracy of 88.75% with a standard deviation of 2.22% in EO condition. Subsequently, one-shot test on the fifth subject was applied. In the EO condition, the test accuracy was of 92.5 %. The considered features were i) Difference of absolute powers in high beta band in T3 channel, ii) Difference of absolute powers in gamma band in T3 channel, and iii) Difference of relative powers in gamma band in T5 channel. In the EC condition, the test accuracy was of 100.0 %. The considered features were i) Difference of absolute powers in gamma band in T3 channel, ii) Difference of absolute powers in gamma band in Cz channel, and iii) Difference of absolute powers in gamma band in Pz channel.

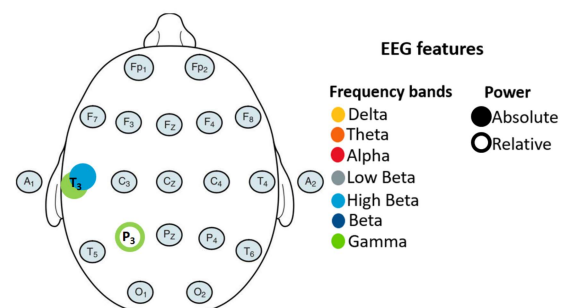


Figure 3: Most informative EEG features in Eye-Open condition for discriminating tDCS- and sham-treated patients.

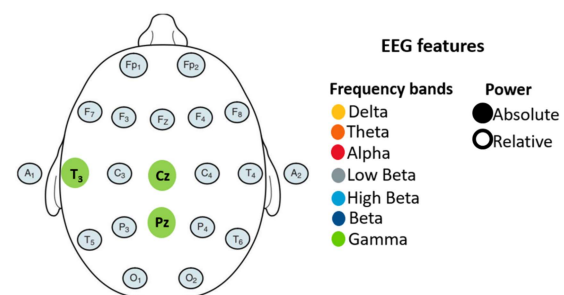


Figure 4: Most informative EEG features in Eye-Closed condition for discriminating tDCS- and sham-treated patients.

## DISCUSSION

Only two studies on EEG-based assessment of tDCS treatment effectiveness in MS patients are reported in literature, namely *Gholami et al.* [31] and *Ayache et al.* [30]. In particular, in *Gholami et al.* study, EEG acquisition with eyes closed in resting state was considered. An increase of Power Spectral Density (PSD) in alpha band and a decrease of PSD in beta and high beta bands over F3, C3, and P3 resulted between pre- and post-tDCS treatment. However, statistical tests did not confirm the significance of the results. The authors hypothesize that the impact of tDCS causes a power shift from beta to alpha frequencies, associated with the improvement in cognitive function, attention, and information processing speed. A limitation arises from exclusively EEG measuring with eyes closed. Indeed, EEG measurements under eyes-open conditions can reveal alterations in the highly myelinated visual structures otherwise not visible in MS patients [36].

For *Ayache et al.*, only the activity of theta band over Fz and Fpz channels during a cognitive task was focused. A statistically significant increase after tDCS with respect to sham treatment was highlighted.

The increased theta activity was associated with analgesic effects on MS symptoms. However, the study focuses only on the theta band within the region of stimulation. Other bands and other regions of the scalp are not explored.

Both studies have restricted EEG assessment to a limited number of frequency bands, thereby omitting evaluations of high-frequency activity such as the gamma band. This omission poses a limitation, considering that demyelination effects induced by MS are particularly evident at high-frequency [37]. In the present study, all EEG frequency bands are explored, revealing the impacts of tDCS treatment on both high beta and gamma bands. Furthermore, for three MS patients undergoing tDCS treatment, a reduction in the mean absolute power difference between pre- and post-treatment is observed in gamma and high beta bands. This result can be linked to the effectiveness of transcranial stimulation in restoring altered EEG features. In fact, alterations in the brain's functional connectivity network due to structural brain damage in MS patients are characterized by increased mean gamma power [29, 38]. This phenomenon is also correlated with partial disconnection of white matter pathways and significant cortical atrophy [39, 40].

The comparison between the presented study and the two aforementioned ones also encompasses the different electrode configuration employed for transcranial stimulation. In all studies, the anode is positioned over the DLPFC area but the cathode positioning is different. In particular, in the *Gholami et al.* and *Ayache et al.* studies, the cathode is positioned over the supraorbital region and on the Fp2 channel, respectively. Conversely, in the presented study it is situated on the right deltoid muscle (Fig 5). By employing this extracephalic refer-

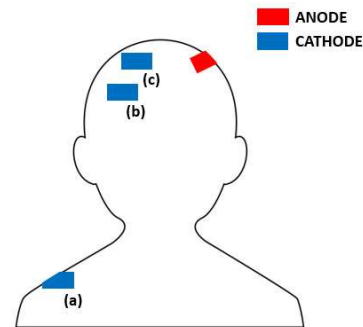


Figure 5: Comparison in tDCS cathode positioning among the different studies. For all the studies, the anode is collocated on the DLPFC area. The cathode is placed on the right deltoid muscle (a) in the present study, on the right supraorbital area (b) in *Ayache et al.* study [30], and on Fp2 channel (c) in *Gholami et al.* study [31].

ence, undesired cephalic hyperpolarization effects from brain areas beneath the reference electrode are mitigated [41, 42]. According to literature, the extracephalic positioning of the cathode in a tDCS treatment contributes to the decrease of absolute power in the high-frequency bands.[43]. However, this electrodes configuration has not been previously studied in the framework of EEG-based assessment of tDCS effectiveness in MS patients.

Results show significant effects of tDCS treatment in brain areas having also a crucial role in ToM framework. Temporo-Parietal Junction (TPJ) is the most involved area in reasoning about the contents of another person's mind [44]. The resulting reduction in gamma-band power on TPJ can be related to a restoration of ToM processes [45, 46]. Analyses on clinical outcomes are ongoing, and early results are very encouraging. Patients treated with tES exhibit improvements in ToM consistent with the electroencephalographic changes noted following combined ToM and tES treatment. Therefore, the identified EEG features can be used to real-time assess the treatment effectiveness and manage adaptation.

Notably, it is not possible to consider the effects of tES treatment and ToM training separately. Consequently, the reported results should be considered only when tES is applied with ToM training, concurrently.

Finally, tDCS treatment has not yet been standardized and variations in anode and cathode placement across studies are frequent. Consequently, comparing results becomes challenging.

## CONCLUSION

Neurocorrelates of transcranial Electrical Stimulation in Multiple Sclerosis patients were investigated. Power differences within high beta in T3 and gamma in T3 and P3 in EO condition and power differences within gamma in T3, Pz, Cz in EC condition were the most discriminative features. The selected EEG features allowed the test patient to be correctly classified as treated with 100.0 % accuracy in EC condition. This study poses the ba-

sis for adaptive tES protocols and stimulation settings according to EEG measurements. In the future, the number of enrolled subjects and the analysed pathologies will be widened to improve the statistical significance of the results. Moreover, a single integrated system allowing both EEG measurement and tES therapy will be developed to improve i) home therapy and ii) the contribution of customized medicine.

## ACKNOWLEDGMENT

This work was part of the "EEG4SmartES" project, financially supported by the Italian Ministry of Economic Development, (MISE). The authors thank Eleanor Schiavone for language assistance.

## REFERENCES

- [1] Ropper A, Samuels M, Klein J, Prasad S. Multiple sclerosis and other inflammatory demyelinating diseases. *Adams and Victor's Principles of Neurology*. 2014;1060–1131.
- [2] Heitz C *et al.* Cognitive and affective theory of mind in dementia with lewy bodies and alzheimer's disease. *alzheimer's research & therapy*, 8, 10. 2016.
- [3] Baron-Cohen S. Theory of mind and autism: A review. *International review of research in mental retardation*. 2000;23:169–184.
- [4] Loma I, Heyman R. Multiple sclerosis: Pathogenesis and treatment. *Current neuropharmacology*. 2011;9(3):409–416.
- [5] McGinley MP, Goldschmidt CH, Rae-Grant AD. Diagnosis and treatment of multiple sclerosis: A review. *Jama*. 2021;325(8):765–779.
- [6] Calabresi PA. Diagnosis and management of multiple sclerosis. *American family physician*. 2004;70(10):1935–1944.
- [7] Medeiros LF *et al.* Neurobiological effects of transcranial direct current stimulation: A review. *Frontiers in psychiatry*. 2012;3:110.
- [8] Yavari F, Jamil A, Samani MM, Vidor LP, Nitsche MA. Basic and functional effects of transcranial electrical stimulation (tes)—an introduction. *Neuroscience & Biobehavioral Reviews*. 2018;85:81–92.
- [9] Yang D, Shin YI, Hong KS. Systemic review on transcranial electrical stimulation parameters and eeg/fnirs features for brain diseases. *Frontiers in Neuroscience*. 2021;15:629323.
- [10] Ghobadi-Azbari P *et al.* Fmri and transcranial electrical stimulation (tes): A systematic review of parameter space and outcomes. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2021;107:110149.
- [11] Das S, Holland P, Frens MA, Donchin O. Impact of transcranial direct current stimulation (tdcs) on neuronal functions. *Frontiers in neuroscience*. 2016;10:550.
- [12] Reinhart RM, Cosman JD, Fukuda K, Woodman GF. Using transcranial direct-current stimulation (tdcs) to understand cognitive processing. *Attention, Perception, & Psychophysics*. 2017;79:3–23.
- [13] Rozisky JR, Antunes LdC, Brietzke AP, Sousa AC de, Caumo W. Transcranial direct current stimulation and neuroplasticity. *Transcranial direct current stimulation (tDCS): Emerging Used, Safety and Neurobiological Effects*. New York: Nova Science Publishers Inc. 2015:1–26.
- [14] Korai SA, Ranieri F, Di Lazzaro V, Papa M, Cirillo G. Neurobiological after-effects of low intensity transcranial electric stimulation of the human nervous system: From basic mechanisms to metaplasticity. *Frontiers in Neurology*. 2021;12:587771.
- [15] Pan S, Chan JR. Clinical applications of myelin plasticity for remyelinating therapies in multiple sclerosis. *Annals of Neurology*. 2021;90(4):558–567.
- [16] Maas DA, Angulo MC. Can enhancing neuronal activity improve myelin repair in multiple sclerosis? *Frontiers in Cellular Neuroscience*. 2021;15:38.
- [17] Mojaverrostami S, Khadivi F, Zarini D, Mohammedi A. Combination effects of mesenchymal stem cells transplantation and anodal transcranial direct current stimulation on a cuprizone-induced mouse model of multiple sclerosis. *Journal of Molecular Histology*. 2022;53(5):817–831.
- [18] Grigorescu C *et al.* Effects of transcranial direct current stimulation on information processing speed, working memory, attention, and social cognition in multiple sclerosis. *Frontiers in Neurology*. 2020;11:545377.
- [19] Ayache SS, Chalah MA. The place of transcranial direct current stimulation in the management of multiple sclerosis-related symptoms. *Neurodegenerative Disease Management*. 2018;8(6):411–422.
- [20] Boonstra TW, Nikolin S, Meisener AC, Martin DM, Loo CK. Change in mean frequency of resting-state electroencephalography after transcranial direct current stimulation. *Frontiers in human neuroscience*. 2016;10:270.
- [21] Mangia AL, Pirini M, Cappello A. Transcranial direct current stimulation and power spectral parameters: A tdcs/eeg co-registration study. *Frontiers in human neuroscience*. 2014;8:601.
- [22] Miller J, Berger B, Sauseng P. Anodal transcranial direct current stimulation (tdcs) increases frontal–midline theta activity in the human eeg: A preliminary investigation of non-invasive stimulation. *Neuroscience Letters*. 2015;588:114–119.
- [23] Maeoka H, Matsuo A, Hiyamizu M, Morioka S, Ando H. Influence of transcranial direct current stimulation of the dorsolateral prefrontal cortex on pain related emotions: A study using electroencephalographic power spectrum analysis. *Neuroscience letters*. 2012;512(1):12–16.
- [24] Schestatsky P, Morales-Quezada L, Fregni F. Simultaneous eeg monitoring during transcranial direct current stimulation. *JoVE (Journal of Visualized Experiments)*. 2013;(76):e50426.

- [25] Song M, Shin Y, Yun K. Beta-frequency eeg activity increased during transcranial direct current stimulation. *Neuroreport*. 2014;25(18):1433–1436.
- [26] Boudewyn M, Roberts BM, Mizrak E, Ranganath C, Carter CS. Prefrontal transcranial direct current stimulation (tdcs) enhances behavioral and eeg markers of proactive control. *Cognitive neuroscience*. 2019;10(2):57–65.
- [27] Salim AA, Ali SH, Hussain AM, Ibrahim WN. Electroencephalographic evidence of gray matter lesions among multiple sclerosis patients: A case-control study. *Medicine*. 2021;100(33).
- [28] Leocani L *et al*. Electroencephalographic coherence analysis in multiple sclerosis: Correlation with clinical, neuropsychological, and mri findings. *Journal of neurology, neurosurgery, and psychiatry*. 2000;69(2):192.
- [29] Shirani S, Mohebbi M. Brain functional connectivity analysis in patients with relapsing-remitting multiple sclerosis: A graph theory approach of eeg resting state. *Frontiers in Neuroscience*. 2022;16:801774.
- [30] Ayache S *et al*. Prefrontal tdcs decreases pain in patients with multiple sclerosis. *front neurosci* 10: 147. 2016.
- [31] Gholami M, Nami M, Shamsi F, Jaberi KR, Kateb B, Jaberi AR. Effects of transcranial direct current stimulation on cognitive dysfunction in multiple sclerosis. *Neurophysiologie Clinique*. 2021;51(4):319–328.
- [32] Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (edss). *Neurology*. 1983;33(11):1444–1444.
- [33] Bechi M *et al*. Theory of mind and emotion processing training for patients with schizophrenia: Preliminary findings. *Psychiatry research*. 2012;198(3):371–377.
- [34] Chang CY, Hsu SH, Pion-Tonachini L, Jung TP. Evaluation of artifact subspace reconstruction for automatic eeg artifact removal. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018, 1242–1245.
- [35] Ververidis D, Kotropoulos C. Sequential forward feature selection with low computational cost. In: 2005 13th European Signal Processing Conference. 2005, 1–4.
- [36] Kropotov JD. Functional neuromarkers for psychiatry: Applications for diagnosis and treatment. Academic Press (2016).
- [37] Krupina NA, Churyukanov MV, Kukushkin ML, Yakhno NN. Central neuropathic pain and profiles of quantitative electroencephalography in multiple sclerosis patients. *Frontiers in Neurology*. 2020;10:1380.
- [38] Vazquez-Marrufo M *et al*. Abnormal erps and high frequency bands power in multiple sclerosis. *International Journal of Neuroscience*. 2008;118(1):27–38.
- [39] Nimrich V, Draguhn A, Axmacher N. Neuronal network oscillations in neurodegenerative diseases. *Neuromolecular medicine*. 2015;17:270–284.
- [40] Sailer M *et al*. Influence of cerebral lesion volume and lesion distribution on event-related brain potentials in multiple sclerosis. *Journal of neurology*. 2001;248:1049–1055.
- [41] Mattioli F, Bellomi F, Stampatori C, Capra R, Miniussi C. Neuroenhancement through cognitive training and anodal tdcs in multiple sclerosis. *Multiple Sclerosis Journal*. 2016;22(2):222–230.
- [42] Fiene M, Rufener KS, Kuehne M, Matzke M, Heinze HJ, Zaehle T. Electrophysiological and behavioral effects of frontal transcranial direct current stimulation on cognitive fatigue in multiple sclerosis. *Journal of neurology*. 2018;265:607–617.
- [43] Marceglia S *et al*. Transcranial direct current stimulation modulates cortical neuronal activity in alzheimer’s disease. *Frontiers in neuroscience*. 2016;10:134.
- [44] Saxe R, Kanwisher N. People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind”. In: *Social neuroscience*. Psychology Press, 2013, 171–182.
- [45] Santiesteban I, Banissy MJ, Catmur C, Bird G. Functional lateralization of temporoparietal junction-imitation inhibition, visual perspective-taking and theory of mind. *European Journal of Neuroscience*. 2015;42(8):2527–2533.
- [46] Mai X *et al*. Using tdcs to explore the role of the right temporo-parietal junction in theory of mind and cognitive empathy. *Frontiers in psychology*. 2016;7:380.

# TOWARDS AUDITORY ATTENTION DECODING WITH NOISE-TAGGING: A PILOT STUDY

H. A. Scheppink<sup>1</sup>, S. Ahmadi<sup>1</sup>, P. Desain<sup>1</sup>, M. Tangermann<sup>1</sup>, J. Thielen<sup>1</sup>

<sup>1</sup>Donders Institute, Radboud University, Nijmegen, the Netherlands

E-mail: jordy.thielen@donders.ru.nl

**ABSTRACT:** Auditory attention decoding (AAD) aims to extract from brain activity the attended speaker amidst candidate speakers, offering promising applications for neuro-steered hearing devices and brain-computer interfacing. This pilot study makes a first step towards AAD using the noise-tagging stimulus protocol, which evokes reliable code-modulated evoked potentials, but is minimally explored in the auditory modality. Participants were sequentially presented with two Dutch speech stimuli that were amplitude-modulated with a unique binary pseudo-random noise-code, effectively tagging these with additional decodable information. We compared the decoding of unmodulated audio against audio modulated with various modulation depths, and a conventional AAD method against a standard method to decode noise-codes. Our pilot study revealed higher performances for the conventional method with 70 to 100 percent modulation depths compared to unmodulated audio. The noise-code decoder did not further improve these results. These fundamental insights highlight the potential of integrating noise-codes in speech to enhance auditory speaker detection when multiple speakers are presented simultaneously.

## INTRODUCTION

People suffering from hearing loss often have great difficulty in scenarios in which multiple individuals are speaking simultaneously, known as the ‘cocktail party scenario’, something which normal hearing persons have no difficulties with [1]. In these scenarios, hearing aids are not able to provide a good solution, as even though they are capable to suppress background noise, they are less capable of suppressing the unattended speakers. Some hearing aids attempt to mitigate this problem by using a heuristic, for example by enhancing the loudest or closest speaker, or the one who stands right in front of the listener. Unfortunately, these heuristics often lead to selecting the wrong speaker in real-life scenarios.

As an alternative, auditory attention decoding (AAD) aims to decode the attended speaker from neural activity, as the synchronization between the listener’s brain signals and the attended speech envelope is stronger than with the ignored speech envelope [2]. This finding laid the groundwork for more research on hearing aids that allow for cognitive control, so called neuro-steered hear-

ing aids [3]. These hearing aids aim to identify the attended speaker from neural activity, and correspondingly enhance this speaker’s audio signal whilst simultaneously suppressing the other speakers and background noise.

The main idea behind such AAD approaches is to match the speech signals to the neural activity which synchronizes with the attended speech signal. For practical reasons typically electroencephalography (EEG) is used. Most AAD algorithms follow a *stimulus reconstruction* approach, also referred to as *backward modeling* or *decoding*. In this approach, the attended speech envelope is reconstructed from the EEG using a neural decoder, before the speaker whose envelope has the highest similarity with the reconstructed envelope is assumed to be the attended speaker [4]. Another approach is *forward modeling* or *encoding*, in which the objective is to predict the neural response from the speech envelopes via an encoder, and to compare these against the EEG [5, 6]. A third approach, sometimes referred to as the *hybrid* approach, combines decoding and encoding, by transforming both the speech envelopes and the EEG to minimize the irrelevant variance [7, 8]. Using such a hybrid approach, Geirnaert and colleagues [3] achieved a remarkable performance using canonical correlation analysis (CCA) to decode the attended speaker. Presenting audio from two simultaneous speakers, a mean accuracy of about 85 % was reached using 30 s decision windows. However, when decreasing the decision window length to about 10 s, the accuracy quickly dropped to below 80 %. This poses a significant limitation for real-world scenarios where fast speaker detection is crucial.

Framing the speaker decoding problem as detecting which of several stimuli a person is attending to, another paradigm from the brain-computer interfacing (BCI) field recently reached remarkable performances. Specifically, the locus of visual attention can be decoded from EEG data using the code-modulated visual evoked potential (c-VEP). A c-VEP is the EEG response to pseudo-random visual stimulation sequences where stimuli are watermarked using noise-codes, a protocol called noise-tagging [9]. These noise-codes are selected or even optimized to be dissimilar, such that attending to one noise-code evokes substantially different brain activity than when attending to another code, facilitating the decoding of the attended stimulus. Such c-VEP BCIs have been reaching state-of-the-art performances up to 100 % clas-

sification accuracy using 1–4 s decision windows [10] or recently even within 300 ms [11] and a high number of stimuli, 29 and 40, respectively.

This study aims to create fundamental insights in the application of the noise-tagging protocol for auditory attention decoding. This can be realized through “watermarking” the speech signal with the pseudo-random noise-codes. To accomplish this, we propose modulating the amplitude of each speech audio with the amplitude of a unique noise-code, effectively embedding the noise information within the speech signal. In turn, not only can we decode the attended speaker based on the speech envelope, but we can also leverage the hidden noise-tags for enhanced speaker identification.

In this pilot study, we make the first step towards auditory attention decoding using noise-tagging. Firstly, we aim to assess the feasibility of decoding the code-modulated auditory evoked potential (c-AEP), the response to auditory noise-tagging. Therefore, we use sequential presentation, i.e., only one stimulus is presented at a time. Secondly, in this work we compare various modulation depths against no modulation. Thirdly, we compare how decoding approaches based on the speech envelope and noise-tag compare in terms of classification accuracy.

Successful implementation could make the step towards improving the decoding accuracy and speed in identifying the attended speaker. Furthermore, this exploration may pioneer a novel research avenue for the application of code-modulated responses in the auditory domain, a domain that has seen limited application compared to the visual modality, as so far only one study attempted this [12].

## MATERIALS AND METHODS

**Participants:** Five participants (aged 19–31 years, average 23 years, 3 females and 2 males) participated in the pilot experiment. Two of these participants were authors of this study. All participants gave written informed consent prior to the experiment. The experimental procedure and methods were approved by and performed in accordance with the guidelines of the local ethical committee of the Faculty of Social Sciences of Radboud University.

**Materials:** The EEG data were recorded at a sample rate of 500 Hz with 64 active electrodes placed according to the 10-10 system and amplified by a BrainAmp (Brain Products GmbH) amplifier. The EEG data were preprocessed with a non-causal FIR notch filter at 50 Hz and a bandpass filter between 1 and 20 Hz before resampling to 120 Hz. For filtering, the default settings were used from the MNE toolbox, version 1.6.1 [13].

The auditory stimuli were two Dutch short stories [14], narrated by two different male speakers and recorded at 44100 Hz. They lasted approximately 6.5 min each and were a subset of the stimulus materials used by Das and colleagues [15]. Periods of silence exceeding 500 ms were truncated to 500 ms. The stimuli were normalized for loudness and presented dichotically to participants via

headphones, with one story consistently presented to the left and the other to the right ear.

We used two 126-bit binary pseudo-random noise-codes from a family of modulated Gold codes [16, 17] to amplitude-modulate the audio. The codes come in sets that are maximally uncorrelated with each other and each time-shifted versions of themselves. The codes were modulated to include only short (‘010’) and long (‘0110’) events. From the available modulated Gold codes, we carefully selected one that started with a 1, ended with a 0, and exhibited an almost uniform distribution of short and long events. The second code was a 61-bit phase-shifted version of the first. In this way, the noise-codes had identical properties, while minimizing autocorrelation at a maximum delay. The codes were presented at a bit rate of 40 Hz, and we always used the first code to modulate audio for the left ear, while the second code was always used to modulate audio for the right ear. The two stories were presented in their original form, or subjected to amplitude modulation using the binary noise-codes, as shown in Fig. 1. At full modulation depth, i.e. 100 percent, further denoted as condition 100, the audio was directly multiplied with the bit sequence, resulting in undisturbed audio when the noise-code is 1 and complete audio suppression when the code is 0. To avoid potential speech unintelligibility, smaller modulation depths were also tested. For instance, at modulation 90, the audio was dampened by 90 % when the code is 0 and otherwise retained. We tested five modulation conditions: 100, 90, 70, 50, and 0. In other words, condition 0 denotes the unmodulated audio. To address the abrupt bit transitions of the binary codes, we smoothened their edges with a raised cosine function, see Fig. 1.

**Experiment:** During the experiment, participants completed five runs, each corresponding to a distinct modulation condition. The order of conditions was randomized over participants. During a run, two trials were presented, one for each of the two stories, starting with the first story delivered to the left ear accompanied by silence on the right, followed by the second story presented to the right ear with silence on the left. The stories were presented sequentially, to assess the feasibility of auditory noise-tagging before testing the more complex parallel case, where stories would be presented simultaneously.

Throughout a run, a fixation cross was displayed on a mean-luminance gray background at the center of a monitor, positioned approximately 70 cm in front of the participant. Each run started with a 5 s rest, followed by a 1 s cue that indicated the to-be attended side, succeeded by the audio presentation. Participants could take self-paced breaks between runs.

**Analysis:** The classification of the attended speech in this work was done using two approaches, the conventional envelope CCA (further denoted as eCCA) in which only the information of the envelope of the speech signals is used, and reconvolution CCA (further denoted as rCCA), in which both the envelope as well as the noise-codes is used. Both approaches can be applied for the

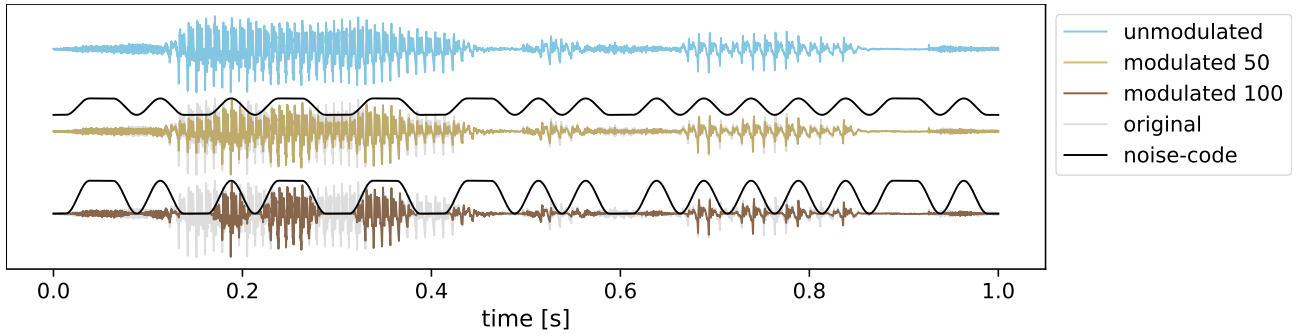


Figure 1: **Visualization of three different modulation depths using noise-tagging.** Depicted is the unmodulated audio, i.e., 0 percent (blue), and 50 (gold) and 100 (brown) percent modulated audio. Additionally, shown are the smoothened noise-tags used for modulation (black). Audio was amplitude-modulated by multiplying with the noise-code, retaining full audio amplitude when the code is 1, while only a percentage when it is zero. Therefore, the noise-code for 50 percent modulation ranges between 0.5–1, instead of 0–1 for 100 percent modulation. To ease comparison, we added the original audio (light gray) at the back of the modulated audio.

modulated audio and corresponding data, however only the eCCA approach can be applied on the unmodulated audio condition as here no code information is available for rCCA. The eCCA is based on the implementation by Geirnaert and colleagues [3]. The use and implementation of rCCA follows the work by Thielen and colleagues [10].

For demonstrating full selective attention, a parallel presentation of stimuli would be required. However, in this pilot study we aimed to make the first step towards auditory noise-tagging, hence the stimuli were presented sequentially. Specifically, when a stimulus was presented to the left ear there was no stimulus presented to the right ear, and vice versa. To evaluate the models, we did however simulate as if the other stimulus of the same condition was presented to the unattended ear, leading to a two-class problem.

Both methods make use of envelope information, which were obtained as follows: the modulated and unmodulated speech signals were each first filtered by a gammatone filterbank resulting in 15 frequency subbands for each speech stimulus [18]. Then, for each subband the absolute value was taken, followed by a power-law compression of 0.6. Next, each subband was lowpass filtered at 20 Hz and resampled to 120 Hz, which were a factor and multiple of the 40 Hz bitrate of the noise-code, respectively. Finally, for each stimulus all subbands were summed, with equal weights, to obtain the stimulus-specific envelopes.

For model evaluation within a single condition, we created four partitions of the data. Specifically, we divided the two 6.5-minute trials into four chronological segments and allocated one segment of each trial to one partition. These four partitions were used for 4-fold cross-validation. This resulted in training and test sets of 9.45 min and 3.15 min, respectively, that had a balanced label distribution.

During the testing of a particular model for a condition, a sliding window was moved over the test data of one class with a stride of 2 samples. To test the sensitivity to data availability, we tested increasing decision window

lengths of  $\tau$  1, 2, 5, 10, 20, 30 and 60 s. The analysis yielded results per sliding window and folds, both of which were averaged to obtain the accuracy of one model in one condition at one decision window length for each participant.

In the following two sections, we explain the two decoding approaches, starting with the eCCA method followed by the rCCA.

**Envelope CCA:** The eCCA approach aims to find a direct correspondence between the EEG data and the speech envelope to detect the attended speaker. Let's assume EEG data  $\mathbf{X} \in \mathbb{R}^{C \times T}$  of  $C$ -many channels (here  $C = 64$ ) and  $T$ -many samples (here  $T$  is one segment size). Additionally, let's assume the speech envelope  $\mathbf{A}_i \in \mathbb{R}^{L \times T}$  for the  $i$ th speaker with  $L$  time-lagged envelopes (here  $L = 60$  for 500 ms at 120 Hz) of  $T$ -many samples each. Then, eCCA optimizes a spatial filter  $\mathbf{w} \in \mathbb{R}^C$  and a temporal filter  $\mathbf{r} \in \mathbb{R}^L$  such that the projected data and envelope are maximally correlated. Let's assume we have a labeled training dataset  $\{(\mathbf{X}_1, \mathbf{y}_1), (\mathbf{X}_j, \mathbf{y}_j), \dots, (\mathbf{X}_J, \mathbf{y}_J)\}$  with  $J$  segments. Then, CCA optimizes the following correlation  $\rho$ :

$$\arg \max_{\mathbf{w}, \mathbf{r}} \rho(\mathbf{w}^\top \mathbf{S}, \mathbf{r}^\top \mathbf{Z}) \quad (1)$$

where  $\mathbf{S} = [\mathbf{X}_1, \mathbf{X}_j, \dots, \mathbf{X}_J]$  are the concatenated training EEG segments, and  $\mathbf{Z} = [\mathbf{A}_{y_1}, \mathbf{A}_{y_j}, \dots, \mathbf{A}_{y_J}]$  are the accompanying concatenated speech envelopes.

To classify new data  $\mathbf{X} \in \mathbb{R}^{C \times T}$  (here  $T = \tau$  the decision window length), eCCA chooses the candidate speech envelope that maximizes the correlation  $\rho$  between the spatially filtered EEG data and the projected speech envelopes:

$$\hat{y} = \arg \max_i \rho(\mathbf{w}^\top \mathbf{X}, \mathbf{r}^\top \mathbf{A}_i) \quad (2)$$

Instead of using the first component only, as in Eq. 2, using multiple components can improve classification accuracy but requires an additional classification model, e.g., a linear discriminant analysis (LDA) [3]. Specifically, CCA can deliver  $K = \min(C, L)$  orthogonal components, ordered on decreasing canonical correlation. The  $k$ -th

component contributes a spatial filter  $\mathbf{w}_k$  and temporal filter  $\mathbf{r}_k$ , and delivers a Pearson's correlation coefficient  $\rho_{ki}$  following Eq. 2. These correlation coefficients across  $K$  components (here  $K = 3$ ) are collected in a vector  $\boldsymbol{\rho}_i$  for speaker  $i$ , and a feature vector  $\mathbf{f}$  is created by subtracting the speakers' canonical correlation vectors,  $\mathbf{f} = \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2$ . The low-dimensional feature vector  $\mathbf{f}$  can then be classified using a vanilla LDA, solving a binary classification problem of whether speaker 1 or speaker 2 was attended.

**Reconvolution CCA:** The rCCA approach consists of a template-matching classifier that predicts the attended speaker given the neural response evoked by the binary noise-code. The reconvolution model is based on the superposition hypothesis, stating that the response to a sequence of events is the linear summation of the responses evoked by the individual events [17].

For the reconvolution, the event time-series  $\mathbf{E}_i \in \mathbb{R}^{E \times T}$  for  $E$ -many events and  $T$ -many samples (here  $T$  is one segment size) denotes the onsets of the  $E$  events for the  $i$ th noise-code. In this work, we used  $E = 2$  events being the short and long events in the noise-codes.

The event matrix is mapped to a structure matrix  $\mathbf{M}_i \in \mathbb{R}^{M \times T}$  for  $M$ -many event time points and  $T$ -many samples. This matrix maps each event to an impulse response function. Specifically, this matrix is Toeplitz-like and describes the onset, duration, and importantly the overlap of each of the events. Assuming both events evoke a response of identical length  $L$ , then  $M = E * L$  (here  $L = 60$  for 500 ms at 120 Hz).

In this work, we extend the standard rCCA model from Thielen and colleagues [10] to incorporate envelope information. This is a crucial step, because a 1 in the code does not necessitate that there was audio in the stimulus. By incorporating the envelope and combining these with the events in the structure matrix, it can be avoided that an event is expected even though there was an audio amplitude of zero in the speech signal at that time. This is achieved by element-wise multiplying the event matrix  $\mathbf{E}_i$  by the amplitudes of the envelope  $\mathbf{A}_i$ , before mapping the event matrix to a Toeplitz-like structure matrix.

Let's assume we have a training dataset  $\{(\mathbf{X}_1, y_1), (\mathbf{X}_j, y_j), \dots, (\mathbf{X}_J, y_J)\}$  including the labeled EEG data for  $j \in \{1, \dots, J\}$  trials with the EEG data  $\mathbf{X} \in \mathbb{R}^{C \times T}$  of  $C$ -many channels and  $T$ -many samples and the associated binary label  $y \in \{0, 1\}$ . To find the optimal spatial filter  $\mathbf{w}$  and temporal response vector  $\mathbf{r}$ , a CCA maximizes the correlation  $\rho$  in the projected spaces:

$$\arg \max_{\mathbf{w}, \mathbf{r}} \rho(\mathbf{w}^\top \mathbf{S}, \mathbf{r}^\top \mathbf{D}) \quad (3)$$

where  $\mathbf{S} = [\mathbf{X}_1, \mathbf{X}_j, \dots, \mathbf{X}_J]$  are the concatenated single trials and  $\mathbf{D} = [\mathbf{M}_{y_1}, \mathbf{M}_{y_j}, \dots, \mathbf{M}_{y_J}]$  are the concatenated accompanying structure matrices.

To classify new data  $\mathbf{X} \in \mathbb{R}^{C \times T}$  (here  $T = \tau$  is the decision window length), rCCA maximizes the correlation  $\rho$  between the spatially filtered data and the projected structure matrix that contains the speech envelope:

$$\hat{y} = \arg \max_i \rho(\mathbf{w}^\top \mathbf{X}, \mathbf{r}^\top \mathbf{M}_i) \quad (4)$$

In this work, for rCCA, we only used the first CCA component for classification similar to the application of rCCA in the visual domain [10].

The code for the reconvolution CCA approach is available at <https://github.com/thijor/pyntbci>.

## RESULTS

This work aimed to investigate fundamental insights in the application of noise-codes for auditory attention decoding. Two different methods were studied: eCCA which used the speech envelope; and rCCA which leveraged the noise-codes. In total, five conditions were used, from audio without modulation (0), to those with increasing modulation depths (50, 70, 90), to audio with full amplitude modulation (100). To investigate the speed of the models, sliding decision windows of length  $\tau$  ranging from 1 to 60 s were used during testing. The mean classification accuracy for all conditions and both models per decision window length is shown in Fig. 2. For an overview of the decoding accuracy for several decision window lengths, see Tab. 1.

Table 1: **Classification accuracy.** Listed are the grand average accuracy for the modulation conditions and both methods at decision window lengths  $\tau$  of 1, 10, 30, and 60 s. Bold values indicate which method (eCCA or rCCA) reached a higher absolute accuracy within that decision window for a particular condition.

	$\tau$	1 s	10 s	30 s	60 s
0	eCCA	60.4	80.7	93.2	98.2
50	eCCA	58.3	80.2	<b>89.9</b>	96.4
	rCCA	<b>61.7</b>	<b>80.4</b>	89.6	96.4
70	eCCA	<b>63.7</b>	83.0	96.1	<b>100.0</b>
	rCCA	61.7	<b>84.7</b>	<b>96.4</b>	99.6
90	eCCA	<b>66.7</b>	<b>89.4</b>	<b>95.7</b>	<b>97.1</b>
	rCCA	59.1	75.6	86.0	90.8
100	eCCA	<b>69.2</b>	<b>94.2</b>	<b>99.6</b>	<b>100.0</b>
	rCCA	59.6	81.0	94.7	99.0

The eCCA method applied to the unmodulated condition (0), which was the baseline in this study, reached a 60 % decoding accuracy for a 1 s decoding window, 80.7 % for 10 s, 93.2 % for 30 s, and 98.2 % for 60 s. Also using eCCA, but with the 100 modulation, the highest absolute decoding performance was reached for all decision windows; 69.2 %, 94.2 %, 99.6 %, and 100.0 % for 1, 10, 30, and 60 s respectively. Modulation condition 90 performed better than 0 for all decision window lengths except 60 s, where it reached 97.1 %. The 70 modulation performed better than 0 modulation for all decision windows. Lastly, modulation condition 50 reached lower performances than 0 modulation for all decision windows. When using rCCA and comparing absolute values, it was observed that the 70 modulation condition performed best for all decision window lengths, whereas 90 modulation performed the worst. The 100 modulation performed worse than 70 modulation but overall better than 50 modulation. For example at  $\tau = 30$  s and ordered at increasing

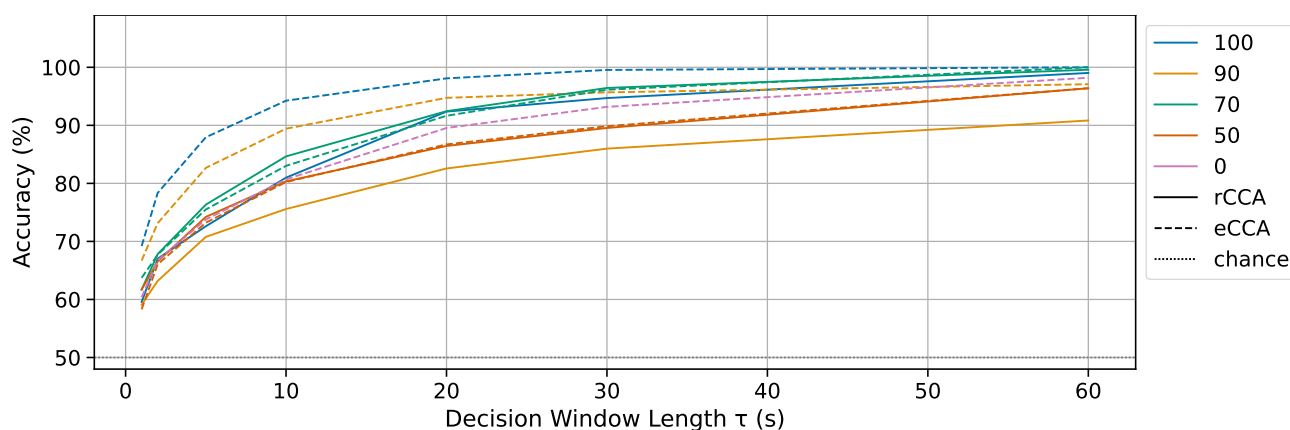


Figure 2: **Decoding accuracy across decision window length and modulation depth.** Depicted is the grand average classification accuracy across decision window length  $\tau$ . Colored lines represent the five modulation conditions: 100 (blue), 90 (orange), 70 (green), 50 (red), and 0 (pink). Solid lines show the performance of rCCA and dashed lines for eCCA. The dashed horizontal gray line depicts theoretical chance level accuracy (50%).

accuracy, 90 modulation reached 86.0 %, 50 modulation 89.6 %, 100 modulation reached 94.7 %, and 70 modulation achieved a performance of 99.6 %.

When comparing the absolute performance values for eCCA against rCCA, it can be observed that the 100 and 90 modulation conditions reached higher performances with eCCA, for all decision windows. For modulation conditions 70 and 50, rCCA reached overall on par performances as eCCA.

When comparing the absolute performance values of eCCA on the unmodulated condition (0) and the modulation conditions using rCCA, it can be observed that the 70 modulation condition (with rCCA) achieved higher performances than the unmodulated condition (with eCCA) for all decision window lengths. For example, for  $\tau = 10$  s, 80.7 % was reached for 0 modulation with eCCA and 84.7 % for 70 modulation with rCCA, and for  $\tau = 30$  s 93.2 % and 96.4 %, respectively. The 100 modulation with rCCA reached higher performances than 0 modulation for decision window lengths of  $\tau \geq 10$  s. The 90 modulation condition with rCCA performed worse than 0 modulation with eCCA for all decision window lengths. For the 50 modulation condition with rCCA, the 0 modulation with eCCA reached higher performances, although for some decision windows rCCA reached an on par or somewhat higher performance, for example for  $\tau = 1$  s 50 modulation with rCCA reached 61.7 % and 0 modulation with eCCA 60.4 %.

In general, using longer decision window lengths  $\tau$  was beneficial for the decoding performance of both methods for all conditions.

## DISCUSSION

This preliminary work aims to contribute to our understanding of fundamental protocol design decisions when noise-tagging is integrated to perform AAD. We argue that noise-tagging may provide additional information to the audio which could enhance AAD performance. In this

pilot study with sequential presentation of the stimuli, we studied five modulation conditions for the speech signals and compared two CCA-based decoding approaches. The results showed that the envelope CCA (eCCA) method achieved higher performances with full modulation than without modulation, while the alternative reconvolution CCA (rCCA) preferred a 70 percent modulation intensity to reach peak performance.

For all decision window lengths, the 100 and 70 modulation conditions performed better than the unmodulated condition for eCCA. It could be speculated that this is due to the modulation adding distinctive uncorrelated high-frequency content. Generally, a 1-9 Hz range was found most informative for cortical tracking of speech envelopes [15]. Arguably, adding the noise-tags at a 40 Hz bitrate increases the envelope's frequency range, which could increase the speed with which the envelopes become distinctive for eCCA. Therefore, we used a 20 Hz lowpass and a higher sampling frequency.

The lowest modulation depth that reached on par or higher performance with the unmodulated condition was a modulation of 70, both for using eCCA and rCCA. For both methods, a 50 modulation depth was able to reach similar performances as the 0 modulation, for decision window lengths of  $\tau \leq 10$  s. Future work could assess the perception thresholds for modulation depths, to obtain the least intrusive protocols for high usability.

Overall, the performance of rCCA matched that of eCCA, or was lower. However, differences exist between these methods. Firstly, eCCA emphasizes global and higher-order activity associated with the speech envelope, while rCCA may focus more on early sensory responses evoked by the noise-codes. Secondly, eCCA uses multiple CCA components and an LDA classifier. While such enhancements could potentially also benefit rCCA, we stayed close to existing literature for this pilot study. Future investigations should evaluate rCCA's decoding performance upon integrating these additional optimizations.

A strong characteristic of applying amplitude modulation

using noise-codes is that the resulting decoding is less limited by the distinctiveness of the envelope of the audio signal, for example in speech. The method could therefore also be applied more broadly to any type of audio signal, the envelope of which may be more or less correlated, such as music.

As this work is a preliminary study, a number of inherent limitations should be noted. First, the experiment only included five participants, of which two were authors with substantial experience as BCI users. This is a small sample size, which included motivated participants, and needs to be enlarged in future studies. Second, in order to assess the feasibility of decoding a c-AEP response, this pilot used sequential stimulation instead of presenting the two speakers in parallel. Future studies need to investigate, if our observations generalize also to parallel stimulus presentation protocols.

Future work could include various other improvements, such as the noise-codes optimized to maintain speech qualities. Additionally, the noise-codes could be shortened, as the number of classes typically may be lower than 63 in an auditory attention scenario.

## CONCLUSION

Our work showed that adding noise-tags to a speech signal in a sequential paradigm can enhance the decoding performance compared to decoding the original unmodulated speech signal. Specifically, for shorter decision window lengths all higher modulation depths (100, 90 and 70) performed better than the unmodulated condition. The rCCA method on the 70 modulation condition also performed better than the unmodulated condition. Lastly, for the modulated conditions, the eCCA method performed better than or on par with the rCCA method. Overall, these results show the potential of using noise-tags in the auditory modality, and is the first step towards using the noise-tagging protocol for auditory attention decoding.

## ACKNOWLEDGMENTS

We thank K. van der Heijden for helpful insights into AAD.

## REFERENCES

[1] Cherry EC. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*. 2005;25(5):975–979.  
[2] Ding N, Simon JZ. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*. 2012;109(29):11854–11859.  
[3] Geirnaert S *et al*. Electroencephalography-based auditory attention decoding: Toward neurosteered hearing devices. *IEEE Signal Processing Magazine*. 2021;38(4):89–102.

[4] O’Sullivan JA *et al*. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral cortex*. 2015;25(7):1697–1706.  
[5] Lalor EC, Foxe JJ. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*. 2010;31(1):189–193.  
[6] Ding N, Simon JZ. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*. 2012;107(1):78–89.  
[7] de Cheveigné A, Wong DD, Di Liberto GM, Hjortkjær J, Slaney M, Lalor E. Decoding the auditory brain with canonical component analysis. *NeuroImage*. 2018;172:206–216.  
[8] Dmochowski JP, Ki JJ, DeGuzman P, Sajda P, Parra LC. Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage*. 2018;180:134–146.  
[9] Martínez-Cagigal V, Thielen J, Santamaria-Vazquez E, Pérez-Velasco S, Desain P, Hornero R. Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): A literature review. *Journal of Neural Engineering*. 2021;18(6):061002.  
[10] Thielen J, Marsman P, Farquhar J, Desain P. From full calibration to zero training for a code-modulated visual evoked potentials for brain-computer interface. *Journal of Neural Engineering*. 2021;18(5):056007.  
[11] Shi N *et al*. Estimating and approaching the maximum information rate of noninvasive visual brain-computer interface. *NeuroImage*. 2024;120548.  
[12] Farquhar J, Blankespoor J, Vlek R, Desain P. Towards a noise-tagging auditory BCI-paradigm. In: *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course 2008*. Graz, Austria. 2008.  
[13] Gramfort A *et al*. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*. 2013;7(267):1–13.  
[14] Radioboeken voor kinderen. 2007. [Online]. Available: <https://deburen.eu/radioboeken>.  
[15] Das N, Biesmans W, Bertrand A, Francart T. The effect of head-related filtering and ear-specific decoding bias on auditory attention detection. *Journal of Neural Engineering*. 2016;13:056014.  
[16] Gold R. Optimal binary sequences for spread spectrum multiplexing. *IEEE Transactions on Information Theory*. 1967;13(4):619–621.  
[17] Thielen J, Broek P van den, Farquhar J, Desain P. Broad-band visually evoked potentials: Re(con)volution in brain-computer interfacing. *PLOS ONE*. 2015;10(7):e0133797.  
[18] Biesmans W, Das N, Francart T, Bertrand A. Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2017;25(5):402–412.

# TOWARDS GAZE-INDEPENDENT C-VEP BCI: A PILOT STUDY

S. Narayanan<sup>1</sup>, S. Ahmadi<sup>1</sup>, P. Desain<sup>1</sup>, J. Thielen<sup>1</sup>

<sup>1</sup>Donders Institute, Radboud University, Nijmegen, the Netherlands

E-mail: jordy.thielen@donders.ru.nl

**ABSTRACT:** A limitation of brain-computer interface (BCI) spellers is that they require the user to be able to move the eyes to fixate on targets. This poses an issue for users who cannot voluntarily control their eye movements, for instance, people living with late-stage amyotrophic lateral sclerosis (ALS). This pilot study makes the first step towards a gaze-independent speller based on the code-modulated visual evoked potential (c-VEP). Participants were presented with two bi-laterally located stimuli, one of which was flashing, and were tasked to attend to one of these stimuli either by directly looking at the stimuli (overt condition) or by using spatial attention, eliminating the need for eye movement (covert condition). The attended stimuli were decoded from electroencephalography (EEG) and classification accuracies of 88 % and 100 % were obtained for the covert and overt conditions, respectively. These fundamental insights show the promising feasibility of utilizing the c-VEP protocol for gaze-independent BCIs that use covert spatial attention when both stimuli flash simultaneously.

## INTRODUCTION

A brain-computer interface (BCI) records its users' brain activity and translates it into a computer command, opening a novel non-muscular channel for communication and control [1]. Typically, a BCI records brain activity with electroencephalography (EEG) because it is affordable, practical, and non-invasive.

One of the fastest BCIs for communication uses the code-modulated visual evoked potential (c-VEP) as measured in the EEG [2]. The c-VEP is observed during visual stimulation of the user with a pseudo-random sequence of flashes. As each of the presented symbols concurrently flickers with a random but unique sequence of flashes, specific brain activity is evoked when the user attends to one of the symbols. Subsequently, machine learning algorithms infer the attended symbol from the users' evoked brain activity. Such a visual BCI speller allows its user to select symbols or commands and as such communicate, bypassing most of the motor system [3].

Unfortunately, an important limitation of a standard visual BCI speller is the requirement of the users' eyes to shift their gaze towards (i.e., fixate on) a target symbol. Because BCI control is fully dependent on eye movements, this poses a major challenge and quickly renders the BCI uncontrollable for people who have lost volun-

tary control of their eye movements, i.e., people living with late-stage amyotrophic lateral sclerosis (ALS).

In the visual domain, several studies have attempted to develop a gaze-independent BCI. For instance, Blankertz and colleagues developed a BCI speller called the 'Hex-o-Spell' that used motor imagery (imagined right hand and right foot movement, i.e.,  $N = 2$  classes) of the user to aid the selection of characters from six hexagonal fields [4]. They reported a typing speed of 2.3–5 char/min and 4.6–7.6 char/min, for their two participants respectively. Interestingly, Treder and Blankertz showed that visual covert spatial attention can also be used to operate the 'Hex-o-Spell' and the 'Matrix' speller using the P300 event-related potential (ERP) [5]. This covert 'Hex-o-Spell' outperformed the covert 'Matrix' speller, with a classification accuracy of 60 % ( $N = 36$  classes) and 40 % ( $N = 30$  classes), respectively.

Furthermore, work by Treder and colleagues compared the P300-based 'Hex-o-Spell', the 'Cake Speller', which is similar to the former, and a 'Center Speller', where unique geometric shapes with different colors were closely surrounded by characters, and presented centrally on the screen in a sequential fashion [6]. A classification accuracy of 91.3 %, 88.2 %, and 97.1 % was reported for the three spellers, respectively ( $N = 30$  classes). Similarly, Chen and colleagues [7] used an extension of the P300 oddball paradigm, namely, rapid serial visual presentation (RSVP). The authors used two versions: a colored circles paradigm (CCP), and a dummy face paradigm (DFP). The average performances obtained from the CCP and DFP paradigms were in the range 51.6 % and 73.5 %, respectively.

Additionally, Treder and colleagues, in another instance, focused on using changes in alpha band activity induced by covert attention shifts to classify the direction in which attentional shifts occurred [8]. The authors showed that a classification accuracy of 73.65 % was obtained ( $N = 2$  classes). These results indicate the potential of using alpha activity as a feature for spatial attention decoding in gaze-independent BCIs.

Furthermore, Kelly and colleagues designed a gaze-independent BCI for communication by combining features from the steady-state visual evoked potential (SSVEP) and alpha band modulations to decode covert spatial attention [9]. The authors reported an average performance of 70.3 %, 72.8 % and 79.5 % when using the SSVEP, alpha band, or both features in their analysis

pipeline, respectively ( $N = 2$  classes). Similarly, Egan and colleagues [10] aimed for a hybrid gaze-independent speller using the P300 ERP and alpha in addition to the SSVEP. Importantly, adding the P300 response and alpha as additional features in their classification pipeline improved the performance by 17 % to an overall 79 % when compared to the performance using only the SSVEP, achieving 62 % ( $N = 2$  classes).

In this pilot study, we work towards a gaze-independent BCI. The gaze-dependent c-VEP has recently demonstrated exceptional performance, surpassing other evoked paradigms like ERP and SSVEP [11]. Another study revealed the reliable decoding of c-VEP from peripheral stimulation (away from fixation) compared to direct foveal stimulation (at fixation) [12].

Our objective is to acquire fundamental insights on the feasibility of decoding the c-VEP in a fully gaze-independent manner. Specifically, participants will use covert spatial attention to concentrate on stimuli, eliminating the need for direct eye movements to foveate on them. In this pilot work, the stimuli were presented sequentially, to assess whether the c-VEP can be decoded from the far periphery, before testing the more complex parallel stimulation case, where stimuli would be presented simultaneously. If successful, this study provides the first steps to a gaze-independent c-VEP BCI, potentially providing a high-speed neuro-technological assistive device for individuals who may not have reliable control of their eye movements.

## MATERIALS AND METHODS

**Participants:** Five participants (all male, mean age 31 years, range 24-50 years) were included in this study after obtaining written informed consent. Two participants were authors of this study. A pre-screening procedure excluded any participants with a history of epilepsy or brain injury. All participants had normal or corrected-to-normal vision and reported no central nervous system abnormalities. This study was approved by the Ethical Committee of the Faculty of Social Sciences at the Radboud University Nijmegen.

**Materials:** EEG data from 64 Ag/AgCl active electrodes placed according to the international 10-10 system were recorded at 512 Hz amplified by a Biosemi ActiveTwo amplifier. The data were preprocessed using a notch filter at 50 Hz and a bandpass filter with a lower cutoff at 1 Hz and a higher cutoff at 40 Hz. Subsequently, the data were sliced to trials starting at 500 ms before stimulus onset until 20 s after stimulus onset. Finally, the data were downsampled to 120 Hz, and the 500 ms pre-stimulus that may have captured filter artefacts due to initial slicing and subsequently filtering were removed.

The stimulus protocol (see Fig. 1) was displayed on a 27 in Corsair Xeneon 27QHD240 OLED screen at a  $1920 \times 1080$  px resolution with a 120 Hz refresh rate. The participants were seated at a 60 cm distance in front of the display. A black fixation cross was presented at the cen-

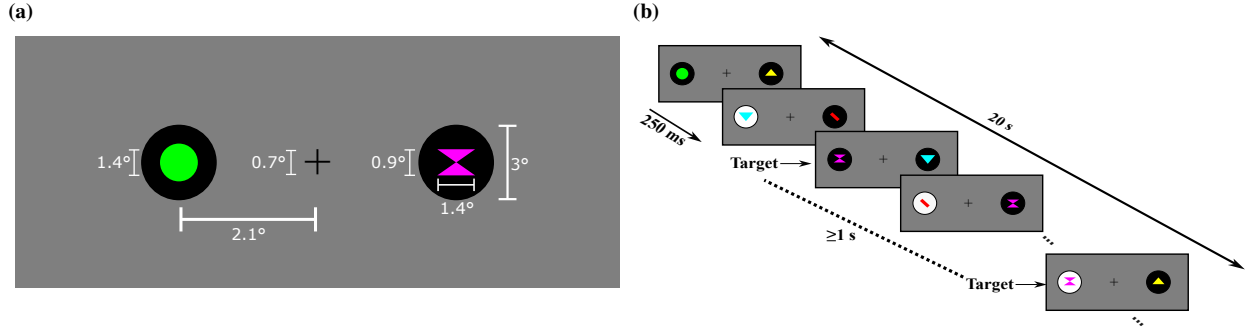
ter of the screen on a mean luminance gray background. To each of the sides of the fixation cross at a distance of  $2.1^\circ$ , two circles with a  $3^\circ$  diameter were presented.

The circles' background color was luminance modulated with binary pseudo-random noise-codes, such that ones represent a white and zeros a black background. We used 126-bit modulated Gold codes [13, 14], which contained only short flashes of 16,67 ms (bit sub-sequence '010') and long flashes of 33,33 ms (bit sub-sequence '0110') at a presentation rate of 60 Hz. From the available modulated Gold codes, we carefully selected one for the left side. For the right circle a 61 bits phase-shifted version of the left code was used. This was done such that the noise-codes' properties were identical, but still had a near-zero correlation at a maximum delay.

Inside the circles ( $3^\circ$  diameter), five colored shapes were presented with a maximum possible height and width of  $1.4^\circ$  each. The shapes and their colors are as follows : a green circle ( $1.4^\circ$  diameter), magenta hourglass ( $0.9^\circ \times 1.4^\circ$ ), cyan inverted triangle ( $0.9^\circ \times 1.4^\circ$ ), red rectangle ( $1.5^\circ \times 0.5^\circ$ , rotated by  $45^\circ$ ), and the yellow triangle ( $0.9^\circ \times 1.4^\circ$ ). All shapes had the same brightness and were sequentially presented in random order at a rate of 4 Hz (see Fig. 1). Participants were asked to count the number of times that the magenta hourglass, i.e., the target shape, occurred on the cued side, to facilitate sustaining their attention and to evaluate the behavioral performance of attending to each of the sides. Within a trial, the temporal distance between the presentation of two target shapes was at least 1 s, the target shape could not be presented on both sides simultaneously, and the number of times the target shape was presented differed for the two sides within a trial.

In this pilot study, we used sequential stimulation in both overt and covert runs to make the first step towards gaze-independent c-VEP BCI. That is, only the circle on the attended (cued) side underwent alternating background changes based on the pseudo-random noise-code, while the unattended side retained a constant black background. Notably, both sides featured distinct shape sequences despite this sequential stimulation protocol.

**Experiment:** During the experiment, participants completed five runs: four runs required covert attention and one required overt attention, the order of which was randomized across participants. Each run consisted of 20 trials, 10 for each of the two classes, in random order. At the start of a run, a 5 s period was used to let the participant prepare for the upcoming trials. At the start of a trial, a 1-second cue was presented to indicate the to-be-attended side using an arrow. Subsequently, for a duration of 20 s, the cued circle flashed according to its bit sequence while the uncued circle remained static, while both circles showed their distinct shape sequences. At the end of a trial, participants were given a maximum of 5 s to enter the number of target shapes they counted on the attended side using a keyboard, after which they received feedback for a period of 1 s on the correctness of their response. Finally, before continuing to the next trial, a



**Figure 1: Stimulus protocol.** In (a), a graphical representation of the stimulus interface is depicted, featuring two stimuli positioned at  $2.1^\circ$  on either side of a fixation cross. The stimuli took the form of circles measuring  $3^\circ$  in both height and width. The fixation cross was  $0.7^\circ$  for each side. The shapes presented were bound to a maximum height and width of  $1.4^\circ$  each. The shapes' heights and widths were as follows: green circle ( $1.4^\circ$  diameter), inverted cyan triangle and yellow triangle ( $0.9^\circ \times 1.4^\circ$ ), magenta hourglass ( $0.9^\circ \times 1.4^\circ$ ) and the red rectangle ( $1.5^\circ \times 0.5^\circ$ , rotated by  $45^\circ$ ). In (b), a graphical representation of the stimulus protocol is depicted comprising two crucial components: first, the background of the stimuli underwent alternating black-and-white transitions following a binary pseudo-random sequence; second, diverse-colored shapes were presented within the stimuli. The stimulus background could dynamically change with each frame of 16.67 ms (60 Hz), while the shapes within the stimuli changed every 250 ms (4 Hz). A trial took 20 s, within which target shapes (the magenta hour glass) appeared randomly in the sequence with at least 1 s distance. Participants engaged with the stimuli by counting the number of target shapes on the attended side. In this pilot study, we adopted a paradigm where only the background of the attended stimulus alternated, while the background of the unattended stimulus remained constant. A left-attended trial is shown in (b).

1 s blank inter-trial interval was presented. At the end of a run, the behavioral accuracy of correct responses was shown on the screen. Participants could take self-paced breaks in between runs.

In summary, we gathered 20 trials for each participant in the overt condition, whereas the covert condition involved the recording of 80 trials per participant. In both conditions, the labels (left and right) were balanced.

**Analysis:** We used a template-matching classifier to predict the attended side (left or right) given the recorded brain activity. Specifically, we used the 'reconvolution' method [14], which assumes that the evoked response to a stimulus sequence can be described by the linear superposition of the responses to the individual flashes in that sequence. The reconvolution approach can substantially reduce the number of parameters while increasing the number of samples to train these parameters, which effectively can limit the required training data [15].

In reconvolution, the event time-series of the  $i$ th stimulus sequence are listed in the event matrix  $\mathbf{E}_i \in \mathbb{R}^{E \times T}$  for  $E$ -many events and  $T$ -many samples. This matrix describes the onset of each of the events in a sequence. In this work, the events were defined as the onset of the stimulation sequence in each trial, and one event for each of the flash durations (short '010' and long '0110'), for a total of  $E = 3$  events.

The event time-series are subsequently transformed to a structure matrix that not only describes the onset, but also the modeled length and importantly the overlap of the transient responses for each of the events in the event matrix. Assuming that the transient response length can be limited to  $L$  samples without losing relevant data, the structure matrix of the  $i$ th stimulus sequence is a Toeplitz-like matrix  $\mathbf{M}_i \in \mathbb{R}^{M \times T}$  for  $M = E * L$  event time points.

Let's assume we have a training dataset  $\{(\mathbf{X}_1, y_1), (\mathbf{X}_j, y_j), \dots, (\mathbf{X}_J, y_J)\}$  including labeled EEG data for  $j \in \{1, \dots, J\}$  trials with the single-trial EEG  $\mathbf{X} \in \mathbb{R}^{C \times T}$  of  $C$ -many channels and  $T$ -many samples and the associated binary label  $y \in \{0, 1\}$ . With this data, we can learn a spatial filter  $\mathbf{w} \in \mathbb{R}^C$  and a temporal response vector  $\mathbf{r} \in \mathbb{R}^M$  by maximizing the following correlation  $\rho$  as part of a canonical correlation analysis (CCA):

$$\arg \max_{\mathbf{w}, \mathbf{r}} \rho(\mathbf{w}^\top \mathbf{S}, \mathbf{r}^\top \mathbf{D}) \quad (1)$$

where  $\mathbf{S} = [\mathbf{X}_1, \mathbf{X}_j, \dots, \mathbf{X}_J]$  are the concatenated single trials and  $\mathbf{D} = [\mathbf{M}_{y_1}, \mathbf{M}_{y_j}, \dots, \mathbf{M}_{y_J}]$  are the concatenated accompanying structure matrices.

Having learned the spatial filter and temporal response vector, we can now predict the label of a new trial  $\hat{y}$  by maximizing the following Pearson's correlation  $\rho$ :

$$\hat{y} = \arg \max_i \rho(\mathbf{w}^\top \mathbf{X}, \mathbf{r}^\top \mathbf{M}_i) \quad (2)$$

Here,  $\mathbf{w}^\top \mathbf{X}$  is the spatially filtered data and  $\mathbf{r}^\top \mathbf{M}_i$  is the predicted response template for the  $i$ th stimulus sequence. To evaluate the performance of the reconvolution CCA on the overt and covert data, we used a chronological 4-fold cross-validation within each condition. The classification accuracy was averaged across folds. Note, the c-VEP stimulation was only applied on the attended side, while the unattended side remained a black background color. In the decoding analysis, we simulated as if the unattended side had been flashing with the noise-code other than the one presented on the attended side.

Code for the reconvolution CCA approach is available at: <https://github.com/thijor/pyntbci>.

## RESULTS

As this study presents the initial step to decode c-VEP from peripheral stimulation, aiming towards covert spatial attention, it is imperative to study how classification accuracy is influenced by the modeled transient response length. Given the potential for distinct transient responses between conditions, we assessed the mean accuracy across transient response lengths spanning from 0.1 to 0.9 s, for all participants (S1-S5) and both conditions (see to Fig. 2).

In the covert condition, mean accuracy fluctuated from 55 % to 99 % across participants, whereas in the overt condition, mean accuracy remained consistently at 100 % for all participants across all transient response lengths.

In the covert condition, participants S3 and S4 achieved a peak accuracy of 85 % and 86 % respectively, observed at a transient response length of 200 ms. Participants S1 and S5 reached a highest accuracy of 88 % and 89 %, respectively, at a transient response length of 300 ms. Participant S2 demonstrated a peak accuracy of 99 % at 400 ms. Notably, the mean accuracy across participants in the covert condition was highest at a transient response length of 300 ms. Hence, for subsequent analysis, we use a transient response length of 300 ms.

Tab. 1 shows the classification accuracy for a transient response length of 300 ms. The scores obtained in the covert condition for S1-5 were 88 %, 98 %, 84 %, 81 % and 89 %, respectively, leading to an average of 88 %. The overt condition performed better for all participants (100 %). All individual scores in Tab. 1 are significantly higher ( $p < .001$ ) than chance level (50 %) as verified by a permutation test using 1000 permutations.

**Table 1: Mean classification accuracy.** The table shows the classification accuracy using a transient response length of 300 ms, for each participant and the grand average, for both overt and covert conditions. All classification results for both conditions and all participants individually were significantly higher than chance (50 %) as verified by a permutation test with 1000 permutations ( $p < .001$ ).

	S1	S2	S3	S4	S5	Avg
<b>Overt</b>	1.00	1.00	1.00	1.00	1.00	1.00
<b>Covert</b>	0.88	0.98	0.84	0.81	0.89	0.88

To investigate the differences in characteristics of the spatial activity patterns and transient responses, we computed these at a transient response length of 300 ms for both conditions. Fig. 3 shows an example of the spatial pattern and transient responses for S4. Across participants, we observed that the spatial activity pattern for the overt condition was more focally distributed, whereas it was more lateralized for the covert condition.

## DISCUSSION

Our pilot study provides fundamental insights into the plausibility of a c-VEP-based stimulation paradigm for

decoding covert spatial attention, thereby potentially eliminating the need for the ability to make eye movements to control a c-VEP BCI. We implemented a two-class paradigm, requiring participants to attend on a stimulus either to the left or to the right of their fixation point. The stimuli background flashed following pseudo-random noise-codes, while their foreground simultaneously presented a random sequence of five distinct shapes with an infrequent target shape. Participants were tasked with counting the occurrences of the target shape amidst the shape sequence (see Fig. 1). In this pilot study, we used sequential stimulation to assess the feasibility of covert c-VEP, before moving to the more complex parallel stimulation requiring covert spatial attention.

In our experiment, participants engaged with the stimuli through either overt means, involving eye movements to foveate on the target, or covertly, relying on spatial attention to focus on a target. In the overt condition, we reached a decoding performance of 100 % for all participants. In the covert condition, we achieved an average accuracy of 88 %. To the best of the authors' knowledge, this marks the first evaluation of a c-VEP BCI using covert attention, although here we still rely on sequential stimulation. Our study highlights the feasibility of such a design for developing gaze-independent BCIs that can be used by people with ALS.

In the overt condition, all participants achieved 100 % accuracy, likely caused by the large data availability, low number of classes, and sequential stimulation. Specifically, this study used 5.3 min of data for training and 20 s for testing, while 1 min training and 1-2 s testing would suffice [15]. In the covert condition, we employed 16 min of data for training, achieving a decoding accuracy of 88 %. This result underscores the lower SNR in the covert condition compared to the overt scenario. Nevertheless, although using sequential stimulation, the attained performance surpasses the 62 % accuracy reported in a similar SSVEP study that used parallel stimulation [10], offering evidence for the potential performance of gaze-independent c-VEP.

It is essential to approach the results of our study on gaze-independent c-VEP BCI with caution and consider two important limitations. Firstly, this preliminary study involved a small cohort of five highly motivated participants. Secondly, the c-VEP protocol employed sequential stimulation, where only the stimulus on the attended side alternated its background based on the pseudo-random noise-code. In practical online usage of the BCI, simultaneous stimulation on both sides is necessary. While our study offers valuable fundamental insights into the feasibility of gaze-independent c-VEP BCI, it is imperative to acknowledge these limitations. Further research, including a larger sample size and parallel stimulation, is crucial to fully unveil the potential of this approach.

Additionally, it is important to acknowledge that stimulation paradigms outside the visual domain have been explored as well for developing independent BCIs. For instance, Schreuder and colleagues developed the P300-

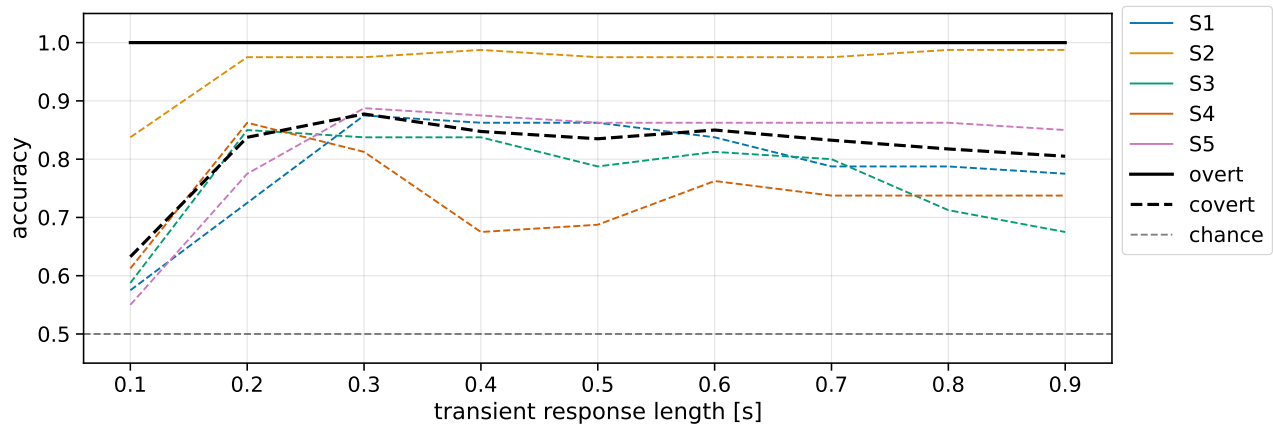


Figure 2: **Classification accuracy across modeled transient response lengths.** Depicted are the participant-specific classification accuracies for both overt (solid lines) and covert (dashed lines) conditions across transient response lengths ranging from 0.1 s to 0.9 s. The grand average over participants is shown in black. Please note, that for the overt condition, the classification accuracy was 100 % for all transient response lengths and all participants. The dashed gray line indicates theoretical chance level (50 %).

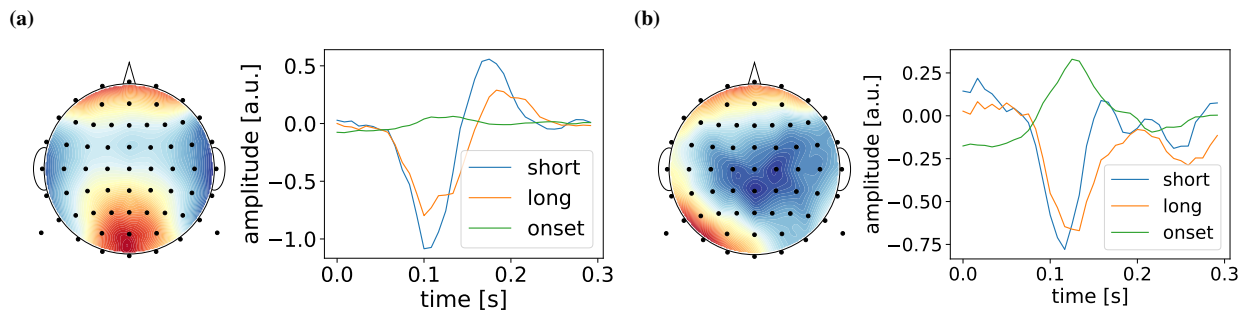


Figure 3: **Spatial activity pattern and transient responses of participant S4.** (a) and (b) show the spatial activity pattern and transient responses of S4 for the overt and covert conditions, respectively. For all participants, the spatial activity for the overt condition was more focally distributed as compared to the more lateralized distribution seen for the covert condition. The spatial pattern  $\mathbf{a} \in \mathbb{R}^C$  was estimated as  $\mathbf{a} = \mathbf{w}^T \mathbf{\Sigma}$ , where  $\mathbf{\Sigma} \in \mathbb{R}^{C \times C}$  is the spatial covariance matrix.

based auditory multi-class spatial ERP (AMUSE) interface reaching a classification accuracy of about 85 % ( $N = 6$  classes) [16]. Similarly, Brouwer and van Erp designed a P300-based BCI using vibro-tactile feedback around the waist with an accuracy of 58 % ( $N = 6$  classes) and 73 % ( $N = 2$  classes) [17]. Moreover, Van der Waal and colleagues [18] used tactile stimulation on the finger tips reaching a classification accuracy of 82 % ( $N = 6$  classes). These results may also highlight the potential to explore the pseudo-random stimulation protocol in the auditory and tactile domain.

Our studies' design enables the use of two additional features in the analysis pipeline, possibly further improving the accuracy. Firstly, the stimulus protocol used in the study was designed such that the infrequent occurrence of the target events within the shape sequence could potentially evoke a P300 response. Hence, the P300 response could be used alongside the c-VEP to decode the attended side, similar to P300 response that was used alongside the SSVEP by Egan and colleagues [10]. Secondly, the alpha-band modulations are expected to be lateralized with respect to the attended side [19]. Specifically,

covertly attending to a stimulus on one side suppresses visual alpha-activity in the contra-lateral (task-positive) hemisphere, while it increases alpha in the ipsi-lateral (task-negative) hemisphere [20]. Hence, visual alpha oscillations can also be used as an additional feature, again similar to the alpha response used alongside the SSVEP in earlier work [10]. Thirdly, aligning with the anticipated lateralization in the alpha-band, we also anticipate lateralization in the c-VEP itself during the covert condition. In our current application of reconvolution CCA, a single spatial filter was employed to decode the attended side. This method can be extended by incorporating distinct spatial filters for each side, a concept referred to as an 'ensemble' decoder [21]. Finally, in the present study, we employed only two stimuli positioned on either side of the fixation point, using luminance modulation with two 126-bit Gold codes. Given the limited number of classes, there is potential to explore shorter codes, which could lead to faster decoding. Furthermore, alternative codes, such as the m-sequence or Golay sequence, may be considered, as they have shown promise in enhancing classification accuracy [22].

## CONCLUSION

Our study shows the feasibility and high performance of a novel covert BCI design based on c-VEP. Our design eliminates the dependence on gaze, which is an essential feature if BCIs are to be used by people that have no voluntary control over their eye movements, such as people living with late stage ALS. Further, the design of the study makes it possible to use additional measures of brain activity to improve classification performance, which is a potential fruitful avenue for future work to improve the efficacy of the gaze-independent c-VEP BCI. Overall, our results suggest the potential for a high-speed BCI that does not rely on any overt behavior.

## ACKNOWLEDGEMENTS

This work was part of the project ‘Obtaining fast brain-computer interfacing without eye movements for communication and control’ with project number OCENW.XS23.1.127 of the research programme ‘Open Competitie ENW XS’ which is financed by the Dutch Research Council (NWO).

## REFERENCES

- [1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*. 2002;113(6):767–791.
- [2] Martínez-Cagigal V, Thielen J, Santamaría-Vázquez E, Pérez-Velasco S, Desain P, Hornero R. Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): A literature review. *Journal of Neural Engineering*. 2021;18(6):061002.
- [3] Verbaarschot C *et al.* A visual brain-computer interface as communication aid for patients with amyotrophic lateral sclerosis. *Clinical Neurophysiology*. 2021;132(10):2404–2415.
- [4] Blankertz B *et al.* The Berlin brain-computer interface presents the novel mental typewriter Hex-o-Spell. 2006.
- [5] Treder MS, Blankertz B. (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behavioral and Brain Functions*. 2010;6:1–13.
- [6] Treder MS, Schmidt NM, Blankertz B. Gaze-independent brain-computer interfaces based on covert attention and feature attention. *Journal of Neural Engineering*. 2011;8(6):066003.
- [7] Chen L, Allison BZ, Zhang Y, Wang X, Jin J. An online gaze-independent BCI system used dummy face with eyes only region as stimulus. In: *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*. Springer International Publishing, 2016, 26–34.
- [8] Treder MS, Bahramisharif A, Schmidt NM, Van Gerven MA, Blankertz B. Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention. *Journal of Neuroengineering and Rehabilitation*. 2011;8(1):1–10.
- [9] Kelly S, Lalor E, Reilly R, Foxe J. Visual spatial attention tracking using high-density SSVEP data for independent brain-computer communication. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2005;13(2):172–178.
- [10] Egan JM, Loughnane GM, Fletcher H, Meade E, Lalor EC. A gaze independent hybrid-BCI based on visual spatial attention. *Journal of Neural Engineering*. 2017;14(4):046006.
- [11] Shi N *et al.* Estimating and approaching the maximum information rate of noninvasive visual brain-computer interface. *NeuroImage*. 2024:120548.
- [12] Waytowich NR, Krusienski DJ. Spatial decoupling of targets and flashing stimuli for visual brain-computer interfaces. *Journal of Neural Engineering*. 2015;12(3):036006.
- [13] Gold R. Optimal binary sequences for spread spectrum multiplexing. *IEEE Transactions on Information Theory*. 1967;13(4):619–621.
- [14] Thielen J, Broek P van den, Farquhar J, Desain P. Broad-band visually evoked potentials: Re(con)volution in brain-computer interfacing. *PLOS ONE*. 2015;10(7):e0133797.
- [15] Thielen J, Marsman P, Farquhar J, Desain P. From full calibration to zero training for a code-modulated visual evoked potentials for brain-computer interface. *Journal of Neural Engineering*. 2021;18(5):056007.
- [16] Schreuder M, Rost T, Tangermann M. Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. *Frontiers in Neuroscience*. 2011;5.
- [17] Brouwer. A tactile P300 brain-computer interface. *Frontiers in Neuroscience*. 2010.
- [18] Waal M van der, Severens M, Geuze J, Desain P. Introducing the tactile speller: An ERP-based brain-computer interface for communication. *Journal of Neural Engineering*. 2012;9(4):045002.
- [19] Worden MS, Foxe JJ, Wang N, Simpson GV. Anticipatory biasing of visuospatial attention indexed by retinotopically specific  $\alpha$ -band electroencephalography increases over occipital cortex. *The Journal of Neuroscience*. 2000;20(6):RC63–RC63.
- [20] Jensen O, Mazaheri A. Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*. 2010;4:186.
- [21] Gembler FW, Benda M, Rezeika A, Stawicki PR, Volosyak I. Asynchronous c-VEP communication tools—efficiency comparison of low-target, multi-target and dictionary-assisted BCI spellers. *Scientific Reports*. 2020;10(1):1–13.
- [22] Thielen J. Effects of stimulus sequences on brain-computer interfaces using code-modulated visual evoked potentials: An offline simulation. In: *Advances in Computational Intelligence*. Springer Nature Switzerland: Cham, 2023, 555–568.

# APPROXIMATE UMAP ALLOWS FOR HIGH-RATE ONLINE VISUALIZATION OF HIGH-DIMENSIONAL DATA STREAMS

Peter Wassenaar<sup>1\*</sup>, Pierre Guetschel<sup>1\*</sup>, Michael Tangermann<sup>1</sup>

\*These authors contributed equally to this work.

<sup>1</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

E-mail: pierre.guetschel@donders.ru.nl

**ABSTRACT:** In the BCI field, introspection and interpretation of brain signals are desired for providing feedback or to guide rapid paradigm prototyping but are challenging due to the high noise level and dimensionality of the signals. Deep neural networks are often introspected by transforming their learned feature representations into 2- or 3-dimensional subspace visualizations using projection algorithms like Uniform Manifold Approximation and Projection (UMAP) [1]. Unfortunately, these methods are computationally expensive, making the projection of data streams in real-time a non-trivial task. In this study, we introduce a novel variant of UMAP, called approximate UMAP (aUMAP). It aims at generating rapid projections for real-time introspection. To study its suitability for real-time projecting, we benchmark the methods against standard UMAP and its neural network counterpart parametric UMAP [2]. Our results show that approximate UMAP delivers projections that replicate the projection space of standard UMAP while decreasing projection speed by an order of magnitude and maintaining the same training time.

## INTRODUCTION

The recording of neural signals offers a window into understanding brain activity, with potential applications in various fields. However, a considerable challenge lies in the fact that these signals, particularly electroencephalograms (EEG), are high-dimensional and very susceptible to noise. Consequently, this situation requires the development of specialized analysis techniques to describe and eventually understand the underlying neural processes.

*Introspectability* deficiency is an issue for various use cases. An example is the brain-computer interfaces (BCI) field, where providing feedback to the BCI user is key to either improve the BCI's performance [3–5] or aid in rehabilitation therapies [6]. Another example is the investigation of novel experimental protocols. To evaluate if a new BCI paradigm is suited, it must be determined if the resulting brain signals contain discriminative information related to the task and if this information is sufficient to accomplish control over an application. Both of these examples could benefit from data introspection in an online environment, as providing feedback immediately may enable a BCI user to adapt on the spot, and experimenters

to investigate novel paradigms using a rapid prototyping approach.

*Interpretability* challenges of brain data can be tackled by extracting higher-level features, such as the embeddings of a neural network, from the data. These features are typically less noisy and have a lower dimensionality, even though they are still too high-dimensional for a human to capture. Obtaining such higher-level features, also known as latent features, is at the core of most machine learning methods in a BCI system. For providing feedback to a human and investigating novel paradigms, these features must be visualized. This may be done by transforming a set of latent features into a 2D or 3D representation and visualizing these features in an image. This process is known as *projecting*.

*Projecting* may be achieved by numerous methods, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Uniform Manifold Approximation and Projection (UMAP) [1], t-distributed Stochastic Neighbor Embedding (t-SNE) [7], and Isometric Mapping (ISOMAP).

Unfortunately, not all of these methods are well-suited for online projections. For example, while ISOMAP is known to deal well with noisy data, it has a high computational complexity for larger datasets [8], which increases the model training time and may stagger the projection rate. Alternatively, PCA may be used for rapid projecting, yet it can not account for complex non-linear structure in the data [9]. To determine if a projection method is a good fit for online projecting, we defined the following four criteria: 1) The produced projections should be a sufficiently accurate lower dimensional (2D or 3D) representation of the input data. 2) The time it takes to train/fit a model should be relatively brief, i.e., in the range of minutes for the typical data dimensionalities encountered in BCI. 3) Projecting a novel data point into an existing 2D/3D representation should be fast, i.e., take tens to one hundred milliseconds at most, 4) Optimally, the method should be lightweight to avoid strain on the hardware that may impact projection and/or training time, and to avoid requiring specific hardware or technical knowledge to run.

*Uniform Manifold Approximation and Projection (UMAP)* [1] is a good candidate for an online projection. It comes with benefits such as utilizing a math-

Table 1: Summary description of the datasets used to evaluate the accuracy of approximate UMAP.

Dataset	Classes	Samples total	Dimensionality	Features
Iris plants	3	150	4	Real positive numbers
hand-written digits	10	1797	64	Integers in [0-16]
breast cancer Wisconsin	2	569	30	Real positive numbers

ematical model to solve a clearly defined optimization problem, making it lightweight. Additionally, UMAP is widely adopted in the field, as such there are numerous variations of the method. However, UMAP projection times are slow, possibly conflicting with the third criterion. To overcome this potential drawback, parametric UMAP (pUMAP) is a possible alternative [2]. One of the advantages of pUMAP over standard UMAP is that it generates projections faster, due to utilizing a neural network. However, pUMAP is less lightweight, which may conflict with criterion 4.

*Approximate UMAP* Motivated by the limitations of existing approaches, we introduce approximate UMAP (aUMAP), a novel alternative that drastically reduces the projection time of UMAP. Its training procedure is identical to standard UMAP and the projection speed increase is achieved by approximating the standard UMAP projections using a nearest neighbors approach.

*Experiments.* The accuracy of the aUMAP projections, i.e., criterion 1, is evaluated by comparing them with projections obtained by standard UMAP on three datasets. The training and projection times, i.e., criteria 2 and 3, are evaluated for all three UMAP methods. The models will be trained on data and project data characterized by varying dimensionality and sample counts in order to examine the impact of these variables.

*Structure of the paper* After introducing details of UMAP, pUMAP and the proposed novel aUMAP method we provide results related to our research questions. In the final discussion section, we debate if the presented UMAP methods satisfy the conditions for online projecting we proposed earlier and provide our conclusion as to which method is most suited for online projecting.

## MATERIALS AND METHODS

*Approximate UMAP (aUMAP)* is designed to be an adaptation of UMAP that reduces the time it takes to produce projections on new data points. It does not deviate from standard UMAP when fitting a model, however, it does train an additional  $k$ -nearest neighbors model in tandem. Additionally, aUMAP has a different approach for projecting novel data points. It does so by minimizing the summed Euclidian distance between the new projection and the projections of the points that lie closest to the new data point in the input space. The mathematical equation of this is expressed in Equation 1:

$$u = \sum_{i=1}^k \frac{\frac{1}{d_i}}{\sum_{j=1}^k \frac{1}{d_j}} u_i \quad (1)$$

where  $u$  is the projection of a new data point  $x$ ,  $k$  is the number of neighbors considered,  $u_1 \dots u_k$  are the already existing UMAP projections of the  $k$  nearest neighbors  $x_1 \dots x_k$  of point  $x$  in the input space, and  $d_i = \text{distance}(x, x_i)$ .

*Benchmarking Data:* In order to account for the impact of sample size and data dimensionality on training and projection times, we generated a number of mock datasets containing data of varying dimensionality. Each dataset contained an equal number of samples, allowing for the selection of subsets for testing multiple sample counts. The datasets were generated from a multiclass Poisson distribution. These datasets were used for measuring the training and projection times for the varying models.

To measure the accuracy of aUMAP, we selected the sklearn datasets Iris plants, hand-written digits, and breast cancer Wisconsin. These datasets were chosen for two reasons. First, they cover a variety of data parameters which includes the number of classes, dimensionality, and sample count, see Table 1. Secondly, standard UMAP is able to learn an unsupervised solution for each of these datasets that separates all classes of the data.

*Model parameters:* The models used during the experiments were initiated using the default parameters provided by their base implementation. While this choice clearly leaves room for domain-specific optimizations, we opted to make use of the default parameters to maximize generalizability. The most notable parameters are the distance metric, number of neighbors, minimum distance, and number of components used by UMAP and nearest neighbors (knn) models. These parameters default to 'euclidean', 15, 0.1, and 2 respectively. For the default parameters of pUMAP see the official documentation<sup>1</sup>. We diverged from the default parameters on two occasions only. First, we increased the number of nearest neighbors used by the knn model from its default 5 to 15 in order to be consistent with the default of the UMAP model. Secondly, we adapted the parameters of the UMAP and knn models to produce better UMAP projections. For the breast cancer dataset, we set the number of neighbors to 200 and the minimum distance to 1. The minimum distances for the Iris plants and hand-written digits datasets were also increased to 5 and 1, respectively. The remaining parameters were kept the same.

*Approximate UMAP accuracy:* aUMAP only differs from standard UMAP by approximating novel projections instead of calculating them. Both methods seek the

<sup>1</sup>Parametric UMAP documentation: [https://umap-learn.readthedocs.io/en/latest/parametric\\_umap.html](https://umap-learn.readthedocs.io/en/latest/parametric_umap.html)

same solution. As such, a suitable way to benchmark aUMAP is to investigate how closely its projections fit those of standard UMAP using the Euclidean distance. The closer the aUMAP projections are in the latent space to their associated standard UMAP projections, the better aUMAP achieves its goal. Here, we refer to associated projections as *latent points* that have been produced using the same input data. We measure the distance as the mean Euclidean distance in the projection space. As the UMAP projections are arbitrarily scaled, the Euclidean distance by itself may not be informative without a normalization by the standard deviation of the projected test points produced by standard UMAP, which we have included for this reason.

**Runtime Measurements:** Training times were measured for each method across a predefined range of dimensionalities and sample counts. A varying dimensionality was paired with a static sample count, i.e., we selected multiple of the mock datasets, which differed in their data dimensionality, and selected an equal number of training samples across each set. Similarly when varying the sample count, the dimensionality was kept consistent by drawing subsets from a single generated dataset. When measuring the projection times, the models of the previous step were repurposed to produce the projections, maintaining either a varying dimensionality or number of training samples. A static number of test samples was passed to each model for projecting, matching the data dimensionality of the model's training data. To account for data being presented only a few samples at a time in an online environment, we recorded the training time using two approaches. In the first approach, referred to as *one-go*, all data was given to the model at once, requiring only a singular call to the projection method. The second approach, referred to as *batching* and designed to better match an online setting, fed data points to each model in small batches of five points at a time.

**Hardware:** All experiments were run using an AMD

Ryzen 7 5800x 8-core processor and a NVIDIA GeForce RTX 3060 Ti. Windows Subsystem for Linux (WSL) v.2.0.9.0 was used to enable Tensorflow GPU support. All models, apart from GPU-run pUMAP, were run on CPU.

## RESULTS

**Performance of aUMAP:** Table 2 shows the performance of aUMAP given as the average Euclidean distance over all test samples in the dataset in standard deviations between aUMAP projections and their corresponding standard UMAP projection. Figure 1 allows visualising these projections in addition to the projections of the points used to train the UMAP model used by both methods. Overall, aUMAP delivers a set of projections that closely match the standard UMAP projections. There is only a small number of aUMAP projections that do not match the classification of the matching standard UMAP projection. Across all datasets, the mean distance between the projections lies around 0.1 to 0.25 standard deviations. There is a large variance for each mean distance, which is most extreme for the breast cancer dataset. The variance is reflected by various outliers as shown in Figure 1. While the majority of aUMAP projections match the projection space clusters, the occasional projection deviates sharply, sometimes appearing projected closer to a different cluster than its standard UMAP counterpart. Standard UMAP also produces outliers, yet these are fewer and less extreme.

Table 2: Average Euclidean distance between novel standard UMAP and aUMAP projections. Distances are normalized by the standard deviation of the novel standard UMAP projections.

Dataset	Mean distance	Variance
Iris plants	0.256	0.150
hand-written digits	0.083	0.104
breast cancer Wisconsin	0.126	0.211

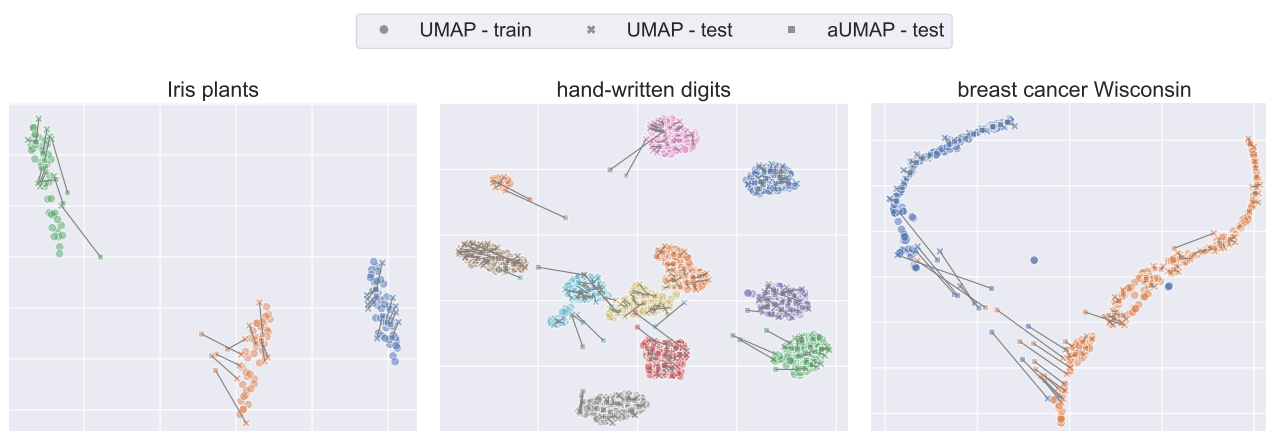


Figure 1: Comparison of UMAP and aUMAP using three datasets. The 2D projections of the training data and test data produced by standard UMAP are displayed in addition to the test set projections produced by aUMAP. The gray lines connect projections of standard UMAP and aUMAP that were obtained from the same test data sample. Colors indicate the classes of the data (not available to the projection methods).

**Training Time:** Figure 2 visualizes the impact of varying data dimensionality and number of training samples, concerning the time required to train a model. When varying the data dimensionality, the number of training examples used for training was fixed to 5000. When varying the number of training examples, the data dimensionality was fixed to 1000.

We observe that the training times of standard UMAP and aUMAP are notably lower than those of pUMAP across any dimensionality or sample count. Training pUMAP on a GPU takes an order of magnitude longer to train than standard UMAP and aUMAP for any dimensionality and sample count. Training pUMAP on a CPU is likewise an order of magnitude slower across all batch sizes and for a dimensionality of 1000, while still being slower than standard UMAP and aUMAP for lower dimensionalities.

The data displays an upward trend across all projection methods as the number of samples increases. This trend is most extreme for CPU-run pUMAP and has a minor effect on GPU-run pUMAP training times until increasing the sample count from 1000 to 5000. Only CPU-run pUMAP is strongly affected by an increase in dimensionality.

**Projection Time:** The projection times are visualized in Figure 3. These results were obtained by projecting 500 test samples either in a single batch (*one-go*) or in multiple sub-batches of 5 samples (*batching*). The experiments were repeated 10 times to obtain an average result. Two outlier training times for specific runs were left out in the final averages due to these measurements deviating more than two standard deviations from the average, whereas the remaining projection times were all within

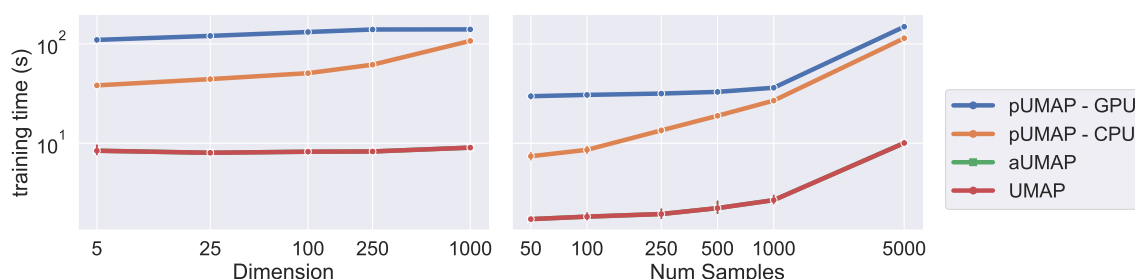


Figure 2: **Training times.** Models were trained on mock datasets generated from a multiclass Poisson distribution. Left: Models were trained on datasets of 5000 samples with varying dimensionalities. Right: Training across varying sample counts was done using subsets of a 1000-dimensional dataset. Standard UMAP and aUMAP models were trained on the CPU. pUMAP models were trained on both, CPU and GPU separately. Note that aUMAP and standard UMAP results are near-identical, causing the line of the latter to be concealed in the graph. All results shown were averaged across 10 repetitions. Error bars indicate the standard deviation across the runs.

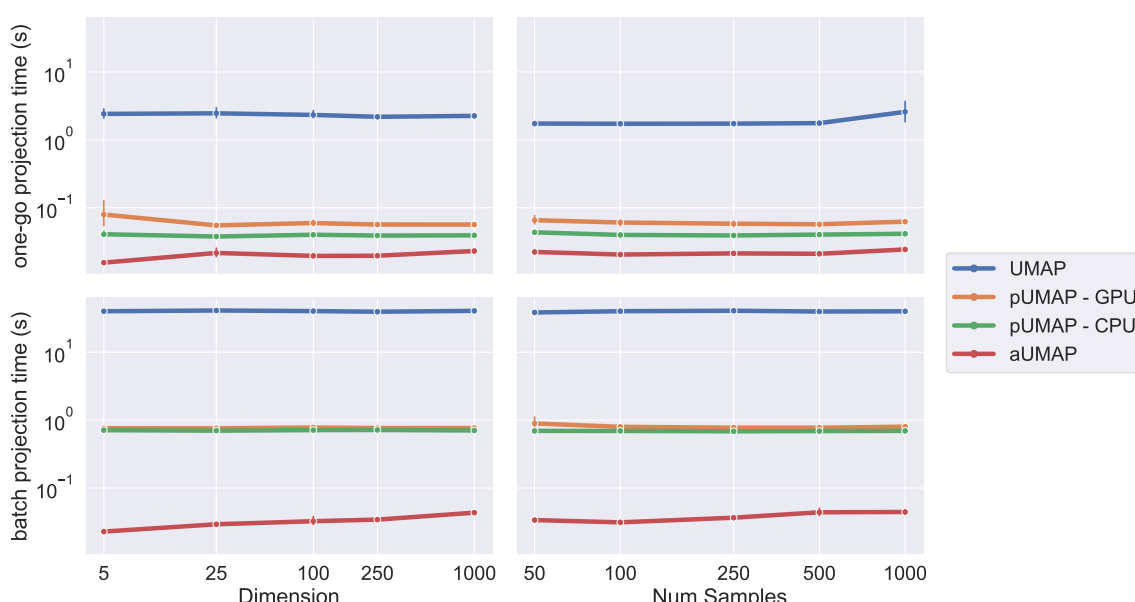


Figure 3: **Projection times.** The models used for projecting were obtained from the training time experiment. For each model and condition, 500 samples from a multiclass Poisson distribution were passed to the models to be projected. Samples were provided either in a singular batch of 500, denoted as *one-go* (upper figures), or in small batches of 5 samples, denoted as *batch* (bottom figures). Standard UMAP and aUMAP models were trained on CPU. pUMAP was trained on both CPU and GPU separately. The results were averaged across 10 repetitions.

one standard deviation. These outliers occurred for CPU-run pUMAP under the *batches* condition, one for dimensionality = 1000 and the other for sample count = 1000. The results show that the projection speed of UMAP is significantly worse compared to the other methods. Under the *one-go* condition, standard UMAP takes over a magnitude longer than aUMAP and pUMAP. This difference remains under the *batch* condition between standard UMAP and pUMAP, while increasing by an order of magnitude against aUMAP.

The effects of an increase in dimensionality or sample count are only notable for aUMAP in both the *one-go* and *batching* conditions. Other methods do not seem affected, except for standard UMAP under the *one-go* condition at the highest sample count tested, however, this increase is paired with a significant increase in variance. Between the *one-go* and *batch* conditions, aUMAP retains similar projecting times. For standard UMAP and pUMAP, projection times are increased by an order of magnitude when comparing *batch* to *one-go*. As a result, aUMAP runs an order of magnitude faster than pUMAP for the *batched* condition, while running at a similar speed for the *one-go* condition.

## DISCUSSION

In the INTRODUCTION, four requirements are presented to determine if a projection method is suitable for an online setting. 1) The produced projections should be an accurate lower dimensional (2D or 3D) representation of the input data, 2) The time it takes to train/fit a model should be relatively brief, i.e. a timescale of minutes, 3) projections should be rapidly producible in an online fashion on a time scale of tens to 100 milliseconds, 4) preferably, the method needs to be lightweight to ensure projection or training times do not suffer due to strain on the hardware and to avoid a need for specific hardware or technical knowledge.

*Approximate UMAP Accuracy:* Our results in Figure 1 show that the proposed aUMAP method upholds the clustering produced by standard UMAP. The average distance between the aUMAP projections and standard UMAP projections remains well below one standard deviation of the standard UMAP projections. Although aUMAP produces quantitatively more and more extreme outliers than standard UMAP, it still reproduces the same clustering in the projection space as standard UMAP. This suggests that the accuracy of aUMAP is close enough to that of standard UMAP to suite online projecting, although more prone to outliers. As such, aUMAP satisfies the Criteria 1.

*Training Time:* Our results show that all tested methods satisfy the training time condition. The maximum of approx. 2 minutes was observed for pUMAP when trained on 5000 samples of 1000 dimensions. This time is in line with the requirement we set out. However, it should be noted that this method is sensitive to an increase in the number of training samples, so it might vio-

late the requirement if more training data were to be used. Unfortunately, we cannot properly assess the runtime of pUMAP when using a GPU. A closer look at the GPU-based pUMAP experiments is presented in a later section. We observe that pUMAP scales with data dimensionality, more so than the other methods, yet to a lesser extent than the input sample size.

Our data shows that GPU-run pUMAP takes more time to train than CPU-run pUMAP. These results go against our prior expectations which expected pUMAP to be trained faster when having access to a GPU. As such, a closer look at GPU-run pUMAP is presented in a later section. This and the following section will focus only on CPU-run pUMAP when discussing pUMAP.

Comparing the training speed of CPU-run pUMAP to the other methods, pUMAP is notably outperformed by both standard UMAP and aUMAP, which have about the same training time due to also fitting a UMAP model which is what dominates the training duration. aUMAP fits a nearest neighbors model in addition, yet this only trivially contributes to the training time complexity. Standard UMAP and aUMAP are trained an order of magnitude quicker than CPU-run pUMAP and are less influenced by larger dimensionality and sample size. As a result, standard and aUMAP would each be the best choices for a projection method according to Criteria 2.

*Projection Time:* The observed projection times display a significant difference between the *one-go* and *batches* condition for all methods except aUMAP, which consistently obtains the lowest projection times. Under the *one-go* condition, the projection time is significantly faster for standard UMAP and CPU-run pUMAP, being an order of magnitude quicker. The projections of aUMAP take longer in the *batch* condition but remains on the same scale.

Given that only a few data points are provided at a time in an online environment, the times given in the *batch* condition are of greater interest. Therefore, the following discussion is based on this condition. Standard UMAP requires a significantly longer time to generate projections compared to the other methods. For every dimensionality and training sample count tested, UMAP requires approximately 40 seconds to project 500 samples, translating to roughly 800 ms per sample, which is significantly longer than the acceptable projection duration we proposed in our third criterion. As such, standard UMAP cannot be regarded as a good fit for online projecting. CPU-run pUMAP is a better fit, as it projects the 500 points in just above one second under its slowest conditions, or 2 ms per data point. This satisfies Criteria 3. aUMAP performs an additional order of magnitude faster, where all 500 projections are generated in less than 100 milliseconds. Both aUMAP and CPU-run pUMAP scale with the number of training samples and dimensionality, however, not at a significant degree. As such, the impact of the training sample count or data dimensionality on projection time may be disregarded unless dealing with values that are of multiple magnitudes larger than the highest

values we tested.

This leads to the conclusion that, according to Criteria 3, aUMAP is most suited for online projecting, however, CPU-run pUMAP is also a feasible option.

*GPU-run parametric UMAP:* Finally, we will discuss GPU-run pUMAP. The results we obtained show that pUMAP runs faster on CPU than on GPU, both when fitting a model and projecting. This is contradictory to our expectations and the results of the paper that introduced pUMAP [1]. Although the pUMAP study does not comment on the effect of CPU or GPU usage on the model's training speed, the authors do compare the projection speed of a CPU-run model to that of a GPU-run model, showing that the projection speed is faster for GPU-run pUMAP. According to that paper, GPU-run pUMAP achieves a projection time that is approximately one magnitude lower than CPU-run pUMAP for three of the six datasets used. Additionally, the study shows the effect of data dimensionality on the projection time, comparing 2 to 64 dimensions, which is in line with our results.

Based on the results of the original pUMAP paper, we may make an inference on how GPU-run pUMAP would compare to the projection times found for CPU-run pUMAP. At worst, there would be an insignificant difference, at best, GPU-run pUMAP would run an order of magnitude faster. Given the best-case scenario, pUMAP would still project slower than aUMAP. This implies that, given Criteria 3, aUMAP would still be a better projection method for online projecting.

Additionally, the pUMAP paper states that the training times of pUMAP are within the same order of magnitude as those of standard UMAP. The study highlights this by showing the cross-entropy loss convergence of pUMAP and standard UMAP, which occurs at around 1 second for two of the three shown datasets and at  $10^2$  seconds for the other dataset. Based on these results, we could speculate that training pUMAP, when having access to a GPU, is as fast as both standard UMAP and aUMAP, making pUMAP an equally valid choice as an online projection method in accordance with Criteria 3.

*Implementation:* The implementation of the aUMAP algorithm can be found online<sup>2</sup>. Additionally, a graphical application for visualizing incoming data stream in real-time and integrating this algorithm was created in the <sup>3</sup> platform.

## CONCLUSION

To conclude, our results suggest that aUMAP can approximate the projection space of standard UMAP sufficiently well for the targeted application. aUMAP may generate projections that lay closer to different clusterings than their standard UMAP counterparts and produces more and more extreme outliers than standard UMAP.

Additionally, our results suggest that pUMAP and aUMAP are good fits for real-time projecting. Standard UMAP, on the other hand, does not meet the criteria to be regarded as a good fit due to its projection times being longer than our proposed acceptable maximum. Overall, aUMAP seems the best option for an online projection tool, having the lowest training and projection times while being more accessible than pUMAP. However, aUMAP is more prone to producing outliers in projection space than standard UMAP. As such, if high accuracy is desired pUMAP may be a better choice.

## ACKNOWLEDGEMENTS

We thank Matthias Dold for his insights into the Dareplane platform which was used to implement the graphical application. Additionally, this work is in part supported by the Donders Center for Cognition (DCC) and is part of the project Dutch Brain Interface Initiative (DBI2) with project number 024.005.022 of the research programme Gravitation which is (partly) financed by the Dutch Research Council (NWO).

## REFERENCES

- [1] McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction. 2018.
- [2] Sainburg T, McInnes L, Gentner TQ. Parametric UMAP embeddings for representation and semisupervised learning. *Neural Computation*. 2021;1–27.
- [3] Tidoni E, Gergondet P, Kheddar A, Aglioti SM. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in Neurobotics*. 2014;8.
- [4] Sollfrank T *et al.* The effect of multimodal and enriched feedback on SMR-BCI performance. *Clinical Neurophysiology*. 2016;127(1):490–498.
- [5] Luu TP, He Y, Brown S, Nakagome S, Contreras-Vidal JL. Gait adaptation to visual kinematic perturbations using a real-time closed-loop brain–computer interface to a virtual reality avatar. *Journal of Neural Engineering*. 2016;13(3):036006.
- [6] Pillette L, Lotte F, N’Kaoua B, Joseph PA, Jeunet C, Glize B. Why we should systematically assess, control and report somatosensory impairments in BCI-based motor rehabilitation after stroke studies. *NeuroImage: Clinical*. 2020;28:102417.
- [7] Hinton G, Maaten L van der. Visualizing data using t-sne. *journal of machine learning research*. 2008;9(11):2579–2605.
- [8] Liang D, Qiao C, Xu Z. Enhancing both efficiency and representational capability of isomap by extensive landmark selection. *Mathematical Problems in Engineering*. 2015;2015:241436.
- [9] Nyamundanda G, Brennan L, Gormley IC. Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics*. 2010;11(1):571.

<sup>2</sup>aUMAP implementation: [https://neurotechlab.socsci.ru.nl/resources/approx\\_umap/](https://neurotechlab.socsci.ru.nl/resources/approx_umap/)

<sup>3</sup>Dareplane: <https://github.com/bsdlab/Dareplane>

## ANALYSIS OF THE EEG RESTING-STATE SIGNALS FOR BCI.

Enrico Mattei<sup>2†</sup> and Daniele Lozzi<sup>1,2†</sup>, Alessandro Di Matteo<sup>1,2</sup>, Costanzo Manes<sup>2</sup>, Filippo Mignosi<sup>2</sup>, Matteo Polsinelli<sup>4</sup> and Giuseppe Placidi<sup>1,3</sup>.

<sup>1</sup>A2VI-Lab, Univ. L'Aquila, L'Aquila, Italy

<sup>2</sup>Dept. of DISIM, Univ. L'Aquila, L'Aquila, Italy

<sup>3</sup>Dept. of MESVA, Univ. L'Aquila, L'Aquila, Italy

<sup>4</sup>Dept. of DISA-MIS, Univ. Salerno, Fisciano, Italy

E-mail: enrico.mattei@graduate.univaq.it

† The first two authors contributed mainly as primary co-authors.

**ABSTRACT:** In the Brain-computer interface (BCI), the recognition of movements is useful for controlling external devices, such as robotic arms, helping people with disabilities or performing remote operations in unsafe places. In this work, we present a new method to build an online BCI for motor execution classification that takes into account not only the movements but also the resting period being essential to recognize when an individual is not engaged in any activity. An artificial intelligence model, EEGnet, was first trained on three classes of left- and right-hand movements, and resting with 0.43 of accuracy. The same type of network was trained on two classes by combining the three classes above, thus having left-right, rest-left, and rest-right, with 0.73, 0.67, 0.63 of accuracy, respectively. Therefore, the 2-classes EEGnet were combined in a network tree that is able to correctly classify not only left- and right-hand movements but also resting signals to improve the accuracy to 0.55 of these three classes.

### INTRODUCTION

The study of brain-computer interface (BCI) focuses also on recognizing movements of the hand, upper limb, wrist, and fingers from human neurophysiological signals to control external devices such as robotic arms. Brain signals are obtained through acquisition procedures that can be characterized as surgical, invasive, or non-invasive [1]. The noninvasive class includes electroencephalography (EEG), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and near-infrared spectroscopy (NIRS). Among these, EEG strikes a good compromise between sensitivity, spatial-temporal resolution, and costs. In the literature, EEG-based BCI is used to decode a user's movement intention known as motor imagery (MI), and as well as to decode real human movements, motor execution (ME), using Artificial Intelligence (AI)[2]. The ME deep learning (DL) algorithms must be trained using EEG datasets. The ME datasets, available in the literature, collect many types of movements, ranging from the most complex to the most intuitive, and DL techniques are applied to classify these

data. The EEG Motor Movement/Imagery Dataset [3] was developed by obtaining signals from the BCI2000 64-channel system by choosing 109 volunteer individuals who imagined opening and closing left or right hand, opening and closing both hands or both feet according to video stimuli, and then they are replicated with the real movement of the subjects. Another dataset is EEG Data for Voluntary Finger Tapping Movement which is a collection of EEG data acquired during voluntary asynchronous index finger tapping by 14 healthy adults. EEG was recorded using a TruScan Deymed amplifier with 19 channels for three conditions: tap of the right finger, tap of the left finger, and resting state, with a sampling rate of 1024 Hz. Each participant performed 120 tests, 40 for each of the three conditions [4]. In the case of on-line BCI, besides classifying movement, it must also be identified the no motion phase, also called resting state. Some work has been published that refers to classifying rest for MI. In [5] a hybrid model that combines convolutional neural network (CNN) and transformer architectures, ConTraNet, was implemented using their strengths to improve classification performance in various EEG applications, including categorization of rest. In [6] the VS-LSTM model was introduced to classify limb MI using EEG signals, with a particular focus on distinguishing between motor and resting states. In the study by [7], novel approaches for ME/MI classifications were introduced, achieving a high level of accuracy. However, the authors did not take into account the rest period for classification. In [8], the authors performed the classification on MI signals, open and close eyes, using Linear Discriminant analysis (LD), Naive Bayes (NB), and Support Vector Machine (SVM) classification algorithms achieving an accuracy of 91.18%, 95.41%, and 99.51% respectively. Moreover, in [9] the authors classified 2, 3 and 4-classes using EEGnet on MI and rest task, reaching and overall accuracy of 82.43%, 75.07%, and 65.07% respectively. Also in [10], the authors classified the MI signals and Rest, reaching an accuracy of 70.64% on 5 classes. This work shows that most of the studies conducted on resting state are carried out for MI, leaving out

ME. In [11], it can be observed that the two cognitive processes MI and ME are different. For this reason, the present work describes the activity of the first stage that focuses on analyzing the response of the rest signal to the real movement of the hand (left and right) for the robotic control arm in rehabilitation [12]. The objective of this research is to determine the optimal method for distinguishing EEG data of REST phases from EEG data of ME by employing a deep learning architecture, EEGnet[13]. This involves decomposing the three-class problem (REST, LEFT, RIGHT) into several binary classifications (REST vs LEFT, REST vs RIGHT, LEFT vs RIGHT), and addressing this through the construction of a *Network Tree* that integrates several EEGnet trained for identify the REST and the two movement-related classes, for enhancing classification performance compared to the EEGnet trained on three classes that shown low accuracy during our previous analysis.

## MATERIALS AND METHODS

Data from 105 people from the Physionet Dataset [3] were used. The original data set consisted of 109 persons, but four subjects were excluded because they performed different numbers of trials with a different sampling rate. Each subject was recorded while performing an execution or the imagined execution of opening and closing the right or left hand for 4 seconds. Before each exercise performed or imagined, there was a 4 seconds of rest. Each person was recorded three times and each recording contained approximately seven movements in each category, plus one rest period for each movement. Of 64 electrodes, we used only the following 15: Fp1, Fp2, F7, F8, Fz, P3, Pz, P4, O1, O2, F3, F4, C3, C4, Cz. This was done to reduce the number of channels and make the system lighter and more portable, given the possibility of using a cheaper EEG device (such as Enobio 20<sup>1</sup> or other light mobile EEG device). After exclusion, the epochs were extracted using one second of signal, starting from the marker, without any baseline reduction. Signals were downsampled the data from 160 to 80 Hz and a high-pass filter to 1 Hz was used.

In Fig 1, the pipeline for preprocessing applied to the dataset and the classification stage are shown.

Our work aims to create a deep learning model capable of recognizing the movement execution of the right hand, the left hand, and the rest period. For this reason, only the EEG data acquired during ME conditions were used. Finally, our dataset was composed of 4725 signals under “ME rest condition”, 2369 signals under “ME left-hand movement” and 2356 signals under “ME right-hand movement”. In each class, the corresponding trials of all selected subjects are grouped. In Fig. 2 the train-validation-test sets splitting is shown. The movement decoding process was carried out using the DL architecture EEGnetV4, from now on EEGnet [13], trained on the EEG signal from the EEG motor move-

ment/image dataset [3]. Preprocessing and deep learning models were performed using Python with MNE library [14] for EEG data analysis, PyTorch [15] for the deep learning and [16] for the EEGnet architecture. Before training, starting from the original data, we performed the transformation using Common Spatial Pattern (CSP) [17, 18] and used the transformed signal as the input for the network. This operation reduced the number of channels from 15 to 2, maximizing the distinguishability between classes. This process was repeated for all EEGnet used. All networks were trained using a balanced training set, so all classes had the same number of samples in both for training and for test. The training hyperparameters, displayed in the Tab. 1, were selected following several experimental trials. We noticed that signal downsampling did not affect performance, so we downsampled the signal to improve training speed. We used the Stochastic Gradient Descending (SGD) to improve the convergence of the training, applying also a dynamic Learning Rate, starting from a value of 0.004, that reduced its value automatically when, after 25 epochs of inactivity, it did not improve. The factor of reduction of Learning Rate was set at 0.2. Moreover, we used a double batch size [19] that changes when there was no more improvement of the network also after reducing the Learning Rate. Finally, we used a decay rate of 0.001 to improve the learning rate of the network. The signals of REST, execution of open close right-hand (RIGHT) and execution of open and close left-hand (LEFT) were considered. As a first step, the EEGnet network was trained on the three classes REST (0), LEFT (L), and RIGHT (R), called 0LR, with unsatisfactory results, as shown in Tab. 2 in the last row. Therefore, comparisons were made between the same EEGnet trained on two classes left-right, rest-left and rest-right classes defined as LR, 0L, and 0R, respectively. After this analysis, we observed that the 0LR network did not correctly distinguish the three classes to the LR, 0L, and 0R networks. For this reason, a *Network Tree* based on previously tested EEGnets was created to improve the classification accuracy and analyze the REST signal. Our “*Network Tree*”, was composed of three different EEGnetV4 [13] individually trained on different subsets of our dataset, to classify different types of signal:

- *EEGnet<sub>LR</sub>*: the network was able to classify correctly the signals labeled as Left or Right;
- *EEGnet<sub>0L</sub>*: the network was able to classify correctly the signals labeled as Rest or Left;
- *EEGnet<sub>0R</sub>*: the network was able to classify correctly the signals labeled as Rest or Right.

To build the *Network Tree*, the classification of the class unknown to each network was tested with the following results:

- a Rest epoch, if given as an input to *EEGnet<sub>LR</sub>*, was classified as Right;

<sup>1</sup><https://www.neuroelectrics.com/solutions/enobio/20>

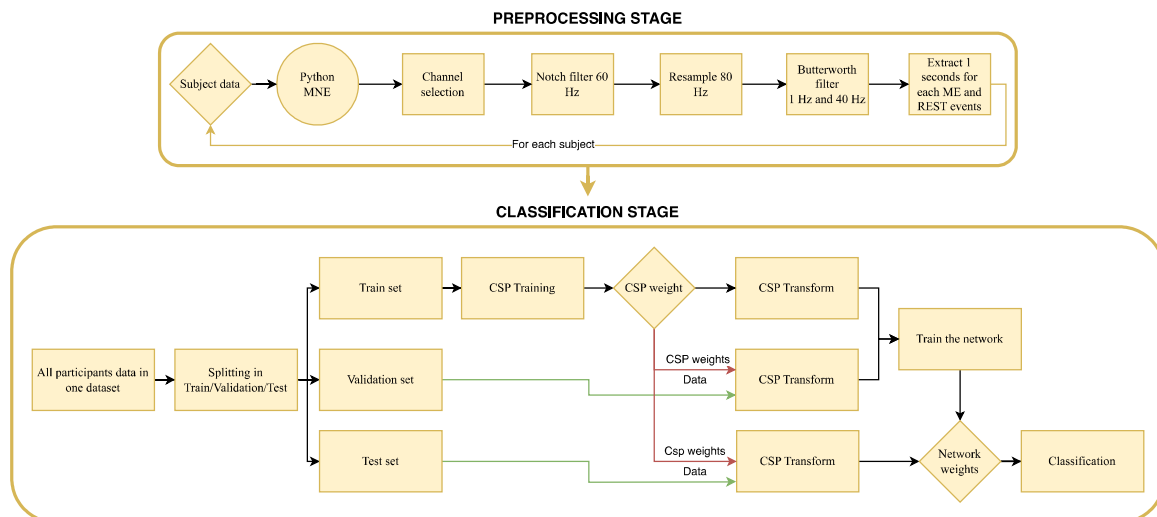


Figure 1: The pipeline used for data preparation and classification stages.

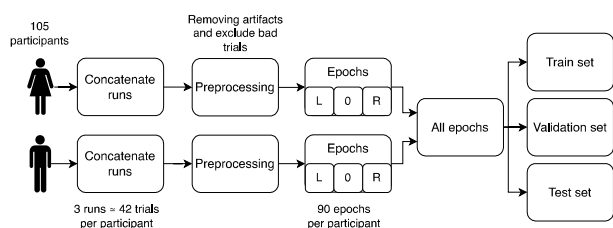


Figure 2: The split of the original set in train, validation and test set after preprocessing and epochs extraction.

Hyperparameter	Value
Sample shape	1, 15, 79
Sampling rate	80 Hz
N. channels	2
N. epochs	2000
Optimizer	Stochastic Gradient Descending (SGD)
Scheduler type	Reduce learning rate when a metric has stopped improving
Learning rate	0.004
Patience	25 epochs
Factor scale	0.2
Batch size	16, 32
Decay	0.001

Table 1: The hyperparameters used to train all EEGnet

- a Right epoch, if given as an input to  $EEGnet_{OL}$ , was classified as Rest;
- a Left epoch, if given as an input to  $EEGnet_{OR}$ , was classified as Right.

Each network started with a CSP transformation, and each CSP transformation was different for each network. Based on the results shown in the Tab. 2, our system behaves as follows: a sample was given as input in  $EEGnet_{LR}$ : if it was classified as Left, the original epoch was given to  $EEGnet_{OL}$  for final classification in Rest or Left movement; If it was classified as Right, then

the original epoch was given as input in parallel to two networks:  $EEGnet_{OL}$  and  $EEGnet_{OR}$ . Based on the results, the truth table in Fig. 3 was applied to classify the epoch.

The network tree architecture is shown in Fig. 3. Tab. 3 shows the results of the classification using the *Network Tree*.

## RESULTS

For evaluating the model, Accuracy, Precision, Recall, and F1 metrics were used. In Tab. 2 the results are shown for all networks trained individually. In the last three rows of Tab. 2, the results for all the 2-class networks shown that the binary classification was more accurate related to the 3-class network (last row). The  $EEGnet_{RL}$ , shown a good discriminability power between Left and Right movements, while when a Rest signal was given as an input to this net, it created a light unbalancing classification result towards the Right class. It was interesting to note that a Right sample was classified as Rest if it was given as input to  $EEGnet_{OL}$ , while a Left sample was classified as Right if it was given as input to  $EEGnet_{OR}$ . Starting from these results, we built the *Network Tree* combining all 2-classes network in a classification cascade. The proposed network achieved an overall accuracy of 0.55, greater than the chance [20] and improving the performance respect to  $EEGnet_{OLR}$ . Moreover, in the  $EEGnet_{OLR}$  the single classes were misclassified frequently, especially Rest and Right, and it was shown looking the Recall metric. It must be emphasised that in the training of  $EEGnet_{OLR}$ , three classes were extracted with the CSP instead of two, so the resulting multivariate signal was composed by 3 channels. For comparison the results for the *Network Tree* are reported, in Tab. 3, showing a more stable behavior of the network in terms of classification. This aspect is well represented by the F1 score, which takes into account both precision and recall.

Network	Class	Support	Precision	Recall	F1	Accuracy	Unknown Class	Prediction
<i>EEGnet<sub>0LR</sub></i>	Rest	574	0.62	0.15	0.24	0.43	-	-
	Left	594	0.39	0.96	0.55		-	-
	Right	606	0.72	0.18	0.29		-	-
<i>EEGnet<sub>LR</sub></i>	Left	600	0.74	0.73	0.74	0.73	Rest	0.46
	Right	582	0.73	0.73	0.73			0.54
<i>EEGnet<sub>0L</sub></i>	Rest	576	0.66	0.64	0.65	0.67	Right	0.73
	Left	609	0.67	0.69	0.68			0.27
<i>EEGnet<sub>0R</sub></i>	Rest	568	0.67	0.46	0.55	0.63	Left	0.42
	Right	614	0.61	0.79	0.69			0.58

Table 2: Prediction metrics table for all networks used in *Network Tree*

Network	Class	Support	Precision	Recall	F1	Accuracy
<i>Network Tree</i>	Rest	591	0.48	0.45	0.46	0.55
	Left	586	0.57	0.61	0.59	
	Right	596	0.58	0.59	0.58	

Table 3: Results of *Network Tree*

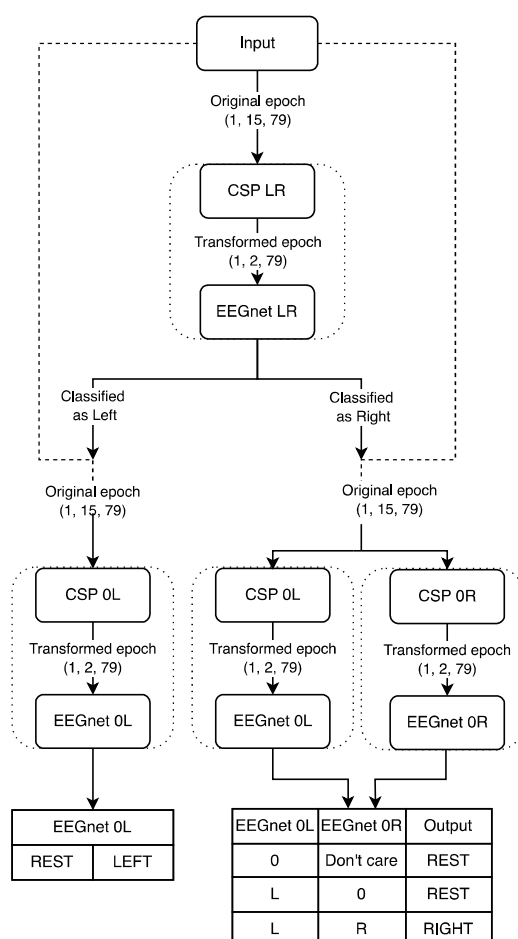


Figure 3: *Network Tree* - the first classification of the signal is performed by the *EEGnet<sub>LR</sub>* and then a classification check is carried out. If L is identified, the input signal is sent to the *EEGnet<sub>0L</sub>* for final classification between 0 or L. In case R is classified, the input signal is sent both to the *EEGnet<sub>0L</sub>* and *EEGnet<sub>0R</sub>*. The result of their combined classification returns the result of the final classification between 0 or R. The input undergoes, before being sent to a network, a transformation by the associated CSP.

## CONCLUSION

This work aimed to present a new method for a motor-based BCI. In an online classification, the EEG signal is recorded while performing L or R hand movement or while having a rest. In this field, there is a literature gap on the non-movement classification, also known as Rest, but this is fundamental in online application since there are many non-movement periods while using a BCI [21]. The novelty lies in the concatenation of a network tree of able of correctly classifying not only left and right hand movements, but also the rest signals. Future work should test the system in real-time by including a pre-processing pipeline capable of quickly cleaning the signals [22, 23]. Furthermore, other tasks, such as motor imagery on emotions [24], needs to be tested, and using different types of features extracted by the EEG data such as connectivity [25], independent components [26], or other features [27]. Moreover, future improvement of the *Network Tree* will include both subject-dependent and subject-independent cross-validation analysis, splitting the subjects into training and test groups before the training.

## REFERENCES

- [1] Saibene A, Caglioni M, Corchs S, Gasparini F. Eeg-based bcis on motor imagery paradigm using wearable technologies: A systematic review. *Sensors*. 2023;23(2798).
- [2] Gwon D, Won K, Song M, Nam CS, Jun SC, Ahn M. Review of public motor imagery and execution datasets in brain-computer interfaces. *Frontiers in Human Neuroscience*. 2023;17:1134869.
- [3] Schalk Gerwin McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. Bci2000: A general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*. 2004;51(51).

- [4] Wairagkar M. Eeg data for voluntary finger tapping movement. University of Reading. 2017.
- [5] Ali O, Rehman M Saif-ur, Glasmachers T, Iossifidis I, Klaes C. Contranet: A hybrid network for improving the classification of eeg and emg signals with limited training data. *Computers in Biology and Medicine*. 2024;168:107649.
- [6] Wang P, Wang M, Zhou Y, Xu Z, Zhang D. Multi-band decomposition and spectral discriminative analysis for motor imagery bci via deep neural network. *Frontiers of Computer Science*. 2022;16:1–13.
- [7] Fan C, Yang B, Li X, Zan P. Temporal-frequency-phase feature classification using 3d-convolutional neural networks for motor imagery and movement. *Frontiers in Neuroscience*. 2023;17:1250991.
- [8] Taşar B, Yaman O. Eeg signals based motor imagery and movement classification for bci applications. In: *2022 International Conference on Decision Aid Sciences and Applications (DASA)*. 2022, 1425–1429.
- [9] Wang X, Hersche M, Tömekce B, Kaya B, Magno M, Benini L. An accurate eegnet-based motor-imagery brain–computer interface for low-power edge computing. In: *2020 IEEE international symposium on medical measurements and applications (MeMeA)*. 2020, 1–6.
- [10] Fadel W, Kollod C, Wahdow M, Ibrahim Y, Ulbert I. Multi-class classification of motor imagery eeg signals using image-based deep recurrent convolutional neural network. In: *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*. 2020, 1–4.
- [11] Bencivenga F, Sulpizio V, Tullo MG, Galati G. Assessing the effective connectivity of premotor areas during real vs imagined grasping: A dcm-peb approach. *Neuroimage*. 2021;230:117806.
- [12] Placidi G, De Gasperis G, Mignosi F, Polsinelli M, Spezialetti M. Integration of a bci with a hand tracking system and a motorized robotic arm to improve decoding of brain signals related to hand and finger movements. In: *International Symposium on Visual Computing*. 2021, 305–315.
- [13] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*. 2018;15.
- [14] Gramfort A *et al*. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*. 2013;7(267):1–13.
- [15] Paszke A *et al*. Automatic differentiation in pytorch. 2017.
- [16] Schirrneister RT *et al*. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*. 2017.
- [17] Koles ZJ, Lazar MS, Zhou SZ. Spatial patterns underlying population differences in the background eeg. *Brain topography*. 1990;2:275–284.
- [18] Blankertz B, Tomioka R, Lemm S, Kawanabe M, Muller Kr. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Processing Magazine*. 2008;25(1):41–56.
- [19] Smith SL, Kindermans PJ, Ying C, Le QV. Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*. 2017.
- [20] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: A closer look on bci results. *International journal of bioelectromagnetism*. 2008;10(1):52–55.
- [21] Di Giamberardino P, Iacoviello D, Placidi G, Polsinelli M, Spezialetti M. A brain computer interface by eeg signals from self-induced emotions. In: *VipIM-AGE 2017: Proceedings of the VI ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing Porto, Portugal, October 18-20, 2017*. 2018, 713–721.
- [22] Placidi G, Cinque L, Polsinelli M. A fast and scalable framework for automated artifact recognition from eeg signals represented in scalp topographies of independent components. *Computers in Biology and Medicine*. 2021;132:104347.
- [23] Pion-Tonachini L, Kreutz-Delgado K, Makeig S. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*. 2019;198:181–197.
- [24] Lozzi D, Mignosi F, Spezialetti M, Placidi G, Polsinelli M. A 4d lstm network for emotion recognition from the cross-correlation of the power spectral density of eeg signals. In: *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. 2022, 652–657.
- [25] Lozzi D, Mignosi F, Placidi G, Polsinelli M. Graph model of phase lag index for connectivity analysis in eeg of emotions. In: *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*. 2023, 348–353.
- [26] Erfanian A, Erfani A. Ica-based classification scheme for eeg-based brain-computer interface: The role of mental practice and concentration skills. In: *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2004, 235–238.
- [27] Jaipriya D, Sriharipriya K. Brain computer interface-based signal processing techniques for feature extraction and classification of motor imagery using eeg: A literature review. *Biomedical Materials & Devices*. 2023:1–13.

# AN ALTERNATIVE TRAINING PROTOCOL FOR A MOTOR IMAGERY BMI BASED ON A COLLABORATIVE APPROACH

Alessio Palatella<sup>1</sup>, Paolo Forin<sup>1</sup>, Stefano Tortora<sup>1,2</sup>, Emanuele Menegatti<sup>1,2</sup>, Luca Tonin<sup>1,2</sup>

<sup>1</sup>Department of Information Engineering, University of Padova, Padova, Italy

<sup>2</sup>Padova Neuroscience Center, University of Padova, Padova, Italy

E-mail: stefano.tortora@unipd.it

**ABSTRACT:** Becoming proficient in the use of brain-machine interfaces (BMIs) represents a challenging task for the subjects, requiring long and intensive training. In this paper, we propose and explore the use of a collaborative BMI (cBMI) as an innovative training protocol that allows two subjects to learn together by cooperating in the control of a real robotic arm. Preliminary results on three pairs of subjects spanning five days of training highlight the promises of the proposed approach in reducing the training time and possibly mitigating the frustration in naive users.

## INTRODUCTION

Motor imagery (MI) brain-machine interfaces (BMIs) represent the most natural approach to control brain-driven robotic devices [1]. Indeed, the endogenous paradigm used in BMIs provides the advantage that subjects can autonomously initiate mental tasks without requiring external stimuli like visual, auditory, or tactile cues, making the interaction with the robotic system more natural [2].

However, becoming proficient using MI BMIs is a challenging scenario that requires a substantial investment of time and effort for the subject [3]. In the last years, researchers have shown that training directly on the final application might be more effective than having subjects engaged in repetitive and artificial mental tasks [4, 5]. However, this approach may not always be practical for naive subjects, given the inherent limitations in the accuracy of their BMIs and the consequent frustration caused by their low performances to perform the given task. This is often attributed to their limited experience, causing them to struggle in effectively generating effective control commands [6].

In this scenario, researchers have recently proposed collaborative brain-machine interfaces (cBMIs) to deal with such limited performance [7, 8]. In this approach, multiple subjects are simultaneously engaged in the same BMI task and the output of their BMIs is combined to enhance the decoding accuracy or increase the number of commands.

One crucial aspect of cBMIs is how the data originating from multiple sources is integrated to generate a single control signal. In the literature, two main approaches

are reported: on the one hand, researchers investigated the possibility to merge the sources at the feature level and then to exploit this unified signal to train a single decoder [9, 10]. On the other hand, it has been explored the feasibility to train a decoder for each subject, and then to combine the output at the decision level to obtain the final command for the external device [7, 8, 11].

In literature, cBMI are typically employed in experiments with BMI based on exogenous visual stimulation or pattern recognition. In a study by Wang et al. [11], 15 subjects were divided into subgroups and they performed animal categorization and single-photograph recognition with a Go/noGo paradigm through a series of flashing pictures. Similarly, Poli et al. [7] instructed 10 subjects to determine whether two subsequent shapes, with the second one masked, were identical and pressing a button consequently. In another study by Valeriani et al. [9], participants were asked to identify the specific geometrical patterns of two horizontal and vertical bars to use a switch to send the decisions.

To date, research on cBMI with MI BMI is limited, and the studies that apply this concept to control a robot in real-life scenarios are neglected. For instance, Yijie et al. [10] independently trained eight subjects in hands and feet motor imagery for moving a point on a grid. They combined the individual subjects' results offline to simulate a collaborative protocol. In a different study, Bonnet et al. [8] developed a game where subject pairs control the movement of a virtual ball using motor imagery, both collaboratively and competitively, to place it inside a net. The game software uses the two decoders' output to decide in which direction the ball should move and change the feedback accordingly. However, the literature predominantly emphasizes subject performance and methods to enhance classification accuracy, and there are no studies reporting on the subject's learning during cBMI. In this work, we aim to investigate this particular aspect by exploring the cBMI as a novel approach to assist subjects during the training of MI BMI. Our training protocol grounded on the concept that when we learn complex skills in our daily life (e.g., cycling, driving), we are not alone. Based on this, we have designed an alternative training protocol utilizing a cBMI where two subjects work together to accomplish the same task of controlling a robotic arm during reaching operations. The rationale

is that the presence of a second subject can help the less proficient one to still conclude the task, decreasing negative mental states due to frustration or inefficiency present in other BMI [12]. We hypothesize that this approach can increase overall performance and promote learning in both subjects [13].

The objective of this work is twofold: firstly, to showcase the viability of MI cBMI for controlling a real robotic device even in the early stage of the training; secondly, to investigate this collaborative approach as a promoter of individual learning of BMI skills.

## METHODS

**Participants:** Six healthy subjects, with an average age of  $25 \pm 2$  years, were recruited and divided into three pairs (G1-C7, G2-D7, C9-C8). Moreover, three of them had no experience with BMI system before (G1, G2, C9). The first three sessions took place over two weeks in December 2023, while the remaining three sessions were conducted over three weeks in January 2024. Each participant provided consent by signing a form detailing the use of recorded data and privacy protocols, adhering to the principles outlined in the Declaration of Helsinki.

**Experimental design:** The experiment comprised six sessions, each consisting of three runs, as illustrated in Figure 1a. The initial session (Session - Day 0) involved the calibration and evaluation of the BMI Gaussian decoder of each subject. The following sessions (Session - Day 1-5) focused on online runs, during which subjects tested the trained decoder, re-training it if its accuracy was below 70% [14] with two new calibrations and evaluations, and subsequently controlled the robotic arm in collaboration with another subject. All sessions are subdivided into runs, which can be of three types: calibration runs for recording the data for training the decoder; evaluation runs to test the decoder performance; control runs where subjects controlled the robotic arm using the proposed cBMI.

During the control runs, subjects were asked to move a robotic arm from a home position to one of the five target objects placed on a circle of radius 50 cm in front of the robot. Subjects were allowed to see the robot and the workspace on a monitor, with superimposed visual feedback as depicted in Figure 1b. At the beginning of each trial, subjects were presented with a target object to pick, indicated by a blue dot, for 1 s (cue period). After that, the robot initiated the motion based on the output of the cBMI. When sufficiently close to an object, the robot autonomously performed the pick, returned in the home position, and a new trial started. For each run, subjects were required to pick each of the five objects twice, hence ten trials in total. Figure 1 (c) shows the experimental setup, comprising the subjects, the robotic arm, and monitors for visual feedback. The combination of the brain-machine interfaces (BMIs) output is also provided as visual feedback using a rotating wheel<sup>1</sup>.

<sup>1</sup>Video: <https://cloud.dei.unipd.it/index.php/s/TgYJ475A9M7tz3x>

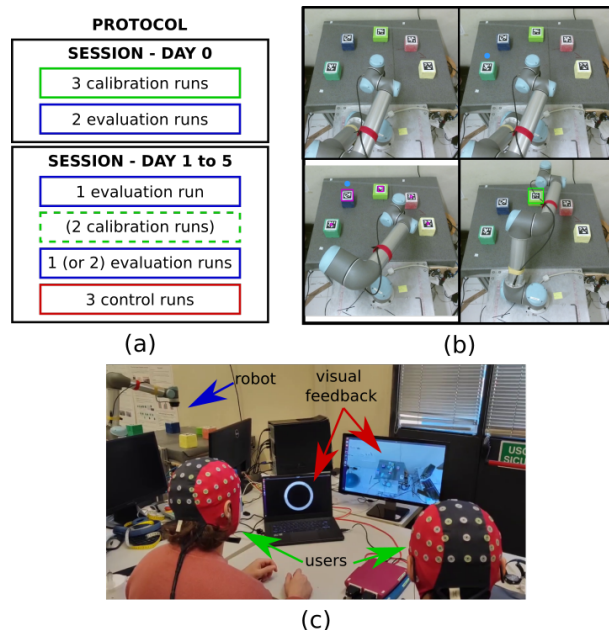


Figure 1: Experimental design. (a) Schematic representation of the experiment division for sessions (days). (b) BMI visual feedback provided to the subject to control the robotic arm. (c) Experimental set-up for the controlling runs in which both subjects controlled the manipulator.

**Collaborative BMI:** We acquired electroencephalography (EEG) data with a 64-channel amplifier (AntNeuro eego sports 64, Netherlands) operating at a sampling rate of 512 Hz. A 2-class MI paradigm was employed where the subjects were required to make the kinesthetic imagination of their feet or their hands. Our method was based on a classical MI BMI already widely evaluated in literature [6, 15]. In particular, we selected a subset of 32 electrodes placed over the sensorimotor cortex (i.e., *FP1*, *FP2*, *FZ*, *FC5*, *FC1*, *FC2*, *FC6*, *C3*, *CZ*, *C4*, *CP5*, *CP1*, *CP2*, *CP6*, *P3*, *Pz*, *P4*, *F1*, *F2*, *FC3*, *FCZ*, *FC4*, *C5*, *C1*, *C2*, *C6*, *CP3*, *CP4*, *P5*, *P1*, *P2*, *P6*) to detect the neural patterns related to MI. We used the power spectral density of the EEG signals with Welch's periodogram (from 4 to 48 Hz every 2 Hz) in 1-second windows sliding every 62.5 ms. The most discriminant features of each subject were identified using Canonical Variate Analysis (CVA) and a Gaussian classifier was trained with these features. The classifiers were trained on EEG data acquired during calibration or evaluation runs. During each calibration run, the subject was asked to perform 10 trials of both hands imagination, 10 trials of both feet, and trials of rest, providing always a positive feedback with the visual interface. During the evaluation, the feedback was controlled by the BMI output and the subject was asked to perform 10 trials of each MI class.

During the robot control, the output of the BMIs decoder of each subject in the pair was fused to allow the collaborative control of the robotic system. To merge the information coming from the two decoders, we applied a weighted mean on the posterior probabilities:

$$pp_{merged} = \frac{W_1 \cdot pp_{s_1} + W_2 \cdot pp_{s_2}}{W_1 + W_2} \quad (1)$$

The weights  $W_1$  and  $W_2$  were used to balance the contribution of each subject to the control according to their performance during the last evaluation run before starting the control. The weights have been computed using the balanced binary focal cross-entropy loss [16]. In particular, the weight for the subject  $s$  is obtained as  $W_s = -1/L_s$  with  $L_s$  calculated as follow

$$L_s = \sum_{Tr=1}^N \sum_{n=1}^{N_{Tr}} \begin{cases} \frac{pp(y_n)^\gamma \cdot \log(1-pp(y_n))}{N_{Tr}} & y_n \in class(0) \\ \frac{(1-pp(y_n))^\gamma \cdot \log(pp(y_n))}{N_{Tr}} & y_n \in class(1) \end{cases} \quad (2)$$

where  $N$  is the number of trials in the evaluation run,  $N_{Tr}$  is the number of samples in each trial and  $pp(y_n)$  is the posterior probability of the decoder given the features  $y_n$ . We fixed  $\gamma = 2$  as focal factor which is used to reduce the impact on the loss of well-classified samples for which  $pp(y \in class(0)) < 0.6$  or  $pp(y \in class(1)) > 0.4$ . Finally, after the fusion process, the merged probability were integrated over time using an exponential smoothing function as  $D_t = \beta \cdot D_{t-1} + (1 - \beta) \cdot pp_{merged}$ , where  $D_t$  and  $D_{t-1}$  are the current and previous cBMI output, respectively, and  $\beta = 0.96$  is the integration coefficient.

**Robot control:** The commands in output from the collaborative BMI were used as input for the robotic manipulator. To this end, the integrated cBMI output  $D_t$  was mapped to an angular direction  $d_R \in \{-90^\circ, 90^\circ\}$ , with  $0^\circ$  corresponding to the forward direction, for the movement of the manipulator on the horizontal plane through the following sigmoid functions [17]:

$$d_R(D_t) = \begin{cases} -90^\circ + \frac{90^\circ}{1 + \exp(-25 \cdot (D_t - \beta_1))} & 0 \leq D_t \leq 0.5 \\ \frac{90^\circ}{1 + \exp(-25 \cdot (D_t - \beta_2))} & 0.5 \leq D_t \leq 1 \end{cases} \quad (3)$$

Parameters  $\beta_1$  and  $\beta_2$  denote the probability values at which the sigmoid intersects  $-45^\circ$  and  $45^\circ$ , respectively. These parameters were initialized respectively to 0.25 and 0.75 in the first session, and then tailored to each pair at the beginning of the control runs based on the performance obtained in the previous session [17]. Given the movement direction, a velocity command was thus delivered towards that direction with constant speed equal to 5 cm/s.

In order to assist subjects in driving the robotic manipulator towards one of the target objects, the robot was controlled using a shared control architecture for teleoperation based on artificial potential fields (APF) presented in [18]. In short, the system firstly computed the probability of each target object to be the goal of the reaching task from the sequence of input velocity commands and robot positions. Then, an attractor point for the APF was generated at each time step as the center of mass of the

objects' position weighted by their probability:

$$[x, y] = \sum_{n=1}^{n_{targets}} P_{(x,y)_n} \cdot p_n \quad (4)$$

where  $[x, y]$  are the coordinates of the attractor point on the table surface,  $P_{(x,y)_n}$  represents the position of the target  $n$ , and  $p_n$  is the probability associated to that target as computed by the shared control system. The probability of each object was also shown in real-time to the subjects through the visual feedback to help them understanding the behavior of the robotic device. For more details on the shared control implementation, please refer to [17, 18].

## RESULTS

**Feature evolution:** Figures 2a-b summarize the evolution of the features, the distribution of the posterior probabilities and the performances of the subjects during the 5-day sessions. In particular, Figure 2b illustrates the distributions of the posterior probabilities for a pair of subjects (G1 on the x-axis and C7 on the y-axis) during the control of the robotic manipulator. It can be seen that, while subject C7 was able to span the whole range of probabilities since the first session, the second subject G1 started from a distribution quite shrank around 0.5, meaning that he was not significantly contributing to turning the robot left-right. Nevertheless, thanks to the training also subject G1 was able to expand the distribution of probabilities covering nearly the entire probability space, particularly around the extreme values during the last session. This improvement is then reflected in the distribution of the fused probabilities (Figure 2b, bottom right), where each curve colour corresponds to the same session in the scatter plots. Initially, on the first day, they were concentrated around 35% and, as sessions progressed, the distribution gradually transformed into a more uniform distribution, encompassing a broader range of possible velocity commands to control the robot. To better understand how this improvement is reflected at the neural level, we analysed the feature maps of each subject during robot control. In order to provide a reliable label (both hands or both feet) to each sample, we calculated the angle  $\varphi_{ee-goal}$  between the current position of the robot's end-effector and the position of the trial's goal with respect to the forward direction. If  $\varphi_{ee-goal} \geq 30^\circ$ , the corresponding EEG sample was labelled as both feet, if  $\varphi_{ee-goal} \leq -30^\circ$  as both hands, otherwise it was not considered for the feature maps analysis. Figure 2a shows for each session the grand-average across all subjects of the feature maps in the  $\alpha$  (8 – 12Hz) and  $\beta$  (18 – 22Hz) bands calculated using the Fisher's score [4]. Overall, it can be noted a clear emergence of discriminant features through the sessions, particularly in lateral channels (C4, C6, FC6, FC4 and C3, C5, FC5, FC3) and in the  $\beta$  band, as expected from our MI paradigm.

**BMI accuracy:** Figure 2c shows the average accuracy and the pick error for each pair over the five days. The pick accuracy is computed by counting the correct trials

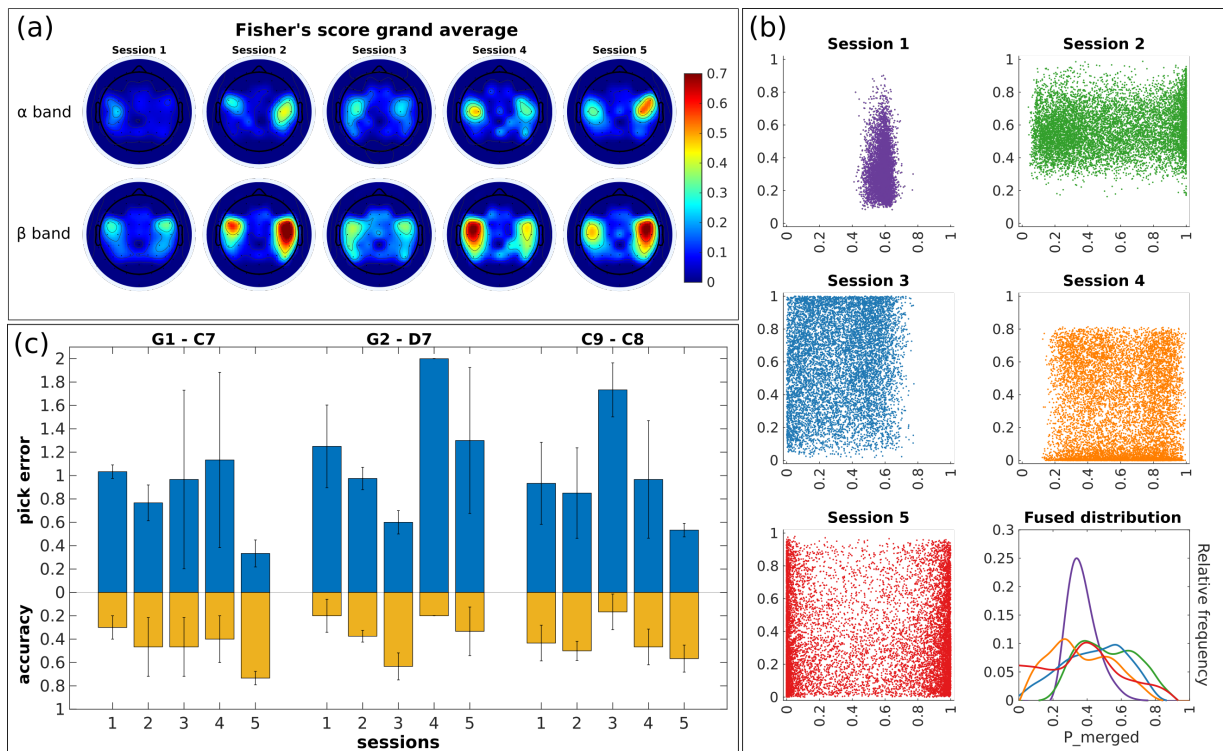


Figure 2: BMI performances. (a) Topographic representations across sessions of the grand-average feature maps of all the subjects computed as the Fischer's score in the  $\alpha$  (8 – 12 Hz) and  $\beta$  (18 – 22 Hz) bands. (b) Scatter plot of the combined posterior probabilities of the pair G1-C7 during the control runs of every session (G1 on the x-axis and C7 on the y-axis). The bottom-right plot represents the distributions of the merged probabilities across the training sessions (same color as the corresponding scatter plot). (c) Pick accuracy and pick error for every pair of subjects across the five sessions.

Session	Robot pick accuracy [%]					Robot pick error [number of objects]				
	1	2	3	4	5	1	2	3	4	5
G1-C7	30 ± 10	47 ± 25	47 ± 25	40 ± 20	73 ± 5	1.0 ± 0.1	0.7 ± 0.1	0.9 ± 0.7	1.1 ± 0.7	0.3 ± 0.1
G2-D7	20 ± 14	37 ± 5	63 ± 11	20 ± 0	33 ± 20	1.2 ± 0.3	0.9 ± 0.1	0.6 ± 0.1	2.0 ± 0.0	1.3 ± 0.6
C9-C8	43 ± 15	50 ± 8	17 ± 15	46 ± 15	57 ± 11	0.9 ± 0.3	0.8 ± 0.3	1.7 ± 0.2	0.9 ± 0.5	0.5 ± 0.1

Table 1: Average pick accuracy (chance level at 20%) and pick error over the sessions for every pair.

(i.e., trials in which the manipulator reaches the correct object) over all trials. Instead, the pick error measures the precision of the control as the number of objects between the picked object and the correct one. For example, if the correct object is the one in the forward-left direction and the users pick the forward object, this error is equal to 1. On the other hand, if they pick the right-most object, the error is equal to 3. If a trial is correct, the error is equal to 0 for this trial. This metric is then normalized by the total number of trials in a single session. The average values of accuracy and pick error for each pair and each session are reported in Table 1.

Overall, all the pairs showed an improvement in performance from the first to the last session, both in terms of pick accuracy and error. The best improvement is shown by the pair G1-C7 with an increase of accuracy from 30% on average to more than 70% in the last session. Moreover, in the last session, the wrong picks are limited only to adjacent objects as highlighted by the average pick error of  $0.3 \pm 0.1$ . The worst improvements are achieved by the pair G2-D7 showing the highest performance in session

three. Nevertheless, they also obtained an increase in the average pick accuracy of more than 10% in the last session with respect to the beginning of the training.

**Control performance:** Figure 3 illustrates the trajectories computed for a sample pair (G1-C7) throughout all sessions, wherein the target reached corresponds to the one prompted by the cue (i.e., correct trials). Notably, there is an observable enhancement in spatial exploration across sessions. In the initial session, the pair was able to reach only the most central target objects. While, in sessions two and three, they managed to reach four targets, and in the subsequent two sessions, they successfully reached all five targets. Figure 3b portrays the Fréchet distance [19] with respect to ideal trajectories, computed by averaging the trajectories of all the correct trials of every pair and every session for each target. This metric is calculated for both correct and erroneous trials, with the mean value computed for each session. Additionally, Table 2 outlines the average Fréchet distance for each session and pair. Except for the pair G2-D7, the subjects showed a reduced Fréchet distance of more than

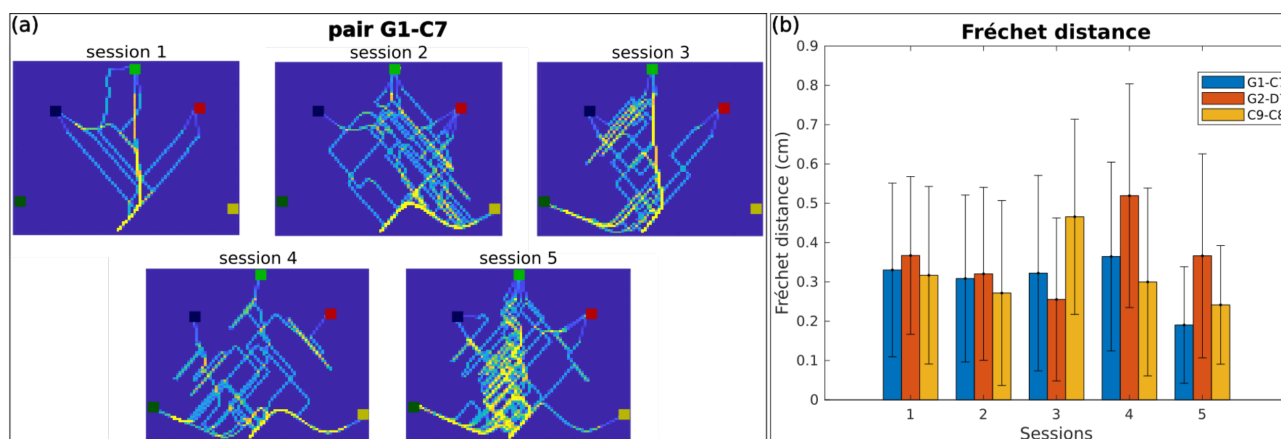


Figure 3: Control performances. (a) Correct trajectories computed by pair G1-C7 for each session. In the initial session, they reached three targets, while in the subsequent two sessions, they achieved four goals. Remarkably, in the last two sessions, they successfully reached all five targets. (b) Fréchet distance (in centimetres, cm) for each session and pair of subjects.

	G1-C7	G2-D7	C9-C8
session 1	33.0 ± 22.1	36.7 ± 20.0	31.6 ± 22.5
session 2	30.8 ± 21.2	32.0 ± 22.0	27.1 ± 23.5
session 3	32.2 ± 24.8	25.5 ± 20.7	46.5 ± 24.8
session 4	36.4 ± 24.0	51.9 ± 28.4	29.9 ± 23.8
session 5	19.0 ± 14.8	36.6 ± 25.9	24.1 ± 15.0

Table 2: Fréchet distance (cm) for each session and pair.

10cm in the last session with respect to the first session, signifying a reduced deviation of the paths traversed to the ideal trajectories.

## DISCUSSION

As anticipated, the aim of this work is to deviate from the literature which uses the cBMI only as a method to increase the BMI performance but rather to exploit it in combination with a real robotic application as an innovative approach to foster the subject's learning. As reported in Figure 2, all the pairs show improvements in controlling the robotic device through the cBMI. Indeed, when the experiment started we hypothesised that each subject had a low capacity to control the robot, due to inexperience with this new approach, and because the modulation of brain rhythms is difficult at the beginning. More in details, at the beginning it was difficult for subject G1 to explore all the probability space of the BMI decoder. Nevertheless, thanks to the cBMI approach the pair was able to have sufficient control to directly train on the real robotic device and subject G1 learnt how to perform the required mental tasks leading to a more uniform distribution of the BMI output in the last sessions. It is worth noting that a broader spanning of the probability space is particularly important in our paradigm as the robotic device was controlled by mapping the BMI output to continuous control commands, rather than exploring a discrete control approach as commonly employed in the literature. By analysing the evolution of the neural correlates over the training sessions, we suggest that these improvements of performance in all the subjects are not only related to a

familiarization of the subjects to the system, but rather to an effective learning of BMI skills. Indeed, the topographic maps of Figure 2(b) displays a visible increase in the number and power of the discriminant features in both  $\alpha$  and  $\beta$  bands, in line with previous works on subject's learning in BMI [4, 5].

A thing to note is that the third session was done immediately after the winter holidays, thus more than two weeks apart from the previous two sessions. This may explain why all the pairs show a drop in the performance around the third and fourth sessions, as it can be seen also in Figure 2(c). Indeed, all the subjects were either completely naive or not proficient BMI users at the beginning of the experiment. Thus, they were not able to stabilize their features and skills in the short-term training before the stopping period. Nevertheless, when the training was restarted they recovered and further boosted the discriminant features. The learning of these BMI skills is also suggested by the performance of robot control. Two out of three pairs (G1-C7, C9-C8) showed an improvement in pick accuracy, pick error and Fréchet distance from the first to the last session, highlighting a more optimal control of the robotic manipulator directed to the desired target object.

Our study suggests that cBMI might be exploited as an alternative training protocol for MI BMIs. However, this work suffers from some limitations. First of all, the small number of subjects participating in the experiment, in fact, a larger pool is required to verify and strengthen our findings. Secondly, the relatively short training period (only 5 sessions) that might be not sufficient for all subjects to acquire stable and robust features. Future work will address these limitations by expanding the number of subjects and by designing a longitudinal study.

## CONCLUSION

In this article, we introduce a novel protocol for collaborative MI BMI with shared control for manipulating a

robotic arm. Our hypothesis posited that with a collaborative BMI, both subjects would acquire proficiency in generating BMI commands and, thus, in controlling the robotic arm across multiple sessions. The outcomes obtained from a sample of three pairs of subjects, either completely naive or not expert in MI BMI, suggested promising results supporting our hypothesis and demonstrating the feasibility of utilizing this protocol to train subjects in acquiring BMI skills while directly engaging with real robotic application since the first days. We believe that the proposed approach of using the cBMI as training protocol will help not only to reduce the time investment required for reaching satisfying BMI performance, but also to mitigate the potential frustration experienced by naive users at the beginning of the training.

## FUTURE WORKS

To further validate the proposed approach, we aim to evaluate in a larger study the improvements introduced by the collaborative BMI on the subject's learning with respect to training the subjects individually. Moreover, we want to investigate how the learning process is influenced by the pairing of individuals. Thus, we will compare the learning of pairs of both naive subjects with respect to pairing a naive subject with an expert BMI user. Finally, the use of the proposed cBMI will be extended to other and more challenging scenarios such as controlling the robotic arm in daily-living tasks or for entertainment in collaborative gaming applications.

## ACKNOWLEDGMENT

The work is partially funded by the Piano Nazionale di Ripresa e Resilienza (PNRR), Project PE8 "AGE-IT" (Spoke 9) and Progetti di Rilevante Interesse Nazionale (PRIN 2022, 2022BCZ52A) - Ministero dell'Università e della Ricerca. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

## REFERENCES

- [1] McFarland DJ, Wolpaw JR. EEG-based brain-computer interfaces. *current opinion in Biomedical Engineering*. 2017;4:194–200.
- [2] Abiri R, Borhani S, Sellers EW, Jiang Y, Zhao X. A comprehensive review of EEG-based brain-computer interface paradigms. *Journal of Neural Engineering*. 2019;16(1):011001.
- [3] Alimardani M, Nishio S, Ishiguro H. Brain-computer interface and motor imagery training: The role of visual feedback and embodiment. *Evolving BCI Therapy-Engaging Brain State. Dynamics*. 2018;2(64).
- [4] Tortora S *et al.* Neural correlates of user learning during long-term BCI training for the cybathlon competition. *Journal of NeuroEngineering and Rehabilitation*. 2022;19(1):1–19.
- [5] Tonin L *et al.* Learning to control a BMI-driven wheelchair for people with severe tetraplegia. *iScience*. 2022;25(12).
- [6] Leeb R *et al.* Transferring brain-computer interfaces beyond the laboratory: Successful application control for motor-disabled users. *Artificial intelligence in medicine*. 2013;59(2):121–132.
- [7] Poli R, Valeriani D, Cinel C. Collaborative brain-computer interface for aiding decision-making. *PloS one*. 2014;9(7):e102693.
- [8] Bonnet L, Lotte F, Lécuyer A. Two brains, one game: Design and evaluation of a multiuser BCI video game based on motor imagery. *IEEE Transactions on Computational Intelligence and AI in games*. 2013;5(2):185–198.
- [9] Valeriani D, Poli R, Cinel C. Enhancement of group perception via a collaborative brain-computer interface. *IEEE Transactions on Biomedical Engineering*. 2017;64(6):1238–1248.
- [10] Yijie Z *et al.* A multiuser collaborative strategy for MI-BCI system. In: 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP). 2018, 1–5.
- [11] Wang Y, Wang YT, Jung TP, Gao X, Gao S. A collaborative brain-computer interface. In: 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI). 2011, 580–583.
- [12] Myrden A, Chau T. Effects of user mental state on EEG-BCI performance. *Frontiers in human neuroscience*. 2015;9:308.
- [13] Roc A *et al.* A review of user training methods in brain computer interfaces based on mental tasks. *Journal of Neural Engineering*. 2021;18(1):011002.
- [14] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: A closer look on bci results. *International journal of bioelectromagnetism*. 2008;10(1):52–55.
- [15] Tortora S, Tonin L, Chisari C, Micera S, Menegatti E, Artoni F. Hybrid human-machine interface for gait decoding through bayesian fusion of EEG and EMG classifiers. *Frontiers in Neurorobotics*. 2020;14:582728.
- [16] Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*. 2017, 2980–2988.
- [17] Tortora S, Gottardi A, Menegatti E, Tonin L. Continuous teleoperation of a robotic manipulator via brain-machine interface with shared control. In: 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA). 2022, 1–8.
- [18] Gottardi A, Tortora S, Tosello E, Menegatti E. Shared control in robot teleoperation with improved potential fields. *IEEE Transactions on Human-Machine Systems*. 2022;52(3):410–422.
- [19] Alt H, Godau M. Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*. 1995;5(01n02):75–91.

## THE CHALLENGE OF DRIVING BCI WITH EMOTIONAL SIGNALS COLLECTED BY EEG

Daniele Lozzi<sup>1,3†</sup> and Enrico Mattei<sup>3†</sup>, Roberta Ciuffini<sup>2</sup>, Alessandro Di Matteo<sup>1,3</sup>, Alfonso Marrelli<sup>6</sup>, Raffaele Ornello<sup>5</sup>, Matteo Polsinelli<sup>4</sup>, Chiara Rosignoli<sup>5</sup>, Simona Sacco<sup>5</sup>, and Giuseppe Placidi<sup>1,2</sup>

<sup>1</sup>A<sup>2</sup>VI-Lab, Univ. L'Aquila, L'Aquila, Italy

<sup>2</sup>Dept. of MESVA, Univ. L'Aquila, L'Aquila, Italy

<sup>3</sup>Dept. of DISIM, Univ. L'Aquila, L'Aquila, Italy

<sup>4</sup>Dept. of DISA-MIS, Univ. Salerno, Fisciano, Italy

<sup>5</sup>Dept. of DISCAB, Univ. L'Aquila, L'Aquila, Italy

<sup>6</sup>Neurophysiopathology Unit, San Salvatore Hospital, L'Aquila, Italy

E-mail: daniele.lozzi@graduate.univaq.it

† The first two authors mainly contributed as primary co-authors.

**ABSTRACT:** The paper describes the important challenges of driving a BCI through EEG signals of emotions. In particular, the complex emotional processing activated by the human brain and the necessity of generating elicitation protocols to synchronize the acquisition of EEG signals from emotions are presented. Besides, the limitations of EEG in dealing with signals from emotions are also discussed. Then, the specific neuropsychological issues related to the use of protocols for eliciting emotions are described. Due to the huge difficulty in managing the uncertainty deriving from the above issues, the surprising results obtained by recently proposed automatic strategies for emotion classification and recognition, also raising doubts about the correctness of the results, are reported and discussed. Finally, suggestions are presented regarding some procedures for uncertainty reduction and for the future complete development of EEG-based emotional BCIs.

### INTRODUCTION

A Brain-Computer Interface (BCI) [1] is a computer-based communication system that collects signals generated by the evoked neural activity of the Central Nervous System (CNS) and its goal is to provide a new channel of output for the brain and requires voluntary adaptive control by the user [2]. Electroencephalography (EEG) is one of the most commonly used techniques to measure neural activity, through electrical signals, by placing electrodes outside the skull [3]. EEG provides high temporal resolution responses, is easy to use, safe, low-cost, and, for these reasons, effective in providing the necessary brain feedback for a BCI. A sketch of a BCI driven by EEG signals is provided in Fig. 1. BCIs, mainly those used as an alternative communication tool for disabled people, are based on event-related signals induced by external stimuli and synchronized with them (an example is the P300

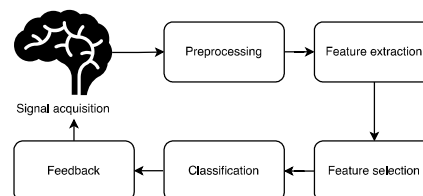


Figure 1: Overview of an online BCI system. The signal undergoes acquisition (not shown), preprocessing, feature extraction, feature selections, and classification before feedback is produced depending on its final classification.

[4]). Another consistent part is based on sensory-motor rhythm amplitudes [5]. For patients with impaired vision, or suffering from seizures attacks caused by too fast visual stimuli such as those used in P300, or that have never experienced the control of the motor part of their body, or whose signals produced by sensory-motor rhythms, mostly at the alpha band, can be easily confused with those due to artifacts caused by involuntary and frequent movements, other paradigms, such as auditory [6] and tactile [7], have been explored. However, cases in which also these paradigms have little or no effect are frequent [6]. Moreover, there are some situations in which some well-known techniques to build a BCI are not possible, like severe brain injuries that affect, for example, P300 elicitation [8]. For this reason, new ways have been explored and one of the most promising is that BCIs are based on the voluntary brain activity produced by emotions, being emotions mostly related to the deeper parts of the brain, mostly unaffected by disabilities [9–14]. Emotions were first explored in the field of affective computing where some fascinating studies are dedicated to making the computer more empathic to the user and involved in the measurement of the user's emotions and representing them into human-computer interaction systems [15]. They aim to find the activation of specific

brain regions in response to specific emotions but, while some regions are more active than others when experiencing specific emotions, no specific region is activated by a single emotion [16–18]. The brain regions most responsible for emotions are the amygdala, insula, anterior cingulate cortex, and orbitofrontal cortex. Through fMRI, it has been found that there exist specific patterns of brain activity, i.e. groups of brain districts, related to specific emotions and that these patterns are common across individuals [18]. A scheme of the brain regions mostly involved in the processing of emotions is indicated in red in Fig. 2. Despite these advances, it remains very difficult to recognize emotions, especially across individuals, because their patterns are very similar and can confuse each other and because they are also subjective (i.e. different individuals can have different ways of dealing with emotions). Moreover, complex multi variable pattern analysis techniques have to be used to identify distributed patterns associated with specific emotions, especially from EEG. Indeed, in EEG the sensitivity of the electrodes is higher for external neurons and decreases for deeper neurons (yellow part in Fig.2). This means that not all neurons equally contribute to the EEG signal, with an EEG predominately reflecting the activity of cortical neurons near the electrodes on the scalp. Deep structures within the brain, mainly involved in the emotional process, are further away from the electrodes and will not contribute directly to an EEG. Furthermore, EEG signals have an intrinsic nonlinear and nonstationary nature. The basic source of the nonstationarity in EEG signal is a reflection of switching of the inherent quasi-stable states of neural assemblies during brain functioning and is not due to the casual influences of the external stimuli on the brain mechanisms. EEG signal recorded from a scalp electrode is influenced by different deeper sources, each 'transmitting' with different and variable intensity, thus making the main source of the registered signal from one brain structure to another. Nonstationarity also arises because of different time scales of the dynamic processes of brain activity. Finally, the signals are affected by noise and artifacts which, in some cases, are difficult to reduce effectively. Despite that, recently proposed nonlinear automatic strategies, learning by data collected during task-related protocols execution, promise to recognize emotions with very high accuracy, starting from EEG signals collected by eliciting protocols. In what follows, we first discuss the complex emotional processing activated by our brain, the disagreement in creating a commonly accepted model for representing emotions, and the necessity of generating elicitation protocols to synchronize the acquisition of EEG signals from emotions. Then we describe the specific neuropsychological issues related to the use of protocols for eliciting emotions. Moreover, we emphasize the surprising results obtained by recently proposed automatic strategies for emotion classification and recognition from EEG signals collected by these elicitation protocols, also raising doubts about the correctness of the results in light of the great uncertainty of the data

and protocols. Finally, we propose some cues for the future of EEG-based emotional BCIs.

## BRAIN EMOTIONAL PROCESSING AND EMOTIONAL MODELS

The neural locations involved in the genesis and processing of emotions are multiple, including the Autonomic Nervous System (ANS), the hypothalamus, the ascending reticular system, the limbic system, some cortex lobes, and the amygdala [19]. In recent years, attention has also been focused on the network of interconnected regions involved in emotional processing, even if neuropsychological data do not confirm the theory of a single emotional network, but of multiple networks that control multiple emotions [19]. In these networks, structures such as the thalamus, somatosensory cortex, somatosensory association cortex, amygdala, insula, and medial prefrontal cortex have been clearly recognized [19]. Although there is no single shared definition of emotion, many researchers define an emotion as a feeling related to an event occurring during a subjective experience characterized by a complex brain function including information acquisition, manipulation, storage, and recall [20]. Ochsner and Gross [21] assumed that emotions are consequences of external stimulations or internal mental representations with valence and well-defined characteristics. Furthermore, some researchers [19] hypothesize that emotions have three components: physiological reaction, behavioral response, and feeling, i.e. a subjective response to emotions. Emotion recognition aims to detect the affective state of a subject directly by brain activity and recent works focalized the attention on the correlation between brain oscillations and emotions [22, 23]. Actually, besides the lack of an univocal definition of emotion, there is also no consensus in the scientific community on a general theory of emotions. Indeed, the theory of emotions is divided into two main models: discrete and dimensional models [24]. Discrete models classify emotions without considering any axis to quantify the specific characteristics of each one. Furthermore, the number of emotions depends on the theoretical framework. Dimensional models define some common continuous features and place each emotion in a point of the space based on the values of the considered features. One of the most accepted dimensional models is the Russel Circumplex Model [25]. It has two dimensions: Arousal (degree of activation) and Valence (degree of pleasure). More complex dimensional models, besides Valence and Arousal, also include Dominance (degree of attention) to produce a Valence Arousal-Dominance (VAD) or Pleasure-Arousal- Dominance (PAD), creating a 3D space of emotions [26]. Valence goes from unpleasant to pleasant, Arousal goes from passive to active, and Dominance goes from submissive to dominant, representing the degree of controllability of a specific emotion. Another model [27] is composed of Valence, Arousal, Dominance, and Predictability, where the last element

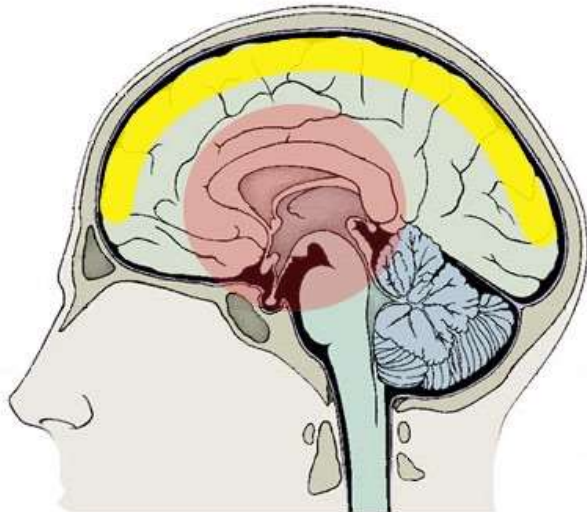


Figure 2: EEG brain sensitivity map (in yellow), with sensitivity decreasing with depth, and the main circuits of the limbic system, where emotions mostly originate (in red).

represents the level of surprise. Among many, EEG is one of the technique to measure brain activity in real-time and in almost normal, minimally invasive, conditions. It has the advantage to have great time resolution but low spatial resolution, and, compared to others, it is cheaper and more portable. However, the accurate study of emotions would require their direct measurement. The measurement of spontaneous emotions in daily-life conditions is challenging, due to many confounding variable that occurs during the recording. Usually, for EEG, an external stimulation (videos, sounds, images, etc.) or memory recall is required to elicit emotions in well-defined experimental sessions and environments. The resulting signals are usually stored in public datasets [28] to allow open-source recovery, processing, and analysis. In particular, many of them use external stimuli such as videos or/and audios [29–31] and some of them use self-stimulation from their memory [32, 33]. Emotion recognition is a particularly difficult domain due to the number of variables influencing emotions, including the elicitation mode, the experimental protocol, the subject's basal emotional state, and the used instrumentation. The main problem with datasets collected through external elicitation is the presence of biases, mainly related to the used protocol. For this reason, in the last years, many signal datasets of emotions have been proposed, each with pros and cons, but no one is completely free from biases.

## ELICITATION PROTOCOLS, EEG DATASETS, AND CHALLENGES

An important aspect necessary for recognizing emotions by EEG is to get specific features characterizing their signals, collected from several analyzed subjects. This is done through the design of specific elicitation protocols and the organization of measurement sessions during which the elicited emotions are synchronized with the

measurement, making it very difficult to measure spontaneous emotions in a real-life context. The protocols produce stimuli in the form of items, events, or conditions that cause a person to elicit emotional responses or behaviors for studying and understanding various psychological processes. Stimuli can include situations, scenarios, or social interactions. Well-known stimuli for eliciting the targeted emotions are virtual reality (VR), images, video games, music, audio/video clips, audio, and/or videos [28, 34–36]. Based on the type of stimulus, various emotions are elicited which are manually ranked to be used by nonlinear processing strategies. Without going into specific details about the elicitation protocols and the related datasets that originated from them, it is worth noting that several criticalities take origin from every stimulation protocol we can define. Overall, significant challenges arise to obtaining reliable and representative EEG data of emotions, with many limitations stemming from elicitation approaches, lack of adequate baselines, reduced number of sensors, equipment instability, etc. In summary, we can observe numerous challenges that must be overcome in order to obtain EEG data that more accurately represent emotions, aiming to surpass current limitations caused by stimulation protocols, lack of adequate baselines, limited number of sensors during acquisition, and instrument instability. The number of EEG channels should always be very high ( $> 64$ ) to allow for high-density acquisition, which can then be reduced during preprocessing to identify the most significant channels, and a multimodal acquisition should always be used rather than a single system (e.g., EEG + ECG). Additionally, environmental parameters should be measured to verify and correct instrumental errors. User interaction with the experiment (e.g., evaluation of each stimulus) introduces numerous artifacts and interruptions that could introduce further bias and noise into the measured signal. Furthermore, a simple limit to overcome concerns acquiring all possible demographic, physical conditions (age, hair, health...) and psychological data to control for potential confounding variables (e.g., lateralization) and using validated psychological scales for assessing emotional states. Additionally, using many different trials for each recorded emotion allows for intra-subject and inter-subject analyses, and the acquisition context should be as neutral as possible to avoid introducing additional biases due to the environment. Finally, the subject's adaptation time to the new experiment should be respected to allow them to feel comfortable, and consideration should be given to the subject's fatigue for long-duration experiments, including assessing their mental workload at another time. Among the current limitations specifics for emotional studies, we found the use of individual emotional elicitation trials for each participant, as well as the significant subjectivity of responses to individual external emotional stimuli, which thus require prior and/or posterior evaluation to assess whether such stimuli are appropriate for the subject. Often, there is a lack of adequate baseline in databases, which should instead be recorded

at numerous points before, after, and during EEG data acquisition and also is needed the acquisition of their personal “ground truth” of the stimuli used, acquiring their evaluation of the stimuli used in the experiment. Finally, there is the ethical problem that must be taken into account during emotional stimulation. In emotional stimulation protocols, images or videos containing emotional situations are often used to elicit emotional reactions in participants that are commonly referred to external situations (e.g. people in critical situations) and not related to the participant’s personal situation. It is important to note that this approach tends to focus mainly on negative emotions, as these cannot be induced directly in participants. In contrast, positive emotions can be evoked either directly or through indirect stimulation that shows third parties experiencing positive emotions. It is crucial to consider that, under normal conditions, participants are in a state of relaxation, which can be interpreted as a positive emotion according to many theories, including Russell’s model, which posits that there are no “neutral” emotions [25]. Finally, the advancement in using EEG for recovering emotions depends on the ability to overcome the above issues effectively, when possible.

#### THE SURPRISING PERFORMANCE OF THE RECENT EMOTION RECOGNITION MODELS

In the last few years, with the advent of non-linear AI-based models, there has been a huge effort to improve emotion recognition by EEG [37, 38] and many more have been added since the above reviews. In particular, deep learning (DL) methods with convolutional neural networks (CNN) [39] and long short-term memory (LSTM) [40] have provided better performance, in terms of accuracy, than traditional deterministic processing algorithms. Indeed, over the years these models have been refined to include intrinsic characteristics of the problem, to get any potential spatial/temporal unknown relationship, and to better fit the data at hand. For example, the dynamic relationship among EEG channels in different regions has been represented graphically through dynamic graph CNN [41] and adaptive graphs [42], for taking into account the spatial correlations of EEG data. Besides, to include the cross-subject variability and non-stationarity of the EEG signals, transfer learning has been introduced to solve the inconsistencies between training signals and the signals used for the test. To this aim, new hypotheses were added to the models regarding brain symmetry/asymmetry, as in the bi-hemisphere domain adversarial neural network [43] and in the bi-hemispheric discrepancy model [44]. Other methods try to reduce subject variability in EEG by introducing specific domain residual networks [45, 46]. These models are similar to residual networks with the advantage that they do not require any prior information during training. Several other model adaptations have regarded the proposal of a dynamic adversarial adaptation network that dynamically learns domain-invariant representations both on a

local and a global scale. One of these methods [46], proposed a joint distribution adaptation to take into account the joint distribution differences across individuals. Another recently proposed model [47] proposed an adversarial discriminative temporal convolutional network for cross-domain (cross-subject and/or cross-dataset) emotion recognition with further achievements in emotion recognition. Without going into further details, several other methods for emotion recognition have been proposed, some of them are here reported [48–54], whose results have allowed to improve the accuracy that, for a three-class emotion dataset [55] (e.g. Positive, Negative and Neutral), ranged from about 60%, from the first moves of DL, to above 90% and, for a four-class emotion dataset [56] (e.g. Neutral, Sad, Fear, Happy), ranged from 38% to above 80%, with a huge improvement in the result stability (standard deviation passed from about 13% of the first DL models, to about 7% of the recently proposed models).

#### DISCUSSION AND CONCLUSION

Recent DL models applied to EEG have become ever more and more complex to include any sort of prior information and/or to highlight any sort of potential relationship among the measured data. In some cases, they also included constrained based on findings provided by other diagnostic tools, like fMRI, to push the performance toward ideal values. Since most of the intrinsic mechanisms of brain behavior are still not completely understood, in particular regarding emotions, and because complementary measurement tools collect different parameters of the same phenomenon, consider different measurement brain places, are differently affected by noise and artifacts, and use different protocols, it could be improper to use the results of one to design constraints for the other. According to EEG, its main limitation on emotion recognition is that its measurements take place far from the source of emotions and that it probably measures the brain processing of emotions more than the emotions themselves. However, this does not mean that EEG measurements are not useful for emotion recognition but just that the task could be difficult. Besides, the EEG signals are intrinsically nonlinear and nonstationary, and, those from emotions, are strongly subject-dependent and related to other brain processing tasks. Last, but not least important, emotions are not collected during real-life experience but during the execution of a task elicited by a specific activation/synchronization protocol. Despite that, the performance of the recent DL models for emotion recognition by EEG is surprisingly good and doubt arises: the designed models could unintentionally use plenty of heuristics, producing unaware data overfitting. The present article aims at giving rise to a reflection on the goodness of these results, to push, for the future, to a substantial verification of the proposed methods, to define further elicitation protocols to collect data to be used in this verification, to apply explainable AI for comprehend

which features, brain regions, and brain connections (spatial and temporal) are mostly involved in the emotional process and, finally, to apply these verified findings for EEG, emotion-driven, BCI.

## REFERENCES

- [1] Yadav H, Maini S. Electroencephalogram based brain-computer interface: Applications, challenges, and opportunities. *Multimedia Tools and Applications*. 2023;82(30):47003–47047.
- [2] Shih JJ, Krusienski DJ, Wolpaw JR. Brain-computer interfaces in medicine. In: *Mayo clinic proceedings*. 2012, 268–279.
- [3] Niedermeyer E, Silva FL da. *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins (2005).
- [4] Farwell LA, Donchin E. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*. 1988;70(6):510–523.
- [5] Neuper C, Müller-Putz GR, Scherer R, Pfurtscheller G. Motor imagery and eeg-based control of spelling devices and neuroprostheses. *Progress in brain research*. 2006;159:393–409.
- [6] Kübler A, Furdea A, Halder S, Hammer EM, Nijboer F, Kotchoubey B. A brain–computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Annals of the New York Academy of Sciences*. 2009;1157(1):90–100.
- [7] Müller-Putz GR, Scherer R, Neuper C, Pfurtscheller G. Steady-state somatosensory evoked potentials: Suitable brain signals for brain-computer interfaces? *IEEE transactions on neural systems and rehabilitation engineering*. 2006;14(1):30–37.
- [8] Placidi G, Cinque L, Di Giamberardino P, Iacoviello D, Spezialetti M. An affective bci driven by self-induced emotions for people with severe neurological disorders. In: *New Trends in Image Analysis and Processing-ICIAP 2017: ICIAP International Workshops, WBICV, SSPandBE, 3AS, RGBD, NIVAR, IWBAAS, and MADiMa 2017, Catania, Italy, September 11-15, 2017, Revised Selected Papers 19*. 2017, 155–162.
- [9] Garcia-Molina G, Tsoneva T, Nijholt A. Emotional brain–computer interfaces. *International journal of autonomous and adaptive communications systems*. 2013;6(1):9–25.
- [10] Nie D, Wang XW, Shi LC, Lu BL. Eeg-based emotion recognition during watching movies. In: *2011 5th international IEEE/EMBS conference on neural engineering*. 2011, 667–670.
- [11] Lozzi D, Mignosi F, Spezialetti M, Placidi G, Polsinelli M. A 4d lstm network for emotion recognition from the cross-correlation of the power spectral density of eeg signals. In: *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. 2022, 652–657.
- [12] Iacoviello D, Petracca A, Spezialetti M, Placidi G. A real-time classification algorithm for eeg-based bci driven by self-induced emotions. *Computer methods and programs in biomedicine*. 2015;122(3):293–303.
- [13] Iacoviello D, Petracca A, Spezialetti M, Placidi G. A classification algorithm for electroencephalography signals by self-induced emotional stimuli. *IEEE transactions on cybernetics*. 2015;46(12):3171–3180.
- [14] Pistoia F *et al*. Eeg-detected olfactory imagery to reveal covert consciousness in minimally conscious state. *Brain injury*. 2015;29(13-14):1729–1735.
- [15] Picard RW, Klein J. Computers that recognise and respond to user emotion: Theoretical and practical implications. *Interacting with computers*. 2002;14(2):141–169.
- [16] Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD. Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*. 2008;42(2):998–1031.
- [17] Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF. The brain basis of emotion: A meta-analytic review. *Behavioral and brain sciences*. 2012;35(3):121–143.
- [18] Kassam KS, Markey AR, Cherkassky VL, Loewenstein G, Just MA. Identifying emotions on the basis of neural activation. *PloS one*. 2013;8(6):e66032.
- [19] Gazzaniga MS, Ivry RB, Mangun G. *Cognitive neuroscience. the biology of the mind*, (2014). 2006.
- [20] Alarcao SM, Fonseca MJ. Emotions recognition using eeg signals: A survey. *IEEE transactions on affective computing*. 2017;10(3):374–393.
- [21] Ochsner K, Gross J. The cognitive control of emotion, *trends in cognitive sciences*. 2005; 9 (5): 242–249.
- [22] Placidi G, Avola D, Petracca A, Sgallari F, Spezialetti M. Basis for the implementation of an eeg-based single-trial binary brain computer interface through the disgust produced by remembering unpleasant odors. *Neurocomputing*. 2015;160:308–318.
- [23] Di Giamberardino P, Iacoviello D, Placidi G, Polsinelli M, Spezialetti M. A brain computer interface by eeg signals from self-induced emotions. In: *VipIM-AGE 2017: Proceedings of the VI ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing Porto, Portugal, October 18-20, 2017*. 2018, 713–721.
- [24] Fujimura T, Matsuda YT, Katahira K, Okada M, Okanoya K. Categorical and dimensional perceptions in decoding emotional facial expressions. *Cognition & emotion*. 2012;26(4):587–601.
- [25] Russell JA. A circumplex model of affect. *Journal of personality and social psychology*. 1980;39(6):1161.
- [26] Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. *Journal of research in Personality*. 1977;11(3):273–294.
- [27] Fontaine JR, Scherer KR, Roesch EB, Ellsworth PC. The world of emotions is not two-dimensional. *Psychological science*. 2007;18(12):1050–1057.

- [28] Kamble K, Sengupta J. A comprehensive survey on emotion recognition based on electroencephalograph (eeg) signals. *Multimedia Tools and Applications*. 2023;1–36.
- [29] Koelstra S *et al.* Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*. 2011;3(1):18–31.
- [30] Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*. 2011;3(1):42–55.
- [31] Katsigiannis S, Ramzan N. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*. 2017;22(1):98–107.
- [32] Onton JA, Makeig S. High-frequency broadband modulation of electroencephalographic spectra. *Frontiers in human neuroscience*. 2009;3:560.
- [33] Bigirimana AD, Siddique NH, Coyle D. Brain-computer interfacing with emotion-inducing imagery: A pilot study. In: the 7th Graz BCI Conference 2017. 2017.
- [34] Marchewka A, Żurawski Ł, Jednoróg K, Grabowska A. The nencki affective picture system (naps): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior research methods*. 2014;46:596–610.
- [35] Lang P, Bradley MM. The international affective picture system (iaps) in the study of emotion and attention. *Handbook of emotion elicitation and assessment*. 2007;29:70–73.
- [36] Sarma P, Barma S. Review on stimuli presentation for affect analysis based on eeg. *IEEE Access*. 2020;8:51991–52009.
- [37] Wang J, Wang M. Review of the emotional feature extraction and classification using eeg signals. *Cognitive Robotics*. 2021;1:29–40.
- [38] Rahman MM *et al.* Recognition of human emotions using eeg signals: A review. *Computers in Biology and Medicine*. 2021;136:104696.
- [39] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*. 2018;15(5):056013.
- [40] Wang Y *et al.* Eeg-based emotion recognition with similarity learning network. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, Jul. 2019.
- [41] Song T, Zheng W, Song P, Cui Z. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*. 2020;11(3):532–541.
- [42] Song T, Liu S, Zheng W, Zong Y, Cui Z. Instance-adaptive graph for eeg emotion recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020;34(03):2701–2708.
- [43] Li Y, Zheng W, Zong Y, Cui Z, Zhang T, Zhou X. A bi-hemisphere domain adversarial neural network model for eeg emotion recognition. *IEEE Transactions on Affective Computing*. 2018;12(2):494–504.
- [44] Li Y *et al.* A novel bi-hemispheric discrepancy model for eeg emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*. 2020;13(2):354–367.
- [45] Ma BQ, Li H, Zheng WL, Lu BL. Reducing the subject variability of eeg signals with adversarial domain generalization. In: *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part I* 26. 2019, 30–42.
- [46] Li J, Qiu S, Du C, Wang Y, He H. Domain adaptation for eeg emotion recognition based on latent representation similarity. *IEEE Transactions on Cognitive and Developmental Systems*. 2019;12(2):344–353.
- [47] He Z, Zhong Y, Pan J. An adversarial discriminative temporal convolutional network for eeg-based cross-domain emotion recognition. *Computers in Biology and Medicine*. 2022;141:105048.
- [48] Zhong P, Wang D, Miao C. Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*. 2020;13(3):1290–1301.
- [49] Cimtay Y, Ekmekcioglu E. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition. *Sensors*. 2020;20(7):2034.
- [50] Chen H *et al.* Personal-zscore: Eliminating individual difference for eeg-based cross-subject emotion recognition. *IEEE Transactions on Affective Computing*. 2021;14(3):2077–2088.
- [51] Cao J, He X, Yang C, Wang Z. Multi-source and multi-representation adaptation for cross-domain electroencephalography emotion recognition. *Frontiers in Psychology*. 2022;12:809459.
- [52] Cui H, Liu A, Zhang X, Chen X, Liu J, Chen X. Eeg-based subject-independent emotion recognition using gated recurrent unit and minimum class confusion. *IEEE Transactions on Affective Computing*. 2022.
- [53] Li Z *et al.* Dynamic domain adaptation for class-aware cross-subject and cross-session eeg emotion recognition. *IEEE Journal of Biomedical and Health Informatics*. 2022;26(12):5964–5973.
- [54] Zhang G, Davoodnia V, Etemad A. Parse: Pairwise alignment of representations in semi-supervised eeg learning for emotion recognition. *IEEE Transactions on Affective Computing*. 2022;13(4):2185–2200.
- [55] Zheng WL, Lu BL. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*. 2015;7(3):162–175.
- [56] Zheng WL, Liu W, Lu Y, Lu BL, Cichocki A. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*. 2018;49(3):1110–1122.

## PROJECT NAFAS: ANNOUNCEMENT AND BRIEF OVERVIEW

Laurens R. Krol<sup>1</sup>, Thorsten O. Zander<sup>1,2</sup>

<sup>1</sup> Zander Laboratories GmbH, Germany

<sup>2</sup> Brandenburg University of Technology Cottbus–Senftenberg, Germany

E-mail: laurens@zanderlabs.com

### ABSTRACT

Funded by the German federal agency *Agentur für Innovation in der Cybersicherheit* - “*Innovation for Cybersecurity*” (Cyberagentur) with a record €30 million, we announce Zander Labs’ Project NAFAS, which aims to integrate Brain-Computer Interface (BCI) technology with Artificial Intelligence (AI). By first addressing the traditional constraints of EEG-based neurotechnology and developing mobile, secure hardware capable of decoding multiple mental states in real time, this project paves the way for a new era of Neuroadaptive Human-Computer Interaction (HCI)—and, ultimately, Neuroadaptive AI. Beyond the project’s scientific aims which we briefly introduce, Project NAFAS itself represents confidence in the ability of the scientific community to solve the critical challenge of transitioning BCIs from theoretical constructs to practical real-world applications, and in the positive impact the resulting BCI technology can have in our daily lives.

### INTRODUCTION

Already in the 1960s, when initial thought came to paper concerning what was termed man-computer symbiosis, it was said to be “likely that the contributions of [humans] and [computers] will blend together so completely [...] that it will be difficult to separate them” [1]. This was more than a decade before Vidal introduced BCI as a unique human-computer communication method [2], and even a few years before Weizenbaum’s ELIZA inspired a large movement in AI research to focus on yet a different kind of “blend” [3]. In many ways it has been this idea of blending, or merging, human cognitive processes with digital computation in various forms that has guided HCI, BCI and AI advancements ever since. Project NAFAS (*Neuroadaptivity for Autonomous Systems*) sees itself in this same tradition, and targets these same technologies. By further developing passive BCI [4] technology, we enable HCI and AI to become neuroadaptive [5], and introduce a more intuitive, more natural, and indeed, more symbiotic form of human-computer or human-AI interaction.

Project NAFAS is Zander Labs’ winning proposal to a tender published by the Cyberagentur in 2022. The project will be executed together with a number of subcontractors, which we cannot yet exhaustively list at the time of submission. With its states goal “to harness

key technologies and breakthrough innovations that help to enable and improve internal and external security” ([6], translation from German by the authors), the Cyberagentur called for “secure neural human-machine interaction” to be developed, with which they thus identified (passive) BCI as a key technology of strategic interest. With € 30 million, to the best of our knowledge, this is the largest single-entity funding ever granted within the European Union. The tender in general, and Project NAFAS’ win in particular, highlights the importance of passive BCI and its related fields, and underscores their public recognition as such. We are happy to announce this project to the community and are looking forward to sharing our results.

### OBJECTIVES

In an era where technology increasingly mediates our interactions with the world and with other human beings, the importance of intuitive and natural forms of human-computer communication has never been more acute. Current BCI and AI technologies, while groundbreaking, often remain confined within the realms of academia and specialized applications, largely detached from the daily experiences of the broader public.

Project NAFAS is a four-year effort that was formulated to address these and other current limitations of neurotechnology in general and (passive) BCI in particular. Project NAFAS will develop safe, practical neurotechnology that allows a new generation of technologies to adapt in real-time to the cognitive and emotional states of users. We believe such a development will make digital interaction, be it with AI or more traditional HCI applications, more fluid, instinctive, and human-centered—or, will even remove the explicit need to ‘act’ altogether.

The safety and practicality of the technology are all elements that are explicitly addressed in this project, starting with new hardware that will be developed. Also the “new generation of technologies” is addressed, by developing Neuroadaptive HCI as a core technology, Neuroadaptive AI as an AI-focused extension, and a series of demonstrators to showcase all of these results.

#### *Safe and Practical: The Mobile EEG Suite*

Central to Project NAFAS is the development of a *Mobile EEG Suite*, which combines mobile, self-applicable electrodes and amplifiers with dedicated BCI

hardware.

The electrodes will be designed with a focus on ease of use and comfort, such that they can be applied by individuals without technical expertise, making the technology accessible to a broad audience. We recognize self-applicability and comfort as a requirement to reduce the current barriers to BCI adoption, and to enable widespread use across various environments. We do not limit ourselves to specific use cases, but rather target usability in the widest possible range, be it personal computing environments, outdoor settings, or sophisticated research laboratories.

Complementing the electrodes, the project will develop lightweight, miniaturized amplifiers that further increase the system's mobility and practicality. These amplifiers will be designed to be compact and efficient, with extended battery life suitable for prolonged use. Together with the electrodes, this will enable the continuous and comfortable monitoring of brain activity in everyday settings.

The monitored brain activity will also be decoded in hardware. A core target of Project NAFAS is to develop *universal classifiers*, allowing for the real-time interpretation of brain activity across different individuals without the need for dedicated calibration sessions. This addresses one of the most significant challenges in BCI technology today. A large-scale data collection project will be set up to provide the basis of these classifiers.

When EEG hardware can be worn anywhere, and calibration is no longer necessary, a person's naturally-occurring brain activity can be recorded and decoded at any time. To address the clear privacy issues arising from such a technology, the universal classifiers will be implemented directly in hardware. This hardware will give the user maximum authority over their data: they will have full, physical control over the decoding process, while the hardware ensures that no unauthorized brain activity can leak out. This follows, among other guidelines, the privacy- and (cyber)security-preserving BCI framework that was previously funded by the Cyberagentur [7].

In short, Project NAFAS will address a number of fundamental issues of present-day neurotechnology by providing mobile, comfortable, and safe hardware capable of decoding brain activity in real time without explicit calibration periods. This will serve as the basis for further developments within the project; specifically: Neuroadaptive HCI and Neuroadaptive AI.

#### *Neuroadaptive Human-Computer Interaction: Passive BCI in the Wild*

On the basis of the above neurotechnological hard- and software solutions, universal passive BCI classifiers can be made available in almost any and all HCI context, enabling *Neuroadaptive HCI*. A Neuroadaptive HCI software framework, which will be developed in this project, will enable systems to implicitly obtain, use, and even learn to anticipate, a person's needs and

preferences, and to adjust their functionality accordingly. Neuroadaptive HCI will thus allow for unprecedented personalization of existing HCI systems, and for the development of novel systems following neuroadaptive principles. For instance, educational software could modify its approach based on the learner's current mental state, enhancing both engagement and comprehension. By also integrating context assessment, the Neuroadaptive HCI framework will help ensure that interactions are not just personalized but also relevant to the user's immediate environment and situation.

The Neuroadaptive HCI framework will provide a core technology to make future interactive applications more intuitive, more natural, and more human-centric on the basis of brain activity.

#### *Neuroadaptive Artificial Intelligence: Empowering AI with Human-Like Understanding*

One of the main innovations of Project NAFAS will be to use all of the above technologies to ultimately expand the capabilities of AI. Project NAFAS' above-mentioned results will enable us to provide AI with real-time insights into human cognitive and emotional intricacies. We believe that this combination can cultivate AI systems that are not only more human-like, but ultimately even empathetic, capable of engaging in interactions that resonate on a human level.

With access to the kinds of subjective human nuances that are not present in any amount of raw data, but can only be obtained directly from implicit mental states, these neuroadaptive AI entities can adjust their responses and actions in real time, effectively anticipating user needs and preferences. Furthermore, we believe that these kinds of insights into the inner workings of human cognition, mental strategies, and decision making, can be used to teach future AIs uniquely human skills that are currently beyond their reach.

The concept of Neuroadaptive AI, then, represents a type of technology that is both uniquely human itself, and genuinely tuned to human idiosyncrasies, bridging the gap between cold computational processes and the dynamic spectrum of human emotion and cognition.

## CONCLUSION

The aims of Project NAFAS are bold: to stride toward the seamless integration of technology with the essence of human experience. But what the project represents, beyond the data, results, developments, and solutions that it targets, may be even bolder: it's a call to action for the global BCI, HCI, and AI research communities to envision and create a future where technology doesn't just serve us—it understands us, adapts to us, and becomes an empathetic extension of our own, human intellect. We see this as our collective chance to not only advance our scientific and technological frontiers, but to also make a profound impact on how we interact with the digital world.

## REFERENCES

- [1] Licklider, J. C. (1960). Man-computer symbiosis. IRE transactions on human factors in electronics, (1), 4-11.
- [2] Vidal, J. J. (1973). Toward direct brain-computer communication. Annual Review of Biophysics and Bioengineering, 2(1), 157–180.
- [3] Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45.
- [4] Zander, T. O., & Kothe, C. A. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. Journal of Neural Engineering, 8(2), 025005.
- [5] Krol, L. R., Zander, T. O., Birbaumer, N. P., & Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. Proceedings of the National Academy of Sciences, 113(52), 14898–14903.
- [6] Agentur für Innovation in der Cybersicherheit GmbH (n.d.). *Was macht die Cyberagentur?* Retrieved May 8, 2024, from <https://www.cyberagentur.de/agency/>
- [7] Kapitonova, M., Kellmeyer, P., Vogt, S., & Ball, T. (2022). A framework for preserving privacy and cybersecurity in brain-computer interfacing applications. arXiv preprint arXiv:2209.09653.

# NEUROFEEDBACK PERFORMANCE UNDER CHALLENGING CONDITIONS: THE THETA-AGENCY INTERPLAY

C. Dussard<sup>1</sup>, L. Pillette<sup>2</sup>, C. Dumas<sup>1</sup>, E. Pierrieau<sup>3</sup>, L. Hugueville<sup>1,4</sup>, B. Lau<sup>1</sup>, C. Jeunet<sup>3,\*</sup>, N. George<sup>1,4,\*</sup>

<sup>1</sup>Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France.

<sup>2</sup>Université de Rennes, CNRS, IRISA, UMR 6074, 35000 Rennes, France.

<sup>3</sup>Université de Bordeaux, CNRS, EPHE, INCIA, UMR5287, 33000 Bordeaux, France.

<sup>4</sup>Institut du Cerveau, ICM, Inserm, U1127, CNRS, UMR 7225, Sorbonne Université, CENIR, Centre MEG-EEG, Paris, France

\* co-last author

E-mail: cdussard@ensc.fr

## ABSTRACT

Neurofeedback (NF) consists in training the self-regulation of some target neural activity. Yet, the neural underpinnings of NF performance remains largely unknown. Here, we investigated Motor Imagery (MI) based NF with EEG, training subjects to regulate motor-related activity in the large  $\beta$  (8-30 Hz) band. We examined the electrophysiological correlates of NF performance across the whole scalp and the frequency spectrum. In addition to the rewarded  $\beta$  activity, fronto-central  $\theta$  activity predicted NF performance. The association was modulated by the participants' sense of agency over the feedback with stronger effects in participants with lower agency. Fostering agency in NF protocols may reduce cognitive effort and reliance on additional rhythms beyond  $\beta$ . Considering these effects could be important for optimizing NF performance.

## INTRODUCTION

Neurofeedback (NF) is a cognitive training procedure aiming to train subjects to modulate a specific neural activity, by providing them real-time feedback (FB) on this activity. The premise is that there is a causal link between neural activity and mental (sensory, cognitive, or motor) functions. Mastering control over a targeted activity may improve or restore the corresponding cognitive or sensorimotor ability. However, NF raises several scientific and technical challenges. One of them is the high percentage of non-responders and the variability of NF performance. This raises the question of the neural mechanisms of NF learning and the neural processes associated with NF performance.

Among the possible processes involved, placebo effects [1,2], non-specific training effects [3], and indirect causality effects are debated [4].

In this study, we focused on motor imagery (MI) NF paradigms using electroencephalography (EEG), to train participants to reduce the motor-related  $\beta$  band activity (here considered between 8 and 30 Hz) by imagining right hand movements.

NF learning has been proposed to rely on reinforcement learning, yet this remains untested [5,6]. Psychological factors such as technology acceptance, attention and spatial abilities are known to influence performance [7]. Besides psychological traits, the dynamic cognitive

processes contributing to NF success were seldom investigated. These may entail cognitive control, attentional processes, and reward processing among others. These processes are underpinned by rhythms different from the rewarded  $\beta$  activity, hindering the specificity of the training and complicating the interpretation of behavioral or clinical effects [4]. In this line, recent studies have demonstrated functional connectivity outside of the expected motor networks in MI-based NF protocols [8]. Moreover, some NF studies targeting  $\alpha$ ,  $\beta$  or  $\gamma$  band activities have reported modulations of electrophysiological rhythms beyond the rewarded frequency band [9,10], while others provided evidence for specificity of NF training [11,12].

In this study, we aimed to move past identifying rhythms whose modulation temporally coincided with NF training. We investigated the neural activities associated to successful NF performance beyond the targeted  $\beta$  activity.

Exploratory studies have shown that psychological predictors of NF performance differ between  $\theta$  and  $\beta$  NF paradigms varies [13]. Additionally, experimental factors such as visual or tactile feedback modality modulate differential rhythms [14]. Thus, psychological and experimental factors may modulate the association between non-specific activities and NF performance.

We used an MI-based NF task rewarding the downregulation of activity on central, motor regions in the 8-30 Hz band. Participants were trained with three different FB conditions and we measured their sense of agency i.e. sense of control over the FB. An initial analysis focused on the relationship between agency and NF performance and showed that the subjective sense of agency over FB predicted NF performance [15]. Here, we focused on the electrophysiological correlates of NF performance beyond the rewarded activity and explored activities in the  $\theta$  (3-7Hz),  $\alpha$  (8-12 Hz) and  $\beta$  (13-30Hz) bands across the whole scalp. We also examined whether FB conditions and/or sense of agency interacted with these electrophysiological correlates.

## MATERIALS AND METHODS

We used Dussard et al.'s dataset [15]. Full materials and Methods details are provided in [15].

## Participants

Twenty-three healthy right-handed participants (mean age  $28 \pm 7$  years, 11 women) were included in this study approved by the CPP IdF VI ethics committee. Participants gave written informed consent and received financial compensation after participation.

## Experimental protocol

The experiment consisted in one NF session. Participants performed the MI-NF task with three different FB conditions (Figure 1A).

The FB conditions were either visual, with 1) a pendulum (PENDUL) oscillating to the right, 2) a clenching virtual hand (HAND) or multimodal, with 3) a clenching virtual hand combined with motor illusion vibrations (HAND+VIB).

For each FB condition, participants performed 2 runs of 5 trials (Figure 1B). We presented the FB in separate blocks to avoid the potential cognitive cost of trial-by-trial FB switching. The order of the FB conditions was counterbalanced across participants.

Each NF trial lasted 24s and featured 16 FB movements lasting 1.5s (Figure 1C). Participants were trained and instructed to perform MI at the pace of the rhythmic visual FB movements.

The sense of agency was measured after each run of 5 trials with a 11-point Likert scale in response to the question: “Did you feel like you were controlling the movements of the pendulum/hand?”.

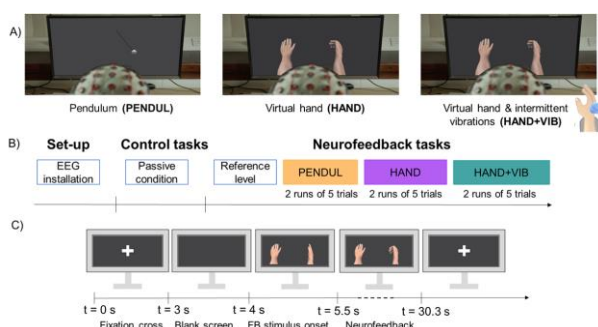


Figure 1. Illustration of the experimental protocol. A. Representation of the different FB conditions, i.e., PENDUL, HAND and HAND+VIB. B. Time course of the experiment. C. Time course of an example trial in the HAND condition

To control for the effect of FB stimulus on EEG activity, participants also underwent control tasks consisting in passive trials where they observed random movements of the pendulum or the virtual hand and eventually received additional vibrations (Figure 1B).

A 2x30s resting state recording with eyes open on a fixation cross established NF reference threshold, based on median 8-30Hz activity on C3 electrode (computed as the Laplacian between C3 and FC1, FC5, CP1, CP5) (Figure 1B). Then, during the NF trials, a 10% reduction of this activity induced positive FB in the form of movement on screen and over 55% reduction triggered maximal amplitude of FB movement (either swing for the pendulum or clenching for the virtual hand). Activity reduction between these upper and

lower limits triggered a linearly proportional movement amplitude. Twenty-one out of 23 participants obtained positive visual FB.

For HAND+VIB FB, vibrations were delivered by a vibrator attached on the right-hand extensor tendons. They were triggered every 6s if the participant maintained an average of 30% reduction in the previous 6s. Tactile FB was intermittent to avoid habituation-related opposite direction movement illusions [16].

## EEG Data acquisition

EEG signal was recorded with an actiChamp Plus system (Brain Products GmbH) using a 32-active electrode cap (ActiCAP snap, Brain Products GmbH). The signal was referenced to Fz electrode. The ground electrode was Fpz. The data were recorded at 1 kHz with a band-pass filter of DC-280Hz. Data were transmitted to OpenViBE 2.2.0.

## Online EEG signal processing

A laplacian filter was computed over the C3 electrode by subtracting the signals from CP5, FC5, CP1 and FC1 electrodes. The signal was epoched into 1s time windows with 0.75s overlap then filtered in the 8-30 Hz band. The signal values were squared and averaged over time in each epoch. These epoch values were streamed to a Unity application using Lab Streaming Layer (LSL) communication protocol. Each FB movement was determined by the mean of four consecutive epochs. This mean was compared to the pre-determined reference threshold. The amount of reduction in  $\beta$  power was conveyed by the amplitude of FB movements.

## Offline EEG signal processing

We performed offline analyses of the event-related desynchronisation/synchronisation (ERD/ERS) during trials with MNE Python. The continuous raw data were filtered with 0.1 Hz high-pass, 90 Hz low-pass, and two zero-phase notch filters (50 and 100 Hz cut-off). The signal was epoched into NF and control trials. We excluded the trials with muscular artifacts. We rejected electrodes around the maxillary regions from analysis due to frequent muscular artifacts. We removed ocular artifacts with independent component analysis. The data were average-referenced and downsampled to 250 Hz. We computed EEG power between 3 and 30 Hz with a Morlet wavelet transform with 1 Hz frequency bins. We averaged the resulting time-frequency data across trials, for each run of each FB condition, in each participant. We normalised power values relative to a 2s fixation cross baseline before the trial onset using a log-ratio. Finally, we averaged the obtained ERD/ERS data across time in each condition.

## Statistical analyses

### ERD/ERS predictors of NF performance

We used a mass univariate approach based on linear mixed-effects regression (LMER) models to explore the relationship between NF performance and ERD/ERS computed between 3 and 30 Hz over the whole scalp. Thus, for each electrode  $i$  and each frequency  $j$ , we computed a model with NF performance as the outcome

variable and the ERD/ERS value at the electrode  $i$  for the frequency  $j$  (ERD\_ERSvalue $_{i,j}$ ) as fixed effect factor. We included runs as a fixed effect covariate and a random intercept of the NF performance across participants.

The models were written in R 4.0.4 with the lme4 package, as follows:  $NF\_performance \sim ERD\_ERS_{i,j} + run + (1 | participant\_id)$

Model parameters were estimated using Restricted Maximum Likelihood and p-values were estimated using Type III ANOVA. Parameter estimates of the fixed effect of ERD/ERS were extracted for each electrode and frequency and tested for significance with false discovery rate (FDR) correction for multiple comparisons applied to the p values ( $n = 756$ : 28 electrodes x 27 frequencies).

This analysis allowed the identification of a fronto-central  $\theta$  activity predictive of NF performance. We investigated further this activity in subsequent analyses.

#### Impact of FB condition on fronto-central $\theta$ activity

To control for the potential effect of vibrations in the HAND+VIB condition, we repeated the initial LMER analysis by excluding the vibration periods from the trial data, before averaging the ERD/ERS data across time. For this, we excluded NF performance values of cycles 4, 8, 12 and 16, which could feature vibrations.

#### Potential modulators of fronto-central- $\theta$ predictor

We tested if FB condition and sense of agency modulated the identified fronto-central  $\theta$  activity. To do so, we averaged the ERD/ERS values that significantly predicted NF performance after FDR-correction on fronto-central electrodes in the  $\theta$  band. We used LMER analysis to assess the interaction between this averaged fronto-central  $\theta$  activity and i) agency, ii) FB condition.

Thus, the model was the following:

$$NF\_performance \sim ERD/ERS\_Theta * Agency + ERD/ERS\_Theta * FB + Run + (1 + FB + Agency + ERD/ERS\_Theta | participant\_id)$$

We chose this random-effects structure to control for Type I error while allowing model convergence.

We ran the same analysis on the left central  $\beta$  cluster (corresponding to the rewarded activity) as a control.

## RESULTS

#### Electrophysiological correlates of NF performance

First, we examined the ERD/ERS patterns that accounted for NF performance by computing LMER models over the scalp and the frequency spectrum.

Decreased power in the rewarded 8-30Hz band over the left central regions was positively associated with NF performance (Figure 2A). This was expected since our design trained participants to reduce this activity and FB (aka. NF performance) was computed on the basis of 8-30 Hz band activity on C3 through OpenVibe. This effect seemed particularly marked in the 13-30 Hz band and extended bilaterally; it extended on parietal electrodes in the high- $\beta$  band. Such activity is typical of MI task [17].

Moreover, power in the low  $\theta$  band (3-4Hz) over fronto-central regions was positively associated with NF performance: increased fronto-central low  $\theta$  power predicted higher NF performance (Figure 2A and 2B).

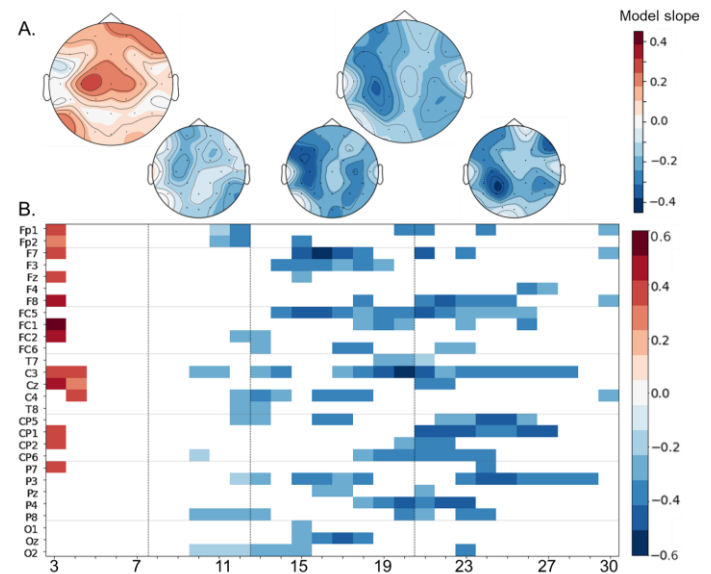


Figure 2. Parameter estimates of the effect of ERD/ERS on NF performance across electrodes and frequencies. Blue indicates negative estimates, that is, decreased power predicts higher NF performance. Red color indicates positive estimates increased power predicts higher NF performance. A. Topographic maps of the parameter estimates averaged in different frequency bands. Topmost maps: left map for 3-7 Hz and right map for 8-30 Hz. Lower maps depict from left to right: 8-12 Hz, 13-20 and 21-30 Hz. B. Electrode-frequency representation of parameter estimates, with electrodes in ordinate (from frontal electrodes on the top to occipital electrodes on the bottom) and frequencies in abscissa. Only statistically significant parameter estimates at  $p < .05$  with FDR correction are displayed.

We then focused on this fronto-central  $\theta$  pattern, which stood out as it was not rewarded in our NF protocol and is not typically associated with MI.

#### Impact of FB condition on fronto-central $\theta$ activity

First, we investigated if this fronto-central  $\theta$  activity was influenced by the FB condition. Our previous findings showed vibration-locked patterns of  $\theta$  synchronisation in the time-frequency representations [16] (see Figure 3A and 3B).

If  $\theta$  activity was a byproduct of the vibratory FB, this could confound our result since the vibrations were by design associated with successful NF performance.

The effect of the vibratory FB on the  $\theta$  band was further illustrated by displaying topographical maps of  $\theta$  activity averaged over the time periods of the vibrations, during the NF trials (Figure 3A top) and the passive, control trials (Figure 3B, top).

The fronto-central pattern  $\theta$  ERS was concomitant of the vibrations. Yet, it was somewhat more central in the passive condition (Figure 3B, top).

The  $\theta$  ERS was short-lived, lasting  $\sim 0.5s$  of the 2s vibration duration. Its amplitude decreased from the first to last vibration in the passive condition.

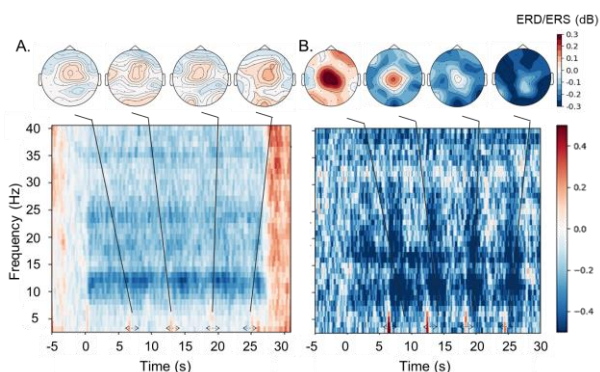


Figure 3. A. Grand average data of the NF HAND+VIB trials. B. Grand average data of the passive condition with virtual hand and vibration stimuli. Top: Topographical maps of the 3-7Hz ERD/ERS during the vibration periods. Bottom: Time-frequency representation of ERD/ERS on C3 electrode.

As an additional control, we re-ran our LMER analysis by excluding the vibration periods in the HAND+VIB trials. The results remained unchanged. This suggests that the fronto-central  $\theta$  effect on NF performance was not attributable to a confounding effect of the vibratory FB.

#### Potential modulators of fronto-central $\theta$ predictor

We then investigated whether FB conditions and/or sense of agency modulated the association between the fronto-central  $\theta$  activity and NF performance. Indeed, our original analysis showed that subjective sense of agency over FB was a significant predictor of NF performance [16].

For this analysis, we extracted and averaged the ERD/ERS values on the electrodes and the low  $\theta$  frequencies (3-4Hz) where a significant effect on NF performance had been found. We then ran an LMER analysis taking into account the effects of FB condition and agency and their potential interaction with the fronto-central  $\theta$  effect.

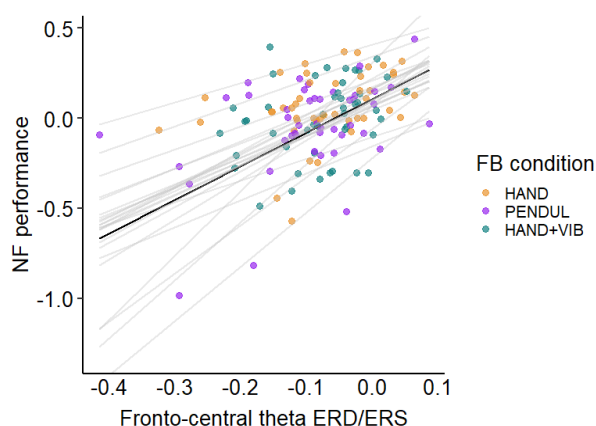


Figure 4. Relation between NF performance and fronto-central  $\theta$  activity. Individual data for each FB condition (HAND in orange, PENDUL in purple and HAND+VIB in turquoise). Thin gray lines represent the individual random slopes and intercepts of the effect of fronto-central  $\theta$  on NF performance. The black thick line represents the estimated fixed effect of fronto-central  $\theta$  on NF performance.

This analysis showed when accounting for the effects of FB condition and agency, the main effect of fronto-central  $\theta$  on NF performance remained significant (parameter estimate: = 1.87, 95% CI [0.93, 2.81];  $F(1, 71.9) = 18.8$ ,  $p < 0.001$ ) (Figure 4). Thus, fronto-central  $\theta$  activity was neither a mere electrophysiological correlate of sense of agency nor of FB condition. There was no significant interaction between fronto-central  $\theta$  and FB conditions ( $F(2, 81.4) = 0.55$ ,  $p = 0.58$ ). Therefore, fronto-central  $\theta$  activity seemed associated with NF performance regardless of FB condition.

In contrast, there was a significant interaction between fronto-central  $\theta$  and sense of agency (parameter estimate = -0.22, 95% CI [-0.34, -0.10];  $F(1, 93.3) = 12.5$ ,  $p < 0.001$ ). To illustrate this interaction, we represented NF performance as a function of fronto-central  $\theta$  values, splitting the data according to different scores of agency (1<sup>st</sup> tercile [0-3] in red, 2<sup>nd</sup> tercile [4,7] in brown and 3<sup>rd</sup> tercile of agency scores [8-10] in green, in Figure 5). We represented the model predictions by plotting the estimated slopes of fronto-central  $\theta$  effect on NF performance at three fixed agency values (2, 6 and 8) (colored lines in Figure 5).

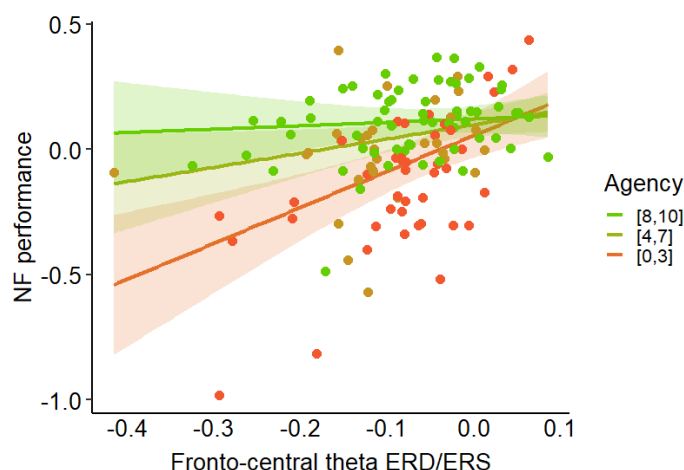


Figure 5. Relation between NF performance and fronto-central  $\theta$  activity as a function of agency. Dots represent the individual data for each run and each FB condition, colored by agency bin (1<sup>st</sup> tercile in red-orange, 2<sup>nd</sup> tercile in light brown, 3<sup>rd</sup> tercile in green). The colored lines represent the estimated slopes of fronto-central  $\theta$  effect on NF performance for the median agency values of each tercile (agency values of 2, 6 and 8), with shaded areas indicating 95% confidence intervals around these slopes.

This showed a positive correlation between NF performance and fronto-central  $\theta$  for low scores of agency. With higher scores of agency, the slope declined gradually, reflecting that NF performance was less associated to fronto-central  $\theta$ .

The same analysis run on the C3  $\beta$  cluster showed a significant main effect of  $\beta$  on NF performance ( $F(1, 53.9) = 8.03$ ,  $p = 0.006$ ). In contrast to the fronto-central  $\theta$  activity, we did not find any interaction between FB and C3  $\beta$  ( $F(2, 57.9) = 1.86$ ,  $p = 0.17$ ) or between sense of agency and  $\beta$  ( $F(1, 79.5) = 1.47$ ,  $p = 0.23$ ).

## DISCUSSION

This study investigated the electrophysiological correlates of NF performance in an MI-based EEG NF protocol. We found significant associations between NF performance and ERD/ERS patterns in the central motor region in the rewarded  $\beta$  band. Additionally, fronto-central  $\theta$  activity consistently predicted NF performance. Diving deeper, we identified an interaction between this fronto-central  $\theta$  activity and sense of agency. NF performance was more strongly associated with fronto-central  $\theta$  activity in participants with lower agency.

### *EEG patterns of NF performance*

Our results showed that NF performance was associated with power reduction in the rewarded 8-30Hz range in left central regions. Left central  $\mu$ - $\beta$  desynchronisation is reliably associated with right-hand MI-BCI performance [18,19]. Our study extends these findings by examining performance correlates across different FB conditions in a within-subject NF design.

In addition to the rewarded left central  $\beta$  activity, fronto-central  $\theta$  power predicted NF performance. In line with this result, both pre-cue [20,21] and on-task  $\theta$  power [22] were shown to predict BCI performance.

Frontal  $\theta$  activity notably emerges in response to perceived conflict [23] such as negative FB in NF. Negative FB processing would enable for adjustment of the NF strategy towards better NF performance. In line with this interpretation, an  $\alpha$  NF study reported that  $\theta$  activity differentiated sham from NF participants. This was seen as a constancy of conflict/error-prediction signals in the sham group, while the NF group reduced conflict by improving NF performance [24].

MI modulates frontal  $\theta$  activity more than motor execution, highlighting increased mental effort [25]. Specifically, frontal  $\theta$  shows higher involvement in kinesthetic than visual MI [26]. Mental demand has been correlated to a  $\theta$ - $\beta$  combination in a BCI paradigm [27]. Studying within-trial dynamics of  $\theta$  activity and NF performance may reveal if  $\theta$  activity arises locally in response to negative FB or if it reflects overall attention and cognitive effort.

### *Impact of FB condition on fronto-central $\theta$ activity*

Vibrations in the HAND+VIB condition were associated with fronto-central  $\theta$  synchronisation in the NF tasks. Midfrontal  $\theta$  oscillations have been proposed to encode the value of tactile delay [28]. This is especially relevant because the vibrations were delivered following the integration of a 6s time segment i.e. they constituted a delayed, asynchronous FB. In contrast, during the passive task,  $\theta$  patterns were more central, potentially reflecting sensory processing. Vibrations modulate activity on central  $\theta$  activity [29]. However, the relationship between NF performance and fronto-central  $\theta$  activity remained significant after excluding vibration segments from the analysis.

### *Agency is a modulator of the fronto-central- $\theta$ predictor*

Both sensorimotor  $\beta$  [30] and fronto-central  $\theta$  [31, 32] relate to sense of agency. Yet, our analysis showed that

both  $\theta$  and  $\beta$  predictors remained significant after accounting for agency.

Further analysis showed that sense of agency modulated the association between  $\theta$  and NF performance: a stronger association was observed in participants with low sense of agency. We suggest that participants reporting high sense of agency may have required less mental effort to achieve NF performance. Our findings resonate with recent reinforcement learning research, which reports differences in  $\theta$  and  $\beta$  in response to positive and negative FB between agent and passive participants [33]. Negative FB elicited more fronto-central  $\theta$  activity than positive FB. Crucially, the difference between positive and negative FB more pronounced difference in agent participants. This contrasts with our observations that  $\theta$  dynamics in the NF context were heightened in participants with a low sense of agency. Altogether, these results suggest a potential role for sense of agency in shaping strategies employed during reinforcement learning tasks, including NF.

## CONCLUSION

This study shed some light on the neurophysiological correlates of NF performance, highlighting the role of  $\theta$  activity and its interaction with sense of agency. All in all, sense of agency allows for better performance, associated with a more specific pattern of modulation. It is important to consider that participants can mobilize different processes to manage NF performance. It is key to control for the activities that are modulated with NF training as they may contribute to non-specific effects. Monitoring such activities could be important for some clinical applications of NF where neurophysiological specificity is of crucial importance.

## REFERENCES

- [1] Thibault, R. T., Lifshitz, M., & Raz, A. (2017). Neurofeedback or neuroplacebo? *Brain*, 140(4), 862-864.
- [2] Kober, S. E., Witte, M., Grinschgl, S., Neuper, C., & Wood, G. (2018). Placebo hampers ability to self-regulate brain activity: a double-blind sham-controlled neurofeedback study. *Neuroimage*, 181, 797-806.
- [3] Sorger, B., Scharnowski, F., Linden, D. E., Hampson, M., & Young, K. D. (2019). Control freaks: Towards optimal selection of control conditions for fMRI neurofeedback studies. *Neuroimage*, 186, 256-265.
- [4] Kvamme, T. L., Ros, T., & Overgaard, M. (2022). Can neurofeedback provide evidence of direct brain-behavior causality? *Neuroimage*, 258, 119400.
- [5] Strehl, U. (2014). What learning theories can teach us in designing neurofeedback treatments. *Front. Hum. Neurosci.*, 8, 894.
- [6] Lubianiker, N., Paret, C., Dayan, P., & Hendler, T. (2022). Neurofeedback through the lens of reinforcement learning. *Trends Neurosci.*45(8), 579-593.
- [7] Jeunet, C., N'Kaoua, B., & Lotte, F. (2016). Advances in user-training for mental-imagery-based

BCI control: Psychological and cognitive factors and their neural correlates. *Prog. Brain Res.*, 228, 3-35.

[8] Corsi, M. C., Chavez, M., Schwartz, D., George, N., Hugueville, L., Kahn, A. E., ... & Fallani, F. D. V. (2020). Functional disconnection of associative cortical areas predicts performance during BCI training. *NeuroImage*, 209, 116500.

[9] Dessy, E., Mairesse, O., Van Puyvelde, M., Cortoos, A., Neyt, X., & Pattyn, N. (2020). Train your brain? Can we really selectively train specific EEG frequencies with neurofeedback training. *Front. Hum. Neurosci.*, 14, 22.

[10] Chikhi, S., Matton, N., Sanna, M., & Blanchet, S. (2023). Mental strategies and resting state EEG: Effect on high alpha amplitude modulation by neurofeedback in healthy young adults. *Biol. Psychol.*, 178, 108521.

[11] Kober, S. E., Witte, M., Neuper, C., & Wood, G. (2017). Specific or nonspecific? Evaluation of band, baseline, and cognitive specificity of sensorimotor rhythm-and gamma-based neurofeedback. *Int. J. Psychophysiol.*, 120, 1-13.

[12] Grosselin, F., Breton, A., Yahia-Cherif, L., Wang, X., Spinelli, G., Hugueville, L., ... & George, N. (2021). Alpha activity neuromodulation induced by individual alpha-based neurofeedback learning in ecological context: A double-blind randomized study. *Sci. Rep.* 11(1), 18489.

[13] Kikkert, A. (2015). Predictors of neurofeedback efficacy: An exploratory study to the influence of personality and cognitive characteristics on the efficacy of theta and beta neurofeedback training. Masters, Psychology, Leiden University.

[14] Vukelić, M., & Gharabaghi, A. (2015). Oscillatory entrainment of the motor cortical network during motor imagery is modulated by the feedback modality. *Neuroimage*, 111, 1-11.

[15] Dussard, C., Pillette, L., Dumas, C., Pierrieau, E., Hugueville, L., Lau, B., ... & George, N. (2024). Agency accounts for the effect of FB transparency on motor imagery neurofeedback performance. *bioRxiv*.

[16] Taylor, M. W., Taylor, J. L., & Seizova-Cajic, T. (2017). Muscle vibration-induced illusions: review of contributing factors, taxonomy of illusions and user's guide. *Multisensory res.*, 30(1), 25-63.

[17] McFarland, D. J., Miner, L. A., Vaughan, T. M., & Wolpaw, J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.*, 12, 177-186.

[18] Vasilyev, A., Liburkina, S., Yakovlev, L., Perepelkina, O., & Kaplan, A. (2017). Assessing motor imagery in brain-computer interface training: psychological and neurophysiological correlates. *Neuropsychologia*, 97, 56-65.

[19] Rimbart, S., & Lotte, F. (2022). ERD modulations during motor imageries relate to users' traits and BCI performances. 44th Annual International Conference of the IEEE (EMBC).

[20] Bamdadian, A., Guan, C., Ang, K. K., & Xu, J. (2014). The predictive role of pre-cue EEG rhythms on MI-based BCI classification performance. *J. Neurosci. Methods*, 235, 138-144.

*Methods*, 235, 138-144.

[21] Kang, J. H., Youn, J., & Kim, J. (2021). Effects of frontal theta rhythms in a prior resting state on the subsequent motor imagery brain-computer interface performance. *Front. Neurosci.*, 15, 663101.

[22] Trambaiolli, L. R., Dean, P. J., Cravo, A. M., Sterr, A., & Sato, J. R. (2019). On-task theta power is correlated to motor imagery performance. In 2019 IEEE International Conference (SMC) (pp. 3937-3942).

[23] Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends Neurosci.*, 37(9), 480-490.

[24] Davelaar, E. J., Eatough, V., Etienne, M., & Ozolins, C. (2018). Mid-frontal theta oscillations discriminate between sham-control and neurofeedback training manipulations: a signal detection analysis. *Curr. Neurobiol.*, 9(3), 95-100

[25] Van der Lubbe, R. H., Sobierajewicz, J., Jongsma, M. L., Verwey, W. B., & Przekoracka-Krawczyk, A. (2021). Frontal brain areas are more involved during motor imagery than during motor execution/preparation of a response sequence. *Int. J. Psychophysiol.*, 164, 71-86.

[26] Menicucci, D., Di Gruttola, F., Cesari, V., Gemignani, A., Manzoni, D., & Sebastiani, L. (2020). Task-independent electrophysiological correlates of motor imagery ability from kinaesthetic and visual perspectives. *Neuroscience*, 443, 176-187.

[27] Vourvopoulos, A., & Bermúdez i Badia, S. (2016). Motor priming in virtual reality can augment motor-imagery training efficacy in restorative brain-computer interaction: a within-subject analysis. *J Neuroeng Rehabil*, 13, 1-14.

[28] Alsuradi, H., Park, W., & Eid, M. (2022). Midfrontal theta power encodes the value of haptic delay. *Sci. Rep.*, 12(1), 8869.

[29] Altamira, G. H., Fleck, S., Lécuyer, A., & Bougrain, L. (2023). EEG Modulations Induced by a Visual and Vibrotactile Stimulation. In 2023 IEEE SMC

[30] Buchholz, V. N., David, N., Sengemann, M., & Engel, A. K. (2019). Belief of agency changes dynamics in sensorimotor networks. *Sci. Rep.*, 9(1), 1995.

[31] Jeunet, C., Albert, L., Argelaguet, F., & Lécuyer, A. (2018). "Do you feel in control?": towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE transactions on visualization and computer graphics*, 24(4), 1486-1495.

[32] Zito, G. A., de Sousa Ribeiro, R., Kamal, E., Ledergerber, D., Imbach, L., & Polania, R. (2023). Self-modulation of the sense of agency via neurofeedback enhances sensory-guided behavioral control. *Cereb. Cortex.*, 33(24), 11447-11455.

[33] Weismüller, B., Kullmann, J., Hoenen, M., & Bellebaum, C. (2019). Effects of feedback delay and agency on feedback-locked beta and theta power during reinforcement learning. *Psychophysiology*, 56(10), e13428.

# PERIPHERAL NERVE STIMULATION AND AUDITORY SIMULATION CLOSED LOOP SYSTEM FOR SENSORY DECISION MAKING IN TRANSHUMERAL AMPUTEES

G. Soghoian<sup>1</sup>, A.R. Biktimirov<sup>2,3</sup>, N.S. Piliugin<sup>1</sup>, M.Y. Sintsov<sup>3</sup>, M.A. Lebedev<sup>4,5</sup>

<sup>1</sup> Vladimir Zelman Center for Neurobiology and Brain Rehabilitation, Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>2</sup> Laboratory of Experimental and Translational Medicine, School of Biomedicine, Far Eastern Federal University, Vladivostok, Russia

<sup>3</sup> Research Center of Motorica LLC, Moscow, Russia

<sup>4</sup> Moscow State University, Moscow, Russia

<sup>5</sup> Sechenov Institute of Evolutionary Physiology and Biochemistry, Saint Petersburg, Russia

E-mail: gsogoyan98@gmail.com

## ABSTRACT:

Peripheral nerve stimulation (PNS) is a key method for restoring sensory feedback in upper-limb prostheses, yet the necessity of invasive feedback for sensory decision-making remains uncertain. In this study, two transhumeral amputees underwent sensory restoration in their phantom limbs via PNS. They performed an active exploration task using a tablet and closed-loop feedback system to assess their sensory decision-making abilities. In the task patient needed to differentiate among three hidden objects using PNS-based tactile feedback or auditory feedback. Interestingly, one patient successfully completed the task only in PNS trials, while the other demonstrated improved speed and accuracy with auditory stimulation. These findings suggest varying responses to different feedback modalities in different subjects. They indicate the potential significance of personalized approaches in designing sensory feedback systems for prosthetic users.

## INTRODUCTION

Phantom limb pain (PLP) affects 80% of individuals who have undergone amputations [1,2] and can be managed through neuromodulation techniques like peripheral nerve stimulation [3]. These methods not only help reduce phantom limb pain but also have potential applications in enhancing sensory feedback for neuroprosthetics [4].

Invasive and noninvasive methods for restoring somatic sensations have their advantages and disadvantages. For instance, vibromotors have a limited stimulating range, leading to restricted sensations [5]. On the other hand, the higher spatial resolution can be achieved with invasive techniques compared to noninvasive methods. Neurostimulation systems integrated as part of bidirectional brain computer interface (BCI) enable the discrimination of object size and texture, improving

prosthesis embodiment and enhancing motor control [4,6].

Nevertheless, the necessity of invasive feedback for sensory decision-making remains uncertain. This question was explored in our previous study [4], where we demonstrated that prosthetic systems utilizing transcutaneous electrical nerve stimulation (TENS) may offer comparable efficacy to PNS-based systems. Additionally, feedback in bidirectional BCI can be delivered through alternative sensory modalities such as auditory cues [10]. Although these systems do not elicit tactile sensations in the phantom or residual limb, they may reduce cognitive load [11] and enhance performance [9]. Despite extensive research on various forms of sensory feedback, it remains unvalidated whether invasively delivered somatotopically matched feedback can augment sensory decision making in amputees compared to auditory stimulation.

To address this question, we conducted a study with two transhumeral amputees completing a sensory decision-making task under two conditions. In a part of trials, they relied on auditory feedback and in the other part they relied on a PNS-based feedback that projected to their phantom limb as somatic sensations. One patient successfully completed the task relying on PNS feedback, whereas the second patient exhibited greater speed and accuracy when utilizing auditory stimulation.

## MATERIALS AND METHODS

Two individuals with amputations took part in the research, both experiencing phantom limb pain (PLP). The study received approval from the Ethical Committee of the Biomedicine School at Far East Federal University (FEFU) under Protocol #4 on April 16, 2021. Prior to their involvement in the experiments, each patient provided informed consent. The study is registered as a clinical trial on <https://clinicaltrials.gov/>

under #NCT05650931.

Participants S12 and S13 underwent the electrode implantation in their left residual limb on a level of shoulder. The implantation surgeries were conducted at the Medical Center of FEFU. Eight-contact electrodes (Directional Lead for the St. Jude Medical Infinity™ DBS System; Abbot; USA) were implanted in the median nerve of all patients while under endotracheal anesthesia.

After the surgery, both patients underwent though the sensory mapping procedure where the electrode sites that caused sensations in phantom hand were discovered. The details about sensory mapping protocol can be found in a subsequent research [4]. Since sensations evoked by stimulation were stable among seven days, both patients were involved in the experiment to access their ability to sensory decision making. During the sensory mapping, S12 tended to report that his sensations were of high naturalness.

In our previous study we introduced an active exploration task where participants explored invisible objects using artificial tactile sensations provided by TENS and PNS. Here, we used a similar protocol, but with the use of auditory feedback instead of TENS. Thoroughly, patients used their intact limb to scan the tablet surface using a stylus and searched for an invisible object. Whenever the stylus made contact with the object, sensory feedback was provided through sound or electrical stimulation. The objects could be in the shape of a square, circle, or pentagon and they were randomly selected for each trial along with two types of sensory feedback: PNS or Auditory. To increase the difficulty of object recognition, the rotation angle of the object varied for each trial (Fig. 1).

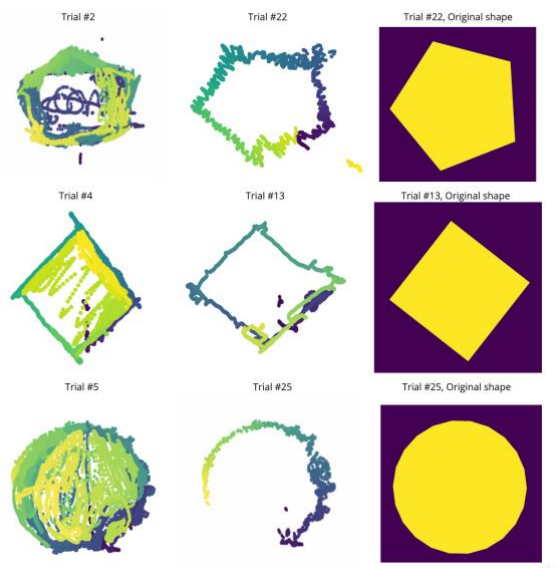


Figure 1. Different shapes of active exploration task.

For trials with auditory feedback, each time the stylus touched an invisible object, 1000-Hz sound was turned on. For PNS trials, two modes of stimulation were employed, namely baseline and target, with identical frequency and pulse width but varying amplitudes. The

amplitude in the baseline mode remained below the sensory threshold while it surpassed the threshold in the target mode. The NimEclipse stimulator was linked to the laptop, which was connected to the tablet. Upon tapping the screen area that corresponded to the invisible shape, the Python script transitioned from baseline to target stimulation mode, thereby enabling the subject to perceive the shape. S12 and S13 utilized a stylus held in the intact limb to interact with the tablet. For both S12 and S13, stimulation settings were chosen individually to elicit tactile sensations in the fingers of the phantom limb.

The active-exploration sessions were held on post-surgery days 8 and 20 for each subject. On each of these days, the experiment consisted of two sessions: a learning session and an evaluation session. During the learning session, participants saw the history of their touches of the screen as black lines. For learning session completion, it was required that the subject to correctly guess each object twice for both auditory and PNS feedback. Each time when an object was recognized correctly, the respectful trial was eliminated from the list of unguessed yet trials. Next trial was randomly selected from that list.

An analysis of variance (ANOVA) was performed to compare the trial durations of patients based on two factors: the feedback type (PNS or Auditory) and the experimental day (day 8 or day 20) for each subject separately.

## RESULTS

Both subjects could complete the following task using both types of feedback. Patients completed the active exploration task by scanning the tablet with a stylus. During the first trials, subjects attempted to differentiate between objects attempting to draw the entire figure; the trajectory of their movements can be seen in the first column of Fig. 1. Since the experiment operator did not instruct patients on the best approach to resolve this task, it took several trials for both patients to discover the so-called “border strategy”. In this strategy it is expected that a participant discovers objects’ border by detecting the moments of stimulation on and stimulation off switch.

During day 8, both subjects met the requirements for completing the learning part of the task. S13 needed 7 trials to complete the training with only one mistake made (Fig. 2a). S12 completed the learning session after performing 20 trials. During the evaluation part of the session: S12 performed with an accuracy of 17%, and S13 with an accuracy of 67%. The chance level was 33%.

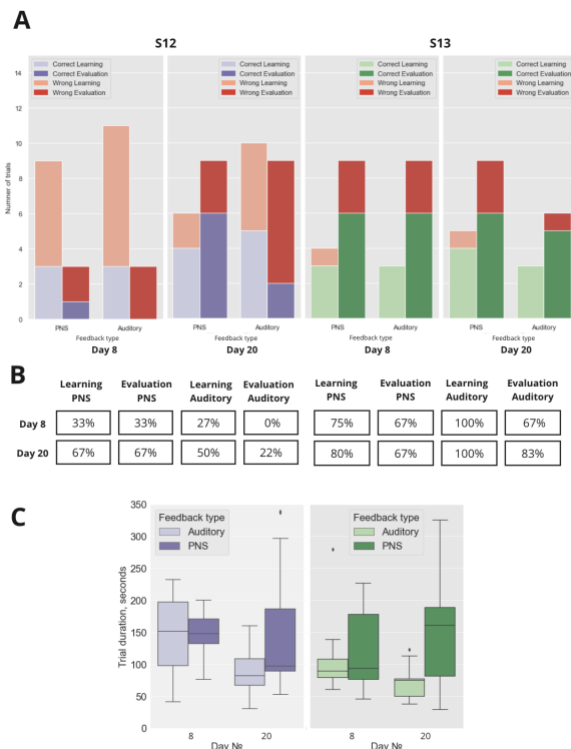


Figure 2. Active exploration task results. (A) - Each panel represents number of trials conducted during day 8 and day 20 for subjects S12 and S13. Bars in each panel represent the number of correctly recognized objects (depicted in purple and green) and erroneously recognized objects (depicted in red) under both PNS and Auditory feedback conditions. Paired bars differentiate the number of trials in learning and evaluation sessions, as S12 and S13 required different numbers of trials to transition to the evaluation session. (B) - Percentage of correctly recognized objects in each experimental session (C) - Variation of trial duration for PNS and Auditory conditions during sessions in day 8 and day 20

During day 20, S12 completed the training session in 16 trials. S13 completed the training session in 8 trials. During the testing session, performance improved in both subjects compared to day 8. Performance accuracy was 44% and 75% in S12 and S13, respectively. In S12, accuracy was 22% when using auditory feedback and 67% when using PNS feedback, while in S13 accuracy made up 67% and 83% respectively (Fig. 2b). S12 mentioned that he felt as if his phantom limb touched the table screen when the PNS-based feedback was used.

We conducted a comparison of the number of seconds taken by patients to complete the trial (trial duration) using ANOVA, considering two factors: the type of feedback (PNS vs Auditory) and the day of the experiment (day 8 vs day 20). The analysis for S12 revealed near significant difference among means of trial duration for two feedback types ( $F_{(1)}=4.167568$ ;  $p_{\text{value}}=0.045927$ ; two-way ANOVA), for different days ( $F_{(1)}=3.646682$ ;  $p_{\text{value}}=0.061309$ ; two-way ANOVA) and for the factor interaction ( $F_{(1)}=3.393133$ ;  $p_{\text{value}}=0.070766$ ; two-way ANOVA) (Fig. 2c). For S13 the mean trial time was different for

different feedback types ( $F_{(1)}=4.681291$ ;  $p_{\text{value}}=0.035964$ ; two-way ANOVA) and day-feedback type interaction ( $F_{(1)}=3.345151$ ;  $p_{\text{value}}=0.074186$ ; two-way ANOVA), no significant difference was observed for different days ( $F_{(1)}=0.008421$ ;  $p_{\text{value}}=0.927299$ ; two-way ANOVA).

After the pairwise comparison analysis, we revealed that S12 needed less time to complete auditory feedback trials during the day 20 in comparison with auditory feedback trials in day 8 (Mean diff.=-58.4792 ;  $p_{\text{adj}}=0.0495$ ; Tukey HSD); PNS trials in day 20 (Mean diff.=59.4498 ;  $p_{\text{adj}}=0.039$ ; Tukey HSD); PNS trials on day 8 (Mean diff.=-57.5598 ;  $p_{\text{adj}}=0.0719$ ; Tukey HSD). Similarly, S13 completed auditory feedback of day 20 faster than in PNS trials of day 20 (Mean diff.=-73.937;  $p_{\text{adj}}=0.0354$ ; Tukey HSD).

Eventually, S12 had lower accuracy of object recognition in auditory trials, despite the increased speed. By contrast, S13, having a shorter duration of auditory trials, completed the active exploration task with the higher accuracy in both types of feedback.

## DISCUSSION

In this study sensations in phantom limb of two transhumeral amputees were restored with the use peripheral nerve stimulation. To estimate their capabilities of sensory decision making, they completed active exploration task with the use of tablet and closed loop feedback system. One of patient was able to complete the task only with the use of PNS feedback, while the second one was faster and more accurate when used auditory stimulation.

In a previous study, it was demonstrated that active exploration tasks can be performed with comparable accuracy using PNS and TENS feedback [4]. In this experiment, under similar conditions, it was found that these tasks could be successfully completed with auditory feedback. Notably, participant S13 exhibited higher accuracy in shape recognition with auditory stimulation, completing trials more quickly compared to PNS feedback trials.

In the same exploration task, another participant, S12, achieved a higher score in PNS trials than in auditory feedback trials. While the study was limited to two patients, indicating caution in drawing broad conclusions, a distinct difference between the two subjects was observed. This variance may be attributed to two main factors. Firstly, S12, who performed better in PNS trials, likely had greater familiarity with PNS stimulation due to its inclusion in their treatment regime. Secondly, S12 perceived PNS as more natural during sensory mapping. Though, these two observations can be associated, because previously it was shown that long-term PNS stimulation in neuroprosthetics has been associated with an enhanced sense of naturalness [7].

Since PNS is a primary method for sensory restoration in upper-limb prostheses [6], it is of great interest to understand the benefits and limitations of PNS

compared to other stimulation approaches. PNS can evoke natural tactile feedback, enhance embodiment in upper-limb prosthetic devices, and alleviate PLP [4,6,8]. These preliminary findings suggest individual variability in response to different feedback modalities, albeit within the constraints of a small sample size. They underscore the potential importance of personalized approaches in designing sensory feedback systems of bidirectional BCI systems and prosthetic users particularly, while acknowledging the need for further research with larger and more diverse cohorts

## CONCLUSION

PNS is an efficient approach to provide feedback to amputees in sensory decision-making tasks and allow them to differentiate between different objects relying on tactile information. However, at least for some patients, alternative sensory feedback devices could offer upper-limb amputees the opportunity for feature recognition without the need for surgery and associated risks.

## ACKNOWLEDGMENTS

This study was supported by the Russian Science Foundation under grant no. 21-75-30024.

## REFERENCES

- [1] De Nunzio, et al. (2018). Relieving phantom limb pain with multimodal sensory-motor training. *Journal of Neural Engineering*, 15(6), 066022.
- [2] Flor, H. (2002). Phantom-limb pain: characteristics, causes, and treatment. *Lancet Neurology*, 1(3), 182–189.
- [3] Kumar, K., & Rizvi, S. (2014). Historical and present state of neuromodulation in chronic pain. *Current Pain and Headache Reports*, 18(1), 387.
- [4] Soghoyan, G., Biktimirov, A., Matvienko, Y., Chekh, I., Sintsov, M., & Lebedev, M. A. (2023). Peripheral nerve stimulation enables somatosensory feedback while suppressing phantom limb pain in transradial amputees. *Brain Stimulation*, 16(3), 756–758.
- [5] Muijzer-Witteveen, H., Guerra, F., Sluiter, V., & van der Kooij, H. (2016). Pneumatic Feedback for Wearable Lower Limb Exoskeletons Further Explored. *Haptics: Perception, Devices, Control, and Applications*, 90–98.
- [6] Raspopovic, S., Valle, G., & Petrini, F. M. (2021). Sensory feedback for limb prostheses in amputees. *Nature Materials*, 20(7), 925–939.
- [7] Cuberovic, I., Gill, A., Resnik, L. J., Tyler, D. J., & Graczyk, E. L. (2019). Learning of Artificial Sensation Through Long-Term Home Use of a Sensory-Enabled Prosthesis. *Frontiers in Neuroscience*, 13, 853.
- [8] Soghoyan, G., Sintsov, M., Biktimirov, A., Chekh, I., & Lebedev, M. (2022, September). Peripheral nerve stimulation for tactile feedback and phantom limb pain suppression. In *2022 Fourth International Conference Neurotechnologies and Neurointerfaces (CNN)* (pp. 162-164). IEEE.
- [9] J. W. Sensinger and S. Dosen, “A Review of Sensory Feedback in Upper-Limb Prostheses From the Perspective of Human Motor Control,” *Front. Neurosci.*, vol. 14, p. 345, Jun. 2020.
- [10] P. Svensson, U. Wijk, A. Björkman, and C. Antfolk, “A review of invasive and non-invasive sensory feedback in upper limb prostheses,” *Expert Rev. Med. Devices*, vol. 14, no. 6, pp. 439–447, Jun. 2017.
- [11] J. Gonzalez, H. Soma, M. Sekine, and W. Yu, “Psycho-physiological assessment of a prosthetic hand sensory feedback system based on an auditory display: a preliminary study,” *J. Neuroeng. Rehabil.*, vol. 9, p. 33, Jun. 2012.

# VALIDATING NEUROPHYSIOLOGICAL PREDICTORS OF BCI PERFORMANCE ON A LARGE OPEN SOURCE DATASET

D.Trocellier<sup>1,2</sup>, B. N’Kaoua<sup>3</sup>, F. Lotte<sup>1,2</sup>

<sup>1</sup> Inria center at the University of Bordeaux, Talence, France

<sup>2</sup> LabRI (CNRS - Univ. Bordeaux - Bordeaux INP), Talence, France

<sup>3</sup> Université de Bordeaux, Bordeaux, France

E-mail: david.trocellier@inria.fr

## ABSTRACT

Brain-computer interfaces (BCI) are systems that process brain activity to decode specific commands from it such as motor imagery patterns generated when users imagine movements. Despite the growing interest in BCI, they present significant challenges, notably in decoding distinct neural patterns, due to considerable variability across and within users. The literature showed that various predictors were correlated with subject’s BCI performance. Among these indicators, neurophysiological predictors appeared to be the most effective, although studies generally involved small samples and results were not always replicated, thus questioning their reliability. In our study, we used a large dataset with 85 subjects to analyse the relationship between different predictors identified in the literature and BCI performance. Our findings reveal that only four of the six predictors tested could be replicated on this dataset. These results underscore the necessity of validating literature findings to ensure the reliability and applicability of such predictors.

## INTRODUCTION

Brain-computer interfaces (BCI) are devices that measure brain activity, typically through electroencephalography (EEG), to extract specific commands for computer input. A particularly widespread paradigm of BCI is motor imagery (MI), which involves decoding EEG patterns characteristic of movement imagination (typically signals from 8-13 Hz located over the motor and sensory motor cortices). MI leads to patterns similar to motor execution, characterized by an event-related desynchronization (ERD), i.e, a diminution of the amplitude, within the mu band (8-13 Hz) and beta band (13-30 Hz) of the contralateral sensory motor cortex, followed by an event-related synchronization (ERS), i.e an augmentation of the amplitude of the signal within the beta band after the imagined movement ends. A primary limitation of current MI-BCI technology is decoding accuracy. Producing clear neurophysiological signals that can be decoded by existing classification algorithms is not a competence that all BCI users have. In their article, Blankertz et al. [1] demonstrate that using a common spatial pattern (CSP) with a

linear discriminant analysis (LDA) classifier, the average BCI accuracy is  $74.4 \% \pm 16.5 \%$ , with significant variability among participants, ranging from perfect classification (100 %) to performance equivalent to chance level (50 %). More recently, Dreyer et al. [2] published a large database of 87 first-time BCI users performing MI-BCI, reporting a mean accuracy of 63.53 % with a large variability of performances (std = 17.61 %). This variability may be due to users’ inability to produce clear and distinguishable patterns that are strong enough to be classified by current algorithms. It is considered that for effective BCI control, performance should exceed 70 % [3]: users below this threshold are deemed "BCI illiterate" or the BCI "BCI deficient". Understanding the parameters explaining differences in user control of such devices has been an important research question for the past 20 years. A better comprehension of those predictors is essential for developing better BCIs, e.g., to later identify the best BCI type for each user or to create BCIs that consider those predictors in their design and into classification algorithms. The literature identifies a broad spectrum of predictors that can be categorized into four main groups: personality traits, cognitive profiles, demographic factors, and neurophysiological patterns [4]. Traits are "stable and enduring, caused by internal circumstances" whereas mental states, as defined by Chaplin et al. [5], are "temporary, brief, and caused by external circumstances". Demographic characteristics correspond to personal characteristics (age, gender, etc.), habits, and environment-related factors. Neurophysiological predictors are predictors from the EEG signal during MI tasks, pre-cue MI tasks, or during a resting state, serving as markers of the user’s mental states, such as attention [6] [7], fatigue [8], or initial capacities for producing the pattern to be decoded [1].

Among personality traits, cognitive profiles and demographic characteristics correlated with BCI performance. Jeunet et al. [4] identified 3 major elements : user relationship with technology, attentional capacities, and spatial abilities. More recently, Leeuwis et al. [9] re-evaluated these predictors in an experiment with 55 subjects and found that MI-BCI performance was significantly correlated with vividness of visual imagery, and the personality traits of orderliness and autonomy. How-

ever, Benaroch et al. suggested that personality traits, cognitive profiles and demographic characteristic might not be sufficient to predict MI-BCI performance using statistical models [10]. They conducted a follow-up experiment incorporating into the models neurophysiological predictors measured during a two-minute baseline at rest [11], and measured their predictive capabilities. They found significant predictability with neurophysiological predictors ( $p < 0.01$ ) while traits and demographic information led to a predictability that was not better than chance level ( $p = 0.88$ ).

The most robust neurophysiological predictor of BCI performance is the sensory motor rhythm (SMR) predictor, with a correlation coefficient ( $R$ ) of 0.53 validated with 80 participants, proposed by [1]. This predictor is computed from a 2-minute EEG recording during an open-eye baseline period, wherein the participant is not engaged in MI tasks associated with BCI activities. The SMR-predictor encapsulates the participant's capability to modulate their SMR. The effectiveness of predictions based on the SMR was confirmed by subsequent studies [11, 12], which also highlights the efficacy of SMR-based predictors. Specifically, Tzdaka et al. [11] introduced additional predictors based on the estimation of the power spectral density (PSD) during the open-eye baseline. With 56 participants, they showed that the mean performance during the session was significantly correlated with the number of frequency peaks necessary to model the PSD at electrodes C3 and C4 ( $R = 0.351$ ) and with the temporal variance in the amplitude of the peak within the mu band during baseline ( $R = -0.477$ ). This also highlights the efficacy of baseline SMR-based predictors of MI-BCI performances.

Other predictors have also been correlated with BCI performance. Grosse-Wentrup et al. identified in a dataset of 10 participants a correlation between the high gamma (55-85 Hz) rhythm in centro-parietal and frontal regions during the MI-BCI task and the SMR quality score, a metric of BCI classification performance ( $r = 0.0786$ ,  $p = 9.998 \times 10^{-5}$ ) [13]. Here, the gamma rhythm was thought to be related to attentional networks active during MI-BCI tasks [7]. Foong et al. demonstrated that for stroke patients ( $n = 11$ ), the mean relative beta power (12-30 Hz) before the cue across all trials was correlated with session performance [8]. This correlation was present for the relative beta power in the frontal (F3,Fz,F4) ( $r = 0.251$ ,  $p = 0.0005$ ) and central (C3,Cz,C4) ( $r = 0.181$ ,  $p = 0.0130$ ) brain regions but not in parietal-occipital regions ( $r = 0.033$ ,  $p = 0.6486$ ). It is important to note that this study made the hypothesis (not empirically verified) that relative beta power was a neural correlate of fatigue. Ahn et al. [14] divided the users in two different groups based on their performances during MI-BCI training, and have shown that during the baseline the efficient group has significantly higher  $\theta$  and lower  $\alpha$  than the inefficient one. Based on these results, they proposed a predictor of BCI performance, named PPfactor, that combined  $\theta$ ,  $\alpha$ ,  $\beta$  and  $\gamma$  power which had a strong correlation with BCI

performance ( $R = 0.59$ ) in their 61 subject dataset. Interestingly, although no significant correlations were found between BCI classification performance and either  $\beta$  or  $\gamma$  power, they were nonetheless included in the PPfactor equation.

Despite the extensive literature in the field, identifying strong and robust predictors of MI-BCI remains a significant hurdle. While the SMR predictor efficacy has been replicated in other datasets [11, 12], it is not the case for the other predictors. Botrel et al. have highlighted the issue, where numerous studies introduce new potential psychological or neurophysiological predictors of performance, yet often fail to replicate the findings of earlier research [15]. Moreover, several of these predictors were identified on small data sets, which can question their reliability. There is thus a pressing need to consolidate these predictors within a single study, on a different data set than the one on which they were identified, to assess their validity and reliability comprehensively. Therefore, in this paper, we aim to evaluate these diverse neurophysiological predictors together using a large ( $n = 85$ ) open source dataset to attempt to confirm their correlations with BCI performance.

This paper is organised as follows. In the Materials and Methods section, we present the dataset used and the methodology for extracting the neurophysiological predictors. The Results Section analyses the correlation between BCI performance and the neurophysiological predictors extracted. The Discussion Section finally compares those results with the one obtained by the original articles.

## MATERIALS AND METHODS

In order to validate those different neurophysiological predictors on a large and open source data set, we need to first extract those predictors from the dataset and then correlate them with BCI performance. In this section, we will first briefly present the dataset, then we will detail the 6 different neurophysiological predictors and how to compute them, lastly we will develop the statistical analysis that we performed.

*Dataset:* The dataset used in this analysis is sourced from an open-access EEG database with 87 participants, collected during a single day of MI-BCI experiments [2]. The experimental protocol was organized into a single session of motor imagery, divided into six runs. Before the first runs the participants were asked to perform two three minutes baseline recordings, one with open eyes, the other with closed eyes, where the participants were asked to fix a cross and relax. Participants were then required to engage in 20 motor imagery tasks (trials) for each hand per run. The first two runs corresponded to a calibration phase, where the feedback provided does not reflect the participants' actual motor imagery performance, but was a sham feedback. Based on the EEG data from these two first runs, a linear classifier employing three pairs of CSP spatial filters and a LDA classifier is

trained. In the subsequent four runs, participants received real-time feedback, visually represented by a horizontal bar whose length varied in proportion to the accuracy of the classifier predictions. The goal for participants was to maximize the length of this bar through their motor imagery efforts. In this analysis, we used epoch-wise accuracy (EAcc) as classification performance metrics. This metrics is expressed in ratio of correctly classified epochs compared to the total number of classification. We divided each MI trial in 1 seconds epoch with 1/16 overlap from 0.5 to 4.5 seconds after the visual cue.

Two participants, identified as A40 and A59, were excluded from the statistical analysis due to missing trials that could potentially impact the data analysis pipeline in subsequent steps.

#### Neurophysiological predictors:

In this paper we focused on six distinct predictors, four of them are extracted during the two-minute eyes-open (OE) baseline, one during the motor imagery (MI) trial, and one across MI trials. The extraction of these neurophysiological predictors was conducted as delineated in their respective foundational studies. The following subsections detail the extraction process for each predictor.

**The SMR-predictor** [1] is computed from a 2-minute EEG recording, taken during an open-eye baseline period when the user is relaxing. The signal is filtered between 4 and 40 Hz using a Laplacian filter around the C3 and C4 electrodes. The SMR-predictor is computed based on an estimation of the power spectral density (PSD) composed by the sum of two functions, as shown in Equation 1. The function  $g_1$ , given in Equation 2, models the noise floor of the signal, while  $g_2$ , outlined in Equation 3, estimates the peaks in the  $\mu$  (8-15 Hz) (with  $\mu_1$  and  $\sigma_1$ ) and  $\beta$  (15-30 Hz) (with  $\mu_2$  and  $\sigma_2$ ) bands in order to best estimate the PSD during the baseline. The SMR predictor is defined as the maximum difference within the mu band between the noise floor and the estimated peak  $g_2$ . From the PSD estimation, Tzdaka et al. [11] found two other predictors correlated with BCI performance: **The sum of the number of peaks** in C3 and C4 for each PSD reconstruction  $g$  (either 0, 1, or 2 for each channel) and **the variance across time of the peak frequency of the SMR predictor during the baseline**. For the latter, the two-minute baseline signal is divided into 10-second epochs with a 3-second overlap. For each epoch, the PSD of the signal is estimated with  $g$ . The variance is computed across all the estimations of the SMR predictor for C3 and C4, and then averaged between the two sensors.

$$g(f; \lambda, \mu, \sigma, k) = g_1(f; \lambda, k) + g_2(f; \mu, \sigma, k) \quad (1)$$

$$g_1(f; \lambda, k) = k_1 + \frac{k_2}{f^\lambda} \quad (2)$$

$$g_2(f; \mu, \sigma, k) = k_3 \phi(f; \mu_1, \sigma_1) + k_4 \phi(f; \mu_2, \sigma_2) \quad (3)$$

where  $k = (k_1; k_2; k_3; k_4) \in \mathbb{R}^4$ ;  $\lambda \in \mathbb{R}$  correspond to the steepness of the noise floor and  $\phi(\cdot, \mu, \sigma)$  indicates the

probability density function of a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ .

**The PPfactor** is calculated using data from a two-minute baseline period with open eyes [14]. During this phase, after applying a common average reference and a notch filter centred at 50 Hz to mitigate power line interference, the power spectrum of the EEG signal at electrodes C3 and C4 is analysed across various frequency bands:  $\theta$  (4-8 Hz)  $\alpha$  (8-13 Hz),  $\beta$  (13-30 Hz) and  $\gamma$  (30-70 Hz). The power values for each band are then normalized by the total power across the entire spectrum (4-70 Hz) to account for individual differences in overall brain activity levels. The PPfactor is subsequently calculated using Equation 4

$$PPfactor = \frac{\alpha + \beta}{\theta + \gamma} \quad (4)$$

**The high gamma predictor** is a marker of attention, extracted during the MI task [7]. During the entire epoch of the MI task, only the electrodes from the Frontal, Central, Central Parietal, and Occipital regions are considered. The EEG signal from these regions is filtered between 55 and 85 Hz to isolate the gamma band. The signal is then processed as follows: The filtered EEG signal is squared to calculate the power of the signal at each time point within the epoch. The mean power is computed by averaging the squared signal over all time stamps within the epoch. The power values are then log-transformed to normalize the distribution. Finally, the mean across all selected channels is computed to obtain a single value representing the mean log power for the gamma band during the MI task. It is important to note that the signal used for classification of MI is filtered between 8 and 30 Hz. Therefore the gamma band is not used as a classification feature.

**The Relative beta power at rest** is derived from a 3-second pre-MI task EEG segment, without explicit rest instructions, using frontal electrodes (F3, Fz, F4) [8]. As presented in equation 5 it is computed from the power  $\omega$  in the broad band (4-50 Hz) and the power  $\beta$  in the beta band (12-30 Hz) and is thought to reflect the level of fatigue of the user during the task.

$$RelativeBeta = 10 * \log_{10}(\frac{\beta}{\omega}) \quad (5)$$

*Statistical analysis:* We conducted correlations between the predictors and the offline classification performances. Our approach varied depending if the predictors is compared with the overall performance during the session or if this predictor is compared with the evolution of performance during the session. In the first case we performed Spearman correlation and in the second multi repeated correlation analysis [16]. The detailed analysis is described below:

- For predictors obtained during the baseline phase, we explored the correlation between each subject's overall performance in a single session and the predictor. We used Spearman correlation tests for this analysis. Spearman correlation was chosen for its ability to capture

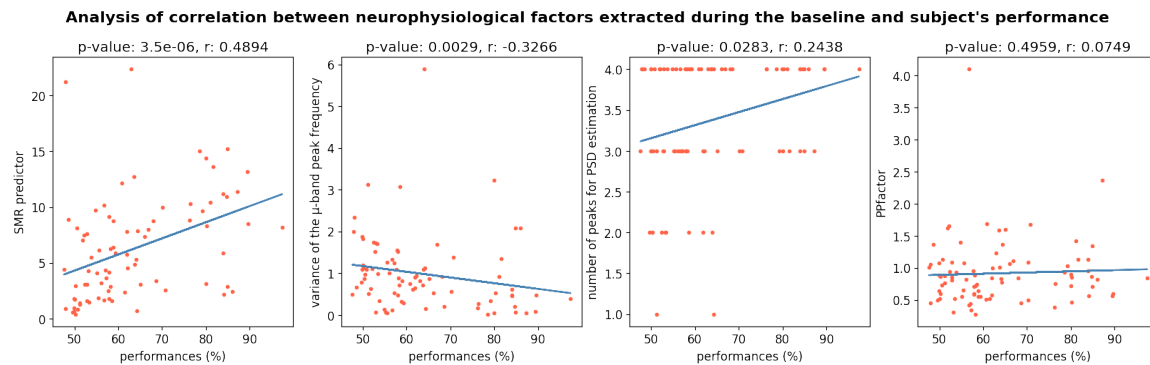


Figure 1: Correlation between subjects' BCI performance and neurophysiological predictors extracted during baseline. The scatter plots from left to right represent: SMR-predictor, the number of PSD peaks in the sensorimotor channels C3 and C4, the variance in the peak of  $\mu$ -band frequency during baseline, and the PPfactor. Each plot displays individual subject performances against the respective neurophysiological predictor, with the fitted linear regression line illustrating the trend. Significance levels (p-values) and correlation coefficients (r) between each predictor's and the performance are indicated based on Spearman correlation.

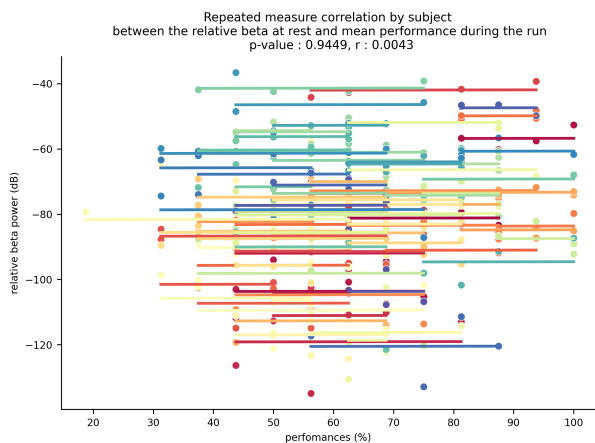


Figure 2: repeated measure correlation analysis between individual subjects' BCI performances and their average relative beta power during rest phases across four runs (n=85 subjects). The lines correspond to the regression for each subject where the slope is fixed and correspond to the repeated measure correlation coefficient ( $r = 0,0043$ ).

monotonic relationships without assuming linearity, for possibly being more robust to outliers and for reflecting better neurophysiological dynamics [17].

- For relative beta power at rest, we assessed the multi repeated correlation for each subject [16] between performance during the run and the mean relative beta power during resting phases before the MI trials. Repeated measures correlation is a statistical technique for determining the common within-individual association for paired measures assessed for multiple individuals, in this case the mean performance and the relative beta power at rest. Note that the original paper [8] found correlations between beta power at rest and BCI performance across sessions while we assess it across trials for a single session.

- For high gamma predictor, we investigated the correlation between the SMR quality score and high gamma predictor during MI. The SMR quality score, as defined in the work of Grosse-Wentrup and colleagues [7], is the



Figure 3: repeated measure correlation analysis between SMR quality score and the high gamma predictor power during the 40 trials of MI (n=85 subjects). The lines correspond to the regression for each subject where the slope is fixed and correspond to the repeated measure correlation coefficient ( $r = 0,1056$ ).

output of the LDA classifier (negative for left-hand MI and positive for right-hand MI) multiplied by the sign of the actual class. Consequently, correct trials yield a positive SMR quality score, whereas incorrect ones result in a negative score. It is essential to recognize that the magnitude of the LDA output (positive or negative) directly relates to the model's confidence in its classification, indicating that higher positive SMR quality scores denote accurate predictions made with high confidence, and higher negative scores denote incorrect predictions made with similar confidence. Correlations were conducted with repeated measures correlation per subject to appropriately analyse these relationships.

## RESULTS

In the analysis of neurophysiological predictors extracted during the baseline phase, we investigated the correlation between each predictor and the subjects' performance during the entire session. Figure 1 illustrates the Spear-

man correlation coefficients with the linear regression being also added to illustrate the correlation. The most significant predictor of performance was found to be the SMR predictor, which exhibited a positive and significant Spearman correlation ( $\rho = 0.49$ ,  $p < 0.01$ ). The second most significant predictor was the variance of the peak frequency necessary to reconstruct the PSD, as defined in Equation 3. This predictor showed a significant negative Spearman correlation ( $\rho = -0.33$ ,  $p = 0.0029$ ). The third predictor evaluated was the number of peaks required to model the PSD during the baseline, demonstrating significant Spearman correlation ( $\rho = 0.24$ ,  $p = 0.028$ ). The PPfactor did not exhibit any significant correlation with performance ( $\rho = 0.07$ ,  $p = 0.50$ ).

The analyse of the correlation between the mean relative beta power at rest and the corresponding performance, depicted in Figure 2, reveals no significant relationship ( $r = 0.0043$ ,  $p = 0.94$ ). In contrast, the correlation between gamma predictor and SMR quality scores, as shown in Figure 3, is weak but very statistically significant correlation ( $r = 0.1056$ ,  $p = 9 \times 10^{-15}$ ).

## DISCUSSION

The identification of robust and reliable neurophysiological predictors of BCI performance is a challenging and promising research issue. In this study, we successfully validated only four out of six proposed neurophysiological predictors using a large dataset comprising 85 subjects. Among these validated predictors, three were extracted during the baseline and demonstrated correlations with the mean session performance. Interestingly, the high-gamma predictor showed a correlation with the SMR quality score, which allows a trial-wise analysis to understand performance variations. In contrast, the predictive values of PPfactor and relative resting beta were not replicated in this dataset. Two important differences between our study and the initial study showing a correlation with relative beta [8] could explain these discrepancies. Firstly, that predictor was identified in stroke patients, a group often affected by fatigue [18]. Since relative beta power was used as a neurophysiological indicator of fatigue, it may have greater predictive value in this patient population rather than in the general population. Secondly, in their research, the repeated measure correlation was conducted across multiple sessions, while in our study, it was assessed across runs within a single session with only four runs per subject. Therefore, relative beta power may serve as a better indicator of overall fatigue rather than reflecting the evolution of fatigue within a session. Further analysis of this predictors will be necessary to confirm this relationship. Regarding the PPfactor, its lack of reproducibility cannot be attributed to differences in the experimental setup. Previous studies attempting to replicate this predictor only found significant correlations in one out of two datasets they tested [19]. Notably, when they applied a Laplacian filter around the C3 and C4 electrodes before extracting the PPfactor, this

enhanced the correlation between this predictor and performance metrics as compared to using C3 and C4 without Laplacian filters. Therefore, incorporating this filtering step into the PPfactor computation may be essential for enhancing the predictor's significance.

While previous studies have suggested that demographic information and personality traits [10, 15] may not be reliable predictors of performance, our study show the reliability and the reproducibility of most neurophysiological predictors.

## CONCLUSION

In this paper, we compiled various studies that identify neurophysiological predictors of MI-BCI performance across and within different subjects. We noted that these predictors have predominantly been explored in small datasets and lacked widespread replication on independent data sets. Therefore, we sought to replicate the predictive value of six distinct indicators within a large ( $n=85$ ), open source dataset. Of these, we managed to replicate the results of 4 of them: the SMR predictor, the number of peaks necessary for estimating the PSD during the baseline, the variability across time in the peak frequency within the mu band during baseline, and the high gamma predictor during the trial. However, the PPfactor and the relative beta power at rest were not successfully replicated. This study reinforces the significance of the SMR predictor as a robust indicator of BCI performance and highlights the critical need for replication of results reported in the BCI literature.

Future research needs to explore the influence of classification algorithm on predictors of BCI classification performance and validate that the predictors that we validate in this article are reliable predictors independently of the classifier used.

## REFERENCES

- [1] B. Blankertz *et al.*, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, 2010.
- [2] P. Dreyer, A. Roc, L. Pillette, S. Rimbart, and F. Lotte, "A large EEG database with users' profile information for motor imagery brain-computer interface research," *Scientific Data*, 1 2023.
- [3] B. Z. Allison and C. Neuper, "Could Anyone Use a BCI?" In *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*, ser. Human-Computer Interaction Series, D. S. Tan and A. Nijholt, Eds., London: Springer, 2010.
- [4] C. Jeunet, B. N'Kaoua, and F. Lotte, "Chapter 1 - Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates," in *Progress in Brain Research*, ser. Brain-Computer Interfaces:

- Lab Experiments to Real-World Applications, D. Coyle, Ed., Elsevier, 2016.
- [5] W. F. Chaplin, O. P. John, and L. R. Goldberg, "Conceptions of states and traits: Dimensional attributes with ideals as prototypes," *Journal of Personality and Social Psychology*, 1988.
  - [6] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu, "The predictive role of pre-cue EEG rhythms on MI-based BCI classification performance," *Journal of Neuroscience Methods*, 2014.
  - [7] M. Grosse-Wentrup and B. Schölkopf, "High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces," *Journal of Neural Engineering*, 2012.
  - [8] R. Foong *et al.*, "Assessment of the Efficacy of EEG-Based MI-BCI With Visual Feedback and EEG Correlates of Mental Fatigue for Upper-Limb Stroke Rehabilitation," *IEEE Transactions on Biomedical Engineering*, 2020.
  - [9] N. Leeuwis, A. Paas, and M. Alimardani, *Psychological and Cognitive Factors in Motor Imagery Brain Computer Interfaces*, 2021.
  - [10] C. Benaroch, C. Jeunet, and F. Lotte, "Are users' traits informative enough to predict/explain their mental-imagery based BCI performances ?," presented at the 8th Graz Brain-Computer Interface Conference 2019, 2019.
  - [11] E. Tzdaka, C. Benaroch, C. Jeunet, and F. Lotte, "Assessing The Relevance Of Neurophysiological Patterns To Predict Motor Imagery-based BCI Users' Performance," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020.
  - [12] C. Sannelli, C. Vidaurre, K.-R. Müller, and B. Blankertz, "A large scale screening study with a SMR-based BCI: Categorization of BCI users and differences in their SMR activity," *PLOS ONE*, 2019.
  - [13] M. Grosse-Wentrup, B. Schölkopf, and J. Hill, "Causal influence of gamma oscillations on the sensorimotor rhythm," *NeuroImage, Multivariate Decoding and Brain Reading*, 2011.
  - [14] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, "High Theta and Low Alpha Powers May Be Indicative of BCI-Illiteracy in Motor Imagery," *PLOS ONE*, 2013.
  - [15] L. Botrel and A. Kübler, "Reliable predictors of SMR BCI performance — Do they exist?" In *2018 6th International Conference on Brain-Computer Interface (BCI)*, 2018.
  - [16] J. Z. Bakdash and L. R. Marusich, "Repeated Measures Correlation," *Frontiers in Psychology*, 2017.
  - [17] G. Rousselet and C. Pernet, "Improving standards in brain-behavior correlation analyses," *Frontiers in Human Neuroscience*, 2012.
  - [18] M. Acciarresi, J. Bogousslavsky, and M. Pacioni, "Post-stroke fatigue: Epidemiology, clinical characteristics and treatment," *European Neurology*, 2014.
  - [19] C. Benaroch, "Contribution to the understanding of mental task BCI performances using predictive computational models," These de doctorat, Bordeaux, 2021.

# THE GOOD, THE BAD, AND THE UGLY OF IEEG SIGNALS: IDENTIFYING ARTIFACTUAL CHANNELS USING CONVOLUTIONAL NEURAL NETWORKS

Z.V. Freudenburg<sup>1</sup>, Y. Zhing<sup>1</sup>, M. P. Branco<sup>1</sup>, N. Ramsey<sup>1</sup>

<sup>1</sup> Brain Center, University Medical Center Utrecht, Utrecht, The Netherlands

E-mail: Z.V.Freudenburg@umcutrecht.nl

**ABSTRACT:** Intracranial electroencephalography (iEEG) signals have established themselves as a key tool for studying human brain function due to its distinct combination of high spatial and temporal precision. The use of both cortical surface and stereo-EEG in effective epilepsy treatment has allowed researchers to study electrophysiology throughout the brain in relatively large numbers of subjects. This provides an opportunity to overcome, the sparse and varied nature of the brain tissue sampling inherent to the clinical use of iEEG by aggregating data across many subjects. Essential to the success of large-scale data aggregation is the efficient and robust identification of recording channels that are dominated by ‘noise’ or artifacts introduced by the recording environment or hardware failure. Here we test the effectiveness of training a convolutional neural network (CNN) for this purpose across multiple types of iEEG recordings. We conclude that a small CNN trained on hand labeled data from a small set of subjects can be applied to identify artifactual channels.

## INTRODUCTION

Electroencephalography (EEG) allows for the recording of electrical signals generated by brain function and as such provides a precise measure of the temporal dynamics of brain function. However, extra-cranial EEG presents many challenges in terms of precisely locating independent neural sources of this activity. In the past decades the need to localize brain activity at the spatial resolution of tens of millimeters and with ms temporal precision to facilitate the localization seizure focus sites for medication resistant epilepsy treatment has led to the intracranial implantation of electrodes (iEEG) either on the cortical surface, often referred to as Electrocorticography (ECoG), or along shafts probing cortical and subcortical areas, often referred to as stereo-EEG (sEEG) [1]. The iEEG’s use in epilepsy treatment and increased use in awake craniotomies for functional localization during brain tumor resection offers a unique opportunity to study the brain function of many humans performing a variety of motor and cognitive tasks. However, due to the nature of the clinical setting in which iEEG is often recorded, the locations that are measured from only sparsely sample the brain and vary widely in number and location between subjects. Hence, showing

reproducibility of results over humans on a whole brain scale for iEEG requires the aggregation of data across tenths or hundreds of subjects.

To facilitate this scale of iEEG data aggregation a robust and efficient method for identifying iEEG signals that are dominated by artifacts or noise introduced by hardware failure or fed by the environment is needed. Often such noise screening relies on the evaluation of experienced iEEG clinicians and researchers. This process is generally quite labor intensive, subjective, and not standard between centers or experts. Here we explore the effectiveness of training a deep learning model to do this. Multiple groups have also attempted to use deep learning for noisy EEG channel selection. One approach is to apply thresholds to certain statistics computed from the signals. For example, APP [2] uses correlation and dispersion, FASTER [3] utilizes correlation, variance, and the Hurst exponent, Automagic [4] employs an independent component analysis-based artifact correction method, CTAP [5] calculates log relative variance and compares it to the median, and so on. Additionally, there are unsupervised methods. For instance, the Local Outlier Factor algorithm [6] identifies bad channels relative to the local neighboring channels, while the bad-by-RANSAC method [7] uses good channels to predict other channels and deems the channel poorly predicted by others as bad. Furthermore, supervised neural networks have also been utilized [8].

Yet, the transference to iEEG of these methods seems to be limited, with fewer reports about bad channel detection. The common method is to calculate statistics over the signals and input these statistics into machine learning methods. For example, in [9], the ensemble bagging classifier was applied to sEEG data, achieving the best accuracy of 99.77% across 110 subjects. In [10], multiple machine learning methods were tested on pigeons’ ECoG data, including the K-Nearest Neighbors Algorithm (kNN), Support Vector Machine (SVM), Random Forest (RF), and others, with the best F1-score of 0.9089 achieved using RF and Synthetic Minority Oversampling Technique (SMOTE) to address the imbalanced dataset.

We chose for a Convolutional Neural Network (CNN) architecture because of its proven ability learn EEG and iEEG signal filters at the lower level of more complex deep networks such as HTNet [11]. This allows us not to rely on predefined derived signal features

while keeping the network relatively simple, since we are not interested in differences in electrophysiological patterns or there spatial distribution on the cortex, but single channel level identification of non-electrophysiological (noisy or artifactual) signal. We apply this model to the three different iEEG recording modalities discussed above (clinical-ECoG, high-density (HD)-ECoG, HD, and sEEG) from a large number of individuals.

## MATERIALS AND METHODS

**Data:** Data from 47 patients implanted with iEEG electrodes for the purpose of drug resistant epilepsy treatment at the University Medical Center in Utrecht were used in this study. The study was approved by the Medical Ethical Committee of the University Medical Center Utrecht in accordance with the Declaration of Helsinki (2013). The patients had either sEEG or clinical scale ECoG electrodes implanted according to clinical needs and gave written informed consent to participate in research tasks and will be referred to as subjects in this work. A subset of the subjects gave additional consent to have HD-ECoG implanted solely for research purposes. Data from and additional 3 patients undergoing an awake craniotomy for tumor removal, in which HD-ECoG grids are briefly placed on the exposed cortical surface, who also consented to performing brief research tasks were also included.

Data from clinical-ECoG implants were recorded from implanted grids and strips of evenly spaced platinum electrodes with an inter-electrode distance of 10 mm and a 2.3mm exposed recoding surface. Implanted sEEG props had 8-15 platinum-iridium cylinder contact points of 0.8 mm diameter and 2mm height with a 1.5 mm inter-contact distance. The HD-ECoG grids used were equally spaced grids of 32-128 platinum with 1.3 mm exposed surface diameter and an inter-electrode distance of 3 or 4 mm.

A total 96 data sets from 50 subjects performing one of 19 different cognitive tasks. Tasks range from simple relaxation without movement to overt and imagined movements to overt and covert speech. Subject data were organized into groups and based on the type of iEEG implant used to facilitate exploration of implant type on noisy signal detection (see Table 1). Furthermore the Clinical-ECoG group was split into an adult group (Ca) and a child group (Cc) test for an age effect on noise detection and the HD-ECoG group (HD) was divided by the recoding setting since a subset of this data was recorded in the operating room (OR) and not in outside the Intensive Epilepsy Monitoring Unit (IEMU) like the remaining subjects because it is known the OR has more noise sources and recordings are made while the electrode grids are still exposed to the air. In the case of the IEMU recordings are made after the skull has been replaced. The sEEG group (sE) was not further subdivided.

Table 1: Data Groups

Grid type	Sub-group	Subject count	Channel count*	Ratio good:bad
Clinocal-ECoG (Ca + Cc)	adult	11	1500	44:1
	child	13	1900	104:1
sEEG (sE)	adult	10	2700	43:1
	child	4 (14)		
HD-ECoG (HDe + HDor)	IEMU	9	2000	33:1
	OR	3 (12)		

(\* = rounded to 100s)

**Preprocessing and labeling:** The recordings are from over a span of 20 years and have different frequencies, ranging from 512Hz to 2048Hz. To ensure that the bad channels exhibit similar patterns, all the recordings are down sampled to the lowest frequency, 512Hz.

For every subject, 10 minutes of recordings are included. For some subjects 5 minute recordings from the beginning are taken from two task files.

In this study we considered as noisy (bad) channels those that are clearly distinct from others in terms of signal content and that would likely distort the signals of non-noisy (good) channels when included in common average re-referencing (CAR). Bad channels were identified by visual inspection by two independent people (author 2 and author 3). For that we visually inspected both the raw-voltage signals and the power-spectrum (1/f, after removal of line-noise and it's harmonics) of every channel in one data file. Channels that had a deviant voltage amplitude compared with other channels in the same file (average voltage amplitude lays between -500 and 500 mV), excessive amount of line-noise, or recurrent large voltage fluctuations throughout the 5 minutes of data, were labelled as 'bad'. Besides determining bad channels, we also identified borderline-bad channels, that would not be considered as 'bad channel' by an expert but could potentially be labelled as such by the algorithm. These channels were labelled as 'maybe'. The remain channels were labeled as 'good'.

In this work a binary classification model is used, since we want the model to learn patterns of good and bad channels and be able to classify uncertain channels afterwards. Therefore, the 'maybe' channels are not included in the training or testing set.

The ratio of good vs bad channels in the data is very high, ranging from 44:1 to 104:1 (see Table 1). During the training, such severely imbalanced data harms the performance of the model. To reduce such effects, a down-sampling method was used in the training set to reduce the ratio of good vs bad channels to 2. The average number of good channels needed per subject need to achieve the 2:1 with the number of bad channels in the training set is calculated. Then random sampling is done to reach the designated amount of good channels per subject. By down-sampling, the amount of data in the training set also decreases, reducing the training time drastically.

According to clinical expertise, a window of 30 seconds in a channel contains enough information to show baseline patterns and classify channels as good or bad. In

addition, the level of noise can change over time due to head movements or medical operations. Splitting a channel into windows allows the model to classify windows in the same channel differently to account for possible changes in characteristics over time. Thus, each recording is chopped into around 10 30-second windows. These windows do not overlap to prevent data leakage from the training set into the testing set. For training all windows from a channel are given the label of that channel.

**Model and Training:** The model is a shallow Convolutional Neural Network (CNN) consisting of 2 convolutional layers with 55 weights followed by 2 linear layers with around 500k weights. To prevent overfitting, Batch-normalization, MaxPool, and Dropout layers with a rate of 0.2 are included. The input of the model is the 30-sec window, i.e., 30 seconds \* 512Hz = 15360 nodes. The structure of the model was derived from experimentation with a training set containing one or two subjects from the Ca group and applied for the rest of the analysis.

**Performance Evaluation:** The model produces a prediction for each window of data. However, the prediction performance of the model for each window is not evaluated window-wise. Instead, the prediction of a channel is calculated by thresholding on the percentage of its windows that are predicted to be bad.

We chose the Matthews Correlation Coefficient (MCC) as our performance evaluation metric because it has been shown to be more robust and reliable among common metrics for imbalanced datasets [12]. MCC is a metric that summarizes a confusion matrix and computes the correlation between ground truth and predictions. The MCC is bounded from -1 to 1, with -1 indicating all predictions are wrong, 0 implying nearly random predictions from the model, and 1 indicating all predictions are correct.

Since the channel level performance of the model will depend on the chosen windowing threshold discussed above a metric for choosing this threshold is needed. In this context this comes down to choosing the best dividing line between the distribution of percent of windows labeled as bad for the set of bad channels labeled by the experts as bad and those labeled as good. For this we use Brent's method [13] because it works better than other methods for finding the border between two distributions when those distributions are multi-modal. Brent's method is a numeric method to find the local minimizer of a certain function. Here, the parameter to minimize is the threshold, bounded by 0 and 1. The function calculates the negated MCC given the threshold and labels of the channels. Brent's method returns the threshold that maximizes MCC.

In this work we performed both within group and between group training-testing comparisons. To evaluate within group performance a leave-one subject-out (LoO) training testing approach was used. For this all channels from a single subject are left out of the training set and included in the test set and this is repeated to gain prediction results for every subject in the testing group.

This means that when we reports results for training and testing on different sub-groups, if the training group sub-groups overlap with subgroups in the testing group the LoO method was used for subjects within the overlapping sub-groups.

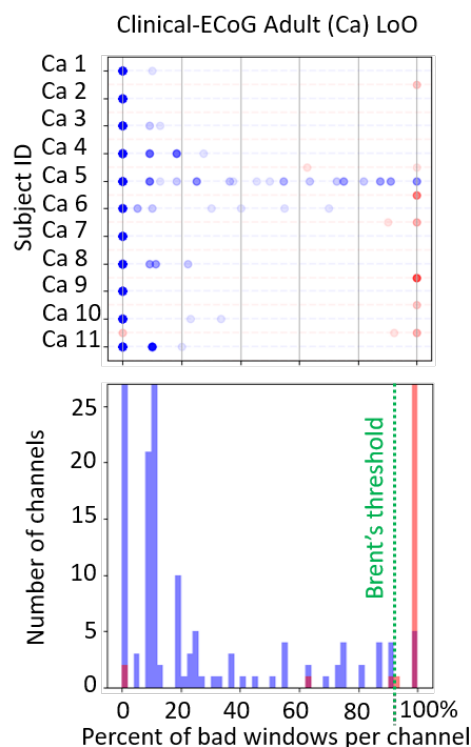


Figure 1: Leave-one subject-out (LoO) results for the Clinical-ECOG Adult (Ca) group. Top: Plot of % of windows classified as noisy (x-axis) for all channels of each subject (y-axis). Red and blue bad and good labeled channels respectively. Bottom: Histogram of number of channels (y-axis) with a certain percentage of windows predicted as bad (x-axis). The red and blue bars represent channels labeled as bad good respectively. The green vertical dotted line indicates the threshold found with Brent's method.

## RESULTS

**The distinction between good and bad varies over iEEG groups** As Figures 1 and 2 illustrate the distinction between the good and bad channels in terms of the percentage of windows classified as bad varies considerable between iEEG date groups. When channels are clearly bad or good most of the time and the model is able to learn a clear distinction between the two signal types you would expect to see the distributions found for the Ca group in Figure 1. Here we see that almost 100% of the windows are classified as bad for most of the channels labeled as bad and often not more than 30 % of the windows from good channels are classified as bad. This means that a wide range of thresholds (30%-90%) will give similarly good MMC scores. In this case a very

conservative threshold of 92% that allows for good channels to have a lot of bad windows is found by Brent's method to give the optimal MMC score of 0.84 (see Table 2, first row). This effectively means that only 5 good channels 4 channels labeled as noisy (out of ~1500) are misclassified. However, as Figure 2 shows, in the case of the sEE group both the distributions of good and bad labels are broader. Meaning that either the network has a harder time distinguishing good windows from bad windows or that the amount of noise in the sEEG signals fluctuates over time causing good channels to have noise at times and bad channels to be less noisy at times. In this situation the optimal threshold really need to balance the false positives (FPs) and false negatives (FNs) to reach the optimal MMC, which was found to be 0.32 at 68% (Table 2) in this case. It should be noted that this MMC is still well above 0 and the good and bad distributions for sE are still distinct from each other allowing for thresholds to be set that can reduce either the number of FPs or FN to almost 0, but not both.

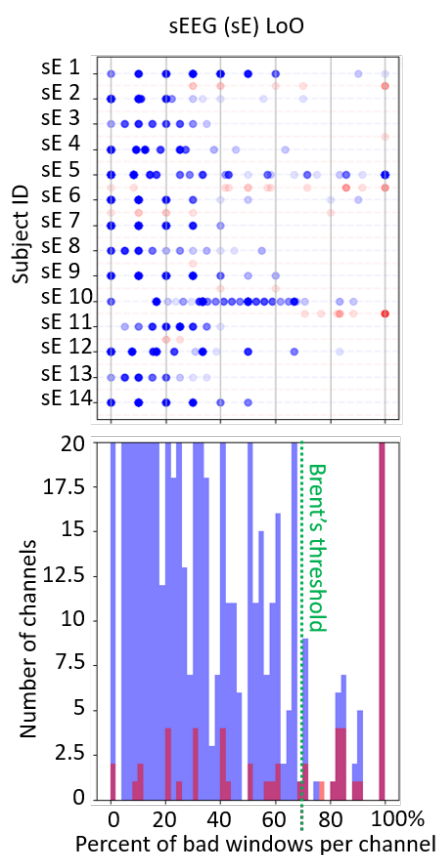


Figure 2: Leave-one subject-out (LoO) results for the Clinical-EEG group. Top: Plot of % of windows classified as noisy.

*In general bad electrode classification works well for Clinical-EEG* As can be seen in Table 2 all analysis that involved training and testing on Clinical-EEG groups performed well. The Cc within group test showed and even higher MMC score (0.94) that that of Ca and even when training on Ca and testing on Cc an MMC score of

0.89 was achieved. This indicates that there is little difference between the good and bad signal characteristics between channels implanted in adults and children. This is further supported by the fact that training on Ca and Cc also gave high MMCs of 0.92 and 0.82 when testing on Ca and Cc respectively.

Table 2: MMC and Brent threshold across different training and test groups combinations.

Test Group	Train Group	MCC	Brent's threshold
Ca	Ca	0.84	92%
Cc	Cc	0.94	85%
sE	sE	0.32	68%
HD	HDe+HDor	0.10	48%
HDe	HDe+HDor	0.43	94%
Cc	Ca	0.89	85%
Ca	Ca + Cc	0.92	70%
Cc	Ca + Cc	0.82	70%
sE	Ca + Cc	0.66	76%
HD	Ca + Cc	0.23	77%
sE	Ca+Cc+sE	0.41	85%

*Distinguishing the bad from the good in sEEG is harder but promising* While the sE group performance discussed above is lower than that of the Clinical\_ECoG groups introducing data from these groups into the model training does improve its performance on the sE group to an MMC of 0.41 (bottom row, Table 1). In fact, only training on the Ca and Cc groups improve the performance on the sE group even further to 0.66. This is a marked improvement and inspection of the percentage bad windows distributions (Figure 3) indicates that by training on channels with clearer bad vs. good signal distinctions the model was able to more clearly separate good and bad sE channel windows. While there are now many bad channels with none of their windows classified as bad, the number of bad channels with >80% of windows classified as bad was relatively unaffected and the number of good channels with >50% of their windows classified as bad decreased greatly.

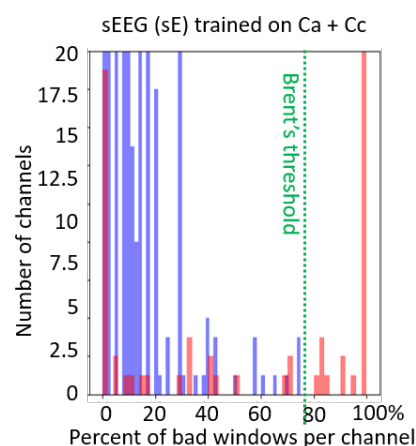


Figure 3: sEEG window classification based on training

on Clinical-ECoG.

*HD-ECoG presents a challenge to automatic noisy channel detection* While still above 0 the MMC for the HD group was only 0.1 (Table 2) and the fact that the optimal threshold was around 50% (48%) and the top plot in Figure 4 indicate that the good and bad distributions were very mixed. In fact, most of the good and the bad labeled channels had around 50% of there windows classified as bad. This could indicate that the amount of noise in the HD-ECoG channels changes a lot over time. It is worth noting that this is especially the case for the sub-group of subjects recorded in the operating room setting (HDor). When excluding the HDor subjects for testing, the performance increases to and MMC of 0.43. While training on the Clinical-ECoG groups does improve performance on the HD group as a whole similar to the sE group, this improvement is mostly due to the better distinction on the HDe sub-group (Figure 4, bottom plot). In the case of HDor subjects, 0% of almost all channels are classified as bad. This is surprising since the HD group has the lowest good to bad channel ratio (Table 1) and the HDor sub-group has a the majority of the bad channels in this group, as can be seen by the large number of red circles for the HDor subjects in the top plot of Figure 4.

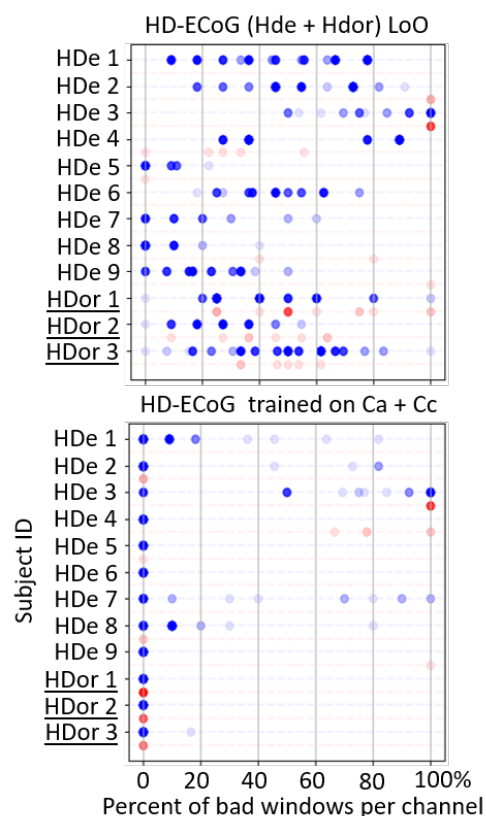


Figure 4: HD-ECoG noisy window classification for within group vs. Clinical-ECoG group training. Top: Plot of % of windows classified as noisy. Bottom: Plot of % of windows classified for the HDe and HDor groups after training on the Ca + Cc groups.

## DISCUSSION

In general we found that our simple model worked well for identifying noisy artifactual signal in clinical-ECoG data and that there is no need to treat signals from children as different from those of adults.

In addition, while we found the our approach was less effective regarding sEEG signals, we did see that performance on sEEG channels can be improved by applying a model trained on clinical-ECoG channels, where there is a larger separation between channels with artifactual signals and those with clean electrophysiological signals.

The results suggest that the difficulty in identifying noisy sEEG channels could be due to a larger variance in electrophysiological signal properties found in sEEG signals. When training on sE the model showed a broad distribution for both bad and good channels. It is known that the electrophysiology of sub-cortical regions differs from that of the cortical surface and the model has no knowledge of this while the expert labeler does. One trend is that the amplitude in signal generally decreases with the depth of the implanted electrode. The expert may consider this when judging a channel to be good even if it has low amplitude. This would indeed make it harder to train on sEEG channels as there will be a larger variance in the types of patterns associated with channels labeled as good. However, the fact the model trained on clinical-ECoG greatly increase the number of sEEG channels with < 20% of windows labeled as bad without decreasing the number of labeled bad or increasing the number of labeled good channels with > 80% of widows predicted as bad shows that relative to clinical-ECoG most sEEG channels contain similar data. This is promising for goal of agrigating sEEG data into clinical-ECoG data sets.

HD-ECoG presented the hardest challenge to accurately predicting bad channels. This was especially true for the sub-group of subjects recorded in the operating room. As opposed to sEEG this is likely due to an increase in the variety of types of artifactual signals in these recording compared to clinical-ECoG. This group had the lowest good to bad channel ratio and hence the largest number of labeled bad channels. The fact that the HD-electrodes are smaller means they will have larger impedences and potentially pick up more external noise though the wire connecting them to the amplifiers. In addition it is known that the recordings from the operating room in the HDor group will be more sensitive to external noise because during the recording the skull is open, unlick in the IEMU where the skull has be replaced and the wound sealed, before recording. This provides more direct exposure to external electrical sources. In addition the fact that most labeled bad and good channels have around 50% of their windows predicted as bad suggests that there may be artifacts that are transient in nature and not consistent throughout the recording. Taken together this suggests that predicting artifactual signal in HD-ECoG will likely require larger set of correctly labeled data to train on. It should be noted that often HD-ECoG implants are aim

towards BCI use and thus longer term implantation where additional work in identifying and labeling artifactual signal is justified.

*Limitations* The main limitation of this work is that ground truth labeling was based on subjective evaluation of channel as a whole. This is sufficient for large scale studies, but not a true separation of electrophysiological signals from external noise. Each channel is not purely noise or brain signal at any one time point. Thus, it could also be beneficial to allow for prediction of a percentage or probability of noisiness as the recorded signal is almost always a combination of both.

However, this proportion is hard to compute and label. This is a limitation inherent to all work on distinguishing artifactual signal from electrophysiological signal in that work on understanding the true ground truth of what parts of recorded iEEG signals are pure reflections of electrophysiology is still very much ongoing.

One approach to overcome this would be to train with as good as possible pure electrophysiological signals and add known amounts of simulated noise. In this way there would at least be ground truth for the known added signal artifacts and knowledge as bout simulating artifacts caused by known noise sources is much better due to the vastness of the field of electronics.

Furthermore, the models could be improved by allowing for additional types of labels for channels such as the cortical or subcortical region where the electrodes are located and/or the type of cognitive task the subject is performing as these factors are known to influence iEEG signal features. In this way, a model trained to identify noisy signal characteristics could also be used to specify what signal features constitute 'normal' iEEG signal from different parts of the brain. This suggests another possible avenue for future work, which would be to explore the signal features encoded in the deeper layers of such a CNN.

## CONCLUSION

In conclusion we feel this work is encouraging for studies aimed at large scale data aggregation over many subjects and multiple institutions in that it shows the feasibility of automating the identification of channels that can be safely included in the analysis and which ones should be excluded.

## REFERENCES

- [1] Herff C, Krusienski DJ, Kubben P. The Potential of Stereotactic-EEG for Brain-Computer Interfaces: Current Progress and Future Directions. *Front Neurosci.* 2020;14:123
- [2] da Cruz JR, Chicherov V, Herzog M, Figueiredo P. An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics. *Clin. Neurophysiol.* 2018; 129:1427-1437
- [3] Nolan H, Whelan R, Reilly RB. FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *J Neurosci Methods.* 2010;192:152-62.
- [4] Pedroni A, Bahreini A, Langer N. Automagic: Standardized preprocessing of big EEG data. *Neuroimage.* 2019;200:460-473
- [5] Cowley BU, Korpela J. Computational Testing for Automated Preprocessing 2: Practical Demonstration of a System for Scientific Data-Processing Workflow Management for High-Volume EEG. *Front Neurosci.* 2018;12:236
- [6] Kumaravel VP, Farella E, Parise E, Buiatti M. NEAR: an artifact removal pipeline for human newborn EEG data. *Dev. Cogn. Neurosci.* 2022; 54
- [7] Bigdely-Shamlo N, Mullen T, Kothe C, Su KM, Robbins KA. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front Neuroinform.* 2015;9:16
- [8] Kumaravel, V.P., Kartsch, V., Benatti, S., Vallortigara, G., Farella, E., Buiatti, M. Efficient artifact removal from low-density wearable EEG using artifacts subspace reconstruction, in: *Proc. of 43rd Ann. Int. Con. of the IEEE Engineering in Medicine and Biology Society*, 2021, 333–336
- [9] Tuyisenge V, Trebaul L, Bhattacharjee M, Chanteloup-Forêt B, Saubat-Guigui C, Mîndruță I, Rheims S, Maillard L, Kahane P, Taussig D, David O. Automatic bad channel detection in intracranial electroencephalographic recordings using ensemble machine learning. *Clin Neurophys.* 2018;3:548-554.
- [10] Li M, Liang Y, Yang L, Wang H, Yang Z, Zhao K, Shang Z, Wan H. Automatic bad channel detection in implantable brain-computer interfaces using multimodal features based on local field potentials and spike signals. *Comput Biol Med.* 2020;116:103572
- [11] Peterson S, Steine-Hanson Z, Davis N, Rao R, Brunton B. Generalized neural decoders for transfer learning across participants and recording modalities. *J. Neural Eng.* 2021;18:026014
- [12] Chicco D, Tötsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min.* 2021;14:13.
- [13] Brent, P Chapter 4: An Algorithm with Guaranteed Convergence for Finding a Zero of a Function", *Algorithms for Minimization without Derivatives*, Englewood Cliffs, NJ: Prentice-Hall, 1973.

# INTRODUCING THE USE OF THERMAL NEUROFEEDBACK

F. Le Jeune<sup>1,\*</sup>, E. Savalle<sup>1,\*</sup>, A. Lécuyer<sup>2</sup>, M. J.-M. Macé<sup>1</sup>, P. Maurel<sup>3</sup>, L. Pillette<sup>1</sup>

<sup>1</sup>Univ. Rennes, Inria, CNRS, IRISA, Rennes, France

<sup>2</sup>Inria, Univ. Rennes, IRISA, CNRS, Rennes, France

<sup>3</sup>Univ. Rennes, Inria, CNRS, Inserm, Empenn ERL U1228, Rennes, France

\*Co-first authorship. Both authors contributed equally.

E-mail: francois.le-jeune@inria.fr

**ABSTRACT:** Motor imagery-based brain-computer interfaces (MI-BCIs) enable users to control digital devices by performing motor imagery tasks while their brain activity is recorded, typically using electroencephalography. Performing MI is challenging, especially for novices. To tackle this challenge, neurofeedback (NFB) training is frequently used and usually relies on visual feedback to help users learn to modulate the activity of their sensorimotor cortex when performing MI tasks. Improving the feedback provided during these training is essential. This study investigates the feasibility and effectiveness of using thermal feedback for MI-based NFB compared to visual feedback. Thirteen people participated to a NFB training session with visual-only, thermal-only, and combined visuo-thermal feedback. Both visual-only and combined visuo-thermal feedback elicited significantly greater desynchronization over the sensorimotor cortex compared to thermal-only feedback. No significant difference between visual-only and combined visuo-thermal feedback was found, thermal feedback thus not impairing visual feedback. This study outlines the need for further exploration of alternative feedback modalities in BCI research.

## INTRODUCTION

A brain-computer interface (BCI) relies on a neurophysiological acquisition method, often electroencephalography (EEG), to record brain activity that is in turn processed and interpreted as a command to control a digital device. BCIs have been used to control external devices for both non-clinical applications, such as video games [1], and clinical applications, such as motor rehabilitation after stroke [2].

One of the main challenges for BCIs, is for their users to generate a brain activity that is reliably recognizable by the computer. One commonly used mental task that produces consistent brain activity is motor imagery (MI), which has been proven to consistently activate the sensorimotor cortex [3]. However, MI is not an easy task for novice BCI users. They have to learn to modulate their brain activity by exploring different MI strategies, such as imagining different gestures or focusing on dif-

ferent sensations, to reliably activate their sensorimotor cortex. Consequently, BCIs rely on training users to control their brain activity. To this end, neurofeedback (NFB) is mostly used: it consists in a closed-loop technique providing users with feedback on their own brain activity so they can learn to modulate it. Thus, providing users with feedback that they can understand and interpret intuitively is of utmost importance in the learning process, and improving the feedback is a key factor of improvement for BCI efficiency. A feedback can be defined through three main characteristics: (i) its content, i.e., the information that it conveys (ex: neuromarker on which the feedback is based), (ii) its modality, i.e., the way this information is conveyed (ex: haptic feedback using vibrators), and (iii) its presentation timing, i.e., the moment when it is provided (ex: continuous presentation with a refreshing rate at 0.1Hz) [4]. Among those, the modality of feedback is the most investigated characteristic.

Several modalities of feedback are reported in the literature. A majority of the studies displayed visual feedback, probably because vision is the sense on which daily life perception relies the most [4]. Haptic feedback including vibrotactile, functional electrical stimulation and robotic orthosis was also provided during MI-BCI user training (see [5] for a review on haptic NFB). Such feedback could particularly be interesting for MI-BCI as it activates similar cortical structures as the ones involved in MI. Controversially, the use of haptic feedback could also contribute to overtax the cortical structures associated with both its processing and the performance of the MI tasks. Previous MI-BCI experiments involving vibrotactile feedback did not find any significant negative or positive influence on the resulting electrophysiological activity or BCI performances compared to visual feedback [6–8], despite the fact that participants reported perceiving the haptic feedback as more natural than the visual one [6]. Multimodal feedback, involving both visual and haptic vibrotactile stimuli at once, seemed however to improve MI-BCI performances [9, 10].

To our knowledge, the thermal component of haptic stimulation has never been investigated as a potential NFB modality. Yet, thermal feedback appears promising as thermal stimulation is inexpensive to develop and

fairly easy to use. Besides, thermal stimulation could be perceived as natural feedback since external thermal sensation is continuously involved when exploring our environment with our body. Furthermore, such feedback could particularly benefit therapeutic applications as studies showed that thermal stimulation facilitates sensory and motor recovery in stroke patients [11, 12]. Hence, it appears of interest to study the feasibility of including the so far disregarded thermal feedback in MI-NFB training. The goal of the present study was to investigate the effects of thermal neurofeedback on users' ability to control their brain activity and on their user experience, in comparison with a visual feedback. To this aim, we investigated NFB training with thermal feedback only, visual feedback only, and with a combination of both visual and thermal feedback.

## MATERIALS AND METHODS

We investigated the influence of three different feedback modalities on participants' ability to modulate their own brain activity and on their user experience. The three modalities were the following ones: **visual-only feedback (V)** in which participants experienced solely visual feedback, **thermal-only feedback (T)** in which participants experienced solely thermal feedback, and a bi-modal **visuo-thermal feedback (VT)** in which participants experienced both thermal and visual feedback simultaneously. All participants experienced the different feedback and their order of presentation was pseudo-randomized (so that the order of conditions was counter-balanced across participants).

**Participants:** Twenty-four healthy participants completed the study (8 women, 15 men, 1 non-binary, age  $25.8 \pm 3.8$  years). None of them had any history of neurological or psychiatric disorders. All participants provided written informed consent before the experiment in accordance with the Declaration of Helsinki and following amendments. The thirteen initial participants were tested with a thermal feedback slightly different from the others (see Section *Thermal Feedback*). Authors of the study observed that thermal feedback obtained from equation 4 resulted in a wider stimulation range than desired. The equation was updated (5) and eleven additional participants were tested. Results for the second group of participants were similar and conclusions thus identical to the ones of the first group. No significant difference was found between groups. Both datasets were thus merged and analysed as a single one.

**Experimental protocol:** The experiment lasted about two hours. Participants were seated in a comfortable armchair, in front of a monitor placed flat on a table right above their arms (Fig. 1.A). First, participants were asked to fill in two questionnaires regarding general demographic information and their handedness. Participants were then equipped with a wearable thermal stimulation system on their right hand. It consisted of a Peltier cell attached to a heat sink and assembled on a 3D-printed

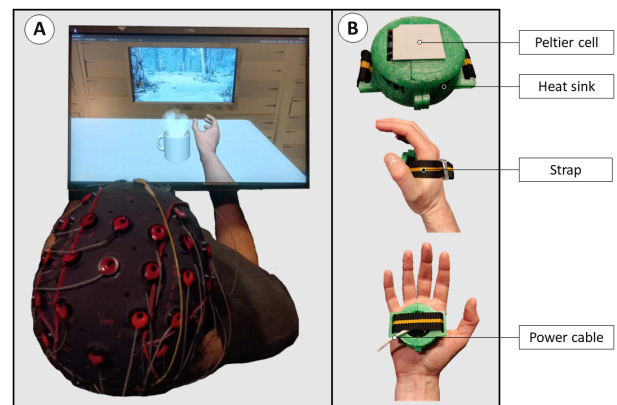


Figure 1: Experimental setup. **A**-Participant seated on a chair in front of the computer screen and wearing the EEG headset. The virtual hand is superimposed above the participant's real hand. **B**-Thermal stimulation system. Top: top view of the system; Middle and Bottom: top and left view (respectively) of the wearable system on the user's hand.

wearable element (Fig. 1.B). The wearable system was strapped to the right hand of participants so that the Peltier cell would be in complete contact with the palm of their hand. Because thermal stimulation varies from user to user, we performed a calibration of the thermal stimulation range to adapt it to each participant (see Section *Thermal Feedback*). Subsequently, participants were equipped with an EEG headset (see Section *EEG Recordings & Signal Processing*) and an electrode was placed on the participants' skin above the anterior proximal part of their forearm to assess hand electromyographic activity (EMG). Afterwards, the participants were given instructions regarding the experimental protocol, including the modalities of feedback and the motor imagery task to perform. They were asked to repetitively imagine closing and opening their right hand while focusing on the sensations related to the movement, such as hand muscle contraction, skin and tendon stretch, and tactile and thermal sensations on the hand. Then, participants were asked during one run to perform the MI task while looking at a fixation cross (no feedback was provided at that point of the experiment). Each run consisted of twenty successive trials of five seconds of rest followed by ten seconds of MI, for a total run duration of five minutes. We calibrated the BCI based on the data from this run by defining a reference ERD ( $ERD_{ref}$ ) set as the 30th percentile of the produced ERDs. Then, all the participants successively experienced the three feedback modalities (pseudo-randomized order across participants). For each modality, we proceeded as follows. First, we asked participants to rest while staring at the center of a white cross displayed on the screen for one minute while we recorded their brain activity as a baseline for future analyses. Second, we asked participants to perform two training runs (separated by a short break) with the feedback modality associated with one of the three modalities. Afterwards, participants were asked to fill in a questionnaire regarding their user experience of the two NFB runs they just per-

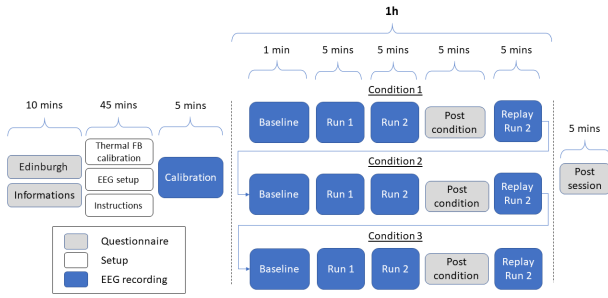


Figure 2: Experimental protocol. The three tested condition (V, T, VT) were performed successively in a pseudo-random order, in a single session.

formed. Finally, during one run (which we called *Replay*, Fig. 2) we replayed the exact stimuli generated during the second user training run while the participants were asked to only pay attention to these stimuli without imagining the movement. A short break was also proposed in-between modalities. At the end of the three modalities, participants were asked to fill in a final questionnaire regarding their preferences in terms of feedback modality. All the equipment was then removed and we made a short debrief (Fig. 2).

**EEG Recordings & Signal Processing:** The EEG data was recorded with 31 active electrodes, using a g.USBamp EEG amplifier (g.tec, Austria). The electrodes were placed over the sensorimotor cortex (at locations Fp1, Fp2, FC5, FC6, F3, F4, FCz, T7, T8, C3, C4, Cz, P3, P4, O1, O2, CP1, CP2, CP5, CP6, FC1, FC2, CP4, C5, C6, FC3, FC4, Pz, C1, C2 and CP3 in the 10-20 system). They were referenced to the left earlobe and grounded to AFz. The data was sampled at 512 Hz, and processed online using OpenViBE 3.4.0 [13].

During the user training runs, the online data used to provide feedback was processed as follows. We first selected the signal from electrodes C3, FC1, FC5, CP1 and CP5, which was then filtered between 8 Hz and 20 Hz. The signals were then passed through a Laplacian filter centered on C3 (with electrodes CP1, CP5, FC1 and FC5). During the resting phases of the runs, the output signal was then epoched using a one second window every 0.1 second. The power over the 8-20Hz frequency band was computed, time-averaged, and the data from the epochs of the last two seconds of the resting time (20 epochs) was averaged. This average was used as the *Rest* value to compute the ERDs of the following MI phase. During the MI phases of the runs, the output signal of the Laplacian filter was epoched using a 0.25 s window every 0.25 s, epochs whose signal's power was computed, time-averaged, and for every epoch, the data from the current epoch and last three epochs were averaged. For every epoch, this average was used as the *Task* value to compute the online ERDs ( $ERD_{on}$ ) as follows:

$$ERD_{on} = (Task - Rest) / Rest * 100 \quad (1)$$

We then used the  $ERD_{ref}$  defined from the calibration run

to compute an ERD score ( $S_{ERD}$ ):

$$S_{ERD} = ERD_{on} / ERD_{ref} * 100 \quad (2)$$

The  $S_{ERD}$  was then used to define the feedback score ( $S_{FB}$ ) according to the following thresholds:

$$\begin{aligned} S_{ERD} < 30\% &\Rightarrow S_{FB} = 0 \\ S_{ERD} \geq 30\% &\Rightarrow S_{FB} = 1 \\ S_{ERD} \geq 60\% &\Rightarrow S_{FB} = 2 \\ S_{ERD} \geq 100\% &\Rightarrow S_{FB} = 3 \end{aligned} \quad (3)$$

For the offline analysis, the EEG data has been pre-processed using MNE-Python [14]. The signal was filtered using a zero-phase shift notch filter with a 50 Hz cut-off frequency and a finite impulse response (FIR) band-pass filter with cut-off frequencies of 1 and 25 Hz and then average-referenced. We used an independent component analysis (ICA) to limit the impact of muscular artefacts. In average, 3 components were removed from the analysis of each participant. We extracted 14 s window epochs from 4 s before the MI instruction cue to 10 s after (one epoch per trial). Epochs with peak-to-peak amplitude greater than 200  $\mu V$  were rejected. In total, an average of 1 epochs out of 20 were removed for each run. The data was then filtered using a Laplacian filter centered on C3 (with electrodes CP1, CP5, FC1 and FC5). Then, we re-sampled our data at 256 Hz, computed time-frequency representation using Morlet wavelets between 8 and 20 Hz, and normalized it with baseline correction by taking the logratio of the signal over the average power during the rest period (-4 s to -1 s before cue). This gave us the power relative to rest period. Then, we averaged this power between 1 s and 9 s post-cue and across trials to obtain  $ERD_{off}$ .

**Visual Feedback:** The visual feedback (developed using Unity 2019.4.18f1) presented to participants during the modalities V and VT consisted of a right virtual hand superimposed over the participants' real right hand. The virtual hand performed wrist rotations to go towards or away from a virtual cup containing a steaming hot beverage placed on a table. The virtual scene would take place in a chalet with a view of a snowy environment. There was a 60° rotation range between the starting position of the hand and the mug. The  $S_{FB}$  i.e., 0, 1, 2 or 3, corresponded to different rotation speeds of the virtual hand, i.e., -2°/s, 4°/s, 6°/s or 10°/s respectively. The virtual hand could not move further back than the starting position. Participants were informed that the better the MI task was performed, the closer to the mug the hand would move and that independently from their brain activity, the virtual hand would continuously open and close (2s period). During the resting period, a white cross was displayed on the screen. During the calibration and for the runs with thermal-only feedback modality, a white frame was additionally displayed on the border of the screen to inform the participants when they should perform MI tasks.

**Thermal Feedback:** To adapt the thermal stimulation range to participants' perception, we calibrated the device

at the beginning of the every experiment using the participant's (i) minimum warmth threshold perception and (ii) potential uncomfortable warmth perception. The temperatures were chosen among the following predefined range [21.5°C (room temperature), 38°C]. This range was chosen based on previous experiments and should not induce painful temperatures [15]. On average (*mean ± standard deviation*), participants defined the thresholds as follows. Lower threshold : 25.9°C ± 1.9°C, upper threshold: 34.8°C ± 3.0°C). The thermal feedback delivered to the participant during the T and VT conditions was then bounded according to their individually defined thresholds, i.e., [minimum warmth threshold perception-1°C, potential uncomfortable warmth perception-1°C]. With the goal of having consistent visual and tactile feedback modalities, the thermal feedback was commanded by the angle between the hand and the mug according to the following functions (see Section *Participants*):

$$T_{com} = T_{com,max} * \theta / \theta_{max} + (T_{com,min} - 11) \quad (4)$$

$$T_{com} = (T_{com,max} - T_{com,min}) * \theta / \theta_{max} + T_{com,min} \quad (5)$$

$$T_{stim} = T_{amb} + T_{com} * 0.15 \quad (6)$$

where  $T_{com}$ ,  $T_{com,min}$  and  $T_{com,max}$  are the current, minimum and maximum stimulation commands (respectively) sent to the Peltier system,  $\theta$  and  $\theta_{max}$  the current and maximum (i.e. 60°) angle between the hand and the mug (respectively), and  $T_{stim}$  and  $T_{amb}$  (21.5°C) the stimulation temperature and room temperature respectively. Participants were informed that the better the MI task was performed, the warmer the thermal stimulation would be. The Peltier cell temperature was controlled through an Arduino Uno R3 (Arduino.cc) receiving the  $\theta$  angle information from the Unity application.

**Variables & Factors:** Our analyses focused on assessing the potential difference induced by the modalities of feedback on the electrophysiological changes resulting from the NFB user training. To that extent, we used three different variables. The previously described  $ERD_{on}$  and  $ERD_{off}$  (see Section *EEG Recordings & Signal Processing*), and a NFB performance variable,  $P_{FB}$ , based on the feedback produced during a run and defined as the sum of all the  $S_{FB}$  produced during a run divided by the maximum total score (i.e. all  $S_{FB} = 3$ ) :

$$P_{FB} = \frac{\sum_{i=1}^N S_{FB,i}}{3 \times N} \times 100 \quad (7)$$

where  $N$  is the total number of online epochs thus of  $S_{ERD}$  computed during a run and  $S_{FB,i}$  the  $S_{FB}$  of the  $i^{th}$  epoch. The statistical analysis of these values consisted of repeated measures ANOVAs, with associated post-hoc analyses using the false discovery rate (FDR) method to correct for multiple comparisons. Our goal was to study the effects of the *modality* of feedback (i.e., V, T and VT) and of the *run* (i.e., Run1 and Run2) on our variables ( $ERD_{on}$ ,  $ERD_{off}$ ,  $P_{FB}$ ).

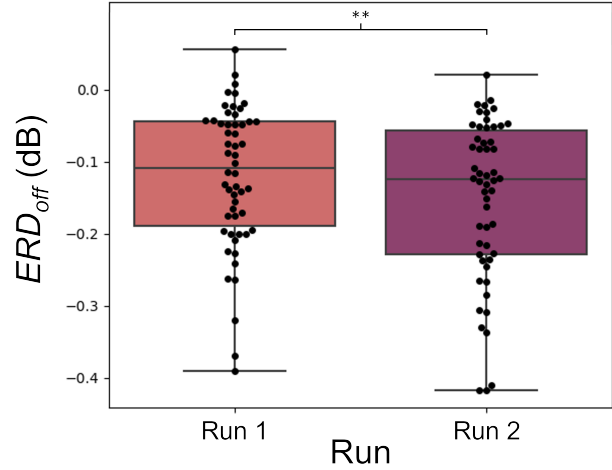


Figure 3: Box and scatter plot of  $ERD_{off}$  for runs *Run1* and *Run2*. Asterisks:  $p < 0.05$ : \* ;  $p < 0.01$ : \*\* ;  $p < 0.001$ : \*\*\*.

## RESULTS

We analysed the potential influence of our conditions on the three variables defined earlier:  $P_{FB}$ ,  $ERD_{on}$ , and  $ERD_{off}$ . As repeated measures ANOVA results are particularly sensitive to outliers, we removed the results from the participants that were over plus or minus two standard deviations relative to the median of this variable. In total, two participants were removed from  $P_{FB}$  analyses, five from  $ERD_{off}$ , and one participant from  $ERD_{on}$  analyses. We tested our data for deviation of normality using a Shapiro-Wilk test. No significant deviation was found. Then, using JASP [16] we computed three 2-way repeated measures ANOVAs to assess if there was an effect of or an interaction between “Modality” (V, T, VT) and “Run” (Run1, Run2) on either of the dependent variables ( $P_{FB}$ ,  $ERD_{on}$ ,  $ERD_{off}$ ).

We found a significant main effect of *run* repetition on  $ERD_{on}$  and  $ERD_{off}$ , but not on  $P_{FB}$  (Tab. 1).  $ERD_{on}$  were found significantly greater during Run2 ( $M = -46.9\%$ ;  $SD = 13.5\%$ ) than during Run1 ( $M = -44.9\%$ ;  $SD = 13.9\%$ ) [ $t = 3.67$ ,  $df = 21$ ,  $p = 0.001$ ].  $ERD_{off}$  were also found significantly greater during Run2 ( $M = -0.15$ dB;  $SD = 0.11$ dB) than during Run1 ( $M = -0.12$ dB;  $SD = 0.10$ dB) [ $t = 3.467$ ,  $df = 17$ ,  $p = 0.003$ ] (Fig. 3). We found a significant main effect of feedback *modality* on  $ERD_{off}$ , but not on  $ERD_{on}$  nor on  $P_{FB}$  (Tab. 1). Post-hoc analyses of the effect of “Modality” revealed **significantly greater  $ERD_{off}$  for V** ( $M = -0.18$ dB;  $SD = 0.12$ dB) **compared to T** ( $M = -0.09$ dB;  $SD = 0.08$ dB) [ $t = -3.993$ ,  $df = 17$ ,  $p < 0.001$ ], and **significantly greater  $ERD_{off}$  for VT** ( $M = -0.14$ dB;  $SD = 0.10$ dB) **compared to T** [ $t = -2.444$ ,  $df = 17$ ,  $p = 0.040$ ]. No significant difference was found between V and VT [ $t = -1.549$ ,  $df = 17$ ,  $p = 0.131$ ] (Fig. 4). P-values were Holm-corrected for multiple comparisons.

## DISCUSSION

The aim of the present study was to investigate the feasibility of introducing thermal stimulation as a new neu-

	$ERD_{on}$	$ERD_{off}$	$P_{FB}$
<i>Modality</i>	$F(2,42)=0.12;p=0.83;\eta^2=0.004$	<b><math>F(2,34)=8.1;p=0.001;\eta^2=0.23</math></b>	$F(2,40)=0.93;p=0.4;\eta^2=0.027$
<i>Run</i>	<b><math>F(1,21)=13.4;p=0.001;\eta^2=0.05</math></b>	<b><math>F(1,17)=12;p=0.003;\eta^2=0.05</math></b>	$F(1,20)=3.16;p=0.09;\eta^2=0.01$
<i>Modality*Run</i>	$F(2,42)=1.61;p=0.21;\eta^2=0.016$	$F(2,34)=0.6;p=0.55;\eta^2=0.005$	$F(2,40)=1.05;p=0.36;\eta^2=0.02$

Table 1: Result of the two-way repeated measures ANOVA for *Modality* and *Run*. Significant main effects are indicated in bold.

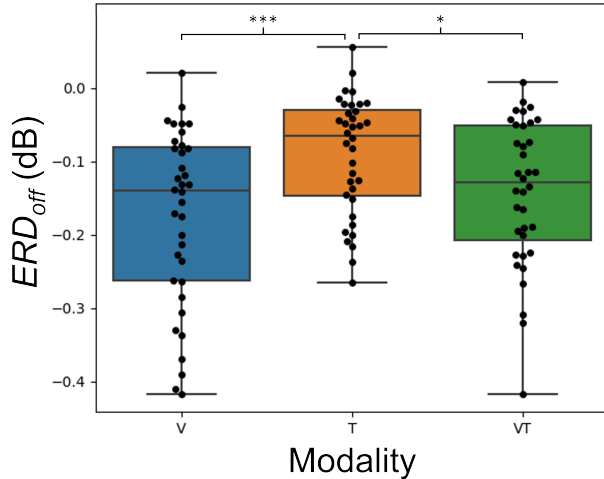


Figure 4: Box and scatter plot of  $ERD_{off}$  for modalities V, T, and VT. Asterisks:  $p<0.05$ : \*;  $p<0.01$ : \*\*;  $p<0.001$ : \*\*\*.

rofeedback modality. To that extent, we compared neurophysiological outcomes of MI-based NFB between a thermal-only feedback modality, a visual-only feedback modality, and a visuo-thermal modality (combining both aforementioned feedback modalities). Through repeated measures ANOVAs, we studied the effects of the feedback *modality* and *run* repetition.

We found that *run* repetition had a significant main effect on NFB performances ( $ERD_{on}$  and  $ERD_{off}$ ), indicating a learning effect. We also found that *modality* had a significant main effect on offline performances (i.e.,  $ERD_{off}$ ) with significantly greater desynchronization over the left sensorimotor cortex for the visual-only and visuo-thermal modalities compared to the thermal-only modality.

The main effect of *modality* of feedback on brain activity modulation was observed on the offline NFB performance ( $ERD_{off}$ ) but not on the online NFB performance ( $ERD_{on}$ ,  $P_{FB}$ ). Two main differences between the online and offline NFB performances most likely explain such differences. Firstly, the offline performances benefit from the ICA muscle activity artifacts correction, which is not applied online. Secondly, during online processing the positive values of  $ERD_{on}$ , i.e., event related synchronization (ERS), were automatically set to 0% not to give negative feedback to users, whereas offline processing took into account both ERD and ERS. Thus, feedback *modality* shows a significant effect when both ERD and ERS are taken into account in the processing ( $ERD_{off}$ ) but not when only ERD are ( $ERD_{on}$ ). This suggests that feedback modality has mainly an effect on ERS, with V and VT modalities resulting in less ERS than their thermal counterpart.

Thermal feedback alone thus appears less effective than

its visual counterpart for MI-NFB learning. A first hypothesis to explain this result may lie in the difference of intuitiveness between both feedback. Considering the visual feedback, participants could see the starting and goal position of the hand (i.e. the mug) and thus appreciate at all times their performance based on the position of the hand between these two positions. On the contrary, during thermal-only NFB runs, participants could use the room temperature as a starting point, but did not have any reference cue related to the goal temperature to reach. Furthermore, thermal noticeable difference, i.e., the minimum temperature difference between two stimuli that a person is able to perceive, may reach a minimum of  $0.75^{\circ}\text{C}$  at  $38^{\circ}\text{C}$ , but rapidly increases for lower temperature ( $1.6^{\circ}\text{C}$  at  $32^{\circ}\text{C}$ ) [17]. Thus, considering the temperature range we used and that a prior characterisation of our thermal stimulation system gave a temperature variation speed of around  $1^{\circ}\text{C/s}$ , the perceived feedback variation experienced by participants during a 10 s task was very likely lower for thermal stimulation than for visual stimulation. This is corroborated by comments from two participants: one claimed to have felt a delay between their MI and the thermal feedback, and the other participant mentioned having a hard time figuring out if they were performing the MI task correctly with the thermal-only feedback condition and even feeling lost without visual feedback. Both aforementioned arguments suggest that visual feedback alone contains a richer information than thermal feedback alone. Despite eliciting lower modulation of brain activity than its visual counterpart and its probable lack of intuitiveness, adding thermal sensations to visual feedback did not decrease the efficiency of the visual-only feedback, as we did not find any significant difference in brain activity modulation between V and VT conditions. A stronger desynchronization could have been expected for the visual and thermal condition as previous studies found that multimodal feedback composed of haptic and visual stimuli has a beneficial influence [9, 10]. This could be caused by an inconsistency between our visual and thermal feedback due to an increased delay in thermal stimulation compared to the visual one. Our results suggest that using thermal feedback coupled with a visual feedback could be an interesting solution, especially for NFB application that would benefit from thermal stimulation such as sensory and motor post-stroke rehabilitation as previously mentioned [11, 12]. Further analyses of the user experience will provide us with complementary information.

Finally, it must be mentioned that including a repetitive grasping movement of the virtual hand in the visual feedback might have brought about the effect of action ob-

servation (AO) from participants and thus enhanced the ERD amplitude in those conditions, as recently reported by Nagai et al. [18]. Indeed, they found that participants produced significantly greater ERDs during MI of fist clenching while synchronously looking at a video of someone else's hand performing the movement compared to pure motor imagery (without AO). Future investigation of the *Replay* runs should enable us to quantify the respective influence of AO and MI on the ERDs amplitude.

## CONCLUSION

In conclusion, this study sheds light on the potential of thermal feedback as a novel modality for motor imagery-based neurofeedback training. We found that it elicits a less pronounced modulation of the left sensorimotor cortex compared to visual feedback alone and to combined visuo-thermal feedback. Nevertheless, adding thermal feedback to other feedback modalities could remain interesting as it does not negatively affect NFB performances when combined with visual feedback. Overall, this study contributes to our understanding of feedback mechanisms in BCI and highlights the importance of considering alternative modalities in pursuit of more intuitive and effective human-computer interaction paradigms. The investigation into thermal feedback opens new avenues for improving neurotechnologies, particularly in clinical applications, such as sensorimotor rehabilitation. By addressing the challenges associated with MI tasks, we can bring further the development of NFB and BCIs and broaden their practical applications. Future research may further refine thermal feedback protocols and explore its potential synergies with other modalities to optimize user experience and BCI performances.

## ACKNOWLEDGEMENT

This work was supported by the GRASP-IT project (ANR-19-CE33-0007, January 2020) and PEPERONI project (CominLabs, 2022-2024). We would also like to thank Justine Saint-Aubert, Jeanne Hecquard and Kyung-Ho Won for their help in the technical development of the setup, as well as all the participants.

## REFERENCES

- [1] Kerous B, Skola F, Liarokapis F. Eeg-based bci and video games: A progress report. *Virtual Reality*. 2018;22:119–135.
- [2] Le Franc S *et al.* Toward an adapted neurofeedback for post-stroke motor rehabilitation: State of the art and perspectives. *Frontiers in Human Neuroscience*. 2022;16:917909.
- [3] Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans. *Neuroscience letters*. 1997;239(2-3):65–68.
- [4] Pillette L. Redefining and adapting feedback for mental-imagery based brain-computer interface user training to the learners' traits and states. Ph.D. dissertation. Bordeaux. 2019.
- [5] Fleury M, Lioi G, Barillot C, Lécuyer A. A survey on the use of haptic feedback for brain-computer interfaces and neurofeedback. *Frontiers in Neuroscience*. 2020;14:524967.
- [6] Cincotti F *et al.* Vibrotactile feedback for brain-computer interface operation. *Computational intelligence and neuroscience*. 2007;2007.
- [7] Gwak K, Leeb R, Millán JdR, Kim DS. Quantification and reduction of visual load during bci operation. In: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2014, 2795–2800.
- [8] Lukyanov M *et al.* The efficiency of the brain-computer interfaces based on motor imagery with tactile and visual feedback. *Human Physiology*. 2018;44:280–288.
- [9] Barsotti M, Leonardis D, Vanello N, Bergamasco M, Frisoli A. Effects of continuous kinaesthetic feedback based on tendon vibration on motor imagery bci performance. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2017;26(1):105–114.
- [10] Pillette L, N'kaoua B, Sabau R, Glize B, Lotte F. Multi-session influence of two modalities of feedback and their order of presentation on mi-bci user training. *Multimodal Technologies and Interaction*. 2021;5(3):12.
- [11] Chen JC, Liang CC, Shaw FZ. Facilitation of sensory and motor recovery by thermal intervention for the hemiplegic upper limb in acute stroke patients: A single-blind randomized clinical trial. *Stroke*. 2005;36(12):2665–2669.
- [12] Wu HC, Lin YC, Hsu MJ, Liu SM, Hsieh CL, Lin JH. Effect of thermal stimulation on upper extremity motor recovery 3 months after stroke. *Stroke*. 2010;41(10):2378–2380.
- [13] Renard Y *et al.* Openvibe: An open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. *Presence*. 2010;19(1):35–53.
- [14] Gramfort A *et al.* Meg and eeg data analysis with mne-python. *Frontiers in neuroscience*. 2013;7:70133.
- [15] Darian-Smith I, Johnson KO. Thermal sensibility and thermoreceptors. *Journal of Investigative Dermatology*. 1977;69(1):146–153.
- [16] JASP Team. JASP (Version 0.18.3)[Computer software]. 2024. [Online]. Available: <https://jasp-stats.org/>.
- [17] Bunk SF, Lautenbacher S, Rüsseler J, Müller K, Schultz J, Kunz M. Does eeg activity during painful stimulation mirror more closely the noxious stimulus intensity or the subjective pain sensation? *Somatosensory & motor research*. 2018;35(3-4):192–198.
- [18] Nagai H, Tanaka T. Action observation of own hand movement enhances event-related desynchronization. *IEEE transactions on neural systems and rehabilitation engineering*. 2019;27(7):1407–1415.

# DOUBLE-BLIND AND SHAM-CONTROLLED AUGMENTED REALITY EEG-NEUROFEEDBACK STUDY

L. M. Berger, G. Wood, S. E. Kober

Department of Psychology, University of Graz, Austria

E-mail: lisa.berger@uni-graz.at

**ABSTRACT:** Traditional Neurofeedback (NF) designs are rather dull and only little engaging, which can negatively influence training performance. NF profits from interesting paradigms implementable through tools such as Virtual and Augmented Reality (AR). AR, however, is still very new in the field of NF and BCI but seems promising in hindsight of less Cybersickness and easier and cheaper usage for tele-rehabilitation, as modern smartphones support AR implementations. However, there are still no sham-controlled AR-based NF studies with larger samples. We propose a one-session sham-controlled and double-blinded NF study comparing AR with 2D feedback. The NF training consisted of sensorimotor rhythm (SMR) up-regulation and we tested 89 healthy participants. Results showed a numerically but non-significant increase in SMR across the NF runs in all four groups. Sham and real feedback groups did not differ in their performance. The study could show that AR is equally viable to 2D feedback and participants were not able to increase SMR within one NF training session.

**Keywords:** Augmented Reality, Neurofeedback, Sensorimotor Rhythm

## INTRODUCTION

Since about 30% of Brain-Computer Interface (BCI) users are not able to alter their own brain activation [1], neurofeedback (NF) and BCI alike profit a lot from engaging and interesting paradigms in order to increase training adherence of the users. Psychological factors such as motivation [2] and attention [3] from the users are positively associated with the NF training success. Virtual Reality (VR) has in respect thereof already shown to be beneficial to increase NF performance [4] over simple traditional 2D paradigms presented on computer screens. Further, stroke patients undergoing VR NF training also reported a high motivation to continue their training and showed high interest [5]. However, VR presents some downsides in its usage as well. About 80% of the users of VR systems are prone to develop symptoms of Cybersickness, such as nausea, oculomotor problems or disorientation [6], which can on the one hand lead to a worse user experience and on the other hand to a worse training outcome [7]. Also, VR is costly on resources to offer for a group of patients on a tele-rehabilitation basis, and there are only little options of combined VR-EEG systems. Another option that is posing some advantages over VR but is still only little

researched in the field of NF and BCI is the application of Augmented Reality (AR). Here, virtual objects are superimposed on real world surroundings. On the reality-virtuality continuum proposed by Milgram and Kishino, AR is classified as closer to reality than VR [8]. The implementation is done by using, e.g., smartphones, webcams, or stereoscopic camera additions to VR goggles. An advantage of AR is that it is less prone to result in the feelings of Cybersickness [9] in the user. They experience less nausea and disorientation during the usage. Further, AR seems more easily applicable for tele-rehabilitation purposes, as most modern smartphone cameras support AR (official list: <https://developers.google.com/ar/devices>). So here no expensive combinational devices would be necessary. AR enables new opportunities to integrate feedback and bodily features can still be visible and can if necessary be included to the feedback. It hence offers the creation of adaptive paradigms that support the training. [10] However, there is a lack of double-blind sham-controlled AR-NF studies and only a handful of studies are using AR in NF settings. In one study from 2014 the researchers created the MindMirror, using a Webcam with a virtual overlay to simulate an AR setting [11]. Participants would see themselves on a computer screen over a webcam with a virtual brain overlay presented on their heads. For the training relevant areas would light up in different colors. Viczko et al. used an AR NF paradigm for a NF meditation application using an Apple iPhone and Emotive headband for the feedback. It showed butterflies hatching from crystals when brain activity reached the desired state and could then be followed with the phone camera as an interactive element [12]. Also, there are several proof-of-concept studies with relatively small samples (5-12 participants in total [13, 14] and no sham-control groups, combining for example steady-state visual evoked potential (SSVEP) based BCI systems [15]. The presented studies offered a first insight into combinational AR-BCI and NF studies. Altogether, studies with big sample sizes and double-blinded sham-control are still missing. Here, we conducted an EEG-based NF study comparing AR feedback with conventional 2D bar feedback design with regard to NF performance (measured as SMR increase over the course of six feedback runs). We hypothesized that participants from the AR group would perform better in a SMR-NF task than participants undergoing a 2D NF with a conventional paradigm. Since VR paradigms have previously been shown to result in better NF performance

compared to 2D paradigms [4, 16] and AR has previously been successfully tested in the field of BCI and rated very positively by participants [11], we expected similar positive NF results for AR-feedback compared to 2D feedback. Also, we assumed participants receiving real feedback would perform better than the feedback group receiving sham feedback. Previous findings in the general field of NF could show beneficial training effects specifically for real feedback groups compared to sham feedback groups [17, 18]. Hence, we expected similar results in our sample with AR-based NF.

## MATERIALS AND METHODS

The study was conducted at the University of Graz. All participants gave written informed consent before the start of the measurement. The ethics committee of the University of Graz, Austria, approved all aspects of the present study in accordance with the Declaration of Helsinki (GZ. 39/119/63 ex 2021/22).

**Participants:** In total, 100 participants were tested (see Table 1). Eleven datasets had to be excluded from the statistical analysis due to bad EEG data quality, problems with the paradigm and drop-outs, hence, 89 datasets survived for further analysis. Forty-four participants performed the 2D NF task, 45 the AR task (see Table 1). All volunteers had normal or corrected-to-normal vision, no neurological, psychological or other severe diseases, as well as no reflex epilepsy. They gave written informed consent and were either paid for their participation (16€) or received research credit hours for their Psychology Bachelor program.

Table 1: Description of the sample.

	AR		2D	
	real	sham	real	sham
N (female)	25 (14)	20 (12)	20 (9)	24 (17)
Mean age	23.76	24.95	24.65	21.79
(SD)	(3.31)	(3.82)	(4.06)	(1.74)
Responder	16	12	10	14
Non-Resp.	9	8	10	10
	(36%)	(40%)	(50%)	(41.6%)

**Neurofeedback-training:** Participants were pseudo-randomized and assigned to one of the four groups: 2D vs. AR feedback, real vs. sham feedback and experimenters just as participants were blinded whether real or sham feedback was given. In the real feedback condition participants got their real brain activation fed back in real time, while in the sham feedback condition the brain activation from another person of another (similar) study was fed back [4]. In the AR condition, participants would see three virtual plants growing out of real plant-pots placed in front of them. The middle one represented the sensorimotor rhythm (SMR; 12-15 Hz) recorded over Cz and the two outer plants Theta (4-7 Hz) and Beta (16-30 Hz), also recorded over electrode position Cz. Participants should make the middle one grow as high as possible and keep the two outer ones as low as possible. The same principle was followed by the

2D paradigm, only they saw three 2D bars equally representing the three frequency bands regularly on a PC screen (see Figure 1). The training consisted of a baseline run and six training runs of three minutes each. In the baseline run, participants were instructed to watch the moving objects without trying to influence it. Afterwards, individual threshold values were calculated based on this baseline activation. For SMR the mean values were calculated and for Theta and Beta the mean plus one standard deviation was calculated. The thresholds were adapted in the paradigm for the training after each run and participants were instructed to be physically relaxed and mentally focused to control the feedback objects.

**Technology:** The AR paradigm was presented via the HTC Vive Pro VR-System. The stereoscopic camera ZED mini from Stereolabs was attached to the VR goggles to enable AR vision. The SDK Unity Plugin version 3.8.0 was used to create the environments in Unity, Version 2020.3.30f1 (see Figure 2). For superimposing the virtual objects in the AR setting via markers, the free Unity trial version of OpenCV (version 2.4.8) was used. For real-time EEG data streaming the LSL4Unity plugin, freely available at <https://github.com/labstreaminglayer/LSL4Unity> was used in combination with OpenViBE, Version 3.3.0. OpenViBE is a free software to stream and preprocess EEG data in real-time. The framerate of both the camera and computer screen were set to 60 FPS. Even though the 2D group got their paradigm presented on a computer screen, they also had to wear the VR-AR system to rule out any group differences related to wearing the system, such as headache or pressure sensations due to wearing the whole system. Here, the camera was simply switched on, so participants would also see their surroundings through the camera.

**EEG recording and Offline EEG data processing:** Data was recorded with the gUSBamp RESEARCH EEG-amplifier from g.tec medical engineering and a sampling rate of 500Hz. We used 16 sintered Ag/AgCl passive ring electrodes, placed according the 10-20 EEG-system, to measure the signal. All electrodes were referenced against left mastoid and the ground was placed at FPz. A right mastoid placed electrode was used to calculate linked mastoid reference during the offline EEG data processing. Brain Vision Analyzer (version 2.2, Brain Products GmbH, Munich, Germany) was used for offline EEG-data processing. At first, a 50 Hz notch filter and a low cutoff filter of 0.01, as well as a high cutoff filter of 100 Hz were applied. Further, big muscle artifacts were excluded and heavy drifts during the raw data inspection. Data was referenced to the linked mastoid reference to rule out hemisphere effects, as the left mastoid was the primary reference electrode. Next, a semi-automatic independent component analysis (ICA) was performed to eliminate blinks and eye movements using a semi-automatic independent component analysis (ICA). Lastly, a second semi-automatic data inspection followed to exclude additional remaining artifacts that survived the other

preprocessing steps (Criteria for rejection: maximum allowed voltage step of  $50\mu\text{V}/\text{ms}$ , maximum allowed difference between values in a segment was  $200\mu\text{V}$ , amplitudes  $\pm 120\mu\text{V}$ , lowest allowed activity in 100 ms intervals was  $0.5\mu\text{V}$ , artifacts were marked 200 ms before and after emergence). Finally, the frequency power bands in the ranges 12-15 Hz (SMR), 4-7 Hz (Theta) and 16-30 Hz (Beta) were extracted using complex demodulation. Data was segmented into 1 s intervals and segments with artifacts were removed.



Figure 1: The used AR-Set-up with the HTC Vive Pro and attached ZEDmini stereoscopic camera. On the table one can see the plant pots and the screen shows the virtual plants that the participant is seeing via the VR-goggles.

**Questionnaires:** In this study we also assessed the user experience of participants with several questionnaires on Cybersickness, technology anxiety, subjective control among others. Results are presented in another study which is currently under submission.

**Statistical Analysis:** To investigate the NF performance (measured as the changes in SMR power across six NF runs) of the four different groups (AR real vs. AR sham and 2D real vs. 2D sham), a linear mixed effect model with three fixed factors (group, condition, feedback runs) was calculated for the dependent variable SMR power over electrode position Cz (Type I Sum-of-Squares Analysis of Variance with Satterthwaite's method). Here, we will only present the findings for SMR. To enable a better interpretation of the results we split the factor runs in two groups, one for the first training half (first three runs) and one for the second training half (runs four to six). The factor subject was included in the model as crossed random effect.

To identify non-responders, we checked whether regression slopes were increasing or decreasing. They were determined by calculating a regression with SMR power as criterion and feedback run number as predictor. Positive slopes indicate a linear increase, showing a successful training and negative slopes a linear decrease, showing an unsuccessful training. To investigate whether non-responders are equally distributed between all four groups we calculated a Chi-Squares test.



Figure 2: Traditional 2D paradigm presented on PC screen.

## RESULTS

The results of the linear mixed effect model showed no group-differences of AR/2D or real/sham feedback groups (for F-statistics see Table 2). Although SMR power increased numerically over the training runs, the main effect Runs was not significant (Table 2, Figure 3).

Table 2: F-statistics of the Linear Mixed Effect Model with Group (AR/2D), Condition (real/sham), and Runs as fixed factors.

Object	<i>F</i>	<i>df</i>	<i>p</i>
Group	0.32	1,234.31	.572
Runs	0.01	1,437.25	.925
Condition	0.01	1,234.73	.928
Group*runs	0.03	1,437.21	.855
Group*condition	0.07	1,233.65	.788
Runs*condition	0.16	1,437.21	.691
Group*runs*condition	0.12	1,437.16	.730

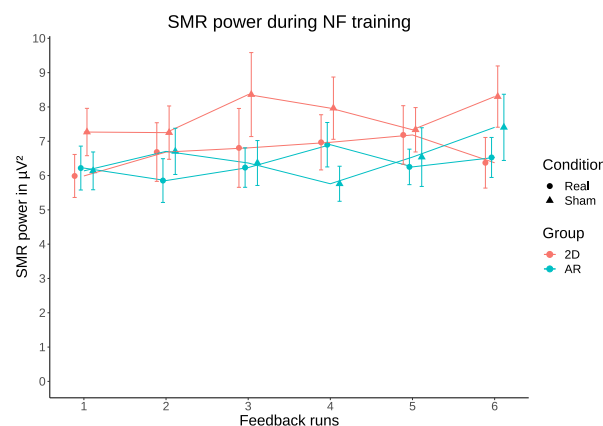


Figure 3: Line graph showing the training performance of all the four groups over the course of the six feedback runs. Error bars are indicating the standard error.

**Responder/Non-responder:** In our sample 41.6% of the users were not able to increase their brain activation across the six feedback runs (see Table 1). All groups had a similar number of non-responders ( $\chi^2=0.93$ ,  $p = .819$ ).

## DISCUSSION

We conducted a study, comparing AR and 2D based

SMR-NF training in a double-blinded pseudo-randomized and sham-controlled study with a large sample size ( $N=89$ ). Statistical analyses revealed no group differences in the NF performance, all four groups showed a comparable small but non-significant increase in the target frequency band over the course of six feedback runs.

In the present study, we did not find any differences in NF performance between the AR and 2D condition within one NF training session. Hence, AR-based NF training had no beneficial effects over conventional feedback designs. Previous studies also found comparable results of conventional and AR NF-paradigms. In a one-session user study from Mercier-Ganady and colleagues (2014) using a webcam to overlay a virtual brain on the users' heads also showed comparable results between the AR and a conventional 2D training, where they used a representation of a temporal gauge. Participants reported that their AR paradigm was less clear but more innovative and original. The EEG results were equal for both groups [11]. Hence, it seems as if only one NF training session was not enough to reveal possible beneficial effects of AR-based NF training. It would be interesting to compare training results over a longer period with trainings consisting of multiple sessions instead of only one to see, whether AR-based feedback might be beneficial for NF training performance over a longer training period.

Further, both of our paradigms were rather similar, each showing three objects growing and shrinking in size. This might explain the similar results for both conditions. Other studies in this field had either no control group [19] or control groups where the outcome or visual feedback differ fundamentally between the two groups. For instance, previous studies investigated the effects of AR-based meditation with and without NF on mood [12], or compared a visual feedback where participants should light up an AR brain-overlay with a temporal gauge as feedback [11]. Future studies should try explore more AR-given possibilities of design and interaction. In a further study one could for example expand the flower idea we used in the current study but for example make the task to turn the whole laboratory into a flowerbed surrounding the participants.

Mohammed et al. speculate that AR-BCI studies would result in a higher cognitive load as it being more cognitively demanding [10]. However, AR still serves less visual distraction than does VR with its surrounding virtual environments. In a review on the impact of AR on tasks performances and cognitive load it could be shown that when using AR to complete different tasks, participants have less or equal cognitive load, as well as a higher performance than those using conventional methods [20]. It remains open, whether these results also reflect AR usage in NF. To overcome the problem of complexity for the participants it might be beneficial to implement introductory sessions where participants could familiarize with the systems.

Further, we found no group differences between real and sham feedback groups. Even though one would expect

better NF performance in real feedback groups compared to sham groups, it is not uncommon to find comparable results in the NF literature especially when performing only one NF training session [21]. Ninaus et al. (2013) could show in an fMRI based NF study that both participants from the real and the sham feedback group showed similar active neural networks [22]. Participants and experimenters were blinded concerning group allocation and were instructed the same way. Hence, both groups were instructed to be mentally focused and physically relaxed. In a review on ADHD and neurofeedback researchers propose that effects from neurofeedback need more time in terms of more training sessions to develop. It is possible that during familiarizing with the task in the first 20-30 minutes, attention and concentration increased naturally, leading to unspecific changes in EEG activity, which are not related to real NF conditions [23].

Finally, the number of non-responders in this study is a bit higher than in other studies, where mostly 30% are reported [1, 24]. However, non-responders of up to 50% can be observed in the literature [25]. The AR group did not have less non-responders than the 2D group and the number also did not differ between sham and real feedback groups. It is difficult to determine non-responders within a single training session, as it is still an intensive learning process and for most participants it was their first NF training. Here unspecific factors might also play a big role, especially with a rather complex set-up used in our study. It would also be interesting to see whether the number of non-responders within the groups would change after several training sessions.

## CONCLUSION

In the current study we did not find any differences in NF performance between an AR NF training and a conventional 2D feedback within one NF training session. Expanding the AR paradigm to a more complex task should shed more light on paradigm differences on the NF performance in future studies. Also, sham and real feedback led to a comparable NF performance.

## References

- [1] Allison BZ, Neuper C. Could Anyone Use a BCI? In: Tan DS, Nijholt A, editors. Brain-Computer Interfaces. London: Springer London 2010; 35–54.
- [2] Kleih SC, Nijboer F, Halder S, Kübler A. Motivation modulates the P300 amplitude during brain-computer interface use. Clin Neurophysiol 2010; 121(7): 1023–31 [https://doi.org/10.1016/j.clinph.2010.01.034][PMID: 20188627]
- [3] Hammer EM, Halder S, Blankertz B, et al. Psychological predictors of SMR-BCI performance. Biol Psychol 2012; 89(1): 80–6 [https://doi.org/10.1016/j.biopsycho.2011.09.006][PMID: 21964375]
- [4] Berger LM, Wood G, Kober SE. Effects of virtual reality-based feedback on neurofeedback training

- performance-A sham-controlled study. *Front Hum Neurosci* 2022; 16: 952261  
[https://doi.org/10.3389/fnhum.2022.952261][PMID: 36034118]
- [5] Kober SE, Reichert JL, Schweiger D, Neuper C, Wood G. Effects of a 3D Virtual Reality Neurofeedback Scenario on User Experience and Performance in Stroke Patients. In: Bottino R, Jeuring J, Veltkamp RC, editors. *Games and Learning Alliance*. Cham: Springer International Publishing 2016; 83–94.
- [6] Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. *Virtual Reality* 2016; 20(2): 101–25  
[https://doi.org/10.1007/s10055-016-0285-9]
- [7] Berger LM, Wood G, Kober SE. Influence of a placebo tDCS treatment on Cybersickness and EEG-Neurofeedback success. *Behav Brain Res* 2024; 114917  
[https://doi.org/10.1016/j.bbr.2024.114917][PMID: 38401602]
- [8] Skarbez R, Smith M, Whitton MC. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Front. Virtual Real.* 2021; 2  
[https://doi.org/10.3389/frvir.2021.647997]
- [9] Hughes CL, Fidopiastis C, Stanney KM, Bailey PS, Ruiz E. The Psychometrics of Cybersickness in Augmented Reality. *Front. Virtual Real.* 2020; 1  
[https://doi.org/10.3389/frvir.2020.602954]
- [10] Si-Mohammed H, Argelaguet F, Casiez G, Roussel N, L, cuyer A. *Brain-Computer Interfaces And Augmented Reality: A State Of The Art*. Verlag der Technischen Universität Graz; 2017.
- [11] Mercier-Ganady J, Lotte F, Loup-Escande E, Marchal M, Lecuyer A. The Mind-Mirror: See your brain in action in your head using EEG and augmented reality. In: *The Mind-Mirror: See your brain in action in your head using EEG and augmented reality*; 2014. IEEE; 33–8.
- [12] Viczko J, Tarrant J, Jackson R. Effects on Mood and EEG States After Meditation in Augmented Reality With and Without Adjunctive Neurofeedback. *Front. Virtual Real.* 2021; 2  
[https://doi.org/10.3389/frvir.2021.618381]
- [13] Benitez-Andonegui A, Burden R, Benning R, Möckel R, Lühns M, Sorger B. An Augmented-Reality fNIRS-Based Brain-Computer Interface: A Proof-of-Concept Study. *Front Neurosci* 2020; 14: 346  
[https://doi.org/10.3389/fnins.2020.00346][PMID: 32410938]
- [14] Huang X, Mak J, Wears A, *et al.* Using Neurofeedback from Steady-State Visual Evoked Potentials to Target Affect-Biased Attention in Augmented Reality. *Annu Int Conf IEEE Eng Med Biol Soc* 2022; 2022: 2314–8  
[https://doi.org/10.1109/EMBC48229.2022.9871982][PMID: 36085716]
- [15] Faller J, Allison BZ, Brunner C, *et al.* A feasibility study on SSVEP-based interaction with motivating and immersive virtual and augmented reality. *arXiv*; 2017.
- [16] Berger AM, Davelaar EJ. Frontal Alpha Oscillations and Attentional Control: A Virtual Reality Neurofeedback Study. *Neuroscience* 2018; 378: 189–97  
[https://doi.org/10.1016/j.neuroscience.2017.06.007][PMID: 28642166]
- [17] Ramos-Murguialday A, Broetz D, Rea M, *et al.* Brain-machine interface in chronic stroke rehabilitation: a controlled study. *Ann Neurol* 2013; 74(1): 100–8  
[https://doi.org/10.1002/ana.23879][PMID: 23494615]
- [18] Schabus M, Heib DPJ, Lechinger J, *et al.* Enhancing sleep quality and memory in insomnia using instrumental sensorimotor rhythm conditioning. *Biol Psychol* 2014; 95: 126–34  
[https://doi.org/10.1016/j.biopsycho.2013.02.020][PMID: 23548378]
- [19] Rajshekar Reddy GS, G.M. L. A Brain-Computer Interface and Augmented Reality Neurofeedback to Treat ADHD: A Virtual Telekinesis Approach. In: *A Brain-Computer Interface and Augmented Reality Neurofeedback to Treat ADHD: A Virtual Telekinesis Approach*; 2020. IEEE; 123–8.
- [20] Buchner J, Buntins K, Kerres M. The impact of augmented reality on cognitive load and performance: A systematic review. *Computer Assisted Learning* 2022; 38(1): 285–303  
[https://doi.org/10.1111/jcal.12617]
- [21] Schabus M, Griessenberger H, Gnjezda M-T, Heib DPJ, Wislowska M, Hoedlmoser K. Better than sham? A double-blind placebo-controlled neurofeedback study in primary insomnia. *Brain* 2017; 140(4): 1041–52  
[https://doi.org/10.1093/brain/awx011][PMID: 28335000]
- [22] Ninaus M, Kober SE, Witte M, *et al.* Neural substrates of cognitive control under the belief of getting neurofeedback training. *Front Hum Neurosci* 2013; 7: 914  
[https://doi.org/10.3389/fnhum.2013.00914][PMID: 24421765]
- [23] Vollebregt MA, van Dongen-Boomsma M, Buitelaar JK, Slaats-Willemse D. Does EEG-neurofeedback improve neurocognitive functioning in children with attention-deficit/hyperactivity disorder? A systematic review and a double-blind placebo-controlled study. *J Child Psychol Psychiatry* 2014; 55(5): 460–72  
[https://doi.org/10.1111/jcpp.12143][PMID: 24168522]
- [24] Thompson MC. Critiquing the Concept of BCI Illiteracy. *Sci Eng Ethics* 2019; 25(4): 1217–33  
[https://doi.org/10.1007/s11948-018-0061-1][PMID: 30117107]
- [25] Khodakarami Z, Firoozabadi M. Psychological, Neurophysiological, and Mental Factors Associated With Gamma-Enhancing

Neurofeedback Success. Basic Clin Neurosci 2020;  
11(5): 701–14  
[<https://doi.org/10.32598/bcn.11.5.1878.1>][PMID:  
33643562]

# INTER-TASK TRANSFER LEARNING BETWEEN UPPER-LIMB MOTOR EXECUTION AND MOTOR IMAGERY

S. Pérez-Velasco<sup>1, 2</sup>, D. Marcos-Martínez<sup>1, 2</sup>, E. Santamaría-Vázquez<sup>1, 2</sup>, V. Martínez-Cagigal<sup>1, 2</sup>,  
B. Pascual-Roa<sup>1</sup>, R. Hornero<sup>1, 2</sup>

<sup>1</sup>Biomedical Engineering Group, E.T.S Ingenieros de Telecomunicación, University of Valladolid,  
Valladolid, Spain

<sup>2</sup>Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina  
(CIBER-BBN), Madrid, Spain

E-mail: sergio.perezv@uva.es

**ABSTRACT:** This study addresses a key challenge in motor imagery (MI)-based brain-computer interfaces (BCIs): improving the decoding accuracy of electroencephalography (EEG) signals. We investigate the inter-task transfer learning potential between motor execution (ME) and MI to enhance the calibration phase of MI-BCIs. Utilizing the *EEGSym* deep learning network, we demonstrate that ME data can effectively train models for MI classification. Additionally, our analysis identifies a significant positive correlation between performances on ME and MI tasks. These findings support the feasibility of a ME-based calibration approach for MI tasks in BCI systems, leveraging the neural and functional similarities between ME and MI. This approach maintains BCI performance and potentially makes it easier to accommodate new users to the MI task while recording ME data during calibration, which could serve as an indicator of the expected MI accuracy. Furthermore, our results suggest that we can exploit the synergies between ME and MI without significantly reducing decoding accuracy of the user's intentions.

## INTRODUCTION

Brain-computer interfaces (BCIs) offer a novel communication channel, directly linking the human brain to external devices [1]. These systems are designed as closed-loop systems with three stages: the recording of brain activity, the processing of this data to interpret the user's intent, and providing feedback to the user. Electroencephalography (EEG) is favored in the recording stage for its non-invasive nature, portability, and excellent temporal resolution [2]. Furthermore, it is more affordable than the alternative techniques used for capturing brain dynamics. An EEG-based BCI system captures the brain's electrical activity using electrodes placed on the scalp. In the processing stage, these signals are analyzed to decode the user's intentions [1]. The processed information then translates into feedback, which could be provided as visual cues on a monitor or the manipulation of a prosthetic limb [3]. Despite EEG's advantages, the technique

faces significant hurdles, such as its inherently low spatial resolution and the challenge of a poor signal-to-noise ratio (SNR). BCIs, therefore, employ various paradigms to generate recognizable brain patterns in the EEG, facilitating the decoding process. Motor imagery (MI), the voluntary simulation of movement without physical execution, is one paradigm that has gained increased research interest. MI activates the primary motor cortex and associated motor regions, mirroring the neural activation patterns observed during motor execution (ME) [4–6]. This neural overlap between MI and ME has critical implications, particularly in rehabilitative contexts. Research demonstrates that employing MI-based BCIs in a closed-loop system, complemented by functional electrical stimulation as feedback, can significantly bolster brain plasticity. Moreover, such targeted interventions have been crucial in enhancing ME capabilities among stroke patients [6].

Nonetheless, one major drawback of MI-based BCIs lies in the difficulty of achieving high enough decoding accuracy from EEG signals. Conventional machine learning (ML) approaches often struggle with BCI inefficiency [7], a phenomenon where BCI systems cannot reliably interpret and extract distinct features from an individual's EEG signals, impacting an estimated 10-50% of users in MI-based BCI applications [8]. Such users fail to attain effective BCI control, a condition that prior research identifies as exceeding a threshold 70% accuracy in binary MI tasks [9, 10]. This inefficiency has been attributed to the shortcomings in the classification stage [11], recording system limitations, or diminished user motivation over prolonged skill acquisition periods [12]. Moreover, there are elusive additional factors that further contribute to BCI inefficiency. Given that classical ML techniques need a calibration stage at the start of each session to address inter-subject and inter-session variability [13], ensuring this calibration phase captures accurate and relevant information becomes critical for the session's subsequent success. However, verifying whether users have correctly comprehended the instructions or are engaging in the MI task poses a significant challenge. An

inadequate calibration run can result in confusing feedback, reducing user motivation and potentially leading to BCI inefficiency. This underscores the importance of having a more robust calibration process that can effectively minimize these issues.

Prior research has identified the calibration phase in BCIs as a significant bottleneck, proposing the use of deep learning (DL) models with strong transfer learning capabilities to overcome the inter-subject and inter-session variability [11, 14, 15]. These DL architectures could be employed in a calibration-less scenario by leveraging MI trials from various users. However, this strategy encounters limitations since the data used to train has been recorded providing feedback of a similar flawed calibration, or it is collected without providing any feedback to the user or the observer on the MI being properly performed. In this work, we investigate a less explored strategy: utilizing inter-task transfer learning not just across users but also between ME and MI paradigms [16–18]. Lee *et al.* [16] and Miao *et al.* [18] demonstrate that a model trained in ME data can effectively translate into an MI task with a minimal amount of MI examples. Shuqfa *et al.* [17] employ data from both ME and MI trials to train their classifiers simultaneously, aiming to improve accuracy due to the similarities between the tasks. We will further evaluate the feasibility of using data from users performing ME to classify MI trials, specifically excluding MI trials in the training set. This approach could enable objective verification of ME activity being performed by visual inspection, thus eliciting discernible brain patterns, which will isolate errors to the recording system. Moreover, it could potentially enhance the preparation for rehabilitation-focused MI-based BCI applications through closer alignment with actual ME brain patterns. Furthermore, leveraging ME data from multiple users could be more reliable to train deep learning models that focus on more relevant brain patterns.

Our research investigates the correlation between upper-limb ME and MI, assessing the extent of transfer learning capabilities of *EEGSym* [11], a DL network previously validated in inter-subject MI classification. Additionally, we aim to elucidate the impact of current EEG recording system limitation on BCI inefficiency by examining the performance correlation between ME and MI tasks. To accomplish this goal, we analyze public database of non-invasive EEG recordings from 109 healthy users performing MI and ME tasks without feedback [19].

## MATERIALS AND METHODS

### *Dataset and preprocessing:*

The Physionet dataset [19] encompasses recordings from 109 healthy participants during one session. These sessions included one run of 42–46 trials focusing on MI without feedback and another run on ME. In both runs, the duration of the imagination or execution phase for each trial was 3 seconds. The 64-channel EEG signal was recorded using the BCI200 system [20]. The majority of

the dataset, covering 105 participants, was recorded at a sampling frequency of 160 Hz, while the recordings from the remaining 4 participants were captured at 128 Hz.

Prior to inputting the dataset into the DL network for classification, we perform a structured preprocessing pipeline, detailed as follows: (1) we apply a notch filter to remove the power line signal, (2) we perform common average reference (CAR) spatial filtering, (3) we do a re-sampling to 128 Hz to homogenize the dataset across the different sampling rates of the input for the DL model, (4) we extract the trials with a time window length of 3 seconds after the onset, and (5) we apply channel-wise z-score standardization on each trial.

### *DL architecture and training:*

The open implementation of *EEGSym* [11] will be used for classification. *EEGSym* introduces a pioneering convolutional neural network (CNN) architecture designed for the classification of MI across different subjects presented in our prior work [11]. Leveraging cutting-edge DL methodologies, *EEGSym* incorporates residual connections, implements data augmentation strategies, employs inter-subject transfer learning, and features a siamese-network design that capitalizes on the inherent symmetry of the brain along the mid-sagittal plane. This CNN has demonstrated significantly improved accuracy in binary MI inter-subject classification, outperforming the performance of four previously established CNNs developed for EEG classification: ShallowConvNet and DeepConvNet [21], EEGNet [22], and EEG-Inception [23]. *TEEGSym* achieved groundbreaking results, setting a new benchmark for accuracy in inter-subject MI classification.

The selection of this network is primarily motivated by its tailored design for inter-subject classification scenarios, which was proven by its superior performance in such tasks. It emerges as one of the better choices to discern and emphasize patterns universally present among users engaged in both MI and ME tasks [7]. This property is expected to also boost transfer learning efficiency across these tasks, thus enhancing the robustness of comparative analyses regarding task performance.

This model was trained on a NVIDIA 3080 Ti GPU, with CUDA 11.2 and cuDNN 8.1.0 in Tensorflow 2.10. For each analysis' training iteration, we allocated 10% of the data from each subject present in the training set for validation, to trigger early stopping. This early stopping mechanism halts the training if the validation loss fails to improve for 10 consecutive epochs.

### *Inter-task transfer learning analysis:*

To assess the transfer learning capabilities across MI and ME tasks, we evaluated the following training schemes:

1. Training the DL model on all subjects within the ME dataset, then evaluating the performance on the MI dataset data, treating left-/right- hand movement imagination as if it was the trained left-/right- hand movement execution. In this training scheme, the ME data from every user, whose MI accuracy is assessed, is included in the training data.

2. Initially, pre-training the DL model on every subjects' trials present in the ME dataset except for one, following a leave one subject out (LOSO) training scheme. Subsequently, the model's accuracy in identifying MI trials for the excluded user is assessed. This process is replicated for every user.

Moreover, we will examine whether including the ME data of the evaluated user significantly impacts the results employing the Wilcoxon signed rank test [24].

#### Task performance correlation analysis:

We evaluate the correlation between the decoding accuracies for ME and MI data. Accuracies are obtained following a LOSO training scheme. For each task, we train the model on every users' data, except for one subject. The excluded user's trials serve as the test set to determine inter-subject ME or MI prediction accuracy [11]. This correlation is quantified using Spearman's rank correlation coefficient, which will describe the monotonic relationship between these inter-subject performances [25].

## RESULTS

Our study yielded several key insights. Firstly, our analysis demonstrated that a DL network, trained on ME trials, is capable of classifying MI trials in the majority of participants with a degree of accuracy ( $\geq 70\%$ ), which is considered sufficient for BCI control in a binary MI task [9–11]. Secondly, a significant and positive correlation was established between inter-subject performances on ME and MI tasks, evidenced by a highly significant  $p$ -value of less than 0.001. Additionally, we found no significant difference in performance between the model trained with ME data, including trials from the target user, and the model trained on MI data from other users.

#### Inter-task transfer learning analysis:

The efficacy of inter-task transfer learning was examined through two distinct training schemes, the results of which are summarized in Table 1. Our findings highlight that incorporating ME data from the target subject into the model's training significantly enhances accuracy ( $p$ -value $<0.05$ ), compared to the model that has not been exposed to ME EEG signal from the evaluated user.

Table 1: Inter-task transfer learning accuracies

Training scheme	Accuracy(%)
ME to MI	85.73 $\pm$ 10.02
ME to MI without subject's ME trials	85.10 $\pm$ 9.93

#### Task performance correlation analysis:

The accuracies of the inter-subject transfer learning for both ME and MI tasks is presented in Table 2 while the correlation between both tasks performances can be observed in Figure 1. The accuracy for the ME task is significantly superior to the accuracy obtained on the MI task. Nevertheless, there is a positive and significant correlation (i.e.,  $p$ -value $<0.001$ ) assessed by Spearman's rank correlation coefficient of 0.6378. Thus, there is a certain expectation of obtaining low or high per-

formances when classifying MI data depending on the accuracy obtained on ME trials. Noteworthy, the accuracy obtained is way above the chance level for the Physionet dataset which is  $50\% \pm 13.86\%$  for individual users and  $50\% \pm 1.40\%$  for the entire dataset, both calculated at a 95% confidence level [26].

Table 2: Accuracies of inter-subject task

Task	Accuracy(%)
Inter-subject ME	87.35 $\pm$ 8.40
Inter-subject MI	85.65 $\pm$ 10.42

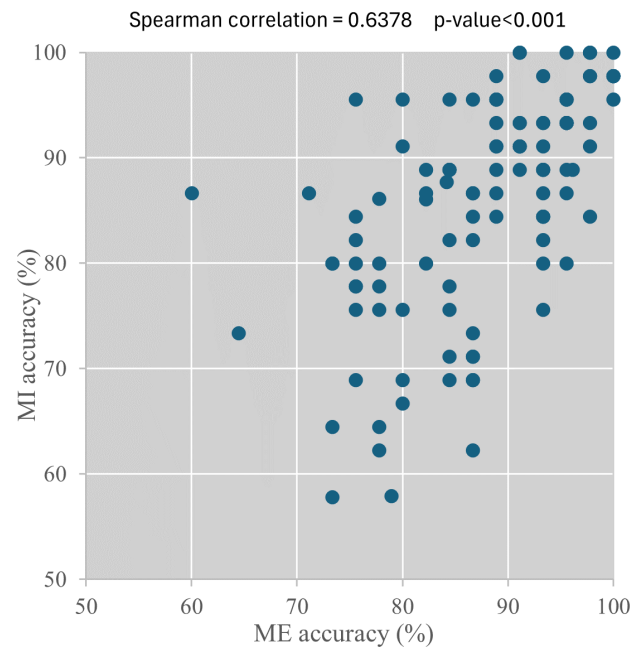


Figure 1: Correlation chart between ME and MI performances

## DISCUSSION

#### ME-based preparation run for MI:

The results obtained in this study have provided a clear picture of the possibilities of transfer learning between the tasks of ME and MI. Of note, we have obtained a comparable performance on decoding MI between a model only trained with ME data,  $85.73\% \pm 10.02\%$ , and the same model trained only on MI data of other participants,  $85.65\% \pm 10.42\%$ . There is a minor, but significant, increase in the MI performance obtained between a model that includes ME data of the final user. Furthermore, the relationship between ME and MI not is only restricted to the possibility of this inter-task transfer learning, but there is also a correlation between ME and MI accuracies as shown in Figure 1.

There have been previous works that have exploited the relationship between ME and MI EEG data [16, 18]. While these works explored this relationship, they have applied it as a previous step to initialize their DL networks without exploring the fully inter-task transfer

learning capabilities. The fully transfer-learning possibility explored in this work, in addition to the correlation between ME and MI performances, could be exploited by establishing a preparation run based on ME where the user can be shown how the instruction and feedback will be presented. There are clear benefits from this ME-based preparation run. For new BCI users, understanding how to perform MI tasks can be difficult. Training users with ME tasks, which are more intuitive and easier to perform, can serve as a stepping stone, helping users learn how to modulate their neural signals effectively before transitioning to MI tasks. This can shorten the learning curve and improve overall BCI control. Furthermore, ME tasks can be performed with less mental effort from the user. Furthermore, the accuracy of this ME-based preparation run could be used to indicate the expected accuracy on MI tasks. Additionally, collecting high-quality MI data can be challenging, especially for BCI users who may struggle with performing consistent MI tasks without physical movement. Collecting ME data can provide a more robust dataset for training BCI algorithms, as ME tasks can be more easily performed and monitored for correctness, leading to higher-quality training data. Finally, in rehabilitation settings, a model trained on ME data could lead a more targeted recovery of the lost functions since it will search for the lost patterns common to the users included in the training ME data.

#### *Limitations and future work:*

While our study offers promising insights into the relationship between ME and MI, as well as the capabilities of inter-task transfer learning, we recognize certain limitations that future research should address. To begin with, our analysis was focused on binary upper-limb classification tasks, which may not encompass the complexity or challenge of distinguishing among more varied types of ME/MI tasks involving movements with less spatially distinct neural activity. Expanding this research to include multi-class classification tasks that incorporate a wider range of movements could offer a more comprehensive understanding of the applicability of our findings. In addition, the analysis was conducted using data collected in a single session from participants who did not receive feedback, limiting our ability to assess the potential for learning or adaptation over time. Investigating the long-term effects of using ME-based preparation run on MI task performance, as well as user satisfaction, could provide valuable insights for the development of more personalized and effective MI-based BCI systems. Moreover, assessing the impact of MI-based rehabilitation, enhanced with feedback from models trained on ME data, in comparison to those trained solely on MI data, would significantly contribute to our understanding of the most effective strategies for leveraging BCIs in rehabilitation contexts.

## CONCLUSION

In this study, we explored the potential of inter-task trans-

fer learning between ME and MI, uncovering that ME data can be effectively utilized to train DL models for MI classification. Additionally, we identified a significant correlation in performance across both tasks. These insights have prompted us to propose an ME-based preparation strategy for MI tasks. By integrating this ME-based preparation run into MI-based BCIs, we introduce a pragmatic solution that leverages the inherent neural and functional similarities between ME and MI. This approach not only maintains BCI performance but also improves accessibility and user experience, making BCIs more intuitive and effective for users. Simultaneously, the ME-based preparation trials offer the opportunity to generate a new corpus of EEG data, which assures the presence of task related information. This enhancement in data quality facilitates the training of deep learning models with improved accuracy.

## ACKNOWLEDGEMENTS

This research has been developed under the grants PID2020-115468RB-I00, PDC2021-120775-I00 and TED2021-129915B-I00 funded by MCIN/AEI/10.13039/501100011033/ and European Regional Development Fund (ERDF) A way of making Europe; under project '0124\_EUROAGE\_MAS\_4\_E' ('Cooperation Programme Interreg VI-A Spain-Portugal POCTEP 2021–2027') funded by 'European Commission' and ERDF; and by CIBER-BBN through 'Instituto de Salud Carlos III' co-funded with ERDF funds. S. Pérez-Velasco and D. Marcos-Martínez were in receipt of a PIF grant by the 'Consejería de Educación de la Junta de Castilla y León'.

## REFERENCES

- [1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*. 2002;113(6):767–791.
- [2] Wolpaw JR, Wolpaw EW. *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press (2012).
- [3] Ramos-Murguialday A *et al.* Proprioceptive Feedback and Brain Computer Interface (BCI) Based Neuroprostheses. *PLoS ONE*. 2012;7(10):e47048.
- [4] Bundy DT *et al.* Contralesional Brain-Computer Interface Control of a Powered Exoskeleton for Motor Recovery in Chronic Stroke Survivors. *Stroke*. 2017;48(7):1908–1915.
- [5] Moldoveanu A *et al.* The TRAVEE system for a multimodal neuromotor rehabilitation. *IEEE Access*. 2019;7:8151–8171.
- [6] Sebastián-Romagosa M *et al.* Brain Computer Interface Treatment for Motor Rehabilitation of Upper Extremity of Stroke Patients—A Feasibility Study. *Frontiers in Neuroscience*. 2020;14(October):1–12.

- [7] Pérez-Velasco S, Marcos-Martínez D, Santamaría-Vázquez E, Martínez-Cagigal V, Moreno-Calderón S, Hornero R. Unraveling motor imagery brain patterns using explainable artificial intelligence based on Shapley values. *Computer Methods and Programs in Biomedicine*. 2024;246(January):108048.
- [8] Alkoby O, Abu-Rmileh A, Shriki O, Todder D. Can We Predict Who Will Respond to Neurofeedback? A Review of the Inefficacy Problem and Existing Predictors for Successful EEG Neurofeedback Learning. *Neuroscience*. 2018;378(January):155–164.
- [9] Lee MH *et al.* EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience*. 2019;8(5):1–16.
- [10] Meng J, He B. Exploring training effect in 42 human subjects using a non-invasive sensorimotor rhythm based online BCI. *Frontiers in Human Neuroscience*. 2019;13(April):1–19.
- [11] Perez-Velasco S, Santamaria-Vazquez E, Martinez-Cagigal V, Marcos-Martinez D, Hornero R. EEGSym: Overcoming Inter-Subject Variability in Motor Imagery Based BCIs With Deep Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2022;30:1766–1775.
- [12] Jeunet C, Glize B, McGonigal A, Batail JM, Micoulaud-Franchi JA. Using EEG-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: Theoretical background, applications and prospects. *Neurophysiologie Clinique*. 2019;49(2):125–136.
- [13] Saha S, Baumert M. Intra- and Inter-subject Variability in EEG-Based Sensorimotor Brain Computer Interface: A Review. *Frontiers in Computational Neuroscience*. 2020;13(January):1–8.
- [14] Dose H, Møller JS, Iversen HK, Puthusserypady S. An end-to-end deep learning approach to MI-EEG signal classification for BCIs. *Expert Systems with Applications*. 2018;114:532–542.
- [15] Zhang K, Robinson N, Lee SW, Guan C. Adaptive transfer learning for EEG motor imagery classification with deep Convolutional Neural Network. *Neural Networks*. 2021;136:1–10.
- [16] Lee DY, Jeong JH, Lee BH, Lee SW. Motor Imagery Classification Using Inter-Task Transfer Learning via a Channel-Wise Variational Autoencoder-Based Convolutional Neural Network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2022;30:226–237.
- [17] Shuqfa Z, Belkacem AN, Lakas A. Decoding Multi-Class Motor Imagery and Motor Execution Tasks Using Riemannian Geometry Algorithms on Large EEG Datasets. *Sensors*. 2023;23(11):5051.
- [18] Miao M, Yang Z, Zeng H, Zhang W, Xu B, Hu W. Explainable cross-task adaptive transfer learning for motor imagery EEG classification. *Journal of Neural Engineering*. 2023;20(6):066021.
- [19] Goldberger AL *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000;101(23).
- [20] Schalk G, McFarland D, Hinterberger T, Birbaumer N, Wolpaw J. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Transactions on Biomedical Engineering*. 2004;51(6):1034–1043.
- [21] Schirrmester RT *et al.* Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*. 2017;38(11):5391–5420.
- [22] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*. 2018;15(5):1–30.
- [23] Santamaria-Vazquez E, Martinez-Cagigal V, Vaquerizo-Villar F, Hornero R. EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-based Brain-Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2020;28(12):2773–2782.
- [24] Wilcoxon F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*. 1945;1(6):80.
- [25] Spearman C. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*. 1904;15(1):72.
- [26] Müller-Putz GR, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: a closer look on BCI results. 2007 1st COST Neuromath Workgroup Meeting, Rome, Italy. 2007;10(1):95–96.

# PURSUING THE IMPLEMENTATION OF A NEUROTUTOR: AN EEG-BASED CLASSIFICATION OF READING TYPES

H. Romero-Morales<sup>1</sup>, J. N. Muñoz-Montes de Oca<sup>1</sup>, A.A. Torres-García<sup>1</sup>, L. Villaseñor-Pineda<sup>1</sup>

<sup>1</sup> Biosignals Processing and Medical Computing Laboratory, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico

E-mail: jenny.munoz@inaoep.mx

**ABSTRACT:** Electroencephalogram (EEG)-based brain-computer interfaces (BCI) emerged as systems to aid impaired people in daily life. Nowadays, the number of applications and target users of BCI has increased, including those for education purposes. An example of these applications, called neurotutor, was posited in 2015 for improving students' learning process. As a first step towards developing a neurotutor, we analyzed the EEG responses related to two types of reading. Specifically, this work assessed whether a machine learning algorithm can distinguish accurately between both classes from features obtained from the signals using one of three wavelet-based techniques. Also, the impact of epoch length on classifier performance was assessed. The method performance was analyzed under two scenarios (intra-subject and inter-subject), outperforming previous work. The best average accuracies were  $94.40 \pm 5.10\%$  and  $54.40 \pm 6.7\%$  for intra-subject and inter-subject classification, respectively. Although the progress obtained for the intra-subject scenario is promising, several steps must be done to effectively implement a neurotutor, especially in inter-subject scenarios.

## INTRODUCTION

BCIs are systems that leverage the neurons' electrical activity, to generate an alternative channel that does not depend on muscular or verbal outputs. Some BCIs' applications are rehabilitation systems, videogames, neuromarketing, and recently in education [1].

On the other hand, Intelligent Tutoring Systems (ITS) are computer assistive systems designed to provide adaptive, personalized content for students [2]. In 2015, a novel ITS and BCI application named *Neurotutor* was elucidated aiming to enhance students' learning experiences and tailoring the content to individual needs. [3].

A fundamental skill in education is reading, which serves as a critical gateway to learning and academic development. Moreover, reading fosters cognitive skills, such as information processing and inference,

self-learning, and analytic thinking.

Previous works have attempted to analyze EEG signals related to the reading process [4, 5]. Also, in [6, 7] a database of EEG signals associated with two reading states was collected and processed with a baseline method to differentiate the reading states. [8] explored this database to characterize the normal reading paradigm looking for patterns of event-related potentials.

In this work, a first step toward the development of an EEG-based neurotutor was made, by analyzing and processing the dataset collected in [6] to distinguish between two normally employed reading strategies (types), comprehension reading (NR), and Task-Specific Reading (TSR aka scanning). Particularly, this study evaluated the capability of machine learning algorithms to accurately differentiate between both reading tasks using three wavelet-based methods. Additionally, the impact of epoch-length on classification performance was assessed.

A *neurotutor* would benefit from assessing reading comprehension to adapt the contents based on the readability of a text. During NR, the student focuses on deriving information about the central themes of the text and drawing inferences. Whereas TSR is a reading strategy in which the reader focuses on specific information (keywords). TSR is usually employed as a pre-reading strategy, in which the user can decide whether a text provides relevant information for the task in question, or after reading, to locate segments of interest. Therefore, recognizing when students engage in one strategy of reading could guide the neurotutor, leading it to adjust to contents that promote deeper comprehension, all while trying to maintain the engagement and motivation of the user.

## MATERIALS AND METHODS

### 1. Dataset description and preprocessing

The Zurich Cognitive Language Processing Corpus 2.0 (ZuCo 2.0) is a dataset of two physiological signals, EEG and eye-tracker, of 18 English native-speaking subjects recorded during two different read-

ing tasks. In the experiment, subjects were asked to read sentences from an annotated Wikipedia corpus (in English), in which each sentence is associated with a specific semantic relation (i.e. Political affiliation, education, founder, wife/husband, job title, nationality, and employer). Participants were asked to read sentences with two different purposes, (1) to find an implicit relation in a text (Task Specific Reading (TSR), 390 sentences) and (2) to fully comprehend the meaning of the sentence (Normal Reading (NR), 349 sentences). To encourage reading comprehension during the NR task, some control questions were randomly presented after some instances. Additionally, each participant was required to do a linguistic assessment (Lexical Test for Advanced Learners of English) to measure their language proficiency [6].

In this work, only EEG signals were employed for classification, seeking to reduce the amount of data needed for the classification. Moreover, EEG signals are currently being researched to derive implicit information about the user's state (memory load, emotions, etc.), this information could aid in the development of a neurotutor. The signals were recorded using a 128-electrode Geodesic Hydrocel System with a sampling frequency of 500 Hz. EEG signals were preprocessed by [6], using Matlab's Automagic and the Multiple Artifact Rejection Algorithm (MARA). Twenty-three electrodes were removed during this stage because of their predominant muscular and ocular information. Furthermore, to reduce data size and computational load, signals were downsampled to 256 Hz.

## 2. Epoch extraction

Since natural and untimed reading was encouraged during both task reading, the duration of reading epochs was variable. To standardize the length of the signals of all participants, epochs of 1, 2, and 3 seconds were selected to analyze the significance of epoch length in classification performance, full-length signals were also analyzed to establish a reference. Records of at least 3 seconds were selected for this analysis, which created a class imbalance; thus a random selection of 100 epochs (samples) per class was applied. Participant YDR was excluded from this study because of insufficient epochs per class. Epochs were extracted from the center region of each sentence, which means the 1-second epoch is contained in the other two epochs, and the 2-second epoch is contained in the 3-second epoch. The whole epoch was also analyzed to obtain a benchmark and evaluate the significance of using a shorter epoch for analysis.

## 3. Channel selection

After the preprocessing stage, a set of 105 channels were kept. Given the inherent characteristics of EEG, certain channels exhibit redundant information; for this reason and aiming to reduce classi-

fication times, our experiments were focused on 31 electrodes taken from a standard 32-electrode setup.

## 4. Feature extraction

Wavelet-based methods have been proven to be accurate techniques to characterize and process biomedical signals [9], which due to their complexity and variability tend to be hard to analyze. In this work, three wavelet-based techniques were analyzed to find the best characterization of the EEG signals related to reading tasks.

### 4.1 Discrete Wavelet Transform (DWT)

This method decomposes the signal using a series of filters. The filtering process is limited by the sampling frequency and the length of the signal. DWT provides n-levels of decomposition, by dividing the signal into a high-frequency component (detail coefficients) and a low-frequency component (approximation coefficient). A second-order Daubechies is used as the mother wavelet, with 6 levels of decomposition. A 2nd-order Daubechies is chosen because of the similarity between the wavelet and EEG patterns; moreover, it has been used successfully to classify EEG signals for seizure detection in epilepsy [9]. The EEG signal, located within 0.5 to 50 Hz, has often been characterized in terms of five brain rhythms: delta (0.5 - 4 Hz), theta (4 - 8 Hz), alpha (8 - 14 Hz), beta (14 - 30 Hz), and gamma waves (higher than 30 Hz). DWT analyzes the signal in the time-frequency domain by decomposing it into sub-bands. Given a 256 Hz sampling frequency, DWT efficiently matches these frequency bands, enabling the extraction of characteristics pertinent to cognitive tasks. Then for each level of decomposition, eleven features were calculated: mean, root-mean square (RMS), kurtosis, median, maximum and minimum amplitude, standard deviation, energy, Instantaneous Wavelet Energy (IWE), Teager Wavelet Energy (TWE), and Hierarchical Wavelet Energy (HWE). In total, 77 features were computed for each channel, resulting in 2387 features per epoch.

### 4.2 Continuous Wavelet Transform (CWT)

Continuous Wavelet Transform highlights the intricate relationship among the frequency, time, and energy of a signal, through a visual representation known as the 'scalogram'. In this study, CWT of each EEG channel was computed using an Analytic Morlet wavelet as the mother wavelet. The scalogram was then divided into the EEG bands described in Section 4.1: delta, theta, alpha, beta, and gamma. From each band, a comprehensive set of 29 characteristics was extracted. These features include the flux at 0, 45, and 90 degrees, as well as the energy of the scalogram, which reflects amplitude variations across the frequency and time axes of the scalogram. Additionally the RMS, mean, standard deviation, skewness, kurtosis, maximum value, entropy, and three key percentiles (75th, 50th, and 25th) were computed. Furthermore, an entropy filter

from Matlab's Image Processing Toolbox was applied to each EEG band, this filter computes the entropy across the image, highlighting the dynamic changes and complexity of the entropy measure within the CWT coefficients. For each entropy-filtered scalogram, features such as the mean, standard deviation, RMS, skewness, kurtosis, and the three percentiles were computed. Subsequently, the average waveform was derived by calculating the mean at each time point across the frequency spectrum of the segmented scalogram. Seven features were then extracted from this averaged signal: mean, median, standard deviation, kurtosis, skewness, RMS, and sample entropy. As a result, 145 characteristics were computed for each scalogram, given that 31 channels were employed within the study, a total of 4495 features were computed per sample.

#### 4.3 Wavelet Scattering Transform

Wavelet Scattering Transform (WST) is a novel wavelet-based method used for the analysis of time series that exhibit non-linear and non-stationary characteristics, such as EEG signals. This advanced mathematical technique yields sparse representations that are invariant to translations and stable to deformations.

In the first level of WST, a decomposition produces a series of coefficients at different scales. A modulus operation is then applied to these coefficients to capture the signal's energy across various frequencies. The resulting modulus wavelet coefficients are subsequently averaged, yielding translation-invariant features of the signal. This operation is recursive yielding higher order coefficients. Typically, first and second-order coefficients capture the majority of relevant frequency information of naturally occurring phenomena. These features are unique to the scattering transform's framework and serve as the foundation for its powerful signal analysis capabilities.

In this work, the scattering time-invariant first-order coefficients are divided into five segments related to brain rhythms (delta, theta, alpha, beta, and gamma). Similar to CWT, WST yields a visual representation often referred to as 'scattergram', which relates time, frequency, and power information of the EEG signal. For each EEG-Band derived coefficients, the 29 descriptors described in Section 4.2 were computed, resulting in a total of 4995 features per epoch.

All three wavelet-based methods were applied to the three epoch lengths (1, 2, and 3 seconds) and for the complete signal.

#### 5. Classification

This study aimed to compare the wavelet-based methods for the classification of EEG signals obtained during two types of reading. Given that EEG signals are highly variably across subjects and even across sessions, an intra-subject approach was pursued, to validate the discrimination power of the

proposed method. Nevertheless, training individualized models requires gathering extensive data, which can be time-consuming, so an evaluation of an inter-subject classification scheme was attempted.

##### 5.1 Intra-subject Classification

For intra-subject classification (IAC), a model was trained for each subject in the dataset. The model was evaluated using a 5-fold validation, with 40 samples (20 of each class) per fold for the testing stage. Three classification algorithms were tested: Support Vector Machine (SVM with a quadratic kernel), K-Nearest Neighbors (KNN with 5 neighbors), and Random Forest (RF with 100 trees). A total of 36 classifiers were trained per subject due to the lengths of the four epochs, three feature-extraction methods, and three classification algorithms were compared.

##### 5.2 Inter-subject Classification

In addition to IAC, an inter-subject classifier (IEC) was trained, using a leave-one-subject-out cross-validation. Prior knowledge of the best epoch duration and feature extraction method was inferred from IAC. 2-second and 3-second epochs were analyzed, using DWT-based features and RF.

## RESULTS AND DISCUSSION

The experiments were carried out to evaluate whether machine learning algorithms, trained on time-frequency representations of EEG signals and with different epoch lengths, could detect differences in brain patterns from subjects engaging in two types of reading: TSR and NR. Furthermore, this rationale was analyzed in two scenarios of classification: intra-subject (personalized models) and inter-subject (generalized models).

##### 1. Intra-subject experiments

Figure 1 shows an analysis of epoch length and its impact on classification performance. In this Figure the global average accuracy for all subjects is taken, regardless of the classification algorithm used, primarily to determine if the length of the signal affects classification outcomes. A trend is observed across all wavelet-based methods; as the epoch size is increased, the performance of the classifier is enhanced. This does not hold when analyzing full-length signals in DWT-derived features. Last, for this and the remaining figures the chance level (50% for two balanced classes) is shown as a dashed line. A non-parametric, Kruskal-Wallis test, with a post-hoc follow-up Dunn's test was performed for each wavelet feature group. Significant differences were found between the one-second epoch and the 3-second epochs in all characterizations. Given the trend observed, and the reduced computational costs in epoch analysis, the two-second and 3-second epochs were further analyzed as promising for TSR and NR classification. A comparative analysis of machine learning algorithms (KNN, SVM and RF) was conducted for

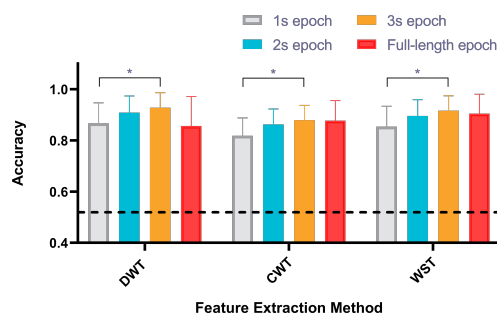


Figure 1: Comparing how epoch length affects classification using wavelet-based features, average accuracy from IAC is shown across KNN, SVM, and RF classifiers. Regardless of classifier type, accuracies are averaged to emphasize epoch length's impact. Significant differences between epoch lengths are marked with asterisks ( $p < 0.05$ , Dunn's test).

all feature extraction methods, using the 3-second epoch, which achieved the highest averaged performance. Figure 2 presents a comparison of the classification algorithm and its effect on global accuracy across all subjects. Significant differences were found in CWT-derived methods, for KNN and RF, as well as for KNN and SVM algorithms. Using WST, differences were found between KNN and RF. On the other hand, the best performances were obtained for DWT and WST regardless of the machine learning algorithm used.

For both wavelet-based methods, the best classification performance was yielded using RF with an accuracy of  $94.40 \pm 5.10\%$ , DWT and  $94.80 \pm 5.10\%$ , WST). Furthermore, DWT-based descriptors exhibited consistent performance across all three machine learning algorithms. DWT has positioned itself as a valuable tool for EEG classification since it provides a time-frequency analysis without information loss or alteration while reducing computational costs [9]. A baseline classifier was trained using the characteristics proposed by [7], and evaluated through the same classification scheme. Baseline characteristics were obtained by filtering each epoch into the relevant EEG bands and obtaining the mean amplitude from each EEG component. To ensure comparable results, the 31 channels selected in this study were also used for benchmark classification.

Even though no significant differences were found within the proposed features and the benchmark, our approach utilizes epochs of 2 and 3 seconds, meanwhile, the average signal length from the original recordings is 5.84 seconds for NR and 4.81 seconds for TSR[6]. Short epochs reduce computational costs and would be more suitable for online applications. Figure 3 shows accuracies obtained for each subject, employing both the DWT-derived features with RF classifier and 3-second epochs. The best accuracy (i.e.  $99.50 \pm 1.11$ ) was achieved by subjects YAK and YMS. Besides that, all subject accuracies were

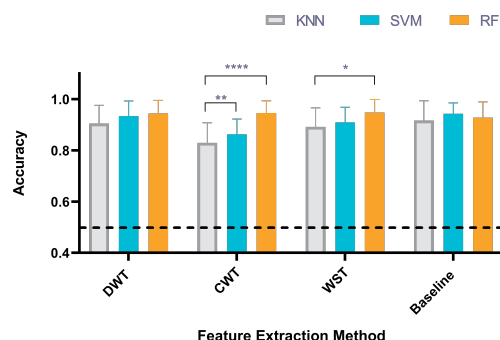


Figure 2: Intra-subject performances compared between wavelet-based and benchmark features using three-second epochs (optimal length). Asterisks indicate significant differences ( $p < 0.05$ , Dunn's test).

greater than 85%. Additionally, the scores provided in [6] for NR control questions (NR scores), and correct semantic text identification (TSR scores) were analyzed to see if task classification was correlated to individual performance in each task. No correlation was found between classification performance and NR/TSR scores (Spearman test). However, a low performance across classifiers and epochs was consistently observed for YAG, achieving one of the lowest performances (i.e.  $87.70 \pm 4.67$ ). This outcome could be explained because YAG also exhibited a low performance on semantic identification. On the other hand, subjects such as YRK, YLS, YMD, YMS, exhibited great performance in both control tests and similarly an accurate classification in the proposed methodology. Although YAK received the lowest score for the set of random questions in NR, the algorithm demonstrated good performance. Since questions were randomly presented, for NR scores it is difficult to assess if scores truly reflect the quality of the task being performed by the user.

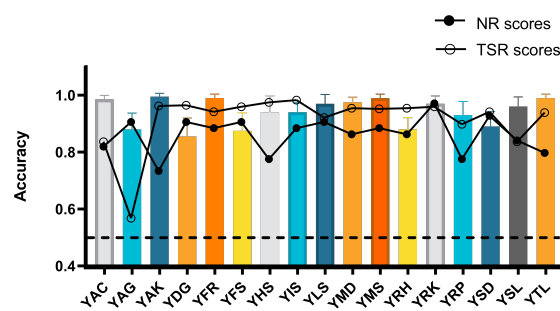


Figure 3: Intra-subject accuracies by RF after 5-fold cross-validation using DWT-based features and 3-second epochs. Also, NR and TSR scores are shown [6]. NR scores show the accuracy of responses to randomly posed comprehension questions, while TSR scores refer to the correct identification of semantic relations within the text.

## 2. Inter-subject experiments

Despite constructing a universal model to interpret highly variable EEG signals being a complex challenge, an experiment for inter-subject classification was also performed on the data, using a leave-one-subject-out validation. For this experiment, only two and 3-second epochs were analyzed, using the most stable characterization (DWT). Additionally, a deep-learning architecture specialized in EEG data (EEGNet) was trained for the identification of NR and TSR, for the 2 and 3 second epochs. The hyperparameters for the network were selected based on the recommendations of the EEGNet developers (128 kernel length) [10]. Also, some parameters were taken from [4] such as Adam optimizer and a batch size of 16 instances. Since that work analyzed EEG signals related to the reading process and obtained promising results. Likewise, 100 iterations were calculated.

A baseline classifier was trained using the characteristics used in [7] and evaluated through the same classification algorithms. Benchmark features were originally calculated using full-length signals. All the models were trained using data from the 31 selected channels. Table 1 presents the performance for all trained models.

The best classification performance,  $54.4 \pm 6.7$ , was achieved by DWT-derived features, using a 2-s epoch. Despite baseline features achieving a global classification accuracy lower than the chance level, all proposed methods and EEGNet slightly outperformed the chance level for the two classes. This could imply different cognitive processes are undertaken in both reading tasks that could be generalized efficiently across subjects through the proposed methodology and EEGNet. Despite no classifier got a higher average accuracy than the empirical chance level for 100 trials per class (58%) [11]; this threshold was overcome for eight subjects.

Table 1: Performance comparison between DWT-based models, EEGNet and baseline descriptors for inter-subject classification.

Method	Average	Median	Max - Min
2-s epoch	$54.4 \pm 6.7$	53.5	68.0 - 45.0
3-s epoch	$51.7 \pm 7.2$	52.0	66.0 - 37.0
EEGNet 2-s	$52.9 \pm 8.8$	52.0	74.5 - 38.5
EEGNet 3-s	$48.06 \pm 11.54$	50.0	66.5 - 27.0
Baseline	$49.5 \pm 12.0$	46.5	80.0 - 31.5

Figure 4 shows the individual test accuracy obtained after leave-one-subject-out validation. Both the proposed methodology and EEGNet implementation generally achieved accuracies surpassing the theoretical random classifier. Specifically, EEGNet showed an accuracy above the 50% threshold for approximately 8 subjects, whereas the proposed method achieved this for 12 out of 17 subjects. For the empirical random classifier, 5 subjects surpassed the

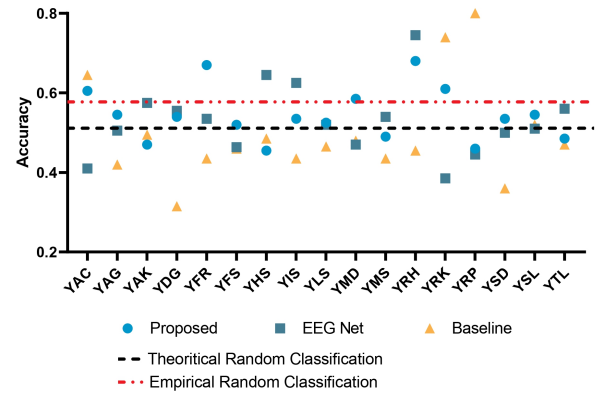


Figure 4: Test accuracy for each subject after leave-one-subject-out validation is shown for three methods: (1) Proposed method using a 2-second epoch, DWT-derived features and RF; (2) EEGNet classifier, and (3) Baseline method trained with the features from [7] and RF. The theoretical and empirical [11] chance levels are indicated by black and red dashed lines, respectively.

58% threshold. This suggests time-frequency features, along with RF might be useful to discern between reading tasks. Moreover, when compared to the baseline classifier leveraging features proposed by [7], this study showed the majority of the participants (9) obtained the lowest classification accuracy, while 13 subjects did not surpass the random classifier. Therefore, baseline characteristics do not seem suitable for the task (even though the overall best accuracy was obtained using them, through subject YRP).

Interestingly, subjects who under-perform in the intra-subject approach (YAG, YDG, and YFS), have similar low results for the inter-subject classifier, obtaining results near random classification (Proposed, EEGNet), or below it (Baseline). Similarly, subjects YAC, YFR, YRK, YMD, overperform both in intra-subject and intersubject analysis.

Reading is a complex task, that requires the activation of various brain sub-processes; beyond language processing and visual decoding, reading evokes responses from attention, working memory load, abstract reasoning, and memory pathways. Consequently, we hypothesized NR and TSR could be differentiated through EEG patterns since cognitive and attention demands are different in each reading strategy. Within this study, DWT-based features allowed the distinction of the two types of reading using EEG data. These differences, represented through the time-frequency domain, result from specific brain processes, such as attention or cognitive load. Additionally, inter-subject classification rates above the random classifier reflect subtle patterns that are generalized across subjects. Research by Hollenstein et al. furtherly support this idea, using eye-tracking. In their study, results indicated subjects focused uniformly on each word when engaging

in NR, on the other hand, in TSR subjects focused on words that determined a specific semantic relation and skimmed through the rest. Reading times were also reduced while engaging in TSR [7]. Despite this, more studies are required to identify features that allow classification, based on cognitive differences.

An ideal neurotutor, would benefit from the detection of different reading states to provide active feedback about the psycho-emotional state of the user. Cognitive overload and academic stress can impair students' well-being and decrease academic performance. Likewise, professors could use this as a tool for monitoring students' progress or evaluating contents in concordance with students' comprehension. Although scanning or TSR is a common reading strategy, useful for navigating through large amounts of information, their extensive use could result in surface level comprehension [12]. A neurotutor could adapt the contents' readability and encourage reading comprehension through different activities, thus helping to reduce the use of scanning and pursuing deeper comprehension in the user.

## CONCLUSIONS

The first approach to NR and TSR classification involved an IAC in which an analysis of epoch length, feature extraction method, and classification algorithm were assessed. Best accuracy (i.e.  $94.40 \pm 5.10$ ) was achieved using 2s and 3s epochs, DWT-derived features, and RF classifier. All subjects performed beyond 85% accuracy. Furthermore, for some cases, a relation was found between classifier performance and control test scores, which could imply that diminished performance in task completion reduces classification outcomes.

IEC proved to be a complex task; nevertheless, both the proposed method (2s epoch, DWT, RF) and EEGNet performed above the random and benchmark classifier. From this, it could be inferred there are brain patterns shared across subjects when performing reading tasks, namely NR and TSR.

Future works will explore the relationship between classification performance and control questions, to determine if removing low-performing subjects could increase classifier performance. Further, techniques of data augmentation could be employed to gain more insight into differences in the reading patterns. Likewise, changes in the inclusion criteria for epoch selection could be performed, even if it results in unbalanced classes. Inter-subject complexity derives from the intricate nature of EEG signals, since they are variable across subjects and even across sessions. A generalized model, enhanced by a limited number of training samples from the new user, could improve classification accuracy while maintaining the benefits of limited training time.

## ACKNOWLEDGMENTS

The authors, H. R-M and J. M-MdO, wish to express their gratitude for the graduate scholarships granted by CONAHCYT, Mexico. These grants have enabled them to conduct the research presented in this work.

## REFERENCES

- [1] Aricò P, Borghini G, Di Flumeri G, Sciaraffa N, Babiloni F. Passive BCI beyond the lab: Current trends and future directions. *Physiological measurement*. 2018;39(8):08TR02.
- [2] Lin CC, Huang AY, Lu OH. Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments*. 2023;10(1):41.
- [3] Müller-Putz G *et al*. The future in brain/neural computer interaction: Horizon 2020. 2015.
- [4] Torres-García AA, Martínez-Santiago F, Montejo-Ráez A, Ureña-López LA. Toward an educative EEG-based neuroIIR system for adapting contents. *International Journal of Human-Computer Interaction*. 2023:1–15.
- [5] Ye Z *et al*. Towards a better understanding of human reading comprehension with brain signals. In: *ACM Web Conference*. 2022, 380–391.
- [6] Hollenstein N, Troendle M, Zhang C, Langer N. ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation. In: *12th LREC. ELRA: Marseille, France, May 2020*, 138–146.
- [7] Hollenstein N, Tröndle M, Plomecka M, Jäger LA, Langer N. The zuco benchmark on cross-subject reading task classification with eeg and eye-tracking data. *Frontiers in Psychology*. 2023;13:1028824.
- [8] Liu X, Cao Z. Enhance reading comprehension from eeg-based brain-computer interface. In: *Australasian Joint Conference on Artificial Intelligence*. 2023, 545–555.
- [9] Chen D, Wan S, Xiang J, Bao FS. A high-performance seizure detection algorithm based on Discrete Wavelet Transform and EEG. *PloS one*. 2017;12(3):e0173138.
- [10] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. Eegnet: A compact convolutional neural network for eeg-based brain-computer interfaces. *Journal of neural engineering*. 2018;15(5):056013.
- [11] Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods*. 2015;250:126–136.
- [12] Elleman AM, Oslund EL. Reading comprehension research: Implications for practice and policy. *Policy Insights from the Behavioral and Brain Sciences*. 2019;6(1):3–11.

# ASSESSMENT OF SEVERAL EEG ACTIVE PARADIGMS IN LOCKED-IN SYNDROME

P.Séguin<sup>1,\*</sup>, E. Maby<sup>1,\*</sup>, R. Bouet<sup>1</sup>, L. Gattaz<sup>1</sup>, A. Querry<sup>1</sup>, L. Rizzo<sup>1</sup>, A. Farnè<sup>1</sup>, J. Mattout<sup>1</sup>

<sup>1</sup> Lyon Neuroscience Research Centre, INSERM UMRS 1028, CNRS UMR 5292, Université Claude Bernard Lyon 1, Université de Lyon, F-69000, Lyon, France

\* Perrine Séguin and Emmanuel Maby contributed equally to this work.

E-mail : perrine.seguin@inserm.fr, jeremie.mattout@inserm.fr

**ABSTRACT:** At first glance, Brain-computer interfaces (BCIs) appear to offer promising solutions for people who have global paralysis and are unable to operate conventional communication devices. However BCI efficacy remains low. To better assess the possible clinical reasons for this lack of efficacy, we conducted a study comparing the performance of patients in three paradigms: motor attempt, sustained auditory attention and spatial selective auditory attention. We included 14 persons with locked-in syndrome (LIS), one person in complete LIS and 27 healthy subjects. Preliminary results show that for the patient in complete LIS and a significant proportion of LIS patients, we could not detect their voluntary modulation of brain signals. Surprisingly, this absence of attentional biomarkers seem more prevalent in brainstem injury than in ALS. We discuss the possible impact of global paralysis on brain signals that are used to control BCIs.

## INTRODUCTION

Brain-computer interfaces (BCIs) could help restoring environmental control and communication for people with severe motor disability. Although their aetiologies differ, these typical BCI end-users share a clinical state of total paralysis resulting from some acquired damage to the cortico-spinal pathway or the peripheral nervous system. The ‘classical’ locked-in syndrome (LIS) is caused by an injury to the ventral pons, most often due to a stroke [1], [2]. The patient is totally paralysed except for vertical eye movements and blinks, which enable them to maintain communication. Others are behaviourally non-responsive because of damage to the third and seventh cranial nerves needed for these movements [2], [3]. This condition can also be encountered in the later stages of amyotrophic lateral sclerosis (ALS), a neurodegenerative disease of the motor neurons in which oculomotor muscles are usually preserved [4], except at a very advanced late stage. Then these patients are often considered to be in the complete locked-in state (CLIS), i.e. conscious, but non-responsive.

Another possible cause of a non-responsive state is severe diffuse brain injury due to stroke or anoxia following a cardiac arrest. After being in a comatose state

for up to four weeks these patients sometimes remain in a state with preserved vegetative functions (e.g. autonomous respiration and eye opening) but no sign of awareness. They are said to suffer from disorders of consciousness (DOC). Some of these patients could be conscious, but a combination of impairments (motor, sensory, cognitive) prevents them from understanding and/or following instructions. Active EEG paradigms that were developed to detect consciousness in these patients are close to the one used in BCI (e.g., motor attempt [5], or attentional focus on sounds [6]).

As a matter of fact, BCIs work poorly with both CLIS [7], [8], [9] and DOC patients [10]. Moreover, there is also a subpopulation of patients with severe motor disabilities who cannot control a BCI [11], [12], [13], [14]. This proportion is higher than for healthy subjects. As visual modality is often used in BCI, it was argued that it is problematic for patient with severe motor disability, as they can present oculomotor impairment [15], [16]. Also, when the motor system is altered, it could impact the robustness of sensorimotor rhythms used in motor imagery BCI. But, more surprisingly, even auditory BCIs turn out to be hard to control for these patients [14]. This may be due to the cognitive impairments and altered electrophysiological signals that some of these patients sometimes present [17], [18], but it is still unclear what factors impact the most BCI performance in this clinical context. Thus, the possible clinical reasons for this lack of efficacy need to be better understood.

We propose here to test three different paradigms: motor imagery, auditory selective attention and auditory sustained attention. None of these paradigms require visual input. We will then confront them to clinical data, hoping to find some predictors of the results.

Our project aims to test the robustness of these three BCI protocols with people in locked-in syndrome, as well as with one patient in CLIS.

## MATERIALS AND METHODS

We are evaluating three active EEG paradigms, oculomotor control and limb motor control. We also assessed the functional impact of paralysis with the ALS-FRS revised scale. Ethical authorizations have been obtained (Clinical trial registration N° NCT02567201).

*Participants:* The subjects in LIS were in need for an augmentative and alternative communication device (i.e. eye-tracking or letter board) due to paralysis. They were expected to have both a score at 0 for the first speech item of the ALS-FRS (on speech), and a score smaller or equal to 1 at the 14th item of the ALS-FRS-EXT scale (Wicks, 2009) (i.e. the patient cannot use fingers to control a communication device). Etiologies encompass Guillain Barré syndrome, ALS or brainstem injury. Their age ranges from 20 years-old to 80 years-old.

The CTRL group was composed of 30 healthy subjects, aged 20 to 80 years-old. We excluded subjects that presented a psychiatric or neurological disease.

*Clinical evaluation:* For patients, we performed a motor assessment thanks to the ALS-FRS scale revised, some items from the ALS-FRS-EXT study (Wicks, 2009), and the BELIS scale (ref). We also realized a clinical oculomotor assessment. The patients with a preserved communication code underwent neuropsychological assessments adapted to severe motor disability (BELIS scale [17]). We collected the medication at the time of the EEG experiments.

*Active EEG paradigms:* We used three previously published EEG paradigms that have been independently validated with other participants. All participants realized the paradigms in the same order: first the auditory BCI, then, after a break, the motor attempt one and finally the Active-Passive auditory protocol.

– The **auditory BCI paradigm** is described in [14]. This paradigm includes one stream of “Yes” sounds delivered to the right ear, and one stream of “No” sounds delivered to the left ear. We used a SOA of 400 ms and a variable number of deviants, that were balanced between conditions. The proportion of deviants was one out of 6, and was fixed for each trial. We then varied the length of the trials. Patients were asked to alternatively count left or right ear deviants. In some trials, randomly, patients were asked to report the result of this count in order to check that they understood the instructions and that they are able to perceive and detect the deviant sounds. There were 36 trials in total.

– The **Active-Passive paradigm** was described in [6]. We performed only the most discriminant conditions: one where the subjects are mentally navigating in their houses when hearing sounds (diverted attention), versus the other condition where they focus on the sounds (focused attention). We could thereby increase the

number of stimulations per condition in order to improve the signal-to-noise ratio.

– The **motor attempt** paradigm is the one that is used in [5]. There are 48 trials, 24 for the left hand and 24 for the right hand. Each attempt lasts five seconds, and is followed by 5 seconds of rest. Patients have to try to move their hand, whereas healthy subjects have to imagine moving their hand.

*Material:* We used a Vamp amplifier (16 channels, BrainProducts), with a sampling rate of 1000 Hz. We recorded EEG (13 channels) with reference on the nose, EOG right and left (2 channels), as well as ECG and breathing with a thoracic belt. For EEG, we included Fp1, Fp2, F3, Fz, F4, C3, Cz, C4, TP9, CP5, Pz, CP6, TP10. This aimed to cover both motor and parietal regions. Temporal electrodes were used to visualize the Mismatch negativity in the Active-Passive paradigm.

### *Extracted variables:*

The signal processing and statistical analysis were similar to the ones described in the original publications. All raw EEG signals underwent a bandpass filter between 1 to 30 Hz. We also used the same measures and decision criteria, namely:

- **Active-Passive:** presence of a “Count” effect, and of a “FOC versus DIV” effect. The “Count” effect reflects the presence of electrophysiological responses to oddball sound when the subject is actively counting deviants. The “FOC versus DIV” effect reflects the attentional modulation of evoked potentials when subjects count the deviant versus when they tend to ignore them, by performing spatial navigation imagery (see Morlet et al 2022 for more details).
- **Motor attempt:** accuracy of the classification between “movement” and “rest” trials. Each of the 48 trial was divided in 3 epochs of 2 seconds for the “moving” condition, and 3 epochs of 2 seconds for the rest condition. Then a cross validation with a SVM was performed, and compared to a permutation test. If less than 5 % of the random permutations gave better results than the real dataset, then the participant was considered as a “responder”.
- **Auditory BCI:** accuracy of the classification between “attended” and “unattended” sounds.

*Statistical analysis:* We employed Generalized Linear Models (GLMs). For all analysis, we used R packages including FactoMineR, lme4, afex, emmeans, and sjPlot.

We first compared demographical characteristics of the CTRL and LIS groups. Our variables to be explained were the group and our predictors were the age and educational level.

We then tested the hypotheses of a difference in EEG based classification accuracy between the two groups.

Therefore, we performed three GLM, one for each of the three active EEG paradigms.

Our main explicative/fixed variables are:

- GROUPS (LIS vs CTRL)
- Age

We used a GLM (Generalized Linear Model).

Whenever one of the above explanatory variables did show a significant effect onto the dependent measure, we conducted post-hoc analyses using t or z tests on the linear predictor scale, with confidence intervals also computed on the linear-predictor scale. P values were corrected for multiple testing using the FDR method.

## RESULTS

The 15 patients and 27 control subjects could realize all the three active EEG protocols.

The results are summarized in Table 1. There weren't any significant differences between the ages of the healthy subjects and the patients (Wilcoxon Rank sum test,  $p = 0.54$ ).

For the auditory BCI, there is a significant different between performances of patients and healthy subjects ( $p < 0.001$ ). The brainstem injury has a particularly strong negative impact on the BCI control: only 2 out of 7 patients (29%) present an attentional modulation, versus 6 out of 8 patients with ALS (75%) and 27 out of 27 healthy subjects (100%). Moreover, none of the patients with brainstem injury has a P300 detected by the automatic pipeline, despite being able to hear the deviant sounds and to count them. Of notes, only two patients could not detect the auditory deviant sounds, and both had ALS at a very advanced stage, with a respirator. None of these two patients could control the BCI. The patient in CLIS did not show any detectable voluntary modulation of brain signal.

Concerning the Active-Passive paradigm, there is also a significantly less detectable attentional modulation for the clinical population ( $p=0.01$ ).

On the contrary, the motor attempt paradigm do not reveal any difference of performances between patients and healthy subjects. Only an impact of age is observed ( $p=0.01$ ).

The clinical data on motor level, neuropsychological abilities and medication are currently being acquired and will be analyzed in the coming months, and confronted to these BCI performances.

Table 1: Populations characteristics and main results at active EEG paradigms

	<i>ALS,</i> <i>N = 8<sup>1</sup></i>	<i>Brainstem,</i> <i>N = 7<sup>1</sup></i>	<i>CTRL,</i> <i>N = 27<sup>1</sup></i>
<b>Condition</b>			
CLIS	1 (13%)	0 (0%)	0 (0%)
CTRL	0 (0%)	0 (0%)	27 (100%)
LIS	7 (88%)	7 (100%)	0 (0%)
Age	58 (54, 61)	49 (28, 59)	52 (40, 68)
<b>EEG protocols results</b>			
<b>Auditory BCI</b>			
Mean	0.93	0.61	0.97
accuracy	(0.69, 0.98)	(0.58, 0.75)	(0.96, 1.00)
P300*	5 (63%)	0 (0%)	25 (93%)
Sensibility	6 (75%)	2 (29%)	27 (100%)
<b>Motor attempt</b>			
Mean AUC at group level	0.70 (0.57, 0.81)	0.67 (0.60, 0.77)	0.70 (0.62, 0.79)
Sensibility	6 (75%)	5 (71%)	25 (93%)
<b>Active-Passive</b>			
Count effect	3 (38%)	3 (43%)	19 (70%)
Focus versus Diversion effect	1 (13%)	1 (14%)	18 (67%)
Sensibility	3 (38%)	3 (43%)	22 (81%)

<sup>1</sup> n (%); Median (IQR); \*:

ALS: Amyotrophic Lateral Sclerosis; CLIS: Complete Locked-in Syndrome; CTRL: Healthy subjects

## CONCLUSION

Our preliminary results confirm that a significant proportion of patients cannot control BCI. The impact of clinical condition is more visible for BCI based on evoked protocols, and strikingly strong in case of brainstem injury in the case of selective auditory attention. These results are surprising in several ways. First, all paradigms are supposed to be gaze independent, but the prevalence of non-responders is striking. Second, ALS, as a neurodegenerative disease in continuum with fronto-temporal dementia, is supposed to induce more cognitive impairments than an injury in brainstem cortico-spinal pathways. The absence of detection of selective attentional modulation in case of brainstem injury is thus surprising, and it is the first time to our knowledge that this specificity is uncovered, especially in comparison with another etiology. The principal limitation of these results is that we rely at this stage on automatic analysis pipelines, whereas patients' brain signals can be very different from the one observed in healthy subjects [19], [20], and hence some of them would probably require a personalized signal processing. However, in a previous pilot study, we observed a strong correlation between BCI results and the presence or absence of classical electrophysiological biomarkers as P300 [14]. An important perspective to better explain these results is the analysis of the possibility of other clinical predictors of BCI performance, as neuropsychological tests results, medication. The functional level of autonomy could also have an impact. Indeed, the possibility to interact physically with the environment is associated to a range of action

preparations that arise automatically with some percepts [21], and the question of the preservation of this action preparation in paralysis [22] and their impact on BCI biomarkers remains open [7], [9].

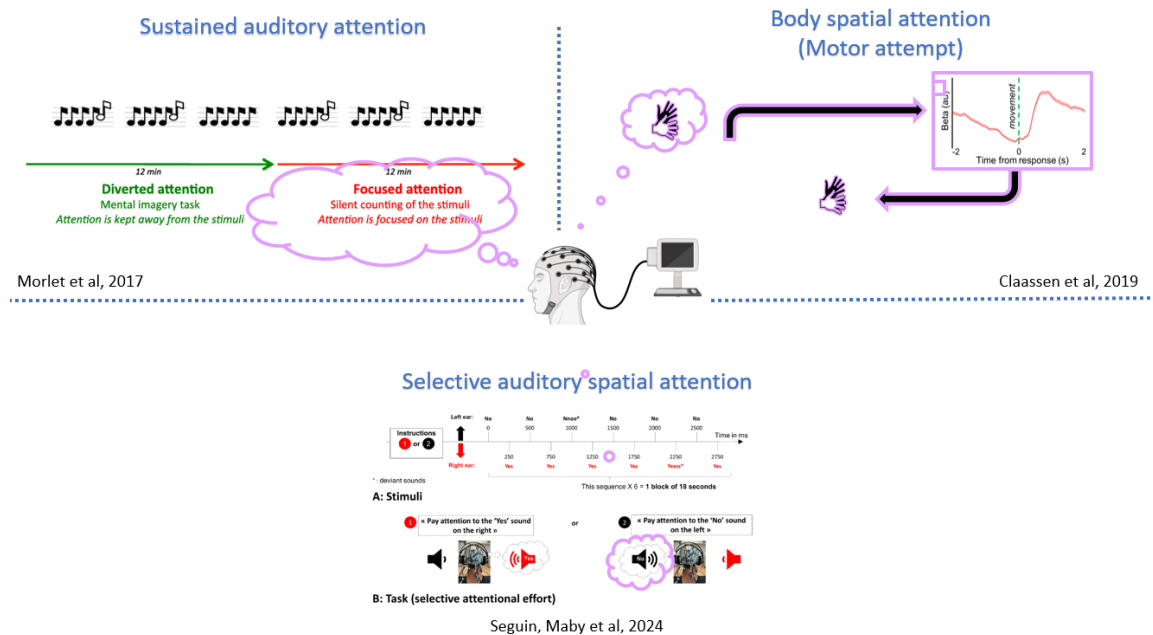


Figure 1: Synopsis of the 3 active EEG paradigms

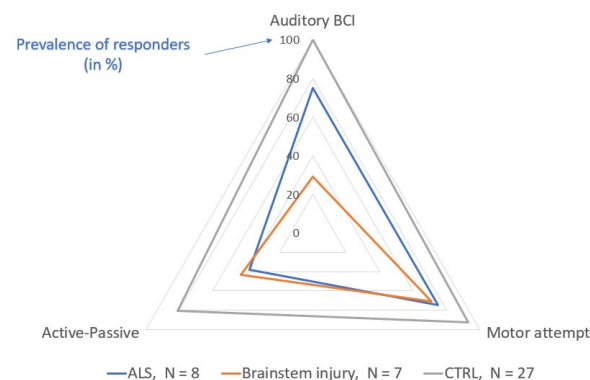


Figure 2: Prevalence of responders for each EEG active paradigm

## ACKNOWLEDGEMENTS

We thank the patients and their relatives for their active participation to this study and for their very valuable feedbacks. We thank Lesly Fornoni and Guillaume Fiard for their help in writing the ethical committee files. We thank Thibaud Lansaman, Benjamin Rohaut, Pascal Giroux, Jacques Luauté, Simon Bertrand, Jean-Philippe Camdessanché, Anne-Laure Kaminsky, Florent Gobert, Marie-Julie Françon, Lydia Oujamaa, Emilien Bernard and Vivien Reynaud for recruiting the patients.

## REFERENCES

- [1] J. León-Carrión, P. van Eeckhout, M. D. R. Domínguez-Morales, et F. J. Pérez-Santamaría, « The locked-in syndrome: a syndrome looking for a therapy », *Brain Inj*, vol. 16, n° 7, Art. n° 7, juill. 2002, doi: 10.1080/02699050110119781.
- [2] F. Plum et J. B. Posner, « The diagnosis of stupor and coma », *Contemp Neurol Ser*, vol. 10, p. 1-286, 1972.
- [3] E. Smith et M. Delargy, « Locked-in syndrome », *BMJ*, vol. 330, n° 7488, Art. n° 7488, févr. 2005.
- [4] J. P. Taylor, R. H. Brown, et D. W. Cleveland, « Decoding ALS: from genes to mechanism », *Nature*, vol. 539, n° 7628, Art. n° 7628, nov. 2016, doi: 10.1038/nature20413.
- [5] J. Claassen *et al.*, « Detection of Brain Activation in Unresponsive Patients with Acute Brain Injury », *New England Journal of Medicine*, vol. 380, n° 26, Art. n° 26, juin 2019, doi: 10.1056/NEJMoa1812757.
- [6] D. Morlet *et al.*, « Infraclinical detection of voluntary attention in coma and post-coma patients using electrophysiology », *Clinical Neurophysiology*, oct. 2022, doi: 10.1016/j.clinph.2022.09.019.
- [7] P. Seguin, E. Maby, F. Perrin, A. Farnè, et J. Mattout, « Is controlling a brain-computer interface just a matter of presence of mind? The limits of cognitive-motor dissociation », arXiv.org. Consulté le: 8 octobre 2023. [En ligne]. Disponible sur: <https://arxiv.org/abs/2310.00266v1>
- [8] S. Silvoni, « Performance of brain-computer communication in Amyotrophic Lateral Sclerosis », Dissertation, Universität Tübingen,

2017. doi: 10.15496/publikation-20805.
- [9] N. Birbaumer, F. Piccione, S. Silvoni, et M. Wildgruber, « Ideomotor silence: the case of complete paralysis and brain-computer interfaces (BCI) », *Psychol Res*, vol. 76, n° 2, Art. n° 2, mars 2012, doi: 10.1007/s00426-012-0412-5.
- [10] V. Galiotta *et al.*, « EEG-based Brain-Computer Interfaces for people with Disorders of Consciousness: Features and applications. A systematic review », *Frontiers in Human Neuroscience*, vol. 16, 2022, Consulté le: 6 octobre 2023. [En ligne]. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.1040816>
- [11] J. R. Wolpaw *et al.*, « Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis », *Neurology*, vol. 91, n° 3, Art. n° 3, juill. 2018, doi: 10.1212/WNL.0000000000005812.
- [12] M. Marchetti et K. Priftis, « Brain-computer interfaces in amyotrophic lateral sclerosis: A metanalysis », *Clinical Neurophysiology*, déc. 2014, doi: 10.1016/j.clinph.2014.09.017.
- [13] Z. R. Lugo *et al.*, « Mental imagery for brain-computer interface control and communication in non-responsive individuals », *Ann Phys Rehabil Med*, avr. 2019, doi: 10.1016/j.rehab.2019.02.005.
- [14] P. Séguin *et al.*, « The challenge of controlling an auditory BCI in the case of severe motor disability », *Journal of NeuroEngineering and Rehabilitation*, vol. 21, n° 1, p. 9, janv. 2024, doi: 10.1186/s12984-023-01289-3.
- [15] M. Graber, G. Challe, M. F. Alexandre, B. Bodaghi, P. LeHoang, et V. Touitou, « Evaluation of the visual function of patients with locked-in syndrome: Report of 13 cases », *J Fr Ophtalmol*, vol. 39, n° 5, Art. n° 5, mai 2016, doi: 10.1016/j.jfo.2016.01.005.
- [16] E. Aust *et al.*, « Impairment of oculomotor functions in patients with early to advanced amyotrophic lateral sclerosis », *J Neurol*, sept. 2023, doi: 10.1007/s00415-023-11957-y.
- [17] M. Rousseaux, E. Castelnol, P. Rigaux, O. Kozłowski, et F. Danzé, « Evidence of persisting cognitive impairment in a case series of patients with locked-in syndrome », *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 80, n° 2, Art. n° 2, févr. 2009, doi: 10.1136/jnnp.2007.128686.
- [18] S. Abrahams, « Neuropsychological impairment in amyotrophic lateral sclerosis—frontotemporal spectrum disorder », *Nat Rev Neurol*, vol. 19, n° 11, Art. n° 11, nov. 2023, doi: 10.1038/s41582-023-00878-z.
- [19] P. Kellmeyer, M. Grosse-Wentrup, A. Schulze-Bonhage, U. Ziemann, et T. Ball, « Electrophysiological correlates of neurodegeneration in motor and non-motor brain regions in amyotrophic lateral sclerosis—implications for brain-computer interfacing », *J. Neural Eng.*, vol. 15, n° 4, p. 041003, juin 2018, doi: 10.1088/1741-2552/aabfa5.
- [20] M. Bensch *et al.*, « Assessing attention and cognitive function in completely locked-in state with event-related brain potentials and epidural electrocorticography », *J Neural Eng*, vol. 11, n° 2, Art. n° 2, avr. 2014, doi: 10.1088/1741-2560/11/2/026006.
- [21] R. J. Bufacchi et G. D. Iannetti, « An Action Field Theory of Peripersonal Space », *Trends Cogn. Sci. (Regul. Ed.)*, vol. 22, n° 12, Art. n° 12, déc. 2018, doi: 10.1016/j.tics.2018.09.004.
- [22] A. Avenanti, L. Annala, et A. Serino, « Suppression of premotor cortex disrupts motor coding of peripersonal space », *NeuroImage*, vol. 63, n° 1, p. 281-288, oct. 2012, doi: 10.1016/j.neuroimage.2012.06.063.

# AN ONLINE SPIKE DETECTION AND MONITORING FRAMEWORK IN IEEG RECORDED USING BRAIN INTERCHANGE DEVICE

Behrang Fazli Besheli<sup>1</sup>, Amir Hossein Ayyoubi<sup>1,2</sup>, Jhan Luke Okkabaz<sup>1,2</sup>, Chandra Prakash Swamy<sup>1</sup>, Michael M. Quach<sup>3</sup>, Kai J. Miller<sup>1</sup>, Gregory Worrell<sup>4</sup>, and Nuri Firat Ince<sup>1,2,5</sup>

<sup>1</sup> Department of Neurologic Surgery, Mayo Clinic, Rochester, MN, USA

<sup>2</sup> Department of Bioinformatics and Computational Biology, University of Minnesota, Rochester, MN, USA

<sup>3</sup> Department of Neurology, Texas Children's Hospital, Houston, TX, USA

<sup>4</sup> Department of Neurology, Mayo Clinic, Rochester, MN, USA

<sup>5</sup> Department of Biomedical Engineering, Mayo Clinic, Rochester, MN, USA

E-mail: ince.nuri@mayo.edu

**ABSTRACT:** In this study, we developed and validated an online analysis framework in MATLAB Simulink for recording and analysis of intracranial electroencephalography (iEEG). This framework aims to detect interictal spikes in patients with epilepsy as the data is being recorded. An online spike detection was performed over 10-minute interictal iEEG data recorded with Brain Interchange CorTec in three human subjects. A pool of detected spikes is then broadcasted using User Datagram Protocol (UDP) to an external graphical user interface for further post-processing and visualization. The real-time spike detector demonstrated a 99% similarity index with the previously published offline detector, identifying interictal spikes. Furthermore, our findings indicated that channels with highest spike rates, captured with Brain Interchange CorTec, were in the epileptogenic focus. By enabling the detection of interictal spikes in an online fashion, this work provides early feedback on the probable seizure onset zone (SOZ) and suggests a promising direction for enhancing SOZ localization accuracy to clinicians, which is crucial for the surgical treatment of epilepsy.

## INTRODUCTION

Epilepsy is a neurological disease marked by recurrent, unprovoked seizures, affecting millions of individuals worldwide [1]. A significant subset of these individuals (around 30%) suffer from medically intractable epilepsy, where seizures are not able to be well-controlled by medication. The localization of the seizure onset zone (SOZ) — the brain area responsible for initiating seizures — is crucial for successful surgical intervention [2,3]. Intracranial electroencephalography (iEEG) has emerged as a fundamental tool in this endeavor, allowing for the precise monitoring of brain activity associated with epileptic discharges [4]. In recent years, interictal spiking activity, a brief transient event, has received considerable attention for SOZ

localization. Although contradicting studies [5] were reported regarding the effect of interictal spikes and ictogenesis, these have been hypothesized by other studies as a potential biomarker for mapping the SOZ [6].

The recent development of implantable devices capable of both recording and stimulating the human brain via iEEG contacts has offered great opportunities for treating neurological diseases [7-8] and has opened a new frontier in the development of brain-computer interfaces (BCIs). In this scheme, we showed the feasibility of recording iEEG with the Brain Interchange (BIC) of CorTec [9-10]. In this study, we expanded the framework with a new feature that can capture interictal spikes in an online fashion while iEEG data is being recorded. Despite the pioneering works [11-12], and advanced machine learning techniques in identifying spikes with high accuracy, the challenge of online detection of these events in a clinical setting remained primarily as an important step.

This study introduces a novel MATLAB Simulink framework designed to automatically detect the spikes in an online fashion. By leveraging the robust recording capabilities of the BIC CorTec amplifier and online signal processing algorithms, our system provides a fully online analysis platform for iEEG data. Unlike previous offline frameworks [13-14] that may have required post-recording analysis, our framework detects interictal spikes as the data is being recorded. We send these detected spikes to an external application using user datagram protocol (UDP) for visualization of the morphology of these events, their spatio-temporal distributions, and further post-processing.

To evaluate this framework, we draw comparisons with a previously published offline spike detector [13], highlighting our framework's capability to achieve a high similarity index in spike detection. Furthermore, our analysis of spike detection rates within and outside the clinically defined SOZ offers compelling evidence of the system's utility in surgical planning.

In summary, our work contributes to technological and clinical advancement in epilepsy research and treatment by providing an online, accurate, and reliable method. Moreover, we opened a new avenue for immediate clinical decision-making and intervention, ultimately aiming to improve the lives of those affected by medically intractable epilepsy.

## MATERIALS AND METHODS

**Patient's demographic:** We recorded iEEG from three patients (two pediatric and one adult) diagnosed with pharmacoresistant epilepsy at Texas Children's Hospital (TCH) of Baylor College of Medicine (BCM) and Mayo Clinic. This study was approved by the Institutional Review Boards (IRBs) of BCM and Mayo

Clinic, ensuring that all experiments and methods were performed in accordance with relevant guidelines and regulations. Furthermore, informed consent was obtained from all participants and/or their legal guardians prior to incorporating their data into this study.

The recordings were acquired in the epilepsy monitoring unit (EMU) using the BIC unit (Fig. 1A), which consists of 32 channels at a sampling frequency of 1 kHz. A subset of these channels was selected based on the clinically defined SOZ, while the remaining channels were chosen from areas outside the SOZ to validate the model. A random 10-minute section of interictal data was selected for further analysis. The clinical team at the affiliated institutes provided relevant medical annotations, including information about the SOZ.

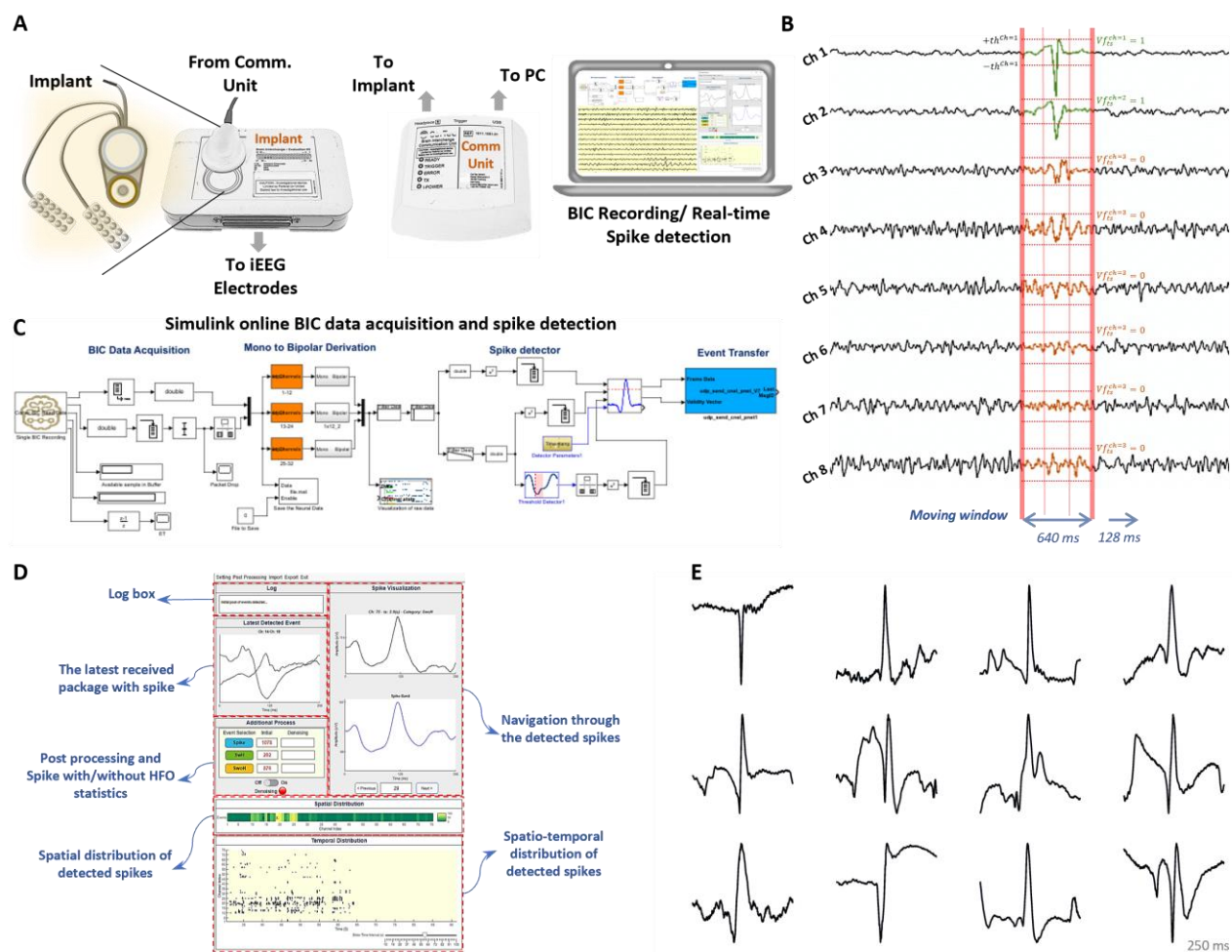


Figure 1: (A) The schematic representation of the BIC CorTec Evaluation Kit, illustrating the components, including the evaluation implant, the communication unit, and the Simulink model designed for efficient data acquisition and online spike detection. (B) The window-based amplitude threshold detector concept to capture interictal spikes in multichannel iEEG recordings. (C) The details of the Simulink model architecture, which includes the data acquisition model, monopolar to bipolar iEEG data conversion to preprocess the iEEG stream for enhanced spike detection accuracy, the spike detection algorithm, and a UDP data transfer block. This block facilitates the transmission of detected events to external software for further post-processing and visualization. (D) Showcases the external GUI developed for the post-processing and visualization of detected spikes. This interface receives the collected events, displays the spatial and temporal distribution of these events, and conducts additional post-processing to distinguish spikes with and without high-frequency oscillations (HFOs). (E) Provides examples of detected interictal spikes.

**Data recording and Online analysis framework:** To duplicate a real-time rapid prototyping environment, we developed a Simulink model for the iEEG data acquisition. The previously recorded data was then fed back to the model at real-time speed (Fig. 1B) to simulate the real data acquisition.

Furthermore, spike detection was conducted on the band-pass filtered data within the spike band range (10-55 Hz), and the detected events pool was generated within the model (Fig. 1C). This pool was then sent to an external graphical user interface (GUI, Fig. 1D) for further post-processing and visualization using UDP. Additionally, spike detection was performed using an offline detector, and the obtained results were utilized as the ground truth to evaluate the performance of the online detection method. Examples of detected spikes

are illustrated in Fig. 1E.

**Wireless data transfer and missing packets recovery:** The BIC unit facilitates wireless data transfer, which is a process inherently susceptible to data loss [9]. In this study, we addressed this challenge by employing linear interpolation to recover missing packets, thereby maintaining signal integrity. It has been demonstrated that this technique effectively restores iEEG data with minimal packet loss (<5%), particularly for spike detection in the frequency band below 80 Hz [9]. The recovered signal is then applied to subsequent analyses.

**Threshold calculation and spike detection:**

To compute the adaptive threshold for spike detection, our model applied a second-order Butterworth high-pass filter at 1 Hz to remove the DC offset. Subsequently, the signal underwent band-pass filtering using a fourth-

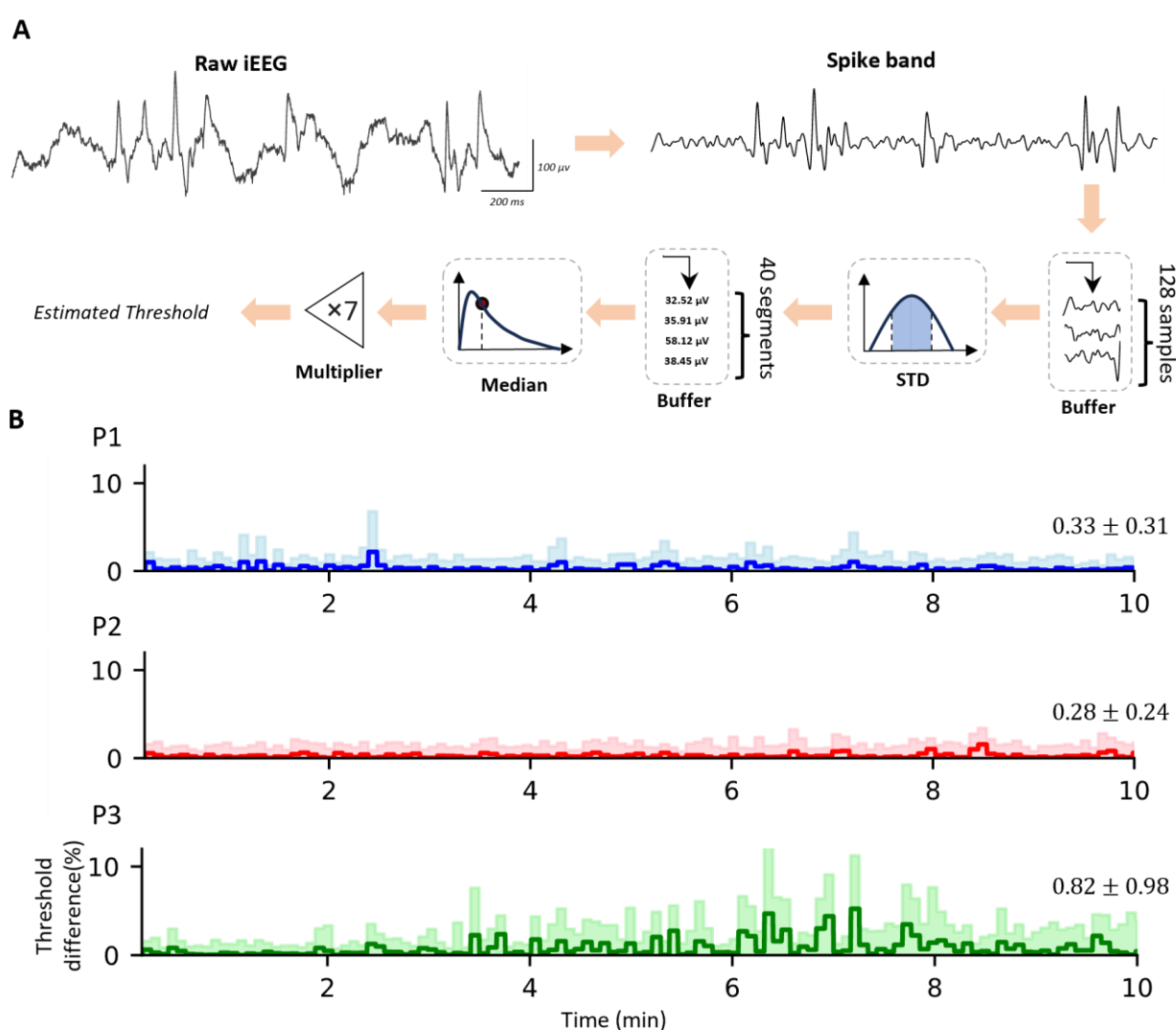


Figure 2: (A) The schematic of the real-time adaptive threshold calculation within the iEEG data analysis framework. Initially, filtered iEEG streams within the frequency range of 10-55 Hz are directed into a buffer block, which captures 128 ms of samples consecutively without overlap and calculates the standard deviation for each buffered segment. Subsequently, a second buffering stage accumulates 40 standard deviation samples without overlap, from which the median value is derived, serving as an estimation of background neural activity. The final step involves applying a multiplier to these median values, thereby generating an adaptive threshold for spike detection across each channel over intervals of 5.12 seconds. (B) The percentage thresholds difference between online and offline calculated across all channels and subjects.

order Butterworth filter with cut-off frequencies of 10 Hz and 55 Hz. The filtered signal was buffered into 128 sample-long segments and the standard deviation (std) was estimated for each frame. The std values were further buffered into 40 segment-long frames, and the median within each frame was calculated as the estimated background activity of iEEG data for each channel. Finally, a multiplier ( $\times 7$ ) was selected based on the previous work [9] to compute the adaptive threshold for the stream of data (Fig. 2A).

In this study, we observed differences in the threshold calculations of the filtered iEEG stream between online and offline analyses. These differences are due to the distinct filters employed in each process. Specifically, for offline processing, we used zero-phase filtering. This non-causal, bidirectional method leverages access to the entire dataset, leading to threshold values that may slightly differ from those generated by the causal filters employed in online processing.

The adaptive threshold is initially computed at intervals of  $128 \times 40$  milliseconds and then transformed into a continuous data stream. We utilized a rate transition block within our Simulink model to modify the sampling rate of the calculated threshold. Thus, aligning with the sampling rate of the iEEG data (1 millisecond). This adjustment ensures that the threshold applies to all iEEG samples and synchronizes with the temporal resolution of the data.

The filtered iEEG and threshold values were then buffered into intervals of 640 sample-long segments with 512 samples of overlap for spike detection using the corresponding estimated threshold. In each segment, we found the points crossing the threshold levels and grouped them as a single event if their distance was smaller than a predefined interval. Furthermore, to ensure accuracy and specificity in spike detection, we implemented a strategy to exclude polyspike components, as discussed in [13]. A spike event is selected for further analysis only if its peak value is positioned at the center of the frame, specifically at 128 samples into the 640-sample frame. An essential step in the detection process involves distinguishing distinct spike events to prevent redundancy. This criterion, aligned with the overlap size, helps in accurate event identification and isolates individual spikes.

The channel information, timestamps, and segments with identified spikes are aggregated into an event pool. This pool is then broadcasted to a secondary computer via UDP for further processing. The separation of initial data acquisition and spike detection from subsequent post-processing and visualization ensures that the recording and primary analysis continue uninterrupted with minimal computational demand. By structuring the methodology in this manner, we maintain a seamless and efficient workflow, allowing for continuous data acquisition and spike detection, followed by detailed event and pool visualization on a separate system.

## RESULTS

In this study, we compared the adaptive real-time threshold with its offline counterpart over all channels across three subjects. This comparison is shown as a shaded plot illustrating the percentage difference between the real-time and offline thresholds for all channels across all subjects (Fig. 2B). Remarkably, in every instance, the difference between these two thresholds remained under 1%, with the maximum difference observed in the last subject (P3) being  $0.82 \pm 0.98\%$ , indicating a negligible difference between real-time and offline threshold calculations.

Further analysis was conducted by deploying the online spike detector on these datasets and comparing its performance with those spikes detected offline (Fig. 3, left panel). This study focused on the rate and spatial distribution of spikes detected in both online and offline methods, as well as their occurrence in clinically defined SOZ across subjects. Our approach to comparing detected spikes involved two individual methods. Initially, we evaluated the cosine similarity between the spatial distributions of spikes detected, discovering the alignment in spike distributions across all channels, with similarity indices surpassing 0.99 and angular differences between the spatial distribution of spike vectors in online and offline analysis measuring  $1.6^\circ$ ,  $2.0^\circ$ , and  $1.8^\circ$ , respectively (Fig. 3, middle panel). Additionally, we employed the Kolmogorov-Smirnov statistical test to compare the rate of detected spikes across all channels in both online and offline analyses. This statistical evaluation revealed no significant difference, with p-values of 0.93, 0.99, and 0.99 for subjects 1-3, respectively (Fig. 3, right panel).

Importantly, our observations highlighted that the rate of detected spikes was consistently higher within the SOZ than outside the SOZ across all three cases. Notably, the initial two contacts exhibiting the highest rate of spikes were identified within the SOZ for all subjects. While this finding confirms previous works [9], it underscores the efficacy of employing the BIC CorTec system for online spike detection and emphasizes its potential in accurately identifying probable SOZ sites. This insight not only reaffirms the precision of our spike detection framework but also demonstrates its utility in enhancing the accuracy of SOZ localization, offering significant implications for the future of epilepsy treatment and management.

## DISCUSSION

The current work introduces a fully online framework designed for the detection of interictal spikes, capable of broadcasting detected events to external applications for subsequent postprocessing and visualization. The methodology is structured around three main components: first, a complete data acquisition module; second, an online spike detection module—both developed as level-2 MATLAB s-functions handling data acquisition and initial spike identification. The

third component features a user-friendly GUI that receives and visualizes the detected events. All essential signal processing blocks have been implemented in Simulink MATLAB to better control the entire framework.

The entire processing pipeline was validated by randomly selecting 10-minute segment of BIC CorTec pre-recorded interictal iEEG from three human subjects streamed in real-time as data playback to illustrate the online spike detection concept. This approach allowed for a comparison with a previously established offline spike detector, revealing that channels with the highest spike rate were associated with the SOZ.

In recent years, there has been a growing interest in spike-guided surgical intervention, referred to as spike-tailored surgery [5-6]. Furthermore, a real-time spike detection is crucial for enabling closed-loop neuromodulation or BCI applications, where timely and

accurate detection of neural activity allows for responsive and adaptive interactions between the brain and external devices. In response to these interests, and as a tool that is essential for the analysis of iEEG recordings, we implemented the online spike detector and added it to the main data acquisition setup. The developed Simulink model holds the potential for adapting to online spike detection from data streams recorded with various biomedical amplifiers, broadening its applicability in future research.

## CONCLUSION

We have successfully demonstrated the feasibility of recording iEEG from human subjects using the BIC CorTec device in a basic rapid prototyping environment within Simulink. In addition, we have integrated a real-time scenario for detecting interictal spikes as a new

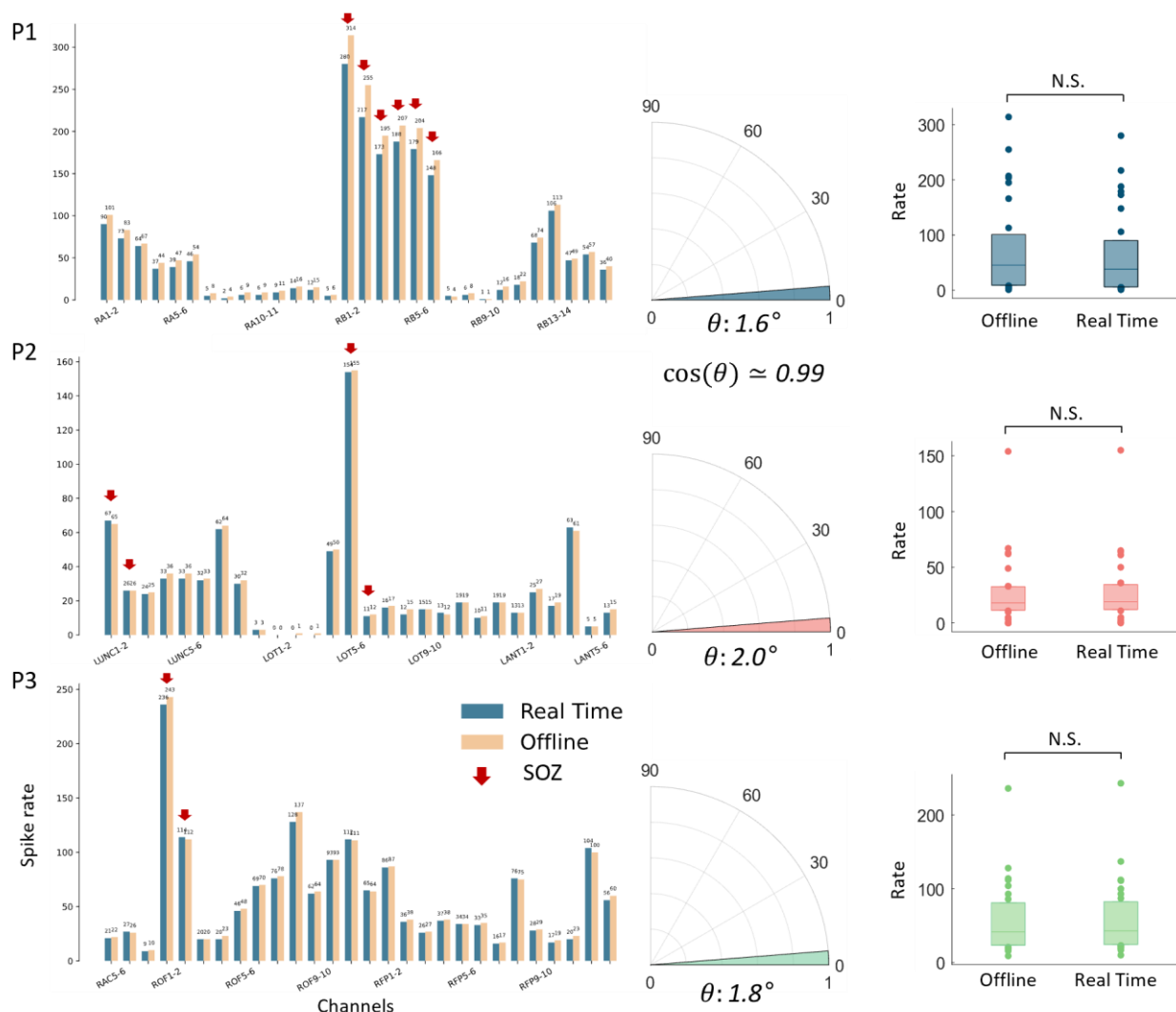


Figure 3: (Left Panel) The comparison of the spike rates across channels, contrasting the performance of online and offline spike detection. (Middle Panel) Illustrates the cosine similarity index between the spatial distributions of spikes detected in online and those identified through offline analysis. It shows the degree of alignment between the two detection methods were more than 99% in all cases. (Right Panel) Presents spike rates obtained from online and offline analyses across channels for each subject. Statistical analysis reveals no significant differences in the rate of spikes detected through online and offline analyses across the subjects, with p-values of 0.93, 0.99, and 0.99, respectively.

feature on top of this framework. Our goal is to continue evolving this framework by incorporating additional functionalities that will allow for concurrent analysis of iEEG data during the recording process.

An important aspect of our approach is the concept of broadcasting initially detected events, i.e., interictal spikes, thus transforming the data acquisition computer into a host. The host then streams the detected events to various clients for further post-processing and visualization. This strategy, when augmented with enhanced functionalities in iEEG, has the potential for iEEG surgical planning in the future.

## ACKNOWLEDGEMENTS

This study was supported by grants R01NS112497 and UH3NS117944 from the National Institutes of Health—National Institute of Neurological Disorders and Stroke. B.F.B. was supported by the Sundt fellowship of the Mayo Clinic Neurological Surgery Department.

## REFERENCES

- [1] Epilepsy [Internet]. World Health Organization; [cited 2024 Mar 13]. Available from: <https://www.who.int/news-room/fact-sheets/detail/epilepsy>
- [2] Rummel C, Abela E, Andrzejak RG, Hauf M, Pollo C, Müller M, Weisstanner C, Wiest R, Schindler K. Resected brain tissue, seizure onset zone and quantitative EEG measures: towards prediction of post-surgical seizure control. *PLoS One*. 2015 Oct 29;10(10):e0141023.
- [3] Vakharia VN, Duncan JS, Witt JA, Elger CE, Staba R, Engel Jr J. Getting the best outcomes from epilepsy surgery. *Annals of neurology*. 2018 Apr;83(4):676-90..
- [4] Schroeder GM, Karoly PJ, Maturana M, Taylor PN, Cook MJ, Wang Y. Chronic iEEG recordings and interictal spike rate reveal multiscale temporal modulations in seizure states. *arXiv preprint arXiv:2201.11600*. 2022 Jan 27.
- [5] Marsh ED, Peltzer B, Brown III MW, Wusthoff C, Storm Jr PB, Litt B, Porter BE. Interictal EEG spikes identify the region of electrographic seizure onset in some, but not all, pediatric epilepsy patients. *Epilepsia*. 2010 Apr;51(4):592-601.
- [6] Karoly PJ, Freestone DR, Boston R, Grayden DB, Himes D, Leyde K, Seneviratne U, Berkovic S, O'Brien T, Cook MJ. Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. *Brain*. 2016 Apr 1;139(4):1066-78.
- [7] Skarpaas TL, Jarosiewicz B, Morrell MJ. Brain-responsive neurostimulation for epilepsy (RNS@ System). *Epilepsy research*. 2019 Jul 1;153:68-70.
- [8] Sun FT, Morrell MJ. Closed-loop neurostimulation: the clinical experience. *Neurotherapeutics*. 2014 Jul 1;11(3):553-63.
- [9] Ayyoubi AH, Fazli Besheli B, Quach MM, Gavvala JR, Goldman AM, Swamy CP, Bartoli E, Curry DJ, Sheth SA, Francis DJ, Ince NF. Benchmarking signal quality and spatiotemporal distribution of interictal spikes in prolonged human iEEG recordings using CorTec wireless brain interchange. *Scientific Reports*. 2024 Feb 8;14(1):2652.
- [10] Schalk G, Worrell S, Mivalt F, Belsten A, Kim I, Klassen BT, Staff NP, Brunner P, Worrell GA, Miller KJ. Toward a fully implantable ecosystem for adaptive neuromodulation in humans: Preliminary experience with the CorTec BrainInterchange device in a canine model. *Frontiers in Neuroscience*. 2022 Dec 19;16:932782.
- [11] Edakawa K, Yanagisawa T, Kishima H, Fukuma R, Oshino S, Khoo HM, Kobayashi M, Tanaka M, Yoshimine T. Detection of epileptic seizures using phase–amplitude coupling in intracranial electroencephalography. *Scientific reports*. 2016 May 5;6(1):25422.
- [12] Medvedev AV, Agoureeva GI, Murro AM. A long short-term memory neural network for the detection of epileptiform spikes and high frequency oscillations. *Scientific reports*. 2019 Dec 18;9(1):19374.
- [13] Janca R, Jezdik P, Cmejla R, Tomasek M, Worrell GA, Stead M, Wagenaar J, Jefferys JG, Krsek P, Komarek V, Jiruska P. Detection of interictal epileptiform discharges using signal envelope distribution modelling: application to epileptic and non-epileptic intracranial recordings. *Brain topography*. 2015 Jan;28:172-83.
- [14] Brown III MW, Porter BE, Dlugos DJ, Keating J, Gardner AB, Storm Jr PB, Marsh ED. Comparison of novel computer detectors and human performance for spike detection in intracranial EEG. *Clinical neurophysiology*. 2007 Aug 1;118(8):1744-52.

# ENHANCING COMPUTATIONAL EFFICIENCY OF MOTOR IMAGERY BCI CLASSIFICATION WITH BLOCK-TOEPLITZ AUGMENTED COVARIANCE MATRICES AND SIEGEL METRIC

Igor Carrara<sup>1</sup>, Theodore Papadopoulos<sup>1</sup>

<sup>1</sup>Université Côte d'Azur, INRIA, Cronos Team, France

E-mail: igor.carrara@inria.fr, theodore.papadopoulos@inria.fr

**ABSTRACT:** Electroencephalographic signals are represented as multidimensional datasets. We introduce an enhancement to the augmented covariance method (ACM), exploiting more thoroughly its mathematical properties, in order to improve motor imagery classification. Standard ACM emerges as a combination of phase space reconstruction of dynamical systems and of Riemannian geometry. Indeed, it is based on the construction of a Symmetric Positive Definite matrix to improve classification. But this matrix also has a Block-Toeplitz structure that was previously ignored. This work treats such matrices in the real manifold to which they belong: the set of Block-Toeplitz SPD matrices. After some manipulation, this set can be seen as the product of an SPD manifold and a Siegel Disk Space. The proposed methodology was tested using the MOABB framework with a within-session evaluation procedure. It achieves a similar classification performance to ACM, which is typically better than – or at worse comparable to – state-of-the-art methods. But, it also improves consequently the computational efficiency over ACM, making it even more suitable for real time experiments.

## INTRODUCTION

In electroencephalography (EEG) based Brain Computer Interfaces (BCI), state-of-the-art algorithms are often built on Riemannian distance based classification algorithms [1]. The basic idea underlying these methods is to treat the spatial covariance matrix (SCM), extracted from the EEG signal, as an element of the Riemannian manifold of Symmetric Positive Definite (SPD) matrices [2].

A recent extension of this work was obtained by using the Augmented Covariance Method (ACM) [3]. ACM relies on the concept of phase space reconstruction of dynamical systems to create an "ACM matrix" (also called ACM) that contains not only an average spatial representation of the signal but also a representation of its evolution in time. Consequently, the amount of information contained in this ACM matrix is increased w.r.t. the standard spatial covariance. As the ACM matrix also turns out to be an SPD matrix, it can be classified using the same Riemannian framework that was so successful for SCMs. However, it also possesses a structural property

of being Block-Toeplitz, that is, a block matrix with constant blocks over all diagonals. Recently, an approach has been proposed to better deal with such Block-Toeplitz SPD matrices [4], with applications in diverse fields such as audio processing or radar signal analysis [5].

The idea of this research is thus to endow the smooth manifold of Block-Toeplitz SPD matrices with a Riemannian metric, thus allowing the ACM matrix to be treated within its true manifold membership. It is actually possible to treat the Block-Toeplitz SPD matrix manifold as the product of an SPD manifold and a Siegel Disk Space, after applying an appropriate conversion of the blocks of the ACM matrix into the Verblusky coefficients [6].

This approach provides a new – more specific – metric to use for BCI classification algorithms. The strength of the approach lies in its ability to deconstruct the manifold into its constituent elements: the Symmetric Positive Definite (SPD) manifold and the Siegel Disk Space. By discerningly analyzing each component within its respective geometrical domain, this method significantly alleviates the computational demand traditionally associated with the ACM methodology. The resulting algorithm achieves performance that is, at worst, comparable with state-of-the-art BCI algorithms, and often provides quite better results (on par with those of ACM). However, it distinguishes itself by achieving this at substantial reduction in computational costs and carbon footprint compared with standard ACM.

The new – Siegel metric based – pipeline was tested and validated against several state-of-the-art algorithms (Machine Learning (ML) and Deep Learning (DL)) on several datasets for motor imagery (MI) classification using several subjects and on a right versus left hand task with the MOABB framework [7], and a within-session evaluation procedure.

## MATERIALS AND METHODS

The EEG signal is represented as a multivariate time series  $\mathbf{X} \in \mathbb{R}^{d \times T}$ , where  $T$  represents the total number of sampled data points, and  $d$  indicates the number of electrodes used in the EEG recording. Since this paper focuses on MI task, we split the EEG signals into smaller sections known as epochs, each representing a snapshot

of brain activity during various tasks or mental states. The core aim of our research is to develop a method that can accurately identify the specific task or mental state associated with these EEG epochs.

The space of SPD matrices is composed by square real symmetric matrices that are positive definite, and this space form a smooth manifold that can be equipped with a Riemannian metric [1]. This space is defined as

$$\text{SPD}_d = \{\mathbf{M} \in \mathbb{R}^{d \times d} \mid x^T \mathbf{M} x > 0 \ \forall x \in \mathbb{R}^d \setminus \{0\}\} \quad (1)$$

ACM [3] extends this methodology by combining it with the phase space reconstruction (PSR) approach that is grounded in the Takens theorem [8]. The ACM matrix thus obtained contains spatial and temporal information of the signal and remains an SPD matrix that can be classified using the same Riemannian metric that was so successful for SCMs. This enrichment with temporal features of the information extracted from the signal allows for an improvement of classification performance.

The idea of using Takens theorem is based on the idea that time series obtained from experimental observations that capture only a fraction of the complex dynamics of the underlying system, can nonetheless be utilized to reconstruct the system's full dynamical behavior. This is achieved using a uniform embedding procedure: consider a time series  $s(n)$  created through a measurement process, the PSR technique generates a point  $\mathbf{s}_E(n)$  of a  $D$ -dimensional space constructed as

$$\mathbf{s}_E(n) = [s(n), s(n - \tau), \dots, s(n - (D - 1)\tau)]^T \quad (2)$$

where  $\tau$  is a positive integer called the embedding delay and  $D$  is the embedding dimension.  $\mathbf{s}_E(n) \in \mathbb{R}^D$  is a uniform embedding of the original phase space.

The ACM matrix (see Fig. 1) is obtained by expanding the original EEG signal using the PSR approach with an embedding dimension  $p$  to get a new  $dp \times T$  time series, parameterized by the fixed delay  $\tau$ . The Augmented Covariance Matrix  $\Gamma_{aug}$  is defined as the autocovariance matrix of this new time series:

$$\Gamma_{Aug} = \begin{bmatrix} \Gamma_0 & \Gamma_{-1} & \Gamma_{-2} & \cdots \\ \Gamma_1 & \Gamma_0 & \Gamma_{-1} & \cdots \\ \Gamma_2 & \Gamma_1 & \Gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \Gamma_{p-1} & \Gamma_{p-2} & \Gamma_{p-3} & \cdots \end{bmatrix}, \quad (3)$$

where  $\Gamma_0$  is the standard spatial covariance matrix and  $\Gamma_i$  is the lagged covariance matrix of the original signal with a delay of  $i\tau$  and  $\Gamma_{-i} = \Gamma_i^T$ . As an autocovariance matrix,  $\Gamma_{aug}$  is symmetric and positive by construction. If not definite, it can be regularized [9], so that we consider it as SPD in the remainder of this article. But, the ACM matrix also has a specific Block-Toeplitz structure, with blocks of dimension  $d \times d$  [5]. More formally,  $\Gamma_{Aug}$  belongs to the space  $\mathbb{B}_{d \times p}$  of Block-Toeplitz and SPD matrices i.e., SPD matrices of size  $dp \times dp$  with constant blocks of size  $d \times d$  along all diagonals. This opens up new possibilities for enhancing the ACM formulation by

mapping the ACM matrix to the most suitable geometric space that fully captures both its Block-Toeplitz and SPD natures.

The blocks of the matrix  $\Gamma_{Aug}$ , have been demonstrated to belong to a specific mathematical space [4]

$$\Gamma_i \in \mathbb{D}_d \quad \mathbb{D}_d = \{\mathbf{M} \in \mathbb{C}^{d \times d} \mid \mathbf{I} - \mathbf{M}\bar{\mathbf{M}} > 0\} \quad (4)$$

with  $\bar{\mathbf{M}} = \mathbf{J}\mathbf{M}^H\mathbf{J}$  where  $\mathbf{J}$  denotes the anti-diagonal matrix and  $^H$  is the conjugate transpose operator<sup>1</sup>. This space has no known Riemannian structure but, by implementing a minor adjustment to the coefficients, it is possible to ensure their belonging within the domain of the Siegel disk [6, 10], defined as

$$\Omega_i \in \mathbb{SD}_d \quad \mathbb{SD}_d = \{\mathbf{M} \in \mathbb{C}^{d \times d} \mid \mathbf{I} - \mathbf{M}\mathbf{M}^H > 0\}. \quad (5)$$

The coefficients that have undergone such modification are also known as Verblunsky coefficients [6].

The transformation allow the following conversion,

$$\begin{aligned} \mathbb{B}_{d \times p} &\rightarrow \text{SPD}_d \times \mathbb{SD}_d^{p-1} \\ \Gamma_{Aug} &\mapsto (\Gamma_0, \Omega_1, \dots, \Omega_{p-1}). \end{aligned} \quad (6)$$

Consider the initial matrix  $\Gamma_{Aug}$  decomposed in its constituent blocks  $(\Gamma_0, \dots, \Gamma_{p-1})$ . The initialization of the recursive transformation is set to  $\mathbf{P}_0 = \Gamma_0$ . The subsequent coefficients are computed with

$$\Omega_{l+1} = \mathbf{L}_l^{-1/2}(\mathbf{R}_{l+1} - \mathbf{M}_l)\mathbf{K}_l^{-1/2}, \quad (7)$$

with  $l = 0, \dots, p-1$  and

$$\begin{aligned} \mathbf{L}_l &= \mathbf{P}_0 - (\Gamma_1, \dots, \Gamma_l)\tilde{\Gamma}_{l-1}^{-1}(\Gamma_1, \dots, \Gamma_l)^H \\ \mathbf{K}_l &= \mathbf{P}_0 - (\Gamma_1^H, \dots, \Gamma_l^H)\tilde{\Gamma}_{l-1}^{-1}(\Gamma_1^H, \dots, \Gamma_l^H)^H \\ \mathbf{M}_l &= (\Gamma_1, \dots, \Gamma_l)\tilde{\Gamma}_{l-1}^{-1}(\Gamma_1^H, \dots, \Gamma_l^H)^H \end{aligned}$$

where  $\tilde{\Gamma}_{l-1}$  denotes the sub-matrix of  $\Gamma_{Aug}$  obtained by keeping only its first  $l-1$  rows and columns. This transformation operates recursively, enabling the foundational blocks of the  $\Gamma_{Aug}$  matrix to be transformed into square matrices that are positioned within the domain of the Siegel Disk.

The smooth manifold of  $\mathbb{B}_{d \times p}$  is thus identified as a Kähler manifold [4], on which is possible to define a Kähler potential  $\Phi$  [11, 12], computed as:

$$\Phi(\Gamma_{Aug}) = -\log(\det(\Gamma_{Aug})) - \log(\pi e) \quad (8)$$

After applying some decomposition properties of the determinant of  $\Gamma_{Aug}$ , it is possible to compute the metric of the manifold simply as the Hessian matrix of the Kähler potential

$$\begin{aligned} ds^2 &= p \text{trace}(\mathbf{P}_0^{-1} d\mathbf{P}_0 \mathbf{P}_0^{-1} d\mathbf{P}_0) \\ &+ \sum_{l=1}^{p-1} (p-l) \text{trace} \left( (\mathbf{I} - \Omega_l \Omega_l^H)^{-1} d\Omega_l (\mathbf{I} - \Omega_l^H \Omega_l)^{-1} d\Omega_l^H \right) \end{aligned} \quad (9)$$

<sup>1</sup> $A > B$  when  $A - B$  is a positive definite matrix.

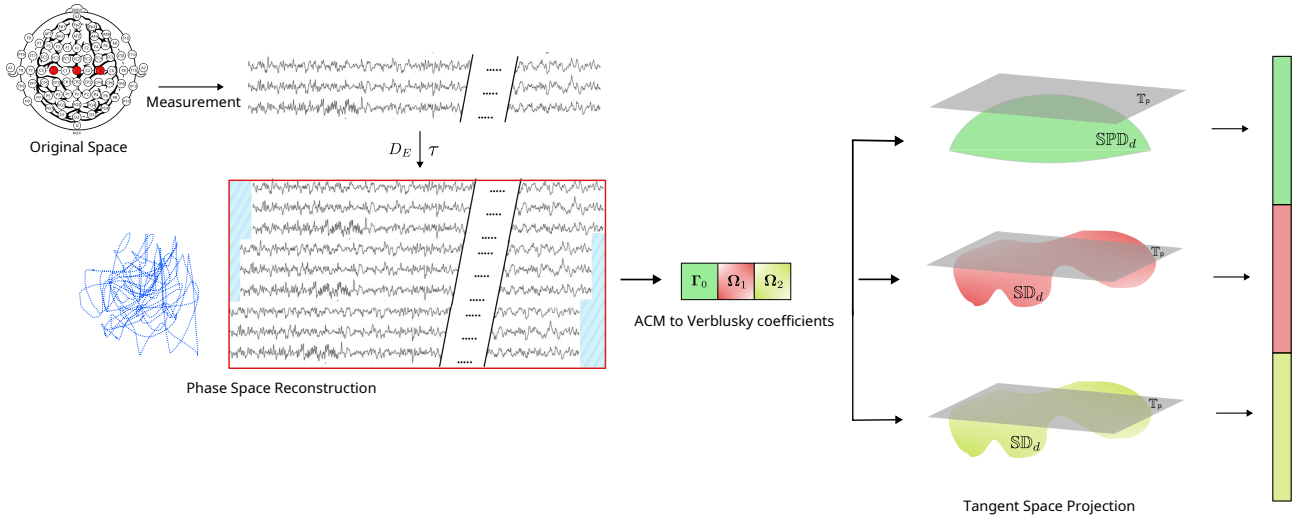


Figure 1: Schematic illustration of the BT-ACM + TG + SVM methodology. The presented example uses only 3 electrodes (in red on the top left plot). The measurement process of the original dynamic system is thus a 3-dimensional time series. The process then begins with the extraction of epoch signal representing left and right hand tasks. We then use the phase space reconstruction process to obtain a dynamic system equivalent to the original one (selection of hyper-parameters are made via grid search using the nested approach). In this figure, we see an embedding corresponding to  $p = 3$  and  $\tau = 10$ . The BT-ACM matrix is computed as the autocovariance of this high-dimensional time series. Subsequently the main blocks are converted in Verblunsky coefficients. Then, each component is mapped to the tangent space using the appropriate Riemannian manifold computations and vectorized. The final step is the application of an SVM-based classification algorithm.

The first term of Equation (9) is identifiable as the metric for the  $\mathbb{SPD}_d$  space. The other term represents the metric for the Siegel disk space  $\mathbb{SD}_d$  repeated  $p - 1$  times, i.e. the space of Block-Toeplitz SPD matrices is equipped with a product Riemannian metric over  $\mathbb{SPD}_d \times \mathbb{SD}_d^{p-1}$ . The resulting algorithm (depicted in Fig. 1) using this new metric using SVM on the tangent space is called BT-ACM+TS+SVM: Block-Toeplitz Augmented Covariance Matrix (BT-ACM) with Tangent Space projection (TS) and SVM classifier.

Using fixed hyper-parameters  $p$  and  $\tau$ , the PSR approach expands the signal ( $p$  and  $\tau$  will subsequently be carefully chosen using a grid-search procedure). The spatial autocovariance matrix of the expanded signal is obtained using regularization through the Oracle Approximating Shrinkage Estimator (OAS) [9]. After the Verblunsky transformation, each component is mapped to the Tangent space using the Logarithmic map of each specific manifold. The final classification step is obtained with a Support Vector Machine (SVM) algorithm.

**Dataset and Evaluation procedure:** To validate the proposed methodology, we use open accessible datasets available from MOABB [7]. We selected 3 datasets with a total of 70 subjects. All the information regarding the considered datasets are presented in Tab. 1.

Table 1: Dataset considered during this study.

Dataset	subjects	channels	sampling rate	trials/class	Epoch (s)
BNCI2014001 [13]	9	22	250 Hz	144	[2, 6]
BNCI2014004 [14]	9	3	250 Hz	360	[3, 7.5]
Cho2017 [15]	52	64	512 Hz	100	[0, 3]

The duration of each epoch within our study is intentionally aligned with the task conditions' length, which

is subject to variation across the datasets employed. On each dataset, we applied a standard band pass filter procedure for the Motor Imagery task, in the frequency range of 8 to 32 Hz.

We use a Within-Session (WS) evaluation procedure as provided in MOABB. This means that our analysis works on each session separately. The implementation is based on a Nested Cross-Validation methodology [16], structured with an outer loop of 5-Fold Cross validation and a inner one composed by a 3-fold Cross Validation. We use statistical tests provided by MOABB to confront the different pipelines, i.e., based on a t-test [17] for datasets with less than 20 subjects, or a Wilcoxon non-parametric signed-rank test [18] otherwise.

The state-of-the-art pipelines used in this research contain both Machine Learning (ML) and Deep Learning (DL) methods. Detail of the pipelines are listed in Tab. 2.

For DL pipelines, we used a standardization step that normalizes every channel to have a zero mean and unit standard deviation. Additionally, we employed a re-sampling procedure to ensure that each architecture integrates a temporal filter aligned with the state-of-the-art techniques' implementations. This procedure was added in order align to the state-of-the-art implementation and avoid the need of redoing hyper-parameter tuning. The DL pipelines are using a Sparse Categorical Cross-Entropy loss function and a standard Adam optimizer using 300 epochs and a batch size of 64. To avoid overfitting, we used an early stopping procedure with a patience parameter of 75.

## RESULTS

Table 2: Pipelines considered in this study are organized into two distinct sections within the table: the first part is dedicated to the traditional classical ML pipelines, while the second part focuses on DL pipelines for MI.

Pipeline	Feature Extraction	Classifier
CSP + LDA [19]	Common Spatial Patterns (CSP) with OAS covariance estimator	Optimized Shrinkage LDA
MDM [1]	Spatial Covariance using OAS	Mean Distance to Mean (MDM)
FgMDM [1]	Spatial Covariance using OAS	Minimum Distance to Mean with geodesic filtering (FgMDM)
TS + EL [20]	Spatial Covariance using OAS mapped to TS	Optimized Elastic Network (EL)
TS + SVM [1]	Spatial Covariance using OAS mapped to TS	Optimized SVM
ACM + TS + SVM [3]	ACM with Sample Covariance Estimator mapped to TS	Optimized SVM
BT-ACM + TS + SVM (Proposed)	BT-ACM with Sample Covariance using OAS mapped to each respectively TS	SVM
ShallowConvNet [21]	Standardized and resample EEG signal at 250Hz	Convolutional Neural Network (CNN)
DeepConvNet [21]	Standardized and resample EEG signal at 250Hz	CNN
EEGNet 8 2 [22]	Standardized and resample EEG signal at 128Hz	CNN with architecture EEGNet

In this section, we describe the results obtained for the Right vs Left hand classification task.

Table 3: Performance (AUC) of Right hand vs Left hand classification. The table contains the results over all subjects (average plus or minus standard deviation).

Pipeline	BNCI2014001	BNCI2014004	Cho2017
CSP+LDA	0.82 ± 0.17	0.80 ± 0.15	0.71 ± 0.15
MDM	0.82 ± 0.15	0.78 ± 0.16	0.63 ± 0.14
FgMDM	0.87 ± 0.12	0.79 ± 0.15	0.73 ± 0.13
TS+EL	0.86 ± 0.13	0.80 ± 0.15	<b>0.76 ± 0.14</b>
TS+SVM	0.87 ± 0.14	0.79 ± 0.15	0.75 ± 0.14
ACM+TS+SVM	<b>0.92 ± 0.10</b>	<b>0.83 ± 0.15</b>	0.74 ± 0.15
BT-ACM+TS+SVM	0.89 ± 0.11	<b>0.83 ± 0.14</b>	<b>0.76 ± 0.14</b>
ShallowConvNet	0.86 ± 0.14	0.72 ± 0.18	0.74 ± 0.15
DeepConvNet	0.82 ± 0.16	0.72 ± 0.19	0.72 ± 0.13
EEGNet	0.77 ± 0.19	0.70 ± 0.20	0.67 ± 0.16

The various approaches of this study are compared in Tab. 3. A detailed picture of the results and the statistical significance of these results is provided in Fig. 2.

Overall, our method scores the best in 2 datasets - BNCI2014001 and Cho2017 - and obtains similar performance with respect to ACM+TS+SVM and TS+EL respectively for the third dataset. Across other datasets evaluated, our approach consistently delivers results that closely rival those of other leading algorithms, with a marginal performance deviation of no more than 1% in the AUC score with the only exception of BNCI2014001 dataset (where only the more costly ACM+TS+SVM does better).

Moreover, a comprehensive analysis across multiple datasets underscores a statistically significant performance enhancement achieved by BT-ACM+TS+SVM compared to all considered algorithms, with the sole exception of TS+EL, where the outcomes are remarkably similar (Fig. 2 (a) and (d)). This evidence collectively affirms the superiority of our method, not only in achieving high-performance benchmarks, but also in maintaining competitive results across different datasets.

We further explored the carbon emission and the computational time. To perform this analysis, we have run all the algorithms on the same hardware, a Dell C6420 dual-Xeon Cascade Lake SP Gold 6240 @ 2.60GHz. Furthermore, in order to conduct a fair comparison, we considered ACM+TS+SVM and BT-ACM+TS+SVM with the same number of parameter to optimize, i.e. we optimize the order and the lag of the augmentation procedure in the range [1 – 10] without any optimization of the SVM parameter. Fig. 2 (b) shows the timings and

estimated carbon footprint for both the ACM+TS+SVM and the BT-ACM+TS+SVM pipelines over the dataset BNCI2014001 composed by 9 subject over 2 sessions. The carbon footprint was estimated using Code Carbon [23] and expressed as gCO2 equivalent emission.

## DISCUSSION

Our analysis across multiple datasets demonstrates (see Fig. 2 (a)) that the BT-ACM+TS+SVM algorithm not only competes with but frequently surpasses the performance of current state-of-the-art methodologies. The only exception to this trend is when compared to the TS+EL algorithm, where our results are statistically indistinguishable (Fig. 2 (d)). Despite their close relationship, the BT-ACM+TS+SVM algorithm shows a significant (even if small) superior performance compared to the ACM+TS+SVM algorithm (Fig. 2 (c)).

We also noticed an improved stability over changes of the SVM parameter, which thus does not need to be optimized as we did in the TS+SVM and ACM+TS+SVM cases (see Table 2).

In addition to its improved classification performance, the BT-ACM+TS+SVM algorithm also exhibits significant advancements in computational efficiency. The difference in computational time is statistically significant as we pass from a mean time of  $(94.58 \pm 2.62)s$  for the ACM methodology to  $(79.71 \pm 0.80)s$  for the BT-ACM+TS+SVM approach. The results have the same level of statistical significance for the carbon emission. Note that the comparison is done using the same number of parameters. The fact that we did not optimize the SVM regularization parameter is not the reason for this improvement.

It is also noteworthy that the ACM methodology exhibits a significant variability in computational times across different sessions and subjects, indicating a fluctuation in performance consistency. In contrast, the BT-ACM approach demonstrates enhanced stability, showcasing a more uniform and predictable computational time.

## CONCLUSION

Throughout this research, we focused on the uses of the Block-Toeplitz Augmented Covariance matrix (BT-ACM) for Motor Imagery classification. This methodology extends the current ACM by comprehensively utilizing the Block-Toeplitz properties of the BT-ACM matrix.

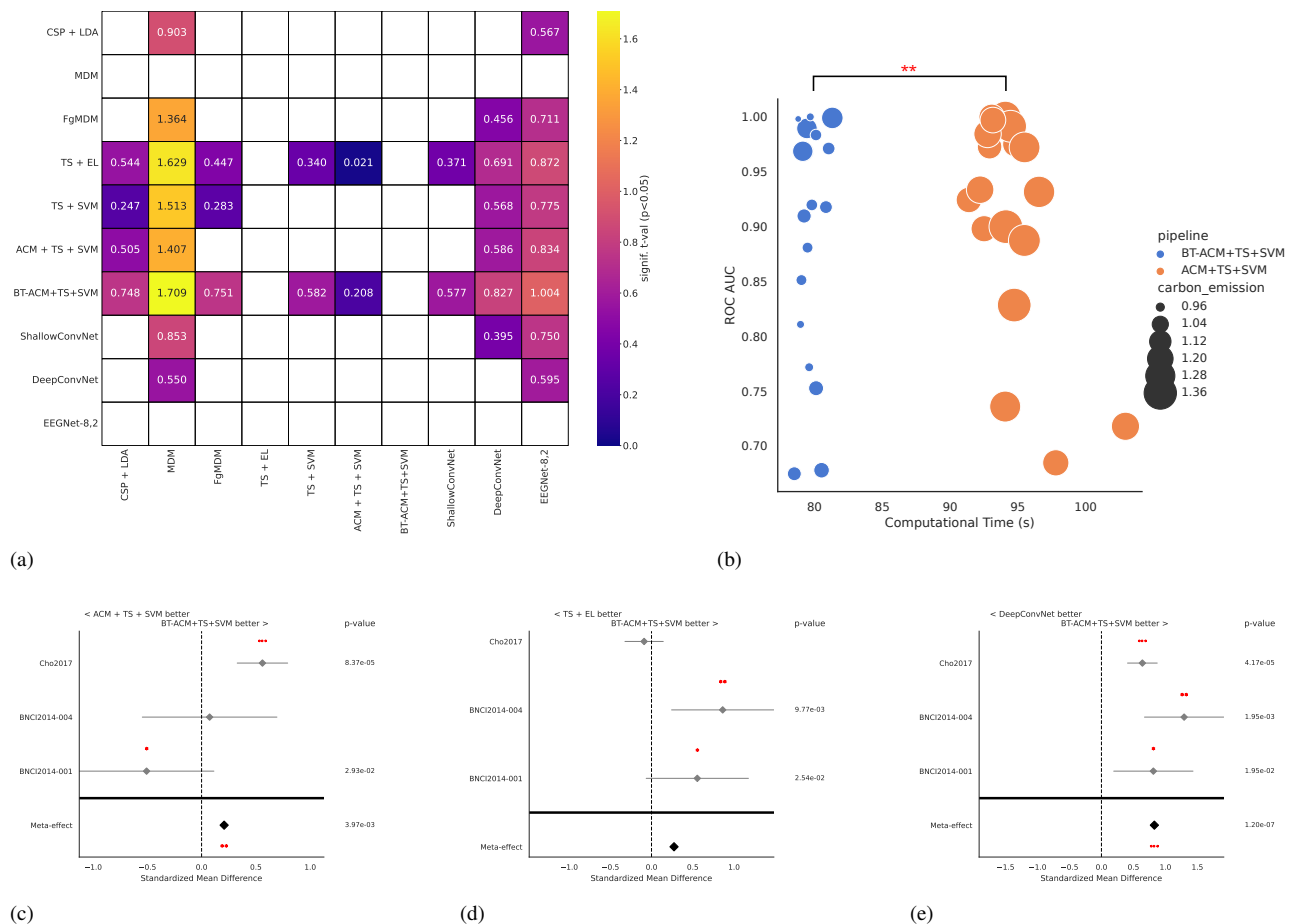


Figure 2: Results for Right vs Left hand classification, using WS evaluation. Plot (a) provides a combined meta analysis (over all datasets) of the different pipelines. It shows the significance that the algorithm on the y-axis is better than the one on the x-axis. The color represents the significance level of the difference of accuracy, in terms of t-values. We only show significant interactions ( $p < 0.05$ ). Plots (b) summarizes the computational time and carbon footprint of ACM+TS+SVM vs BT-ACM+TS+SVM. Plots (c), (d) and (e) show the meta analysis of BT-ACM+TS+SVM against respectively ACM+TS+SVM (Grid), TS+EN, DeepConvNet. We show the standardized mean differences of p-values computed as one-tailed Wilcoxon signed-rank test for the hypothesis given as title of the plot. The gray bar denotes the 95% interval. \* stands for  $p < 0.05$ , \*\* for  $p < 0.01$ , and \*\*\* for  $p < 0.001$ .

This approach transforms the classification challenge, enabling a distinct analysis within the separate domain of SPD and Siegel Disk matrices.

This procedure achieves performance generally superior to the state-of-the-art, or at worst comparable. But, the improvement over the standard ACM is not only in terms of ROC-AUC but also in terms of significant reductions in computational costs and carbon emissions.

An interesting future exploration emerges from the fact that we are not using directly the property of the blocks contained in the BT-ACM matrix since the metric does not belong to any known manifold. In order to treat the problem, we were forced to use the Verblusky coefficient transform that introduces possible errors and complications. This means that it might be interesting to develop the mathematical framework for the direct treatment of such BT matrices.

## ACKNOWLEDGMENT

This work has been partially financed by a EUR

DS4H/Neuromod fellowship. The authors are grateful to the OPAL infrastructure from Université Côte d’Azur for providing resources and support.

## \* Data and Code Availability

The codes used to produce the results of this study are publicly available in this Github repository: [https://github.com/carraraig/BTACM\\_BCI](https://github.com/carraraig/BTACM_BCI).

## References

- [1] Barachant A, Bonnet S, Congedo M, Jutten C. Riemannian geometry applied to bci classification. In: International conference on latent variable analysis and signal separation. 2010, 629–636.
- [2] Förstner W, Moonen B. A metric for covariance matrices. In: Geodesy-the Challenge of the 3rd Millennium. Springer, 2003, 299–309.

- [3] Carrara I, Papadopoulou T. Classification of bci-eeg based on augmented covariance matrix. arXiv preprint arXiv:2302.04508. 2023.
- [4] Jeuris B, Vandebril R. The kahler mean of block-toeplitz matrices with toeplitz structured blocks. SIAM Journal on matrix analysis and applications. 2016;37(3):1151–1175.
- [5] Cabanes Y. Multidimensional complex stationary centered gaussian autoregressive time series machine learning in poincaré and siegel disks: Application for audio and radar clutter classification. Ph.D. dissertation. Université de Bordeaux. 2022.
- [6] Dette H, Wagener J. Matrix measures on the unit circle, moment spaces, orthogonal polynomials and the geronimus relations. Linear algebra and its applications. 2010;432(7):1609–1626.
- [7] Aristimunha B *et al.* Mother of all bci benchmarks. Zenodo. 2023.
- [8] Takens F. Detecting strange attractors in turbulence. In: Dynamical systems and turbulence, Warwick 1980. Springer, 1981, 366–381.
- [9] Chen Y, Wiesel A, Eldar YC, Hero AO. Shrinkage algorithms for mmse covariance estimation. IEEE transactions on signal processing. 2010;58(10):5016–5029.
- [10] Fritzsche B, Kirstein B. An extension problem for non-negative hermitian block toeplitz matrices. ii. Mathematische Nachrichten. 1987;131(1):287–297.
- [11] Ballmann W. Lectures on kähler manifolds. European mathematical society (2006).
- [12] Yang L. Medians of probability measures in riemannian manifolds and applications to radar target detection. Ph.D. dissertation. Université de Poitiers. 2011.
- [13] Tangermann M *et al.* Review of the bci competition iv. Frontiers in neuroscience. 2012:55.
- [14] Leeb R, Lee F, Keinrath C, Scherer R, Bischof H, Pfurtscheller G. Brain–computer communication: Motivation, aim, and impact of exploring a virtual apartment. Neural Systems and Rehabilitation Engineering, IEEE Transactions on. 2007;15(4):473–482.
- [15] Cho H, Ahn M, Ahn S, Kwon M, Jun SC. Eeg datasets for motor imagery brain–computer interface. GigaScience. 2017;6(7):gix034.
- [16] Cawley GC, Talbot NL. On over-fitting in model selection and subsequent selection bias in performance evaluation. The Journal of Machine Learning Research. 2010;11:2079–2107.
- [17] Student. The probable error of a mean. Biometrika. 1908:1–25.
- [18] Wilcoxon F. Individual comparisons by ranking methods. In: Breakthroughs in statistics. Springer, 1992, 196–202.
- [19] Lotte F, Guan C. Learning from other Subjects Helps Reducing Brain-Computer Interface Calibration Time. In: International Conference on Audio Speech and Signal Processing (ICASSP). Dallas, United States, Mar. 2010, 614–617.
- [20] Corsi MC, Chevallier S, Fallani FDV, Yger F. Functional connectivity ensemble method to enhance bci performance (fucone). IEEE Transactions on Biomedical Engineering. 2022.
- [21] Schirrmeister RT *et al.* Deep learning with convolutional neural networks for eeg decoding and visualization. Human brain mapping. 2017;38(11):5391–5420.
- [22] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces. Journal of neural engineering. 2018;15(5):056013.
- [23] Benoit C *et al.* Codecarbon. Zenodo. 2024.

# SYNTHESIZING EEG SIGNALS FROM EVENT-RELATED POTENTIAL PARADIGMS WITH CONDITIONAL DIFFUSION MODELS

Guido Klein<sup>1\*</sup>, Pierre Guetschel<sup>1\*</sup>, Gianluigi Silvestri<sup>1,2</sup>, Michael Tangermann<sup>1</sup>

<sup>\*</sup>These authors contributed equally to this work.

<sup>1</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

<sup>2</sup>OnePlanet Research Center, imec-the Netherlands, Wageningen, Netherlands

E-mail: pierre.guetschel@donders.ru.nl

## ABSTRACT:

Generative models, specifically diffusion models, can alleviate data scarcity in the brain-computer interface field. While diffusion models have previously been successfully applied to electroencephalogram (EEG) data, existing models lack flexibility regarding sampling or require alternative representations of the EEG data. To overcome these limitations, we introduce a novel approach to conditional diffusion models that utilizes classifier-free guidance to directly generate subject-, session-, and class-specific EEG data. In addition to commonly used metrics, domain-specific metrics are employed to evaluate the specificity of the generated samples. The results indicate that the proposed model can generate EEG data that resembles real data for each subject, session, and class.

## INTRODUCTION

One of the most significant challenges of data scarcity in the brain-computer interface (BCI) field is that the acquisition of annotated data is a time-intensive endeavor. The lack of large labeled datasets can be a bottleneck for many machine learning algorithms [1]. Additionally, class imbalances typically found in event-related potential (ERP) protocols which are among the most commonly used EEG-BCI paradigms [2], can be detrimental to classifier performance. Moreover, multiple populations are underrepresented in the current corpus of EEG data [1, 3].

Generative models offer a promising solution to alleviate this data scarcity. Diffusion models, in particular, have shown the ability to generate high-quality data in a variety of domains, including images [4] and audio [5]. Current implementations of diffusion models for EEG data generation are either trained directly on EEG data or use an alternative representation, such as electrode frequency density maps [6], spatial covariance matrices [7], time-frequency maps, and latent representations [8–10].

Models trained on alternative representations, while potentially being easier to train, require an additional pre- and post-processing step, which can hinder their usability. The models trained directly on EEG data are either unconditioned [11, 12], which means that the samples are always generated from the full data distribution, or con-

ditioned, which means that the models are trained on the full data distribution but samples can be generated from a selected part of the data distribution. This conditioning can either be achieved using a classifier [4, 13] or by *classifier-free guidance* [14, 15], which achieves conditioning without the need for a noisy classifier.

Despite the capability of diffusion models to generate high-quality EEG data, there is a lack of proper metrics to quantify the quality of the generated samples. Currently used metrics are either adopted from the image domain, are domain-invariant, or rely on classifier performance [6–11, 13, 15–17]. The metrics from the image domain, the Fréchet inception distance (FID) [4] and the inception score (IS) [4], rely on the activations and output of a standardized trained neural network called Inception V3 [4]. Unfortunately, there is no universally adopted trained network for EEG data, which makes fair and reliable comparison impossible. Additionally, the domain-invariant metrics are incapable of discerning which domain-relevant features are generated well by the model. For example, if there is a low Euclidean distance between the generated and real samples, then that is likely due to a similarity in multiple domain-relevant features, such as amplitude and peak latencies, this makes it unclear which domain-specific features are properly generated. Similarly, metrics based on classifier performance are also unable to disentangle these features. Hence, there is a need for a set of metrics that can capture these domain-specific features.

*Research questions and objectives:* The following research question is investigated: Can we generate artificial ERP examples that are specific to a subject, session, and class using conditional diffusion models with classifier-free guidance?

To answer this question, we train a novel conditional diffusion model to flexibly generate each combination of conditions, i.e., subject, session, and class. An example of generated data can be found in Figure 1. Domain-invariant and image-domain metrics are used to evaluate the quality of generated samples during training. However, as previously noted, these metrics are unable to capture domain-specific features, which makes it impossible to evaluate the ability of the model to generate ERP

data that is specific to a certain condition. Therefore, we aim to introduce domain-specific metrics, which use ERP-related features, to verify the ability of the model to generate ERP data that is specific to each combination of conditions.

## METHODS - DATA DESCRIPTION

The conditional diffusion model is trained on a visual ERP dataset collected by Lee et al. [2]. Visual ERP responses are elicited using a modified oddball paradigm. The most prominent ERP feature is expected to be a positive deflection that occurs approximately 300 ms (referred to as a P300) after being presented with a relatively infrequent target stimulus following multiple non-target stimuli [18]. This dataset is one of three datasets that were recorded to study BCI-inefficiency across three major BCI paradigms: visual ERP, motor imagery, and steady-state visually evoked potential protocols [2]. In the study, fifty-four participants underwent two sessions which were held on different days [2].

During data recording, each visual ERP session was divided into a train and test run. During the train run, trials were not decoded, while during the test run, the trials were decoded and feedback was given to the participant after each trial [2]. Both runs of the same session are combined to train the diffusion model, as they employ the same copy-spelling tasks.

The dataset has been obtained and preprocessed using the Mother of All BCI Benchmarks [19]. It was preprocessed with a relatively simple pipeline. First, 19 channels (Fp1, Fp2, F7, F8, F3, F4, Fz, T7, T8, C3, C4, Cz, P7, P8, P3, P4, Pz, O1, and O2) were selected that provide full scalp coverage. Secondly, the data was bandpass filtered between 1 and 40 Hz with a 4th-order Butterworth filter. Thirdly, the data was downsampled from 1000 Hz to 128 Hz. Fourthly, epochs were constructed as 1-second windows, starting from the stimuli onsets. Lastly, peak-to-peak epoch rejection is applied with a threshold of 150  $\mu$ V. This removed 14.5 % of epochs, equivalent to 63672 epochs out of a total of 438840 epochs. Subject 17 was dropped due to excessive artifacts.

## METHODS - DIFFUSION MODELS

Diffusion models are a generative modeling paradigm where data is progressively destroyed by injecting Gaussian noise, and a neural network is trained to reverse this process [14]. Song and colleagues provide a continuous formulation of such a process, formulated as a stochastic differential equation (SDE), and show how a neural network can be implemented to learn the reverse SDE [20]. In this work, the implementation is based on the variance persevering SDE (VP SDE), which is the continuous equivalent to the noise injection used in the denoising diffusion probabilistic model (DDPM) [20, 21]. Furthermore, “classifier-free guidance” is used to condition the model on the subject, session, and class in parallel [14].

*Implementation details:* The neural network to reverse the destruction of the data is based on the architectures introduced by Torma et al. and Shu et al., called EEG-Wave and diff-EEG [11, 15]. Two key differences were introduced: 1) the timestep embedding was rewritten to be compatible with the VP SDE, because both models use DDPM noise injection, and 2) no normalization is applied to the EEG data.

The model is trained for 900 k steps, with the model being evaluated every 100 k steps. An exponential moving average of the weights is used for sampling and metric calculation. This sampling is done by a predictor-corrector sampler [20]. Preliminary results indicated that an increase in the signal-to-noise ratio (SNR) of the corrector increases the amplitude of the generated EEG data. The number of generated samples matches the number of real samples available for a given subject/session/class combination. This includes the samples used to compute the validation loss but excludes samples removed by the epoch rejection.

For more information about the implementation please refer to: [https://neurotechlab.socsci.ru.nl/resources/generative\\_models/](https://neurotechlab.socsci.ru.nl/resources/generative_models/)

## METHODS - SIMILARITY METRICS

Metrics quantifying the similarity between generated and real data are crucial for model comparison and evaluation. These similarity metrics are divided into four categories: *classifier performance*, *domain-invariant*, *image-domain*, and *domain-specific*. This section will also discuss metric-specific baselines for interpreting the scores obtained on the domain-invariant and domain-specific metrics.

Domain-invariant and domain-specific metrics are computed between the real and generated data within one condition, i.e., a combination of subject, session, and class. These metrics have been computed separately per condition, but their average across conditions is reported.

*Classifier performance:* We compare the performance of a classifier trained on generated data with the performance of a classifier trained on real data. Both conditions use the same test sets which contain only real data. This metric is denoted as the averaged balanced accuracy (**ABA**) and the score obtained when training on the real data is reported as the *within-session baseline*. Specifically, training is subject-specific and implements a within-session five-fold stratified cross-validation. The classifier is a regularized least-squares Linear Discriminant Analysis (LDA) [22]. The LDA is trained on features that represent the average amplitude across channels within non-overlapping time windows, which span between 0.1 to 0.9 seconds and are each 0.1 seconds long.

*Domain-invariant metrics*, such as the sliced-Wasserstein distance (SWD), mean squared error, and Jensen-Shannon divergence, can be used to measure the (dis)similarity between the generated and real data. However, there is no consensus on which metric

to use, with multiple articles implementing different domain-invariant metrics [8, 9, 11, 16]. Therefore, we arbitrarily choose to implement the SWD [23]. The domain-specific metrics introduced later allow for a more nuanced approach to metric selection.

*Image-domain metrics* require a pre-trained classifier, as they are based on its latent activations. Inception V3 [4] is the one used for images but is not suitable for EEG data, so we have trained an EEGNet architecture [24]. EEGNet was chosen as it only has a few hyperparameters, a relatively low number of parameters, and achieves reasonable accuracy. This trained EEGNet model has been made public, allowing future studies to calculate the FID and IS using the same model<sup>1</sup>.

Although the IS is a commonly used metric from the image domain, it is not included in our analysis. The IS leverages the outputs of the classifier to extract information about the quality and diversity of the samples. However, the diversity of samples generated by a conditional diffusion model is, practically speaking, arbitrary. Moreover, the trained EEGNet is biased toward the majority class. Thus, the output of the classifier is unable to accurately measure the diversity of the samples. Therefore, we have decided to exclude it from the analysis.

Fortunately, the other important metric from the image domain, the FID, does not suffer from the same problem as the IS. The FID is calculated by computing the Fréchet distance between a Gaussian fitted to the mean and standard deviations of the activations in the last pooling layer of a trained classifier in response to the real and generated data [25]. The Fréchet distance is small when the features that are picked up by the trained classifier are similar. Unfortunately, it is almost impossible to discern which features contribute to a low FID, because it is unclear which features would cause similar activations. This problem, however, can be tackled by the domain-specific metrics.

*Four Domain-specific metrics* are introduced in the following section. They are designed to address the shortcomings of the previous metrics to evaluate the ability of the model to generate domain-specific features. In particular, their design exploits the stationarity of ERP brain responses by using averaged responses, thus increasing the SNR [22].

The difference between generated and real P300 peaks is measured by the *peak latency delta (PLD)* and *peak amplitude delta (PAD)* metrics. These metrics only consider the channel with the most prominent P300 peak when averaged over all the real target data, which is channel "O1" in the case of the Lee 2019 ERP dataset. The PAD is computed by taking the absolute difference in  $\mu\text{V}$  between the highest peak in the real and generated data at the selected channel. The PLD is the absolute difference in time offset between these peaks, measured in ms. Only target trials are included, as a P300 peak is necessary to compute these metrics.

One of the main downsides of using averaged data is

that information about the diversity of samples is lost. Nonetheless, it is important that the model can generate a variety of EEG data. The final metric, called the *standard deviation Manhattan distance (SD-MD)*, addresses this concern by computing the absolute differences in standard deviation values for both real and generated data for each channel and subsequently averaging these differences over all channels.

*Between-session variability* is reported as a baseline for the domain-invariant and domain-specific metrics. This is achieved by measuring the variability of two sessions of the same subject and the same class. Assuming that the model performs better than the between-session variability, and assuming that variability between subjects is larger than the between-session variability within the same subject and the same class, this test checks if the model can generate data that is specific to only one combination of subject, session, and class.

## RESULTS

One diffusion model is trained and conditioned on all combinations of subjects, sessions, and classes simultaneously. Ideally, this training approach should allow the model to generate data that is specific to a subject, session, and class.

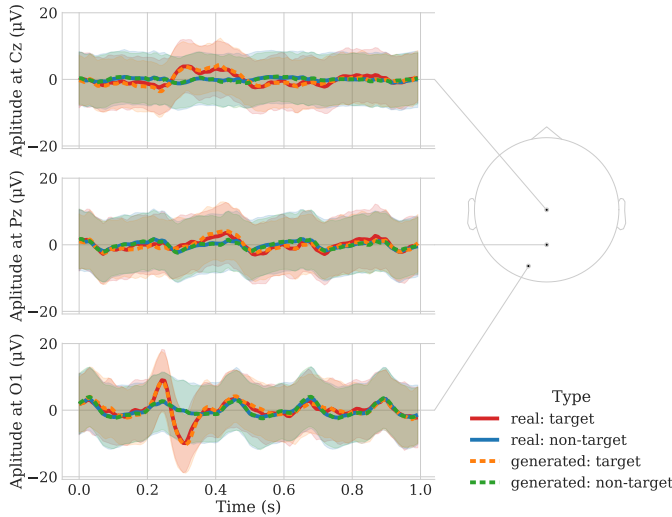
The model's ability to generate data for all combinations is assessed every 100k training steps using the metrics introduced in Methods - Similarity Metrics. The results are visualized in Figure 2. The model of the training step that achieves the highest ABA is further evaluated using the other metrics. These outcomes are provided in Table 1. Additionally, high-resolution plots were created for subject 52 in session 1, as this combination resulted in the worst generated data according to the ABA metric. The plots of this combination are shown in Figure 1.

*Trends in training and model performance:* In general, the model outperforms the between-session variability in the domain-invariant (SWD), and domain-specific (PAD, PLD, and SD-MD) metrics. Using the ABA metric, it performs similarly to the within-session baseline. There seems to be a common trend in the scores obtained by the model. Namely, all metrics show a relatively large increase in performance between 100k and 200k training steps and are relatively stable afterward (see Figure 2). The ABA and PAD do show a slight decrease in performance between the 800k and 900k training steps, but this decrease is not large enough to conclude that the overall performance is decreasing.

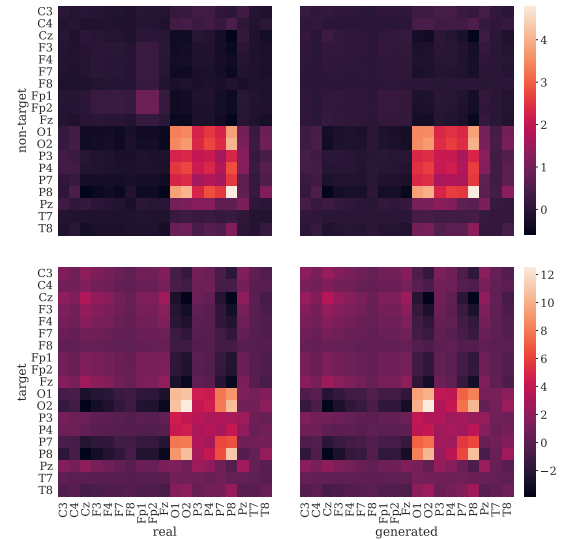
*Classifier performance:* The ABA is 0.818 for the within-session real-data baseline. For generated samples, the highest ABA is 0.817, which is achieved by samples from the model at 600k training steps.

According to ABA, the best subject/session combination is subject 51 in session 2, for which the generated data outperforms the within-session baseline by 0.045. Conversely, the worst combination is subject 52 in session 1, for which the classification score using generated data de-

<sup>1</sup>EEGNet checkpoint: <https://huggingface.co/guido151/EEGNetv4>



(a) Real and generated average temporal **ERP** responses of target and non-target for three selected EEG channels. The error bands indicate the standard deviations of the data.



(b) **Covariance matrices** of averaged real (left) and generated (right) responses for both non-target (top) and target (bottom) responses.

Figure 1: Figures (a) and (b) provide two different comparisons between real and generated data. The figures are based on the averages of EEG data (522 target and 2875 non-target examples), which are sampled for the combination of subject and session combination that resulted in the worst ABA metric (subject 52, session 1). The model with the highest ABA metric over all subjects and sessions (i.e. 600k training steps) was used to generate the samples.

Table 1: Scores of every metric measured on data sampled from the best checkpoint according to the ABA metric compared to a baseline, provided that there is a baseline. The baseline for the ABA is within-subject and within-session, while the baselines for the SWD, PAD, PLD, and SD-MD are computed using the variance between the two sessions of the same subject

	Generated			Baseline		
	target	non-tgt.	both	target	non-tgt.	both
FID ↓	-	-	$93e^{-4}$	-	-	-
ABA ↑	-	-	0.817	-	-	0.818
PAD ↓	0.48	-	-	0.83	-	-
PLD ↓	0.016	-	-	0.042	-	-
SD-MD ↓	3.44	1.33	2.39	7.38	2.94	5.16
SWD ↓	1.20	0.83	1.02	1.69	1.22	1.46

creases by 0.043 compared to the within-session baseline. Thus, even in the worst case, the generated data has LDA features that are very similar to the real data.

**Similar results on SWD and SD-MD:** The domain-invariant SWD of the generated data is better than its between-session baseline. Interestingly, the domain-specific SD-MD shows a strikingly similar pattern. Both with target and non-target data, the offset between the generated data and their respective baseline are almost identical. Thus, it seems that the SWD captures some information about the standard deviation, as this is explicitly measured by the SD-MD. The standard deviation of three channels for the worst subject, according to ABA, can be seen in Figure 1.

**Peak amplitude and latency:** The PLD and PAD of the generated data are also much lower compared to the between-session real-data baseline. The average PAD at the best model, according to ABA, is 0.48  $\mu\text{V}$ . Using

the same model, the maximum PAD is 2.19  $\mu\text{V}$ . This can likely be improved upon by optimizing the SNR of the corrector, either using a hyperparameter search or by fine-tuning using a subject- and session-specific approach. Nonetheless, the fact that the between-session PAD is much larger does indicate that the model can generate amplitudes that are at least in line with a particular session and subject. Additionally, the PLD values suggest that the peak position of the generated data is extremely close to that of the real data, with a difference of 0.016 ms on the same model. This is less than half of the between-session difference. However, there are large inter-subject differences in the PLD of the generated data as can be seen in Figure 3. These can be attributed to the sensitivity of the metric to slight deviations in the generated samples when there are multiple peaks in the selected channel. For example, subject 4 has a high PLD in session 1, however, this does not necessarily mean that the peak of the generated samples is highly dissimilar to the peak in the real data. Instead, the amplitude delta between the highest and the second-highest peak of the real data is quite small, which means that the generated data can have the second-highest peak as the highest peak, thereby disproportionately influencing the PLD. This also influences the PLD of the real data baseline.

**FID performance:** The FID is mostly useful for future comparisons, as it requires other models that are trained on the same dataset. Our model achieves a FID of  $93e^{-4}$ . In the absence of other models, a few baselines are computed to establish a frame of reference. Firstly, the average FID over 20 times computing the FID on two random halves of the real data is  $6.90e^{-4}$ . Secondly, the FID

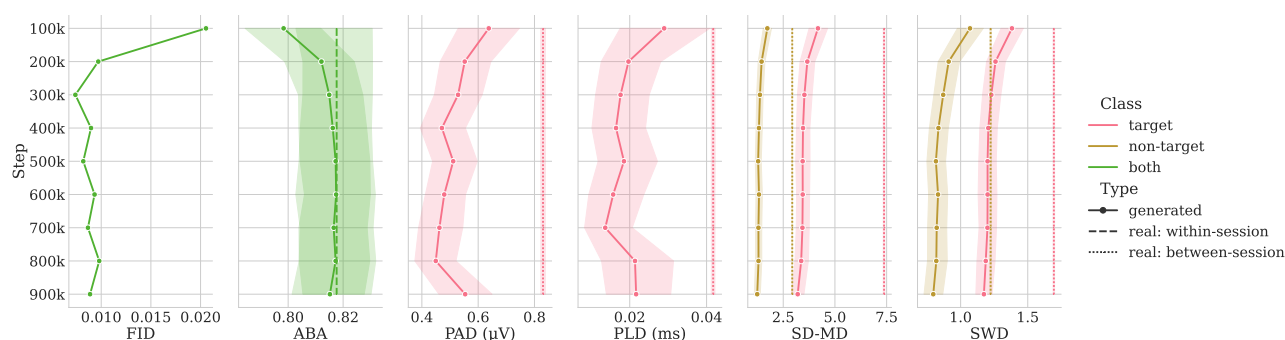


Figure 2: Scores of every metric measured on the generated data of the model at multiple training steps. The “real: within-session” baseline is computed by taking the ABA metric on the real data within a session of the same participant. The “real: between-session” baseline displays the variance between two sessions of the same participant on that particular metric. The band shows the 95 % confidence interval.

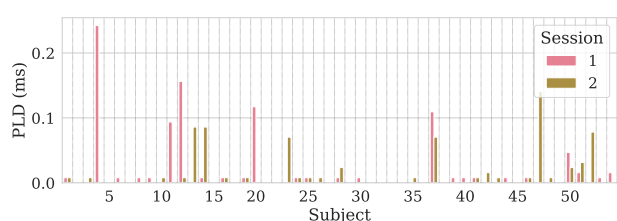


Figure 3: The **PLD** between the P300 peak latency in channel "O1" in the evoked real and generated data.

comparing sessions one and two is  $125e^{-4}$ . Lastly, the FID comparing the first 26 subjects to the last 25 subjects is  $1611e^{-4}$ . Thus, it seems that our model can achieve a FID that is slightly better than the FID between sessions one and two but is nowhere close to the FID computed on two random halves of the real data. However, it should be noted that EEGNet is trained on the real data for which subsequently the FID is calculated. Moreover, it has been established that using a conditioned model, instead of the unconditioned model can decrease the FID [14].

## DISCUSSION

In this work, we aimed to create a conditional diffusion model using classifier-free guidance, that does not lose specificity during sampling. The results indicate that the model can indeed create subject-, session-, and class-specific ERP data that is quite similar to the real data of the Lee 2019 ERP dataset.

*Amplitude, latency, and diversity are well-modeled:* The classifier performance on the real and generated data is highly similar, even for the worst subject. Given that the features on which the classifier is trained are based on amplitude and latency, it is no surprise that the PAD and PLD of the generated data are better than the between-session variability baseline. Furthermore, the diversity, as measured by the SD-MD, also seems to be modeled reasonably well, as it can outperform the between-session variability baseline.

*Limitation of the PLD:* The PLD should be interpreted with caution because it is unreliable when there are multiple peaks of similar height in the data. This can be ad-

ressed by either 1) only computing the PLD when there is only one prominent peak or 2) by computing the peaks of multiple subsets (i.e. 80 %) of the real data and only using the lowest PLD.

*Potential applications:* There are a variety of potential applications for the model presented here. For example, it can be used to alleviate the class imbalance by sampling from the target class. In addition, it can generate additional data to train a classifier, potentially slightly improving the robustness and accuracy of the model. Furthermore, the trained weights of the model can be used for transfer learning, which would allow fine-tuning on a different ERP dataset. Lastly, enlarging datasets with the proposed model may be specifically valuable for the benchmarking of novel algorithms. The diffusion model, however, can not be expected to deliver samples from outside the distribution of the training data.

*Dataset limitations and solutions:* All in all, the ability of the generated data to achieve comparable performance on most metrics to real data, and the high visual similarity of the covariance matrices and ERPs, are promising results for conditional diffusion models that generate EEG data directly. Nonetheless, it should be noted that the model was trained on a rather large dataset, so it remains to be seen how well the results translate to training on smaller datasets. However, the ability to generate data from specific conditions during sampling, while being trained on a full dataset, might make conditioned models more data-efficient compared to unconditioned models.

## CONCLUSION

In this work, we introduce the first diffusion model that is conditioned on subject, session, and classes using classifier-free guidance, that can generate high-quality EEG data for each condition. This enables training on complete datasets, without losing specificity during sampling. Additionally, we introduce multiple domain-specific metrics that can assist in model evaluation and fine-tuning. This conditional diffusion model can now be used to generate high-quality data for all subjects, sessions, and classes present in the Lee ERP dataset.

## ACKNOWLEDGEMENTS

We thank Dr. Luca Ambrogioni for his insights into diffusion models. This work in part is supported by the Donders Center for Cognition (DCC) and by the project Dutch Brain Interface Initiative (DBI2), project number 024.005.022 of the research programme Gravitation which is (partly) financed by the Dutch Research Council (NWO). OnePlanet Research Center acknowledges the support of the Province of Gelderland.

## REFERENCES

- [1] Aldahr RS, Alanazi M, Ilyas M. Evolving deep learning models for epilepsy diagnosis in data scarcity context: A survey. In: 45th International Conference on Telecommunications and Signal Processing. 2022, 66–73.
- [2] Lee MH *et al.* EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience*. 2019;8(5):giz002.
- [3] Choy T, Baker E, Stavropoulos K. Systemic Racism in EEG Research: Considerations and Potential Solutions. *Affective Science*. 2021;3(1):14–20.
- [4] Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. In: *Advances in neural information processing systems*. Curran Associates, Inc., 2021, 8780–8794.
- [5] Huang R *et al.* Make-an-audio: Text-to-audio generation with prompt-enhanced diffusion models. In: *International Conference on Machine Learning*. 2023, 13916–13932.
- [6] Tosato G, Dalbagno CM, Fumagalli F. EEG synthetic data generation using probabilistic diffusion models. *arXiv preprint arXiv:2303.06068*. 2023.
- [7] Ju C, Kobler RJ, Guan C. Score-Based Data Generation for EEG Spatial Covariance Matrices: Towards Boosting BCI Performance. In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Jul. 24, 2023, 1–7.
- [8] Aristimunha B *et al.* Synthetic sleep EEG signal generation using latent diffusion models. In: *Deep Generative Models for Health Workshop NeurIPS 2023*. 2023.
- [9] Sharma G, Dhall A, Subramanian R. MEDiC: Mitigating EEG data scarcity via class-conditioned diffusion model. In: *Deep Generative Models for Health Workshop NeurIPS 2023*. 2023.
- [10] Zhou T, Chen X, Shen Y, Nieuwoudt M, Pun CM, Wang S. Generative AI enables EEG data augmentation for Alzheimer’s disease detection via diffusion model. In: *2023 IEEE International Symposium on Product Compliance Engineering-Asia (ISPCE-ASIA)*. 2023, 1–6.
- [11] Torma S, Szegletes L. EEGWave: A denoising diffusion probabilistic approach for EEG signal generation. *EasyChair Preprint*. 2023;(10275).
- [12] Vetter J, Macke JH, Gao R. Generating realistic neurophysiological time series with denoising diffusion probabilistic models. *bioRxiv*. 2023:2023–08.
- [13] Wang Y *et al.* DiffMDD: A Diffusion-Based Deep Learning Framework for MDD Diagnosis Using EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2024;32:728–738.
- [14] Ho J, Salimans T. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*. 2022.
- [15] Shu K, Zhao Y, Wu L, Liu A, Qian R, Chen X. Data augmentation for seizure prediction with generative diffusion model. *arXiv preprint arXiv:2306.08256*. 2023.
- [16] Habashi AG, Azab AM, Eldawlatly S, Aly GM. Generative adversarial networks in EEG analysis: An overview. *Journal of NeuroEngineering and Rehabilitation*. 2023;20(1):40.
- [17] Neifar N, Mdhaaffar A, Ben-Hamadou A, Jmaiel M. Deep generative models for physiological signals: A systematic literature review. *arXiv preprint arXiv:2307.06162*. 2023.
- [18] Townsend G *et al.* A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*. 2010;121(7):1109–1120.
- [19] Aristimunha B *et al.* *Mother of all BCI benchmarks*. Version 1.0.0. 2023. [Online]. Available: <https://github.com/NeuroTechX/moabb>.
- [20] Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*. 2020.
- [21] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*. 2020;33:6840–6851.
- [22] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components — A tutorial. *NeuroImage*. 2011;56(2):814–825.
- [23] Rabin J, Peyré G, Delon J, Bernot M. Wasserstein Barycenter and Its Application to Texture Mixing. In: *Scale Space and Variational Methods in Computer Vision*. Springer Berlin Heidelberg: Berlin, Heidelberg, 2012, 435–446.
- [24] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces. *Journal of Neural Engineering*. 2018;15(5):056013.
- [25] Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*. 2017;30.

# EEG SINGLE-TRIAL DECODING OF VISUAL ART PREFERENCE

M. Welter<sup>1</sup>, J. Casal Martínez<sup>2</sup>, E. Redmond<sup>3</sup>, J. Baum<sup>1</sup>, T. Ward<sup>3</sup>, F. Lotte<sup>1</sup>

<sup>1</sup>Inria Centre at the University of Bordeaux / LaBRI, France

<sup>2</sup>Valencia Polytechnic University, Spain

<sup>3</sup>Dublin City University, Ireland

E-mail: marc.welter@inria.fr

## ABSTRACT:

Brain-Computer-Interfaces (BCIs) able to decode aesthetic preference could improve user experience in digital spaces by personalizing aesthetic stimuli selection without requiring explicit user feedback that might interrupt aesthetic experience. However, neuroscientific understanding of aesthetic experience remains lacking, while the tried and tested BCI classification algorithms have not yet been applied to decode aesthetic preferences from EEG signals. We thus conducted an experiment in which participants were exposed to visual artworks in a virtual museum and requested to grade their preferences for each of them, all this while their EEG was being measured. Previous neuroaesthetic research suggested that oscillatory modulations in different neural frequency bands could be informative of aesthetic preference. Therefore, we tested a time-frequency feature classification method widely used in BCIs, i.e. Filterbank Common Spatial Patterns feature extraction together with shrinkage Linear Discriminant Analysis, in a 2-class aesthetic preference classification problem. We report promising aesthetic preference decoding accuracies significantly and substantially above chance level.

## INTRODUCTION

Passive Brain-Computer-Interfaces (BCIs) allow implicit and real-time monitoring of cognitive, affective and conative mental and embodied states in human users [1, 2]. Aesthetic experiences are complex experiences that are composed of such states, notably attentional, affective and reward-related states [3]. Humans in the 21st century are exposed to an unprecedented amount of aesthetic stimuli, especially in digital spaces such as social media. In such spaces, presentation of aesthetic stimuli, e.g. visual art or music, relies on recommendation systems that require explicit user feedback. However, giving explicit feedback requires cognitive effort that might interrupt aesthetic experience.

Passive aesthetic preference decoding BCIs, on the other hand, could allow personalization of art presentation in digital spaces without interruption, and thus, improve user experience [4]. Furthermore, aesthetic preference decoding BCIs could be applicable in other domains such as neuro-marketing in order to improve personalized ad-

vertising [5], and it might even help improve positive effects of art exposure on health and well-being [6].

However, to our knowledge, only two studies have investigated single trial aesthetic preference decoding from EEG. One using Deep Learning classifiers [7] and one using Temporal Decision Trees [8]. Neither of them reported any artefact removal procedure which renders the interpretation of their results difficult [9].

Thus, there has been a lack of Electroencephalography (EEG) single trial aesthetic preference decoding studies with validated and effective EEG classification methods. In order to alleviate this lack, this article aims to contribute towards the development of aesthetic preference decoding BCIs by applying validated BCI methods on EEG data recorded in a virtual art museum environment. The different components of aesthetic experience have been shown to be related to oscillatory brain modulations in various frequency bands [9]. Therefore, we used Filter Bank Common Spatial Patterns (FBCSP) [10] and shrinkage Linear Discriminant Analysis (sLDA) [11] in order to decode aesthetic preference from oscillatory EEG features.

In the following sections, we will describe the EEG and subjective aesthetic preference data collection, as well as the offline aesthetic preference decoding pipeline. Then, we respectively report and discuss the aesthetic preference decoding results. Finally, we offer prospects for future research and summarize our findings.

## MATERIALS AND METHODS

### *Participants:*

14 healthy adult participants (7 women, aged  $26.77 \pm 8.5$ ) completed the whole experiment. All participants grew up in Western cultures and, thus, were most familiar with Western art. None of them reported a history of neurological or psychiatric disorder. Participants gave informed consent prior to the study. The study was conducted in accordance with the ethical research guidelines in the Declaration of Helsinki and was approved by Inria's ethics committee, the COERLE (approval number: 2023-11). For one participant, subjective aesthetic appreciation ratings were not saved correctly, due to a bug in the recording. Thus, the following analyses are based on 13 participants.

### Experimental protocol:



Figure 1: A participant taking part in the aesthetic preference decoding BCI experiment

Each participant participated in one session of 2 hours. The session was organized as follows: (1) consent form signature and completion of several questionnaires (around 20 min), (2) installation of the EEG cap (around 20 min), (3) 3 test trials to familiarize themselves with the procedure, (4) 6 runs during which participants were presented artworks (around 60 min in total, including breaks between the runs), (5) completion of post-session questionnaires (around 5 min), and (6) uninstallation and debriefing (around 10 min).

During each run, participants had to perform 20 trials in a virtual museum environment displayed on a computer screen. This Virtual Exhibition Environment (VEE) has been developed through the Unity3D software, which contains the textured 3D models for the visualisation of the environment, artwork and lighting. The first version of the VEE (VEE1) is a desktop application that allows studies in the field of neuroscience, which can be configured through a settings screen (it allows selecting the library of images, modifying the lighting, texturing the walls, adding screens and/or fixation crosses, among others). The application can capture real-time Eye-Tracking data (in a format readable by the OGAMA analysis software), and send signals (using Lab Streaming Layer - LSL [12]) to the OpenVIBE software [13] to synchronise EEG data captured with the experiment's timeline. Figure 1 shows the experimental setup with one participant wearing an EEG cap while gazing at a painting in the VEE. Eye-Tracking was not used in this study.

At the start of each trial, a blank screen was displayed for a randomly sampled duration between 0.5 and 0.8s. After that, a fixation cross appeared for 5s, in order to measure a stable baseline, as well as to washout potential emotions evoked by previous artworks. Then, an artwork stimulus was presented for 10s. Finally, the subject rated their aesthetic appreciation (liking and interest) of the artwork on a scale from 0-100 using a slider. After each run, the participants could rest for a minute. The timeline of one trial of data collection with our experimental protocol is shown in Figure 2. Participants were instructed to gaze naturally at the art work. Instructions were written in advance so that all the participants started with the same

standardized information.

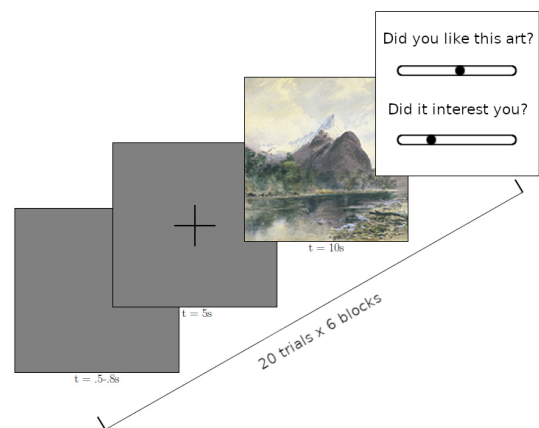


Figure 2: The data collection included 6 blocks of 20 visual art stimuli presentations and subsequent subjective aesthetic evaluations

### Questionnaires:

In addition to the participant's general demographic information, we asked the participants to complete the following questionnaires:

- AREA [14] translated into French, to measure the participant's responsiveness to aesthetic experiences in general. This questionnaire was filled out before the EEG measurements.
- NeXT-Q [15] to measure the participant's mental states before and after the experiment.

### Artworks:

The artworks displayed in the museum were 120 high quality digital reproductions of diverse artworks originating from almost all continents and ranging from pre-historic to contemporary time periods in the public domain. The artworks had a minimum resolution of 450x669, with a mean of 11392920 pixels. We chose artworks that were relatively unknown to a general audience, in order to avoid familiarity effects. The Artworks' brightness levels were normalized in order to avoid different brightness levels affecting the EEG [16].

### EEG Recordings & Signal Processing:

EEG data were sampled at 256 Hz using an ActiChamp amplifier (Brain Products, Gilching, Germany) with 31 active electrodes on a standard 10-20 montage. Electrodes were placed on the following scalp locations: Fp1, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P7, O1, Oz, O2, P4, P8, TP10, CP6, CP2, Cz, C4, T8, FT10, FC6, FC2, F4, F8, Fp2, i.e., on a broad scalp covering. The signal was grounded at Fpz and the reference was placed on Fz during recording. During offline analysis, the data was re-referenced to common average reference. In order to decode aesthetic preference for visual art from the EEG signal, the following signal processing

pipeline was used and validated with a 10-fold shuffled an stratified cross-validation:

First we manually inspected the data to reject bad channels. Then, we applied a 1-100Hz 4th order Butterworth bandpass filter and a notch filter at 50 Hz to remove line noise. Afterwards, we created fixed epochs of 1s length and cleaned these epochs using a local Autoreject [17]. These cleaned epochs were then fed into an Extended Infomax Independent Component Analysis [18]. The resulting independent components were classified with ICLabel [19]. Then, components labeled as artefacts were excluded from the components used to reconstruct a clean signal from the raw EEG.

Then, trial epochs were extracted from 0.1-10s ( $t=0$ s being the start of the artwork display) during stimulus presentation and baseline corrected from -4s to -0.01s before stimulus appearance.

Balanced Like and Dislike classes (for subsequent 2-class classification of aesthetic experience) were determined by partitioning the epochs based on quantilization of subjective ratings inspired by Strijbosch et al. [20]. We chose to include the data from the 45% most liked and the 45% most disliked artworks which resulted in a margin of 10% medium liked artworks that were not included in further analyses. This partitioning procedure resulted in balanced classes with 55-58 epochs per class. Thus, we defined aesthetic preference decoding as a 2-class supervised classification problem with an estimated chance level ( $\alpha=0.01$ ) of 60.9% [21]. Classification accuracies above this threshold can be considered to perform significantly better than random chance.

We extracted discriminant features of the EEG signal for classification with FBCSP-sLDA. During the computation of the spatial filters, covariance matrices were estimated using Oracle Approximating Shrinkage Estimator [22]. A bank of 8 filters was used with 4th order Butterworth bandpass filters in the following frequency bands: 1-4Hz, 4-8Hz, 8-13Hz, 13-16Hz, 16-20Hz, 20-30Hz, 30-50Hz, 50-70Hz.

Note that EEG analyses are often done with a cutoff at 30Hz in order to remove artefacts. However, higher frequencies above 30Hz have been shown to contain discriminative information about aesthetic experience [20]. Therefore, we decided to include higher frequency bands in our analyses. For each of these band-pass filter, 6 CSP spatial filters were learned from the train set and applied on the test set during each global cross-validation fold. After spatial filtering, log-transformed bandpower features were extracted. For each cross-validation fold, optimal features were selected using Recursive Feature Elimination [23] with a local 5-fold shuffled and stratified cross-validation (inner cross-validation) on the training set data of that fold (from the outer cross-validation). The selected features were then fed into a sLDA classifier. Aesthetic preference decoding performance was evaluated by computing the mean test accuracy over global cross-validation folds for each subject. In addition, we also ran a 10-fold shuffled an stratified cross-validated

CSP-sLDA classification for each band-pass filter individually, in order to investigate the discriminatory power of each frequency band.

## RESULTS

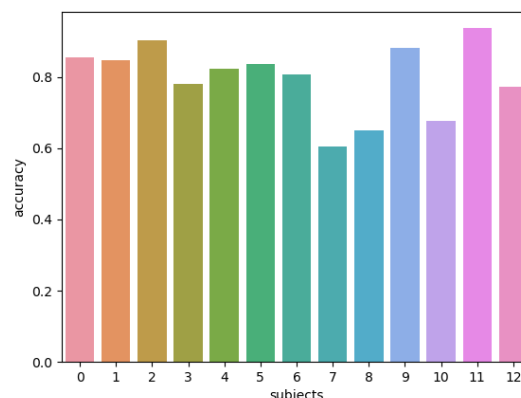


Figure 3: Mean aesthetic preference FBCSP-sLDA decoding accuracy for all subjects

We analysed the aesthetic preference decoding performance in term of mean classification accuracy, for this 2-class BCI. Mean FBCSP-sLDA classification performance with feature selection for all subjects is shown in Figure 3. The overall mean decoding performance was  $0.798 \pm 0.162$  %.

Classification on individual band-pass filters performed much worse and well below the estimated chance level on average with a mean accuracy of  $0.541 \pm 0.052$  and showed a large variability across subjects. Figure 4 shows the variability in mean classification performance across subject for different band-pass filters.

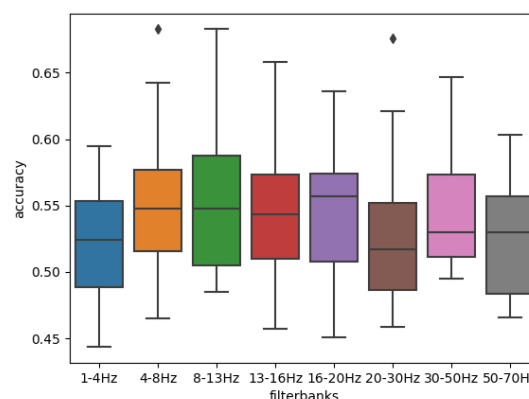


Figure 4: Mean aesthetic preference CSP-LDA decoding accuracy for individual frequency bands across subjects

## DISCUSSION

Overall, FBCSP-sLDA classification yielded surprisingly good results that were better than chance for most participants. Only for one subject, classification performance

was below the estimated chance level of 60.9%. The relatively high classification accuracy suggests that EEG single trial decoding of aesthetic preference is feasible. Furthermore, the findings that combining features from different frequency bands performs best support the idea that aesthetic experiences involve multiple components. Still, it remains possible that such relatively high classification performance relies, at least partly (and despite our artefact correction procedure), on movement artefacts in the EEG signal, as body movements can be informative of aesthetic experience [24]. In the following paragraph we will discuss possible correlates of those frequency bands that performed above chance for at least one subject: Oscillations in the alpha band were most informative which could be related to visual processing, but might also be generated by eye movements. Theta band features also seemed to contain some discriminatory information which might be due to the activation of the Default Mode Network during aesthetic experience [25]. Yet, theta band modulations can also be produced by eye blinks [26]. Low and mid beta frequencies exhibited the next best mean performance and have been implicated in emotional processing during art perception [27]. Finally, features in the gamma bands also performed relatively well for some subjects. Although, gamma bands are frequently excluded from EEG analyses, they have been implicated in aesthetically moving experiences [20]. However, both beta and gamma frequencies are also known to be commonly affected by muscular activity [28].

#### *future work:*

Aesthetic preference decoding with EEG BCIs could potentially be improved by using more advanced BCI classification algorithms, such as Deep Learning [29] or Riemannian Geometry-based classifiers [30]. Furthermore, the inclusion of features from other physiological modalities such as electrodermal activity, heart rate or eye movement might increase classification performance [31].

Although we report good performances for offline single trial aesthetic preference decoding with passive BCI, bridging the gap towards online classification remains challenging. A major limitation remains the development of the calibration phase, as we do not necessarily know a user's aesthetic preferences beforehand which complicates the selection of optimal art stimuli in the training data during calibration. Potentially, statistical image properties of the artworks that have been correlated with subjective ratings [32] could inform stimuli selection in order to train a generalizable aesthetic preference decoder.

#### CONCLUSION

We reported the first neuroaesthetic study using tested BCI algorithms for single-trial aesthetic preference decoding from EEG. Our results revealed better than chance classification accuracies for most subjects in discriminat-

ing preferred vs non-preferred artwork, using a FBCSP-sLDA classification pipeline. Individual bands analyses suggested that the alpha, theta, high beta and gamma bands were the most informative.

Although further work is required to develop online aesthetic preference decoding BCIs, the promising classification results above chance level suggest that decoding of aesthetic experience is feasible with EEG-based BCIs. Future work should focus on improving the accuracies obtained as well as in better identifying the possible contributions of cortical EEG and possibly of muscle or eye artifacts to the obtained decoding accuracies.

#### REFERENCES

- [1] Appriou A, Lotte F. Tools for affective, cognitive and conative states estimation from both eeg and physiological signals. In: Neuroergonomics Conference (NEC). 2021.
- [2] Zander TO, Kothe C. Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general. *Journal of neural engineering*. 2011;8(2):025005.
- [3] Schaeffer J. *L'expérience esthétique*. Gallimard: Paris (2015).
- [4] Welter M, Bouneau A, Lotte F, Ward T. In: Abstracts from the 8th Visual Science of Art Conference (VSAC) Amsterdam, The Netherlands, August 24th–27th, 2022. Brill: Leiden, The Netherlands, 2023.
- [5] Ferrer GG. A neuroaesthetic approach to the search of beauty from the consumer's perspective. *Advances in Marketing, Customer Relationship Management, and E-Services*. 2019.
- [6] Fancourt D, Finn S. What is the evidence on the role of the arts in improving health and well-being? a scoping review. *Nordic Journal of Arts, Culture and Health*. 2019;2:77–83.
- [7] Fraiwan M, Alafeef M, Almomani F. Gauging human visual interest using multiscale entropy analysis of eeg signals. *Journal of Ambient Intelligence and Humanized Computing*. 2023;12:2435–2447.
- [8] Mazzacane S *et al*. Towards an objective theory of subjective liking: A first step in understanding the sense of beauty. *PLOS ONE*. 2023;18(6).
- [9] Welter M, Lotte F. Ecological decoding of visual aesthetic preference with oscillatory electroencephalogram features—a mini-review. *Frontiers in Neuroergonomics*. 2024;5.
- [10] Ang KK, Chin ZY, Wang C, Guan C, Zhang H. Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *Frontiers in Neuroscience*. 2012;6.
- [11] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of erp components — a tutorial. *NeuroImage*. 2011;56(2):814–825.
- [12] Kothe C *et al*. The lab streaming layer for synchronized multimodal recording. *bioRxiv*. 2024.

- [13] Renard Y *et al.* OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain-Computer Interfaces in Real and Virtual Environments. Presence: Teleoperators and Virtual Environments. 2010;19(1):35–53.
- [14] Schlotz W, Wallot S, Omigie D, Masucci M, Hoelzmann S, Vessel E. The aesthetic responsiveness assessment (area): A screening tool to assess individual differences in responsiveness to art in english and german. Psychology of Aesthetics, Creativity, and the Arts. 2020;15.
- [15] Bismuth J, Vialatte F, Lefaucheur JP. Relieving peripheral neuropathic pain by increasing the power-ratio of low- $\beta$  over high- $\beta$  activities in the central cortical region with eeg-based neurofeedback: Study protocol for a controlled pilot trial (smrpain study). Neurophysiologie Clinique. 2020;50(1):5–20.
- [16] Eroğlu K, Kayıkçıoğlu T, Osman O. Effect of brightness of visual stimuli on eeg signals. Behavioural brain research. 2020;382.
- [17] Jas M, Engemann D, Bekhti Y, Raimondo F, Gramfort A. Autoreject: Automated artifact rejection for meg and eeg data. NeuroImage. 2016;159.
- [18] Lee TW, Girolami M, Sejnowski T. Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. Neural Computation. 1999;11:417–441.
- [19] Pion-Tonachini L, Kreutz-Delgado K, Makeig S. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. NeuroImage. 2019;198.
- [20] Strijbosch W, Vessel EA, Welke D, Mita O, Gelissen J, Bastiaansen M. On the neuronal dynamics of aesthetic experience: Evidence from electroencephalographic oscillatory dynamics. Journal of Cognitive Neuroscience. 2022;34(3).
- [21] Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. Journal of Neuroscience Methods. 2015;250:126–136.
- [22] Chen Y, Wiesel A, Eldar YC, Hero AO. Shrinkage algorithms for mmse covariance estimation. IEEE Transactions on Signal Processing. 2010;58(10):5016–5029.
- [23] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine Learning. 2002;46:389–422.
- [24] Kühnapfel C *et al.* How do we move in front of art? how does this relate to art experience? linking movement, eye tracking, emotion, and evaluations in a gallery-like setting. Empirical Studies of the Arts,. 2023;42(1):86–146.
- [25] Vessel EA, Starr GG, Rubin N. The brain on art: Intense aesthetic experience activates the default mode network. Frontiers in Neuroscience. 2012;7.
- [26] Roy RN, Charbonnier S, Bonnet S. Eye blink characterization from frontal eeg electrodes using source separation and pattern recognition algorithms. Biomedical Signal Processing and Control. 2014;14:256–264.
- [27] Herrera-Arcos G *et al.* Modulation of neural activity during guided viewing of visual art. Frontiers in Human Neuroscience. 2017;11.
- [28] Pope K *et al.* Managing electromyogram contamination in scalp recordings: An approach identifying reliable beta and gamma eeg features of psychoses or other disorders. Brain and Behavior. 2022;12.
- [29] Lawhern V, Solon A, Waytowich N, Gordon S, Hung C, Lance B. Eegnet: A compact convolutional network for eeg-based brain-computer interfaces. Journal of Neural Engineering. 2016;15.
- [30] Congedo M, Barachant A, Bhatia R. Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. Brain-Computer Interfaces. 2017;4:1–20.
- [31] Hogervorst M, Brouwer A, Erp J. Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload. Frontiers in Neuroscience. 2014;8(322).
- [32] Geller HA, Bartho R, Thömmes K, Redies C. Statistical image properties predict aesthetic ratings in abstract paintings created by neural style transfer. Frontiers in Neuroscience. 2022;16.

# EXPLORING EOG MARKERS OF FATIGUE DURING MOTOR IMAGERY BCI USE

Pauline Dreyer<sup>1, 2</sup>, Aline Roc<sup>1</sup>, David Trocellier<sup>1, 2</sup>, Marc Welter<sup>1, 2</sup>, Raphaëlle N. Roy<sup>3</sup>, Fabien Lotte<sup>1, 2</sup>

<sup>1</sup>Inria Center at the University of Bordeaux, Talence, 33405, France

<sup>2</sup>LaBRI (Univ. Bordeaux / CNRS/ Bordeaux INP) France

<sup>3</sup>Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, France

pauline.dreyer@inria.fr

## ABSTRACT

Brain-Computer Interface (BCI) performance suffer from various variability sources, including intra-subject factors such as mental fatigue. While frequently measured using subjective reports, mental fatigue can also be assessed via blink parameters extracted from electro-oculography signals. To our knowledge, no study has yet evaluated blink parameters during motor imagery (MI) BCI use to assess the potential development of mental fatigue. In this study, the blinks of 23 MI-BCI participants were analyzed concurrently with subjective reports and BCI performance. Our results showed that blink parameters were correlated with neither MI-BCI performance nor subjective reports. However, they revealed a positive correlation between time-on-task and both blinks number and mean duration. Similarly, subjective fatigue was correlated with time-on-task. This suggests that blinks parameters may be useful for BCI user monitoring, although their relationship with BCI performance and fatigue needs further studies. Altogether, this study paves the way towards a better understanding of mental fatigue during BCI use, and in finding solutions to mitigate it.

## INTRODUCTION

Intra-user Brain-Computer Interface (BCI) performance is known to fluctuate due to the interaction of several potential sources of variability including context, time and day, as well as user engagement and fatigue [1–6]. Mental fatigue<sup>1</sup> has been long known to impact human performance and engagement in general [7], and has started to be studied regarding both active and passive BCI performance in the last decade [8, 9]. The measures that were used to evaluate user fatigue in those previous studies were mostly focused on electroencephalography (EEG) metrics, as well as subjective (i.e. questionnaires) and behavioral metrics.

Yet, to the best of our knowledge, the ocular behavior metrics that can be extracted from the electro-oculography (EOG) signal – and which are widely used

for mental fatigue and vigilance characterization [10, 11] – have never been used to study fatigue during active BCI operation. Such metrics include blink number, blink duration, opening and closing velocity, as well as opening and closing duration [12–14].

Given the lacks identified in the literature, in order to better understand the development of user fatigue during the execution of active BCI tasks, and more precisely during Motor Imagery (MI) BCI tasks, the present study investigates the evolution of blinks and their parameters across MI-BCI runs based on EOG signal analysis. To do so, a standard motor imagery BCI protocol was used, in which participants also had to answer questionnaires to gather subjective reports of fatigue. It was expected that (i) blink number and duration would increase with runs. It was also expected that (ii) blink number and parameters would correlate with BCI performance and subjective reports. The remainder of this paper presents the data used, the analysis performed, the results obtained and their interpretation.

## MATERIALS AND METHODS

### *Participants:*

Twenty-three (23) participants completed the BCI experiment (10 women/13 men), aged  $28.4 \pm 6.2$  y.o. Recruitment was limited to volunteer participants aged 18–60 years old, with no history of neurological or psychiatric disorders, normal (or corrected) vision and naive MI-BCI users, i.e. using a MI-BCI system for the first time. Before participating in each study, participants gave informed consent. The study has been approved and reviewed by Inria's ethics committee, the COERLE (Approval number: 2020-32).

### *Protocol:*

The experiment consisted of 3 experimental MI-BCI sessions (completed on 3 different days) per participant. A brief pre-session questionnaire was assessed at the beginning of each session to measure the participant alertness. However we did not use this questionnaire for

<sup>1</sup>a.k.a., reduced alertness, which arises from growing time-on-task.

this study. Participants were then asked to perform two short working memory tasks to serve as EEG training data for a future offline passive BCI study unrelated to the present paper. The MI-BCI training and use then started.

The training protocol used for this experiment follows the standard left and right hand MI-BCI protocol from TU Graz [15], which comprises two phases: (1) motor imagery practice to collect data for calibrating machine learning algorithms (runs 1–4), and (2) closed-loop user-training with real-time classifier feedback (runs 5–12). After each of the 12 runs, participants were instructed to rate their mental state using a 1-10 scale of selected items from the NASA Task Load Index (NASA-TLX) questionnaire [16] (mental demand, effort and frustration levels), as well as their subjective mental fatigue. Note that other items of the NASA-TLX were not used in order to keep the number of questions to a minimum. The general experiment workflow is illustrated in Figure 1.

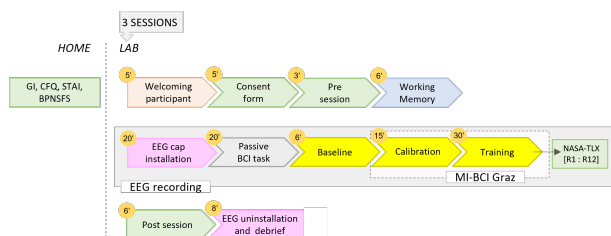


Figure 1: Experimental protocol. The orange circles correspond to the minutes that each part takes. In green rectangles are the questionnaires, in pastel orange rectangles are the experimenter's instructions, in pink rectangles are the technical processes with the cap, in yellow are the EEG recordings.

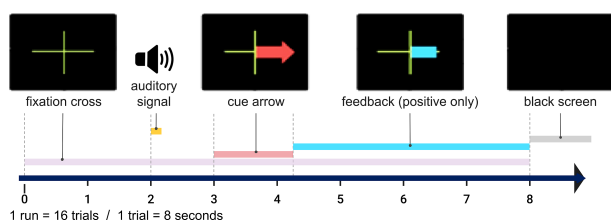


Figure 2: Organisation and timing of a single MI-BCI trial.

During each run, participants performed 16 trials (8 per MI-task, presented in a random order) each trial lasting 8s. First a green cross appeared ( $t = 0s$ ) on the screen, then an acoustic signal ( $t = 2s$ ) announced the appearance of a red arrow ( $t = 3s$ ). The arrow pointed towards the task to be performed. (e.g., towards the left for left hand MI) and remained displayed for 1.25s. From  $t = 4.25s$ , the visual feedback was continuously provided (a blue bar varying in length according to the classifier output). The feedback lasted for 3.75s and was updated at 16Hz, using a 1s sliding window. Positive feedback only was displayed.

Then the screen turned black again after 8 seconds until

the next trial begin, starting randomly between 1.5 to 3.5 seconds later. (see Figure 2). At the end of the session, participants filled-in the post-experiment NeXT-Q questionnaire [17] (around 5 min). Then, the cap was removed and participants were debriefed (around 8 min).

#### EEG and EOG data acquisition:

Participants sat comfortably in a chair in front of a computer screen. EEG data was acquired using 42 active scalp electrodes (i.e., F3, Fz, F4, FC5, FC3, FC1, FCz, FC2, FC4, FC6, C3, C1, Cz, C2, C4, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P3, Pz, P4, AF3, AFz, AF4, FC7, FC8, C5, C6, TP9, TP7, TP8, TP10, PO7, POz, O1, Oz, O2, PO8, 10-20 system), referenced to the left earlobe, the ground electrode being placed in FPz position. The electrooculography (EOG) signals of one eye was recorded using three active electrodes. Two of them were located below and above the eye (EOG1 and EOG3) and one was located on the side of the left eye (EOG2). We also recorded the electromyographic (EMG) signals of both hands using two active electrodes located 2.5 cm below the skinfold on each wrist. Physiological signals were measured using two g.USBamp amplifiers (g.tec, Austria), sampled at 512 Hz, and processed online using the open-source BCI platform OpenViBE [18]. The recording room was dimly-lit. The raw signals were recorded without any hardware filters.

#### Online BCI Performances:

The metric used for quantifying BCI performances is the online Trial-wise Accuracy (TAcc), i.e. the default performance metric provided online in the MI-BCI scenarios of OpenViBE. TAcc measures the accuracy of trial classifications, with each trial categorized as either correctly or incorrectly classified. The classification outcome for each trial is computed by summing the signed classifier outputs over all epochs during the trial feedback period (from  $t = 4.25s$  to  $t = 8s$  of the trial). A trial is considered correctly classified if the sum sign matches the required trial label (negative for left hand MI and positive for right hand MI), otherwise, it is considered incorrect. For this experiment, online classification was performed using Common Spatial Pattern (CSP) (3 filter pairs) band power features in 8-30 Hz and a Linear Discriminant Analysis (LDA) classifier. TAcc for each run was calculated as the percentage of trials accurately classified using this methodology. Notably, this metric utilizes LDA outputs instead of discrete classification outputs for each epoch. Therefore, TAcc also reflects the length of the feedback bar participants observed, as it is proportional to the classifier output. Participants were instructed to train to achieve not only correct classifications but also maximize the length of this feedback bar. Thus, TAcc considers both aspects, providing a comprehensive assessment of BCI performance.

#### EOG signal analysis:

EOG signals were analysed with MNE Python [19],

a popular EEG and EOG data analysis toolbox that provides extensive event detection and feature extraction capabilities, and with NeuroKit2 a Python toolbox for neurophysiological signal processing with advanced artefact detection and removal [20].

The processing of EOG signals, to detect blinks and extract their parameters, was the following, for each run:

- (1) **Bipolar channel:** Creation of a bipolar EOG channel, EOG1-EOG3, focusing on vertical EOG signals which should capture blinks to enhance the blink peak detection with MNE.

- (2) **Cleaning:** Cleaning the bipolar EOG channel using NEUROKIT `eog_clean()` function to prepare for eye blink detection.

- (3) **Event detection:** Detecting EOG events using MNE: with the `preprocessing.find_eog_events()` function, to detect EOG events in the cleaned bipolar EOG channel. While this function typically enables precise identification of eye movements and blinks, its performance may be compromised by the presence of high amplitude artifacts, especially when the signal contains significant noise or bifurcated artifacts. Indeed, such function uses a threshold on EOG amplitude to detect blinks, this threshold being estimated according to the EOG minimum and maximum values. To address this challenge, we implemented a preprocessing step aimed at enhancing the robustness of EOG event detection.

- (4) **Winsorization:** Specifically, we applied winsorization to the raw EOG signal, exclusively on runs that exhibited significant artifacts. Winsorization is a technique that limits the influence of extreme values (here higher than the 95 percentile of the EOG signal) by replacing them with less extreme values (here the 95 percentile), thereby reducing the impact of outliers on subsequent analyses. Doing so, we were able to accurately identify and quantify the overall number blink events within each run, as confirmed with visual analysis.

- (5) **Additional epoching:** In addition to the overall number of blinks per run, we also epoched the runs to estimate the number of blinks occurring only during the motor imagery task. This thus provided us with the number of blinks during MI tasks per run.

- (6) **Blink features:** Then we extracted blink features, describing the characteristics of the blinks, shedding light on underlying physiological processes and potential mechanisms contributing to fatigue. More precisely We used Neurokit `eog_features()` function to extract EOG-related features from each blink of the cleaned bipolar EOG channel. These features included the blink duration and the blink Velocity (of eyes closing velocity denoted as pAVR and eyes opening velocity denoted as nAVR), i.e., the speed at which blinks occur.

*Statistical analyses to study the relationship between blinks parameters, time-on-task, MI-BCI performance and subjective mental states:*

Our goal was to study whether we could identify relationships between the parameters extracted from the blinks (number of blinks, duration and velocity) and MI-BCI performance, time-on-task (here measured as the run index, which increases with time-on-task) and/or subjective mental states, notably mental fatigue.

To assess these potential relationships, we used repeated measures correlation (rmcorr) analyses, to determine the common within-individual association for paired measures assessed multiple times for multiple individuals. [21]. Here the repeated measures per subject were the measures collected across all 8 feedback runs per session (repeated measures across runs). A total of 27 correlation analyses were performed: 4 between the number of overall blinks and the 4 mental states, 4 between MI-BCI performance and the 4 mental states, 4 between the number of blinks (overall and in MI tasks only) and time-on-task or MI-BCI performances, 6 between blinks duration and the 4 mental states, MI-BCI performances and time-on-task, and 8 between the mean blink velocity (pAVR and nAVR), and the 4 mental states and 1 between subjective fatigue and time-on-task. Thus, all p-values for these analyses are reported as corrected for multiple comparison with False Discovery Rate (FDR) across these 27 tests.

## RESULTS

### *Relationship between subjective mental fatigue and time-on-task:*

There was a positive correlation between time-on-task (as measured by the run index) and the subjective mental fatigue, showing that participants tend to report being increasingly more tired as time-on-task with the BCI increases (see Figure 3).

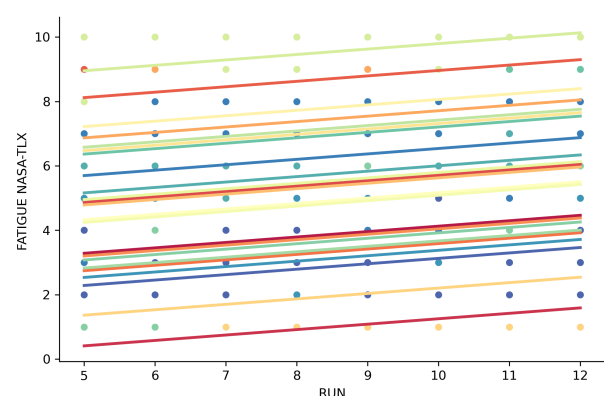


Figure 3: Repeated measure correlation between time-on-task (measured using run index) and the mental fatigue per participant,  $r=0.2$ ,  $p<0.00001$  (one colour per participant).

### *Relationship between BCI performances and mental states:*

No significant correlation was observed between mental fatigue, mental demand, and effort with online performance measures. However, a slight negative correlation

was detected ( $r = -0.138$ ,  $p < 0.005$ ) between the performance and the frustration per participant, indicating that participants with good performance tend to be less frustrated (see Figure 4).

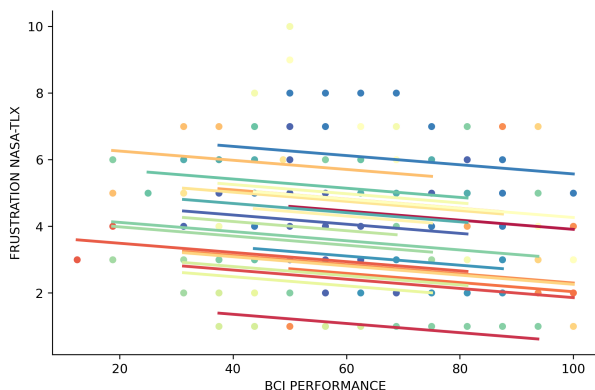


Figure 4: Negative repeated measure correlation per participants between BCI performance in % and the frustration measured by NASA-TLX,  $r = -0.1386$ ,  $p < 0.005$  (one colour per participant).

#### *Relationships between number of blinks and BCI performance or mental states:*

No significant correlations were found between the number of blinks observed during the runs and online performance. Similarly, there were no significant correlations detected between the number of blinks and the four different subjective mental states (i.e., mental fatigue, mental demand, effort, and frustration) assessed.

#### *Relationships between blinks parameters and BCI performance or mental states:*

First, it should be noted that Neurokit was not able to extract automatically the blink parameters from all blinks. Out of 552 runs, Neurokit was able to extract all the blink parameters for all the blinks of 380 of those runs. The subsequent results are thus based on 380 runs.

Our correlation analyses revealed that there were no significant correlations between blink velocity (mean pAVR and mean nAVR) per run and the various mental states. There was no significant correlation between the mean duration of blinks and the different mental state, except with frustration. There was a negative correlation between them ( $r = -0.15$ ,  $p < 0.01$ ), suggesting that blink duration decreases when frustration increases (see Figure 5). There was also no correlation between the mean duration of blinks and MI-BCI performance.

#### *Relationship between blinks parameters and time-on-task:*

A significant positive correlation was observed between the overall number of blinks per run and time-on-task (i.e., with the run index) ( $r = 0.178$ ,  $p < 0.0005$ ), suggesting that for each session, the number of blinks increased with time-on-task, i.e., with the number of MI-BCI runs completed (see Figure 6).

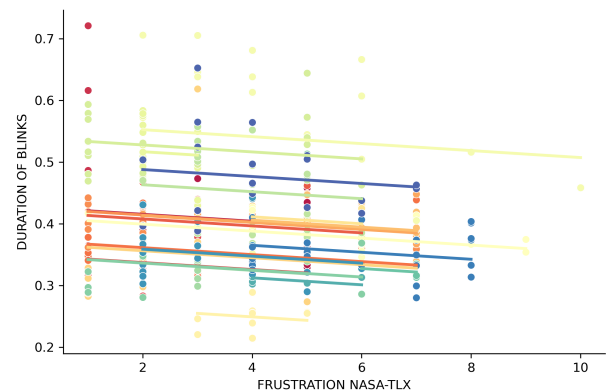


Figure 5: Repeated measure correlation between frustration and the mean duration of blinks,  $r = -0.15$ ,  $p < 0.01$  (one colour per participant).

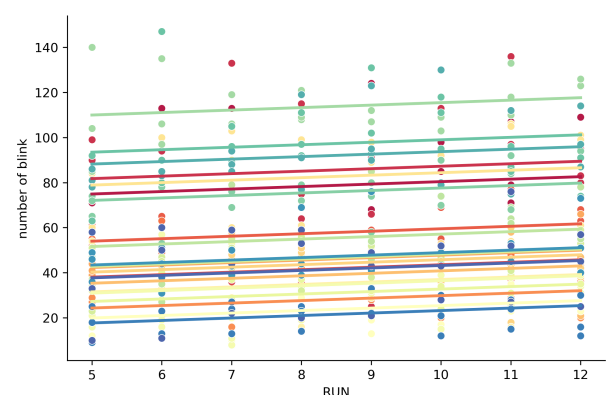


Figure 6: Repeated measure correlation between the overall number of blinks per run and time-on-task (as measured by the run index) per participant,  $r = 0.178$ ,  $p < 0.0005$  (one colour per participant).

Similarly there was a positive correlation between the number of blinks during the MI tasks per run and time-on-task (i.e., run index) ( $r = 0.176$ ,  $p < 0.0005$ ), suggesting as well that the number of blinks during MI tasks per run increased with time-on-task, i.e., with the number of MI-BCI runs completed, in a given session (see Figure 7).

There was a positive correlation between the mean duration of blinks in a run and time-on-task (i.e., with the run index) ( $r = 0.09$ ,  $p < 0.005$ ), suggesting that blinks are becoming increasingly longer with time-on-task, i.e., with the number of runs completed (see Figure 8).

## DISCUSSION

Overall, contrary to our initial hypotheses, we did not find any significant correlation between the number of blinks or any of the blinks parameters (duration, opening and closing velocity) and neither online MI-BCI performance nor with subjective mental states, including fatigue.

However we could find a significant positive correlation between the subjective fatigue and time-on-task, suggesting that a BCI session is increasingly more tiring as time-

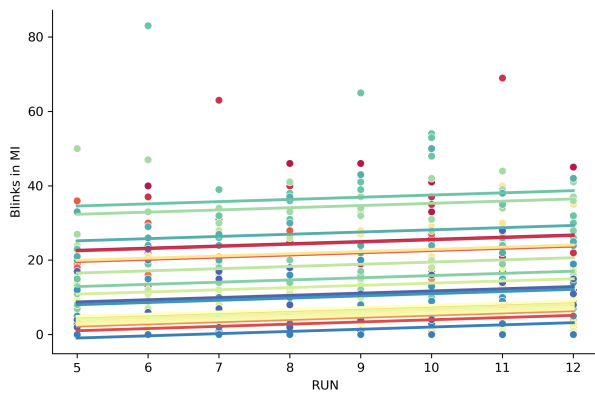


Figure 7: Repeated measure correlation between the number of blinks during the motor imagery task per run, and time-on-task (as measured by the run index) per participant,  $r=0.176$ ,  $p<0.0005$  (one colour per participant).

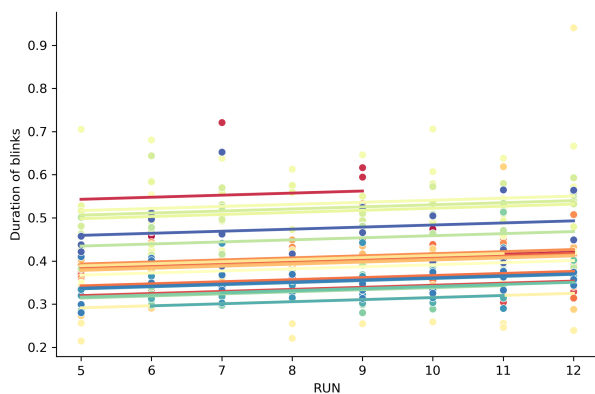


Figure 8: Repeated measure correlation between the duration of blinks during the motor imagery task and time-on-task (as measured by the run index) per participant,  $r=0.09$ ,  $p<0.005$  (one colour per participant).

on-task increases, which confirms subjective reports from our participants.

Interestingly enough, while the number of blinks or the mean blink duration were not correlated with subjective fatigue, they were both significantly correlated (although with a weak correlation) with time-on-task, as subjective fatigue did. In other words, both the number of blinks and the mean blink duration increased with time-on-task with a BCI. In the literature, both parameters have been related with mental fatigue [10, 11], although not only. However, the fact that they were not correlated with subjective fatigue nor with MI-BCI performance raises a number of questions. It may be that the number and mean duration of blinks are rather affected, in this context of MI-BCI, by visual fatigue (the current BCI protocol being based on visual cues and feedback) rather than by mental fatigue. Another possible interpretation could be that MI-BCI performance, blink numbers and duration and subjective fatigue have a more complex relationship, possibly non-linear, that is not captured by the linear correlation analyses we performed. Along these lines, it is interesting to note that subjective fatigue was also not linearly

correlated with MI-BCI performance, even though mental fatigue is known to affect EEG and thus possibly BCI performance [8]. Alternatively, maybe that subjective fatigue and more objective markers of fatigue such as those based on EOG as studied here significantly differ from each other. This will need to be studied in more details in the future.

An unexpected finding was that blink duration significantly decreased with increased subjective frustration during MI-BCI use. A possible interpretation could be that this frustration is most likely due to poor MI-BCI performances (since frustration and MI-BCI performance were also correlated), and that such poor performance motivates the (frustrated) users to focus more on the task, leading to shorter blinks.

## CONCLUSION

In this paper, our aim was to study whether blink parameters, such as their number, their duration or velocity, could be used to monitor fatigue during MI-BCI use, and to study whether they were related to subjective fatigue, MI-BCI performance and time-on-task during MI-BCI use. To do so, we analysed the data of 23 participants, who performed 3 sessions of 12 runs each (including 8 runs with real-time feedback) of MI-BCI training. We studied the (linear) relationships between these participants' blink parameters estimated from their EOG signals and their MI-BCI performance, subjective mental states (including fatigue, measured after each run) and time-on-task.

Altogether, our analyses did not reveal any significant correlation between these blinks parameters and neither MI-BCI performance nor subjective mental states. They did reveal a positive correlation between time-on-task and both the number of blinks and the mean blink duration. Similarly, subjective fatigue significantly correlated with time-on-task.

Overall, while these blink parameters did not prove as accurate as expected to monitor mental fatigue during MI-BCI use, they do reflect time-on-task and may thus still be useful to consider for user monitoring during MI-BCI use. Additionally, the results obtained also call for further studies to better understand the link between MI-BCI performance, subjective measures of fatigue and EOG-related blink parameters.

Future works could consider additional markers of fatigue, e.g., measures of saccade, which may be more reliable for cognitive state monitoring than blinks [14], or non EOG markers, e.g., cardiac markers or even directly EEG markers of fatigue [8]. In conclusion, this study paved the way towards acquiring a better understanding of mental fatigue in the context of MI-BCI use, and therefore in finding solutions to mitigate such fatigue to increase user engagement and performance.

## REFERENCES

- [1] Grosse-Wentrup M, Schölkopf B. A review of performance variations in SMR-based brain-computer interfaces (BCIs). *Brain-Computer Interface Research: A State-of-the-Art Summary*. 2013:39–51.
- [2] Dehais F *et al.* Monitoring pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions. *Sensors*. 2019;19(6):1324.
- [3] Podda J *et al.* Spatial constraints and cognitive fatigue affect motor imagery of walking in people with multiple sclerosis. *Scientific Reports*. 2020;10(1):21938.
- [4] Saha S, Baumert M. Intra-and inter-subject variability in EEG-based sensorimotor brain computer interface: A review. *Frontiers in computational neuroscience*. 2020;13:87.
- [5] Benaroch C *et al.* Long-term BCI training of a tetraplegic user: Adaptive riemannian classifiers and user training. *Frontiers in Human Neuroscience*. 2021;15:635653.
- [6] Roy RN *et al.* Retrospective on the first passive brain-computer interface competition on cross-session workload estimation. *Frontiers in Neuroergonomics*. 2022;3:838342.
- [7] Lal SK, Craig A. A critical review of the psychophysiology of driver fatigue. *Biological Psychology*. 2001;55(3):173–194.
- [8] Roy RN, Bonnet S, Charbonnier S, Campagne A. Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI. In: 2013 35th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013, 6607–6610.
- [9] Talukdar U, Hazarika SM, Gan JQ. Motor imagery and mental fatigue: Inter-relationship and EEG based estimation. *Journal of computational neuroscience*. 2019;46:55–76.
- [10] Roy RN, Charbonnier S, Bonnet S. Eye blink characterization from frontal EEG electrodes using source separation and pattern recognition algorithms. *Biomedical Signal Processing and Control*. 2014;14:256–264.
- [11] Kołodziej M *et al.* Fatigue detection caused by office work with the use of EOG signal. *IEEE Sensors Journal*. 2020;20(24):15213–15223.
- [12] Caffier PP, Erdmann U, Ullsperger P. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European journal of applied physiology*. 2003;89:319–325.
- [13] Schleicher R, Galley N, Briest S, Galley L. Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? *Ergonomics*. 2008;51(7):982–1010.
- [14] Bafna T, Hansen JP. Mental fatigue measurement using eye metrics: A systematic literature review. *Psychophysiology*. 2021;58(6):e13828.
- [15] Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*. 2001;89(7):1123–1134.
- [16] Cegarra J, Morgado N. Étude des propriétés de la version francophone du NASA-TLX. In: *EPIQUE 2009: 5ème Colloque de Psychologie Ergonomique*. 2009, 233–239.
- [17] Bismuth J, Vialatte F, Lefaucheur JP. Relieving peripheral neuropathic pain by increasing the power-ratio of low- $\beta$  over high- $\beta$  activities in the central cortical region with eeg-based neurofeedback: Study protocol for a controlled pilot trial (SMRPain study). *Neurophysiologie Clinique*. 2020;50(1):5–20.
- [18] Renard Y *et al.* OpenViBE: An open-source software platform to design, test and use brain-computer interfaces in real and virtual environments. *Teleoperators and Virtual Environments*. 2010;19(1):35–53.
- [19] Gramfort A *et al.* MEG and EEG data analysis with MNE-python. *Frontiers in neuroscience*. 2013;7:267.
- [20] Makowski D *et al.* NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*. 2021;53(4):1689–1696.
- [21] Bakdash JZ, Marusich LR. Repeated measures correlation. *Frontiers in psychology*. 2017;8:456.

## SHOULD ATTEMPTED MOVEMENTS REPLACE MOTOR IMAGERY IN BCI? THE ISSUE OF COMPATIBILITY WITH GAZE USE

S.L. Shishkin<sup>1</sup>, A.S. Yashin<sup>1</sup>, Y.G. Shevtsova<sup>1</sup>, A.N. Vasilyev<sup>1,2</sup>

<sup>1</sup> MEG Center, Moscow State University of Psychology and Education, Moscow, Russia

<sup>2</sup> Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow, Russia

E-mail: [sergshishkin@mail.ru](mailto:sergshishkin@mail.ru)

**ABSTRACT:** Attempted movements have recently become common in invasive studies as a way to send commands via BCIs and have been successfully employed in some studies of neurorehabilitation using noninvasive BCIs. Nevertheless, they are still far less common in noninvasive BCIs than motor imagery. We proposed a hypothesis that attempted movements can be more compatible with the interaction with the external world than imaginary movements and therefore may help to use BCIs more effectively. The hypothesis was tested in 15 healthy participants who were asked to make prosaccades, which represented an external task, and quasi-movements (movement attempts minimized down to complete extinction of related muscle activation), which were used as a model of attempted movements. Preliminary results of the study were mostly in line with the predictions, although more studies are required for more definite conclusions. The study also may be considered as a new demonstration of the potential of quasi-movements, a very little explored phenomenon, for BCI research.

### MOTOR IMAGERY IN BCI

Motor imagery based brain-computer interfaces (MI BCIs) [1] employ the sensitivity of the EEG sensorimotor rhythms (mu rhythm and sensorimotor beta rhythms) to the imagination of movements and are currently among the most popular noninvasive BCIs. Their accuracy is rather low but can be improved to some extent through training [2]. In addition to assistive technologies, application of MI BCI to neurorehabilitation, especially post-stroke rehabilitation, was addressed in many studies (for reviews, see [3, 4]).

*The case of hybrid eye-brain control* – Among other attempts to improve MI BCI efficiency are combining it with gaze-based control. Recently, the success of the Apple Vision Pro headset demonstrated that the combination of gaze-based control with hand gestures can be effectively used in AR/VR even by healthy individuals [5]. However, gaze-and-MI-BCI control so far was successful mostly when the use of gaze and EEG modalities were relatively independent [6, 7], while their tight integration proved difficult [8, 9].

### MOVEMENT ATTEMPTS IN BCI

Like motor imagery, movement attempts in paralyzed

individuals are accompanied by distinct desynchronization of their EEG sensorimotor rhythms, with spatiotemporal pattern similar to observed when a healthy person makes a real movement. In some early studies of BCI-based neurorehabilitation patients were asked to attempt to make movements, rather than to imagine them (e.g., [10]). This approach has again attracted certain attention recently, when several studies demonstrated better outcomes for attempted compared to imagined movement BCI (see [11, 12] for meta-analyses). Attempts to move were also successfully used by patients in a number of recent high-profile studies of invasive BCIs developed for assistive purposes movement [13, 14, 15, 16] and even in combination with gaze-based cursor control, which was implemented in the first clinical trial with the endovascular BCI [17].

### CONTROVERSY BETWEEN EXTERNAL ATTENTION AND IMAGERY

In active BCIs, either for assistive or rehabilitation purposes, feedback from a BCI plays an important role, informing a patient about the current course of action, helping them to correct control strategies and enabling effective training. However, focusing on imagery means that attentional resources are directed to a mental, internal task, and a BCI user needs to divide their attention between it and feedback that comes from the external world. This need to divide attention remains when not only visual but also haptic or auditory feedback is used, and even though a BCI is normally controlled via kinesthetic imagery (visual imagery is not effective for modulating sensorimotor rhythms). In any case, with the only exception for direct brain stimulation, a BCI is activated by merely mental actions and the feedback is provided via sensory stimulation. Moreover, even when the feedback is mostly haptic (e.g., in exoskeleton-assisted post-stroke rehabilitation), visual attention still may be strongly involved. The controversy seems especially severe when a MI BCI is combined with gaze-based control, where gaze should be intentionally controlled at the same time when motor imagery is executed.

Recent psychophysiological studies indicate that dividing attention between internal and external tasks and instructions for gaze during internal tasks may indeed hinder performance [18, 19] (see also references

in [19]). It was also found that an imagery task took longer time during instructed eye fixation than under free eye movement condition [20]; note that gaze control and external attention are strongly connected.

#### WHY ATTEMPTED MOVEMENTS MAY BE A SOLUTION?

As we noted above, growing evidence indicates that attempted movements may work better in BCI than motor imagery. In the view of the conflict between external and internal tasks, one possible reason could be that attempted movements are not a typical internal task. In an attempt to make a movement a healthy individual is intending an interaction with the external world. A paralyzed individual knows that the attempt will not lead to such interaction, but their intention and effort may not differ dramatically from a healthy person's intention and effort. The ability to imagine a movement may be much different, as it developed in the course of evolution as an ability to simulate reality, not to actually interact with the external world.

#### QUASI-MOVEMENTS AS A MODEL OF MOVEMENT ATTEMPTS FOR STUDIES IN HEALTHY PARTICIPANTS

Studying attempted movements in healthy participants is not a trivial task, since normally attempts lead to actual movements, and related sensory activation changes the EEG dramatically. Constraining a limb (e.g., [21]) changes the pattern of this afferent stream, but evidently cannot exclude sensory activation during attempts. Temporal artificial paralysis is an effective solution [22], but this approach cannot be widely used.

Fortunately, a way to teach healthy participants to make movement attempts without muscle activation exists. Such movement attempts, made by non-paralyzed individuals without actual movement and muscle activation, are called quasi-movements (QM) [23]. They appear when a person is asked to make smaller and smaller movements and to further weaken the attempts to a degree when the electromyogram (EMG) becomes indistinguishable from rest level. Importantly, the EEG activation pattern remains in QM similar to that in overt movements and in IM, and activation in QM is stronger than in IM [23]. QM are not an ideal model of attempted movements, because they model only weak movement attempts, and, even more importantly, require special attention to keep them weak. Nevertheless, due to the lack of good alternatives it might be still important to study this model.

Recently, we showed that QM provides stronger activation than IM independently of residual muscle activation [24]. Moreover, our participants mostly reported that their intention in QM was to make a movement rather than to imagine it [25], which confirms the assumption by [23] that they may serve as a model for attempted movements. However, to our knowledge, no study of possible differences between

QM (as well as other types of attempted movements) and IM from the point of view of internal vs. external orientation of cognitive resources has been undertaken so far.

#### INTERNAL OR EXTERNAL TASK? AN EXPERIMENTAL ASSESSMENT

No standard procedure was adopted so far for assessing whether a mental task is more internal or external. In [19] interference between internal, mental tasks (arithmetic and visuospatial) and an external task (prosaccades, i.e., saccades to a target, in the presence of a distractor) was assessed quantitatively. We decided to use their experimental design, with some modifications, to compare kinesthetic IM and QM in terms of their external or internal nature. More specifically, we asked our participants to make prosaccades at the same time intervals when they performed IM or IQ, to assess the degree of interference in each case (presumably related to the need to divide attention between the tasks).

*Study hypotheses* – We hypothesized that under QM condition, compared to IM condition, subjective difficulty will be lower, accuracy of eye movements will be higher. In addition, for the case if subjective difficulty indeed were lower or same in QM as in IM, we expected that EEG modulation would be more pronounced under QM than in IM. In other words, we expected that IM, as a clearly internal task, will interfere more with prosaccades (an external task) than QM, due to the more external nature of QM compared to IM.

#### METHODS

*Participants* – 15 healthy volunteers (8 female; age 18 to 38, median 23) participated in this study after signing an informed consent. Data from four of them were excluded from the analysis due to eye tracking issues or other technical issues.

*Apparatus and software* – Stimuli were presented at a 60 Hz 24" AOPEN 25XV2Q monitor with 1920x1080 resolution in front of a participant. Gaze data were acquired at 1000 Hz rate with EyeLink 1000 Plus eye tracker (SR Research, Canada). 64-channel EEG, one-channel electromyogram (EMG) from *m. abductor pollicis brevis* and a signal from a photo sensor on the screen (used to precisely synchronize with visual stimuli presentation) were recorded at 1000 Hz sampling rate with 0...300 Hz passband using the NVX136 DC EEG amplifier (Medical Computer Systems, Moscow, Russia). EMG was monitored online as a raw signal and after transforming with the Teager-Kaiser energy operator (to highlight deviations from baseline level). Stimuli presentation, data acquisition, synchronization, online processing and recording were done with *Resonance* platform [26] and additional modules written in Python.

*Experiment design* – Two sessions were run on different days. In the first session, participants were

introduced to the basic movement (right thumb abduction, as in [23], but made in triplets – like in [24], but self-paced) and trained to make IM and QM according to procedure by [23] with modifications described in [24]. The EEG was recorded under single-task QM and IM. The single saccade task (ST) and the dual tasks (ST combined with OM, IM and QM) were in the second session, with the order of IM and QM dual tasks randomized over the group (contrast between these two dual tasks was the main part of the experiment, while the other conditions provided various additional data). In both sessions, the EEG was also recorded under overt movements of the same type and under visual task, to obtain data for CSP spatial filter training; the visual task also helped to regain a baseline sensorimotor rhythm level (see [24], for details).

**Procedure** – All participants were naive to QM. Following [23], we did not reveal to them that they did not actually make movements (their right hand was covered with an opaque case). QM and IM quality was controlled using online EMG control (if EMG increase was observed, participants were asked to relax in IM or to further reduce QM) and using offline EEG analysis. Trial structure is presented in Fig. 1A for single tasks and in Fig. 1B for dual tasks. Targets and distractors for the ST were presented at the same distance from the fixation cross at random positions, but close to each other, following [19].

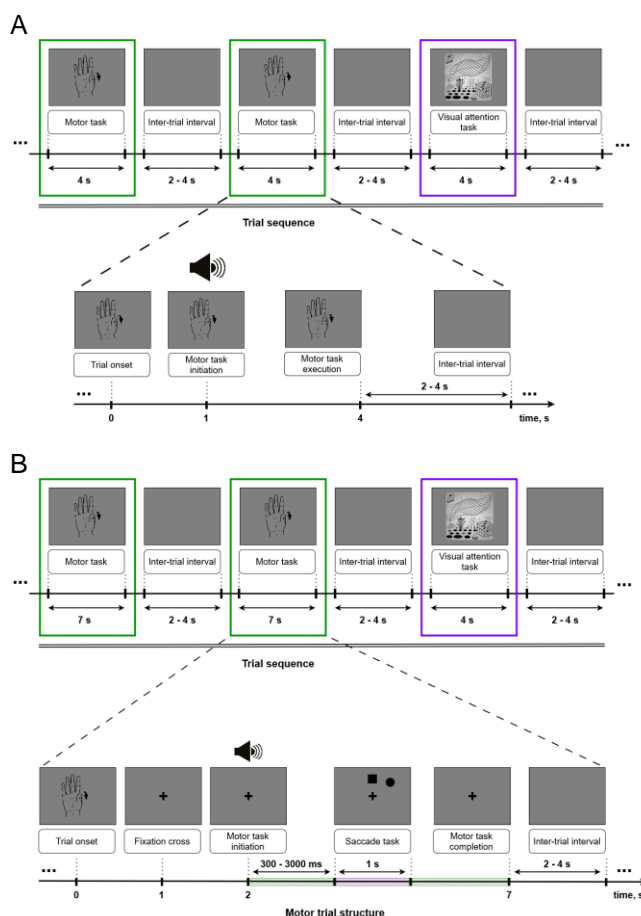


Figure 1. Trial structure in (A) single and (B) dual tasks.

In the beginning of each dual-task trial a fixation cross appeared in the center of the screen. After 2 s a sound signaled to start the sensorimotor task (OM, IM or QM). Participants were asked to fixate on the cross until (after a random time interval) a target and a distractor appeared (for half of the group, they were circle and square, relatively, and for another half, vice versa), when they had to make a saccade to the target as soon as possible. As in [19], they were free to continue the internal (here, sensorimotor) task at this time or to return to it after making a saccade on the target but had to complete it anyway. (See Fig. 1 for additional details).

**EEG and EMG analysis** followed [24]. As the baseline for ERD/ERS computation, however, here we used a 500 ms interval preceding the fixation cross presentation. We also did not use here special procedures for removing possible contribution of residual muscle activation to brain activation, because in the current study the EMG was stricter controlled during the experiment than in [24], and because we already shown in [24] that small residual EMG increases in some trials in this task are not related to any substantial EEG effect. We refrained from assessing performance of BCI classifiers on the EEG data, because it would very likely just mirror the effects observed in the averaged data and because we plan to assess classification performance in an online hybrid BCI experiment, which could serve as a much more relevant model.

**Eye movement analysis** – Fixations were considered maintained on the fixation cross in the dual-task trials if they did not depart from it further than  $1^\circ$  before the saccade. Saccade latency was computed as time between target presentation onset and saccade onset. Saccades were considered as landing on a target/distractor if they ended within  $2^\circ$  from them.

## RESULTS

**Task difficulty** – Participants were asked to indicate whether IM+saccades or QM+saccades condition was more difficult, using a visual analogue scale (VAS). With 0 corresponding to more difficult QM+saccades and 1 to IM+saccades,  $M \pm SD$  was  $0.66 \pm 0.34$ , median=0.83. Only one participant found no difference between the conditions, and four reported QM+saccades as more difficult; importantly, all those four had difficulties in mastering the QM, and three of them reported that the problem for them was avoiding pronounced movements.

**Gaze performance** in dual-tasks with IM was only slightly (insignificantly, according to Wilcoxon paired test) lower than in dual-tasks with QM, although the difference was in favor of QM for all three analyzed indices (Figure 2).

**EEG results** – Group averaged time-frequency plots for strongest individual contralateral alpha band sources of the EEG sensorimotor rhythm are shown in Fig. 3. Stronger alpha band desynchronization was observed in

QM compared to IM dual task conditions (compare  $QM_{ST}$  vs.  $IM_{ST}$  in Fig. 3).

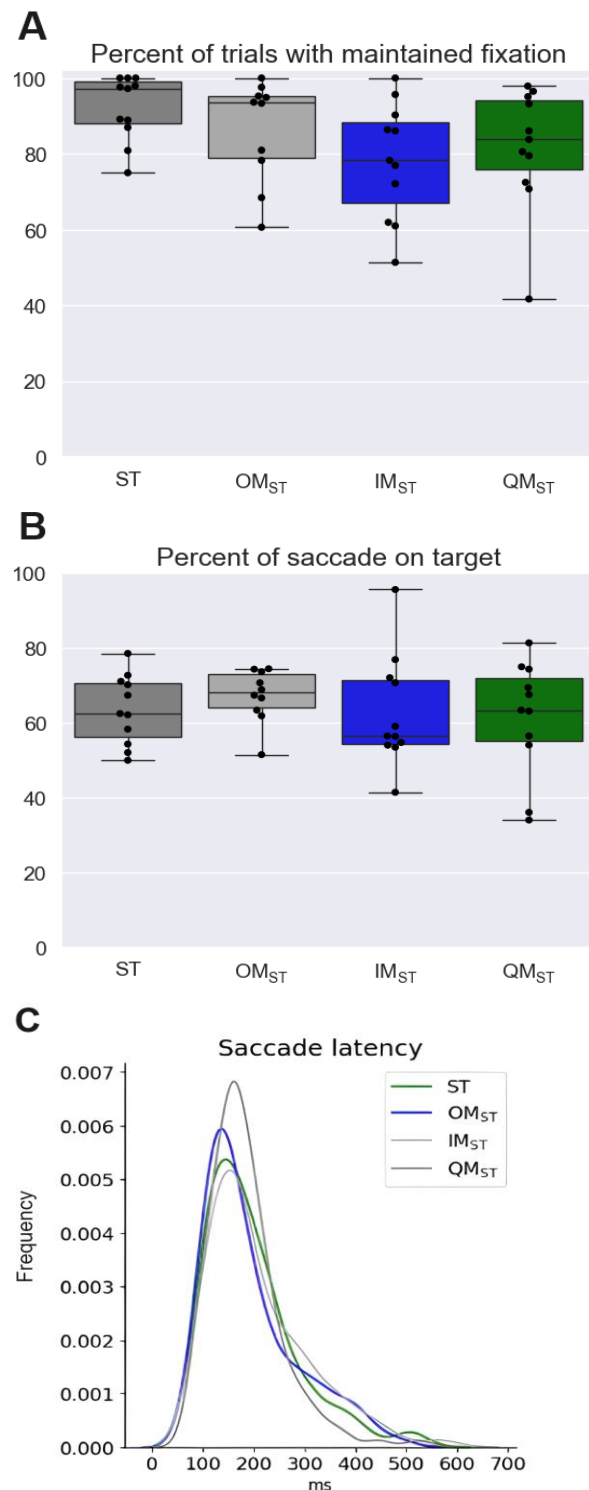


Figure 2. Gaze performance in the saccade task and in the dual task conditions (N=11).

## DISCUSSION

The experimental results reported above should be considered as preliminary, due to the limited number of participants and incomplete analysis. All observed tendencies were in favor of our hypotheses, i.e., were in line with the assumption that attempted movements, modeled here with quasi-movements, are more compatible with intentional gaze use than kinesthetic motor imagery. However, most differences did not reach statistical significance. We are planning to collect data from 10 more participants and to refine the analysis, to get to more definite conclusions. Still, at the current stage attempted movements were confirmed to be at least as effective as motor imagery commonly used in noninvasive BCIs.

Note that quasi-movements are a most minimized form of attempted movements, requiring lowest “motor” effort (apart from additional cognitive control required to prevent movement). To fully understand the potential of attempted movements, it also may make sense to explore them more in paralyzed patients, amputees, and using the constrained movement paradigm [21], where effort can be much stronger. Nevertheless, studies of quasi-movements, featured with an unique combination of attempt to move and absence of any physical effect in people that are able to make movement [23, 25], may significantly enrich the whole picture.

Interestingly, the most common complaint from our participants about the quasi-movements was that it was difficult for them not to make a pronounced movement. Note that in paralyzed patients this is not an issue in most cases, especially in neurorehabilitation, where making a real movement instead of just trying to make it is the goal of training.

Attempted movements have certain features that are helpful from a practical point of view. In particular, they can be more easily explained to many paralyzed individuals than kinesthetic imagery (its training often starts in healthy participants from making overt movements!) and seem to require far less training to produce clear EEG patterns. However, studies are needed to understand if attempted movements, including quasi-movements, can elicit the same or higher EEG effects as imagery after significant time of practice. For quasi-movement, it is also important to study various movements: so far, only thumb abduction was explored in all studies, partly due to the assumed need to precisely control EMG (but this may be not really important, as our previous study showed no contribution of residual muscle activation in quasi-movements to EEG effects [24]).

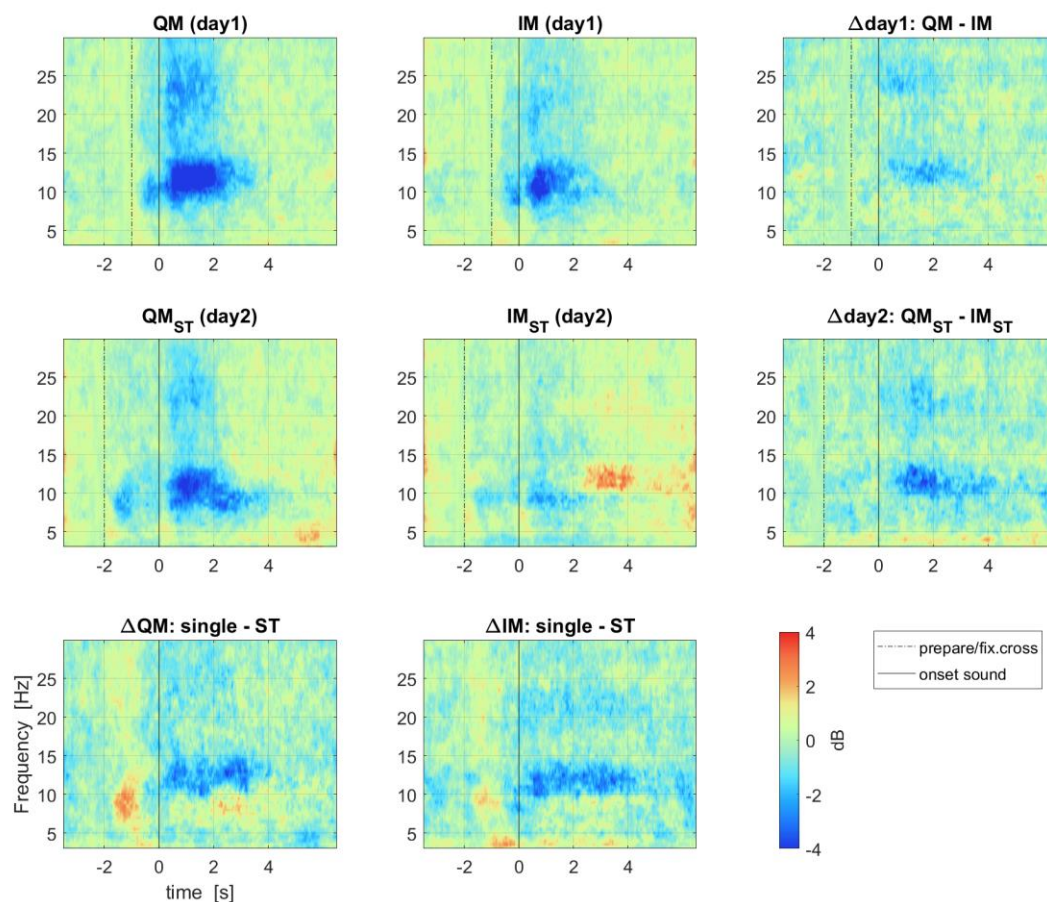


Figure 3. Group (N=11) median-averaged time-frequency plots per condition and differences between them. Individual data were computed for strongest individual CSP-derived contralateral sources for EEG alpha frequency band (see [24] for the details of the analysis).

## CONCLUSION

We proposed a hypothesis that attempted movements can be more compatible with the interaction with the external world than imaginary movements, the mental task commonly used in noninvasive BCIs. Preliminary results of its testing using prosaccades as an example of external task and quasi-movements as a model of attempted movements were mostly in line with its predictions, suggesting that attempted movements should be probably considered at least as an important supplement of imagery in BCIs. More studies, however, are needed for more definite conclusions. The study also may be considered as a new demonstration of the potential of quasi-movements, a very little explored phenomenon, for BCI research.

## ACKNOWLEDGEMENTS

This work supported by the Russian Science Foundation, grant 22-19-00528.

## REFERENCES

- [1] Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. *Proc. IEEE*. 2001;89(7): 1123-1134  
<https://doi.org/10.1109/5.939829>
- [2] Roc A, Pillette L, Mladenovic J, Benaroch C, N’Kaoua B, Jeunet C, Lotte F. A review of user training methods in brain computer interfaces based on mental tasks. *J. Neural Eng.* 2021;18(1): 011002  
<https://doi.org/10.1088/1741-2552/abca17>
- [3] Cervera MA, Soekadar SR, Ushiba J, Millan J, Liu M, Birbaumer N, et al. Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. *Ann. Clin. Transl. Neurol.* 2018;5: 651-663  
<https://doi.org/10.1002/acn3.544>
- [4] Nojima I, Sugata H, Takeuchi H, Mima T. Brain-computer interface training based on brain activity can induce motor recovery in patients with stroke: a meta-analysis. *Neurorehab. Neural Repair.* 2022;36(2): 83-96  
<https://doi.org/10.1177/15459683211062895>
- [5] O’Callaghan J. Apple Vision Pro: what does it mean for scientists? *Nature*, 12 February 2024.

- <https://doi.org/10.1038/d41586-024-00387-z>
- [6] Witkowski M, Cortese M, Cempini M, Mellinger J, Vitiello N, Soekadar SR. Enhancing brain-machine interface (BMI) control of a hand exoskeleton using electrooculography (EOG). *J. Neuroeng. Rehab.* 2014;11: 165 <https://doi.org/10.1186/1743-0003-11-165>
  - [7] Soekadar SR, Witkowski M, Gómez C, Opisso E, Medina J, Cortese M, et al. Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully independent daily living activities after quadriplegia. *Science Robotics.* 2016;1(1): eaag3296 <https://doi.org/10.1126/scirobotics.aag3296>
  - [8] Zander TO, Gaertner M, Kothe C, Vilimek R. Combining eye gaze input with a brain-computer interface for touchless human-computer interaction. *Int. J. Human-Comp. Interaction.* 2010;27(1): 38-51 <https://doi.org/10.1080/10447318.2011.535752>
  - [9] Hou BJ, Bekgaard P, MacKenzie S, Hansen JP, Puthusserypady S. GIMIS: Gaze input with motor imagery selection, in *Proc. ACM Symp. on Eye Tracking Research and Applications*, Stuttgart, Germany 2020, Article No. 18, 1-10 <https://doi.org/10.1145/3379157.3388932>
  - [10] Ramos-Murguialday A, Broetz D, Rea M, Lær L, Yilmaz Ö, Brasil FL, et al. Brain-machine interface in chronic stroke rehabilitation: a controlled study. *Ann. Neurol.* 2013;74: 100-108 <https://doi.org/10.1002/ana.23879>
  - [11] Bai Z, Fong KNK, Zhang JJ, Chan J, Ting KH. Immediate and long-term effects of BCI-based rehabilitation of the upper extremity after stroke: A systematic review and meta-analysis. *J. Neuroeng. Rehabil.* 2020;17: 57 <https://doi.org/10.1186/s12984-020-00686-2>
  - [12] Mansour S, Ang KK, Nair KP, Phua KS, Arvanah M. Efficacy of brain-computer interface and the impact of its design characteristics on poststroke upper-limb rehabilitation: A systematic review and meta-analysis of randomized controlled trials. *Clin. EEG Neurosci.* 2022;53(1): 79-90 <https://doi.org/10.1177/15500594211009065>
  - [13] Metzger SL, Liu JR, Moses DA et al. Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis. *Nat. Commun.* 2022;13: 6510 <https://doi.org/10.1038/s41467-022-33611-3>
  - [14] Lorach H, Galvez A, Spagnolo V. et al. Walking naturally after spinal cord injury using a brain-spine interface. *Nature.* 2023;618: 126-133 <https://doi.org/10.1038/s41586-023-06094-5>
  - [15] Shelchkova ND, Downey JE, Greenspon CM et al. Microstimulation of human somatosensory cortex evokes task-dependent, spatially patterned responses in motor cortex. *Nat. Commun.* 2023;14: 7270 <https://doi.org/10.1038/s41467-023-43140-2>
  - [16] Aflalo T, Zhang C, Revechikis B, Rosario E, Pouratian N, Andersen RA. Implicit mechanisms of intention. *Curr. Biol.* 2022;32(9): 2051-2060 <https://doi.org/10.1016/j.cub.2022.03.047>
  - [17] Oxley TJ, Yoo PE, Rind GS, Ronayne SM, Lee CS, Bird C, et al. Motor neuroprosthesis implanted with neurointerventional surgery improves capacity for activities of daily living tasks in severe paralysis: first in-human experience. *J. Neurointerv. Surgery.* 2021;13(2): 102-108 <http://dx.doi.org/10.1136/neurintsurg-2020-016862>
  - [18] Korda Ž, Walcher S, Körner C, Benedek M. Effects of internally directed cognition on smooth pursuit eye movements: A systematic examination of perceptual decoupling. *Attent., Percept., Psychophys.* 2023;85(4): 1159-1178 <https://doi.org/10.3758/s13414-023-02688-3>
  - [19] Walcher S., Korda Ž., Körner C., Benedek M. The effects of type and workload of internal tasks on voluntary saccades in a targetdistractor saccade task. *PLoS ONE.* 2023;18(8): e0290322 <https://doi.org/10.1371/journal.pone.0290322>
  - [20] Pathak A, Patel S, Karlinsky A, Taravati S, Welsh TN. The “eye” in imagination: The role of eye movements in a reciprocal aiming task. *Behav. Brain Res.* 2023;441: 114261 <https://doi.org/10.1016/j.bbr.2022.114261>
  - [21] Mondini V, Sburlea AI, Müller-Putz GR. Towards unlocking motor control in spinal cord injured by applying an online EEG-based framework to decode motor intention, trajectory and error processing. *Sci. Rep.* 2024;14: 4714 <https://doi.org/10.1038/s41598-024-55413-x>
  - [22] Blokland Y, Spyrou L, Lerou J, Mourisse J, Jan Scheffer G, Geffen GJ Van, et al. Detection of attempted movement from the EEG during neuromuscular block: Proof of principle study in awake volunteers. *Sci. Rep.* 2015;5: 12815 <https://doi.org/10.1038/srep12815>
  - [23] Nikulin VV, Hohlefeld FU, Jacobs AM, Curio G. Quasi-movements: A novel motor-cognitive phenomenon. *Neuropsychologia.* 2008;46(2): 727-742.
  - [24] Vasilyev A. N., Yashin A. S., Shishkin S. L. Quasi-movements and “quasi-quasi-movements”: does residual muscle activation matter? *Life.* 2023;13(2): 303 <https://doi.org/10.3390/life13020303>
  - [25] Yashin AS, Shishkin SL, Vasilyev AN. Is there a continuum of agentive awareness across physical and mental actions? The case of quasi-movements. *Consc. Cogn.* 2023;112: 103531 <https://doi.org/10.1016/j.concog.2023.103531>
  - [26] Nuzhdin YO. Resonance - a BCI framework for working with multiple data sources, in *Proc. 8th Graz Brain-Computer Interface Conf.* 2019, Graz, Austria, 2019, 77–81 <https://doi.org/10.3217/978-3-85125-682-6-15>

# USING A CNN-LSTM ARCHITECTURE WITH DATA AUGMENTATION TO IMPROVE HD-ECOG SPOKEN SYLLABLE CLASSIFICATION

Mehdi Javani Mirehkoohi<sup>1</sup>, Zachary Freudenburg<sup>1</sup>, Amira Neumann<sup>1</sup>, Nick F Ramsey<sup>1</sup>

<sup>1</sup>Brain Center, Department of Neurology and Neurosurgery, University Medical Center Utrecht, Utrecht, 3584 CX, the Netherlands

E-mail: m.javanimirehkoohi@umcutrecht.nl

**ABSTRACT:** Brain-Computer Interfaces (BCIs) have emerged as vital tools in understanding and assisting individuals with LIS due to neurological diseases such as ALS. This study focuses on the and feasibility of recognizing spoken syllables from implanted HD-ECoG signals as a platform for Speech BCIs. We propose a hybrid deep learning model, which uses a modified EEGNet as a feature extractor coupled with an LSTM. A primary challenge in this domain is the limited quantity of ECoG data. To address this challenge, we employ window clipping as a data augmentation technique, effectively increasing the amount of training data available for the model. Using a dataset comprising recordings from six subjects implanted with HD-ECoG, we evaluate our proposed method. Results indicate a notable improvement in classification accuracy achieved through the designed hybrid DL model. Furthermore, our findings elucidate the distinctive impact of data augmentation methods in further enhancing the performance of our designed model.

**Keywords:** HD-ECoG, ECoGNet, CNN, LSTM, Data Augmentation

## INTRODUCTION

The realm of Brain-Computer Interface (BCI) systems has revolutionized human-computer interaction, enabling direct communication pathways between the human brain and external devices. Language BCI represent a frontier in assistive technology, designed to empower individuals with communication disabilities by translating recorded brain activity into language. Electrocorticography (ECoG) due to its capabilities in recording a wide range of frequency and also high density recording of a specific areas of the brain which are responsible for specific cognitive task has been widely used in this regard. In recent years, researchers have made many efforts to leverage various methods to decode language, particularly deep learning methods as the most promising method to this aim ([?], [?], [?]). While various deep learning architectures have demonstrated success in decoding of spoken phonemes, words, and sentences with acceptable performances, the task of syllable decoding poses greater challenges. Unlike words, which vary in length and possess distinct sounds, syllables typ-

ically exhibit uniform length and share acoustic features. Consequently, decoding syllables presents a formidable hurdle, as neural networks cannot rely solely on length or distinctiveness for classification. Also, syllables can involve overlapping combinations of phonemes making them useful building blocks for language but also less distinct. Despite the complexity, decoding syllables is pivotal, serving as a foundational step towards deciphering spoken words. Addressing this challenge necessitates the development of robust and adaptable neural networks capable of enhancing decoding performance, particularly for individuals with limited data. By exploring the potential of such networks, we aim to push the boundaries of language decoding in BCIs, fostering greater inclusivity and effectiveness in communication assistance technologies.

EEGNet efficiently extracts temporal features reflecting short and long-term changes in brain activity. While EEG and ECoG measure brain electrical potential differences using electrodes, they differ in invasiveness and spatial coverage [?]. EEG, non-invasive, captures broader spatial coverage with lower density, while ECoG, invasive, offers higher density with narrower spatial coverage. Despite these differences, both methods share preprocessing and feature extraction techniques, often utilizing frequency analysis. Thus, deep neural networks proficient in extracting frequency information from EEG data could enhance ECoG analysis.

Peterson et al. [?] introduced a modified version of EEGNet tailored for ECoG, incorporating a mapping layer from individual ECoG electrode positions to a 1D input space. While their approach yielded improved results over traditional EEGNet in binary classification tasks, we sought to explore an alternative mapping paradigm. Thus, we directly adapted EEGNet to investigate this alternative approach.

Our mapping concept involves translating the inherent 2D structure of ECoG data into a standardized grid space, aligning native electrode coordinates. Leveraging this spatial input in 2D, we have tailored a modified variant of EEGNet specifically optimized for the unique characteristics of ECoG data. This adapted network, denoted as ECoGNet, maintains a parallel block structure to EEGNet while accommodating the intricacies of ECoG signal processing.

An intriguing progression entails merging CNN and LSTM networks to build hybrid architectures. This synthesis facilitates the concurrent extraction of spatial and temporal features, correspondingly. This novel approach bears substantial potential for augmenting the capabilities of Deep Learning [?, ?, ?].

Utilizing Deep Learning for ECoG signal classification faces challenges due to limited dataset sizes, particularly in Motor Imagery analysis. K-Fold Cross Validation (K-Fold CV) addresses this issue by partitioning data into 'K' subsets, enabling robust model training and evaluation. Hewaidi et al. [?] leveraged K-Fold CV to enhance their methodology, integrating variational autoencoders, deep autoencoders (DAE), and CNNs for EEG motor imagery classification. Recent literature [?] introduces two key K-Fold CV methods: inter-subject and intra-subject, providing insights into model performance across subjects and within individual subjects, respectively.

Data augmentation methods offer a potent solution to the challenge of limited dataset sizes in Deep Learning. By expanding the training data, these techniques bolster classification stability and accuracy, enabling models to generalize better to new datasets [?]. Moreover, data augmentation addresses class imbalance issues, crucial for classification tasks. Techniques such as geometric transformations and noise introduction effectively diversify datasets, enhancing model robustness. The utilization of sliding windows is a prevalent data augmentation technique across various domains. In neonatal seizure detection, O'Shea et al. [?] employed overlapping windows, with 8-second trials and 50% overlap, to augment seizure instances within EEG signals. Kwak et al. [?] explored different shift lengths, ranging from 10 ms to 60 ms within 2-second windows, revealing superior performance with shorter shifts.

In this paper, we present a comprehensive approach to EEG signal classification, leveraging deep learning models and innovative data augmentation techniques. We begin by introducing the analyzed data in the Data and Materials section, followed by an explanation of the used deep learning architectures. Subsequently, we describe our designed model and detail the methods employed to address the inherent challenges posed by limited dataset sizes. Moving forward, the Results section showcases the outcomes of implementing our model, with particular emphasis on the impact of utilizing data augmentation methods. Finally, we conclude by summarizing the project's findings and highlighting avenues for future research and development in EEG signal classification.

## DATA AND MATERIALS

*Data and Preprocessing:* The dataset was collected at UMC Utrecht and comprises recordings from six subjects. Each subject underwent different trials, and the electrode configurations varied among subjects. Some subjects contributed 180 trials, while others had 90, and due to data collection errors, certain trials were elimi-

nated from the valid dataset. Moreover, the number of electrodes differed among subjects, with some recorded using 128 electrodes and others with 64 electrodes. It's worth noting that not all electrodes provided valid signals, as some were too noisy to convey useful information. The properties of the dataset are summarized in table 1.

The locations of the electrode grids for all participants are illustrated in Figure 1. Due to various restrictions and limitations, such as individual anatomical variations and positioning constraints during data collection, the electrode placements vary in their standard Montreal Neurological Institute (MNI) coordinate system locations across subjects. Here the electrode locations are determined based on spherical components (Phi and Theta), with the center of the component aligning with the center of the brain.

Table 1: Summary of the data of all participants

Participant	n. actual/valid trials	Sampling Freq	n. actual/valid Electrodes
S01	180/177	2000	128/128
S02	180/173	512	64/52
S03	90/85	2000	128/125
S04	90/89	2000	128/120
S05	90/86	2000	128/121
S06	90/87	2000	128/109

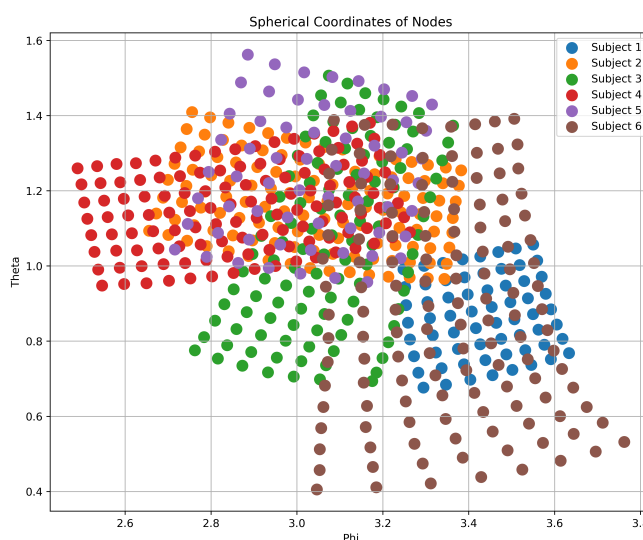


Figure 1: Electrode Grid Placements for Participants

The task entails conducting trials where participants utter one of nine distinct syllables: “mi”, “mu”, “ma”, “ki”, “ku”, “ka”, “zi”, “zu”, and “za”. These syllables exhibit similarities in their articulation and serve as the nine classes we seek to classify. The trial protocol includes randomization of these syllables interspersed with occasional rest trials. Participants are prompted on the screen to perform either 10 or 20 repetitions of each syllable.

The subjects performed the task multiple times leading to a range in trials from 180 trials, to 90 over subjects, and due to data collection errors, certain trials were eliminated from the valid dataset.

Two crucial time points are defined: Cue time marks the initiation of monitoring for the intended syllable, and Voice Onset Time (VoT) indicates when it becomes discernible that the participant starts articulating the syllable. Due to individual differences, participants initiate pronunciation after different durations following the Cue time. Additionally, the duration from VoT varies depending on the syllable and the participant's capabilities.

To standardize trial durations across participants and syllables, a fixed duration of 1 second is set starting from VoT, recognized as the most informative segment of the trial.

**CNN:** Convolutional Neural Networks (CNNs) have garnered considerable acclaim for their adeptness in extracting robust spatial features from images through deep learning. The architectural underpinnings of CNNs ensure spatial robustness, which revolves around three pivotal elements: local receptive fields, convolutional layers, and pooling layers. By employing small receptive fields, convolutional filters adeptly capture fundamental visual features from distinct regions of the input image. These extracted features undergo progressive amalgamation and enhancement across subsequent layers to discern higher-level features. However, the insertion of pooling layers following convolutional layers, while essential for preventing overfitting and reducing spatial dimensions, can potentially entail a loss of precise spatial information—a concern warranting attention.

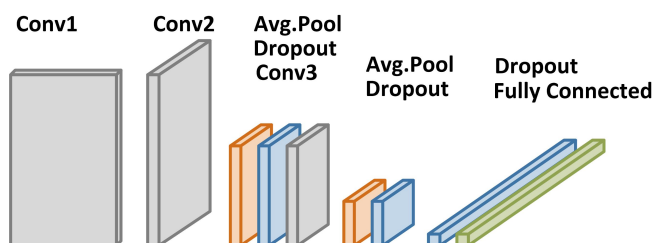


Figure 2: EEGNET Architecture

### LSTM:

The LSTM architecture revolves around three primary states: the cell state ( $C_{t-1}, C_t$ ), the input state ( $X_t$  and  $h_{t-1}$ ), and the output state ( $h_t$ ). Additionally, LSTM incorporates four crucial gates: the forget gate ( $f_t$ ), the input gate ( $i_t$ ), the new memory gate ( $C'_t$ ), and the output gate ( $O_t$ ). These gates play pivotal roles in regulating internal operations within the LSTM.

The cell state serves as a memory reservoir that facilitates information flow across LSTM units. Each LSTM unit features skip connections in the form of gates, which intricately control the inflow and outflow of information to and from the cell state. Specifically, the forget gate discerns which information to retain or discard from the prior cell state, while the input gate governs the integra-

tion of new information.

The new cell state is crafted by merging the previous cell state with inputs from the input gate and the new memory gate. Finally, the output gate oversees the information contributing to the LSTM unit's output. Leveraging these architectural components, including gates and memory units, empowers the network to capture and retain pertinent information essential for effective learning.

The LSTM's prowess in managing long-term dependencies underscores its versatility and efficacy across a spectrum of deep learning applications.

### MODEL ARCHITECTURES (HYBRID CNN/LSTM APPROACH)

To effectively capture the intricate spatial and temporal characteristics inherent in ECoG signals, we propose a sophisticated hybrid neural network architecture that seamlessly integrates Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) networks. This hybrid approach combines the robustness of CNNs in spatial feature extraction with the proficiency of LSTMs in modeling sequential data, thereby augmenting the analysis and classification of ECoG signals.

In this work we have used EEGNET model as the CNN component of the designed hybrid model. we adopt the EEGNET model as the foundation for our CNN component, tailored with necessary modifications to suit the dimensions of our ECoG data (16/8,8,2000/512). This adaptation ensures optimal utilization of the CNN's capabilities in discerning spatial intricacies within the ECoG signals. We call this network ECoGNet.

EEGNet is a Deep Learning model structured with multiple convolutional blocks, outlined in Figure 2. The initial block consists of a standard convolutional layer followed by a batch normalization (BN) layer. Subsequently, a depth-wise convolutional layer is utilized in the following block, succeeded by a BN layer, an Exponential Linear Unit (ELU) activation function, and an average pooling layer. Additionally, a dropout layer is introduced at the end of this block. The third block incorporates a separable convolution, a BN layer, an ELU activation, and another average pooling layer. Notably, dropout layers are applied both before and after flattening the data. For the classification stage, a fully connected layer is employed, followed by a softmax function to classify the data into nine distinct classes.

In tandem with the CNN component, we incorporate an LSTM network to capture the nuanced temporal dependencies inherent in ECoG signals. LSTMs, renowned for their prowess in modeling sequential data, prove instrumental in unraveling the temporal dynamics and long-term dependencies embedded within the ECoG signals. By seamlessly integrating an LSTM network, our model gains the ability to discern intricate temporal patterns, thereby enriching the analysis of ECoG signals.

### Deep Learning Obstacle:

Enhancing the proficiency of designed models often relies on providing them with enough data to be learned. However, in certain applications such as the analysis of brain signals, acquiring a sufficient amount of data can be challenging due to constraints imposed by the nature of the data collection process. This scarcity of data poses a significant obstacle for biomedical specialists seeking to train effective models. In this project we have used a couple of methods to deal with this problem, Cross Validation, and Data Augmentation.

To mitigate the data constraint, K-Fold Cross Validation (K-Fold CV) is commonly employed. K-Fold CV divides the data into 'K' subsets, allowing the model to train on different combinations and reducing overfitting while providing robust evaluation of generalization ability. For instance, Hwaidi et al. [?] utilized K-Fold CV to enhance the performance of their approach, integrating variational autoencoders, deep autoencoders, and CNNs for EEG signal classification. In this work, we have also used this method to not only try to mitigate the data restrictions, but also prevent overfitting. Due to the amount of the data we have, the 5-fold CV is chosen.

The second commonly used method to deal with the obstacle, is the Data Augmentation method. Data augmentation is a highly effective method for addressing the challenge of limited dataset size in deep learning. By increasing the quantity and variety of training data, it enhances classification stability and accuracy, enabling models to be more robust and less biased when handling new datasets [?]. Additionally, data augmentation helps mitigate class imbalance in classification tasks. It employs geometric transformations such as translations, rotations, cropping, flipping, and scaling, along with noise introduction, to expand the dataset and generate new instances. Depending on the type of data involved, various augmentation techniques can be applied. In the realm of biomedical signal analysis, window clipping stands out as a widely utilized method. In our project, we have employed window clipping to augment the available data. Nevertheless, this approach encounters challenges, particularly in determining the optimal quantity of clipped windows and their overlapping ranges. During our experimentation, we conducted tests using different numbers of windows ranging from 1 to 4, with each window having a fixed duration of 1 second. Additionally, we explored various overlapping ranges, spanning from 5% to 50%. These parameters were inherently constrained by the duration of the useful signal.

In our proposed hybrid architecture, the LSTM component follows the CNN component. This architectural arrangement facilitates the seamless flow of information from spatial to temporal domains, as the output of the CNN is meticulously fed into the LSTM network. This cohesive integration empowers the model to discern sequential patterns and dependencies within the ECoG signals, thereby enabling a holistic understanding of both spatial and temporal aspects of the data.

By harnessing the collective strengths of CNNs and

LSTMs, our hybrid architecture endeavors to exploit spatial and temporal information in tandem, thereby enhancing the discriminative power and interpretability of our proposed model. This comprehensive approach facilitates a nuanced analysis and classification of ECoG signals, paving the way for advancements in neuroscientific research and clinical applications.

In this work, we will analyze the results of our designed model from 2 aspect. First, we want to find out how adding LSTM as a classifier to the ECoGNet model may enhance the accuracy percentage, and then we will test the effect of using data augmentation method to the performance of the designed model, and comparing the designed model's performance when we use different numbers of windows in data augmentation.

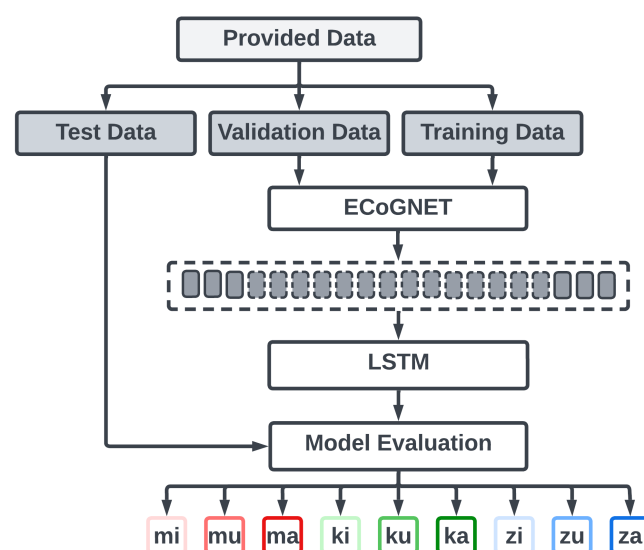


Figure 3: Hybrid Architecture

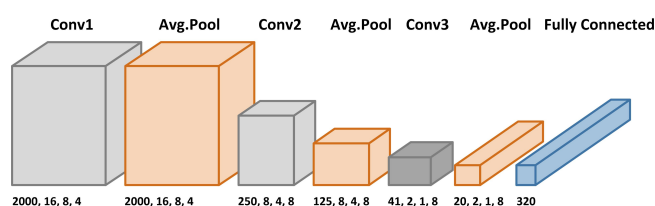


Figure 4: Modified EEGNET Architecture (Imported data shape: 2000 x 16 x 8)

To find out the performance of the designed model, we should provide a baseline model which is strong enough to classify the ECoG data. As a baseline model, we utilized Spatial Match Filters (SMFs), a non-deep learning method commonly employed in BCI research, particularly with ECoG data. SMFs have demonstrated promising results, achieving a classification accuracy of 76% in four phoneme classification tasks [?]. This technique involves a trial-by-trial comparison of activity patterns against mean activity patterns of different conditions. Initially, the signal undergoes wavelet transformation into the time-frequency domain, followed by computation of

mean values across defined frequency bands and time-points for each electrode, focusing on the high frequency band between 65 and 95 Hz. The resulting mean values represent activity patterns for each electrode. Subsequently, correlation analysis is performed between each trial's activity pattern and the mean activity patterns for nine classes. This correlation computation, conducted in a leave-one-out fashion to ensure unbiased estimates, assigns each trial to the class with the highest correlation pattern. Notably, only electrodes with significant functional responses are included in the correlation computation, further enhancing classification accuracy.

The second baseline model which is used to comparison aims is EEGNET that has been introduced.

## RESULTS

The table 2 shows the results of implementing the designed model to classify the 9 syllabus of each participant. the results shows a significant improvement in accuracy. when a LSTM layer is added. This is highlighted by the fact that for all subjects the accuracy is at least 9 percentage points above the theoretical chance level of 11% for the Hybrid DL Model while both ECoGNet and SMF both show 3/6 subjects below or around chance level.

Table 2: Accuracy (%)

Participant	SMF	EEGNet	Hybrid DL Model
S01	39	60	67
S02	24.9	20	20
S03	13.3	12	29.41
S04	0	14.3	33.33
S05	12.5	14.4	21.1
S06	20.2	27.8	27.8
Mean	18.32	24.75	33.11

As it is discussed in previous sections, we have used 5-fold CV and also window clipping method to overcome the limitation of data quantity. During our experimentation, we conducted tests using different numbers of windows and explored various overlapping ranges. These parameters were inherently constrained by the duration of the useful signal.

The results, presented in Table 3, clearly demonstrate the efficacy of data augmentation in improving the model's proficiency in classification tasks.

By leveraging data augmentation techniques, we have successfully enhanced the model's ability to classify biomedical signals. This augmentation strategy not only mitigates the limitations imposed by data scarcity but also contributes to the overall robustness and generalization capability of the model.

Table 3: Accuracy with Different Numbers of Windows

Participant	Windows			
	1w	2w	3w	4w
S01	67	81	62	67.8
S02	20	20	20	20
S03	29.41	32.35	33.3	20.59
S04	33.33	40	43	36.11
S05	21.1	30.6	44.44	47.2
S06	27.8	36.1	27.78	30.56
Mean	33.107	40.01	38.42	37.04

Analyzing the table reveals an intriguing trend: an increase in the number of windows from 1 to 2 correlates with higher accuracy. However, this pattern falters as additional windows are added, resulting in a decline in overall performance by mean. This observation underscores a crucial point: not all segments of each trial contribute equally to classification accuracy. Indeed, the informativeness of added windows varies, with non-informative windows potentially detracting from overall results. Notably, this effect can differ across subjects. For instance, Subject S05 demonstrates an increase in accuracy but experiences a slowdown in processing speed with the introduction of 3 and 4 windows. Conversely, Subject S01 witnesses a decline in accuracy after the incorporation of the third and fourth windows.

## DISCUSSION

Such disparities highlight the nuanced interplay between participant concentration levels, physical capabilities, and data quality. Indeed, individual differences among participants can significantly influence the duration of informative data within each trial. Moreover, the expansion of the number of windows necessitates more extensive data processing, demanding higher computational resources and potentially leading to longer computation times. Consequently, a delicate balance must be struck between model accuracy and computational efficiency. In certain scenarios, such as those where resource constraints are paramount, opting for two windows may represent the more optimal choice.

Thus, a thorough consideration of the trade-offs between performance and computational resources is imperative in maximizing the effectiveness of the classification process while ensuring optimal resource allocation.

## CONCLUSION

During this work, we proposed a novel hybrid deep learning model that combines a modified EEGNet for feature extraction with a LSTM network for temporal analysis. Our approach addresses the challenge of limited ECoG data through the innovative use of window clipping as a data augmentation technique.

Our experiments, conducted on a dataset comprising recordings from six subjects, demonstrate promising results. We observed a significant enhancement in classification accuracy compared to previous models, affirming the effectiveness of our hybrid model in recognizing the syllableness.

Furthermore, our analysis of data augmentation techniques highlights the importance of optimizing the number of clipped windows to balance classification accuracy and computational efficiency. While increasing the number of windows initially improves accuracy, there is a diminishing return beyond a certain point, emphasizing the need for careful consideration of resource constraints and performance trade-offs.

## REFERENCES

- [1] Goli P, Mazrooei Rad E. Advantages of deep learning for ecog-based speech recognition. *The Hearing Journal*. 2019;72(8):10.
- [2] Luo S, Rabbani Q, Crone NE. Brain-computer interface: Applications to speech decoding and synthesis to augment communication. *Neurotherapeutics*. 2022;19(1):263–273.
- [3] Pandarinath C *et al.* High performance communication by people with paralysis using an intracortical brain-computer interface. *eLife*. 2017;6(February):e18554.
- [4] Hashiguchi K *et al.* Correlation between scalp-recorded electroencephalographic and electrocorticographic activities during ictal period. *Seizure*. 2007;16(3):238–247.
- [5] Peterson SM, Steine-Hanson Z, Davis N, Rao RPN, Brunton BW. Generalized neural decoders for transfer learning across participants and recording modalities. *Journal of Neural Engineering*. 2021;18(2):026014.
- [6] Zhang R, Zong W, Dou L, Zhao X, Tang Y, Li Z. Hybrid deep neural network using transfer learning for eeg motor imagery decoding. *Biomedical Signal Processing and Control*. 2021;63:102144.
- [7] Khademi Z, Ebrahimi F, Montazery Kordy H. A transfer learning-based cnn and lstm hybrid deep learning model to classify motor imagery eeg signals. *Journal of Neural Engineering*. 2023;20(2):025004.
- [8] Li H, Ding M, Zhang R, Xiu C. Motor imagery eeg classification algorithm based on cnn-lstm feature fusion network. *Biomedical signal processing and control*. 2022;72:103342.
- [9] Hwaidi JF, Chen TM. Classification of motor imagery eeg signals based on deep autoencoder and convolutional neural network approach. *IEEE access*. 2022;10:48071–48081.
- [10] Wang X, Hersche M, Tömekce B, Kaya B, Magno M, Benini L. An accurate eegnet-based motor-imagery brain-computer interface for low-power edge computing. In: 2020 IEEE international symposium on medical measurements and applications (MeMeA). 2020, 1–6.
- [11] Wang F, Zhong Sh, Peng J, Jiang J, Liu Y. Data augmentation for eeg-based emotion recognition with deep convolutional neural networks. In: *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part II* 24. 2018, 82–93.
- [12] O'Shea A, Lightbody G, Boylan G, Temko A. Neonatal seizure detection using convolutional neural networks. In: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). 2017, 1–6.
- [13] Kwak NS, Müller KR, Lee SW. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PloS one*. 2017;12(2):e0172578.
- [14] Branco MP *et al.* Alice: A tool for automatic localization of intra-cranial electrodes for clinical and high-density grids. *Journal of Neuroscience Methods*. 2018;301:43–51.

# NEURAL CORRELATES OF EXPERTISE DURING KINESTHETIC MOTOR IMAGERY: SHOULD WE REWARD MAXIMUM SMR-ERD?

M. Izac<sup>1</sup>, E. Rossignol<sup>2</sup>, E. Pierrieau<sup>2</sup>, N. Grechukhin<sup>2</sup>, E. Coudroy<sup>2</sup>, B. N’Kaoua<sup>1</sup>, L. Pillette<sup>3</sup>, C. Jeunet-Kelway<sup>2</sup>

<sup>1</sup>Univ. Bordeaux, INSERM, BPH Research Center, UMR 1219, Bordeaux, France

<sup>2</sup>Univ. Bordeaux, CNRS, INCIA, UMR 5287, F-33000 Bordeaux, France

<sup>3</sup>Univ. Rennes, Inria, CNRS, IRISA-F35000 Rennes, France

E-mail: margaux.izac@u-bordeaux.fr

**ABSTRACT:** Athletes practice Kinesthetic Motor Imagery (KMI) for its many benefits. However, lack of feedback impairs regular practice. To optimise KMI efficiency, athletes can use BCIs. Whereas current BCI protocols targeting KMI abilities reward maximum desynchronisation (ERD) of sensorimotor rhythms (SMRs, 12-15Hz), the neural efficiency hypothesis raises the question “what neurophysiological markers should we reinforce?”. We hypothesised that experts’ SMR-ERDs would differ from novices’, in particular when imagining a mastered task. To test this hypothesis, EEG activity was recorded during KMI of bio-mechanically similar tasks: one mastered by experts only and one requiring no specific expertise. Self-reported measures based on validated questionnaires were collected to assess KMI ability and MI frequency of use and to measure their potential impact on SMR-ERD. Experts (basketball players) reported higher perceived KMI abilities than novices, but similar MI practice frequency. In addition, experts showed a stronger SMR-ERD than novices. This effect was only weakly mediated by perceived KMI ability, seeming mainly driven by sport expertise.

## INTRODUCTION

In order to perform, athletes dedicate themselves to both physical and mental training. The latter can take various forms, one of them being Motor Imagery (MI), which can be defined as a “dynamic state during which one simulates an action mentally without any body movement” [1]. Previous research results have shown MI’s positive impact on motor skills, allowing gains in strength [2, 3] or even movement precision [4, 5], especially when practiced in a kinesthetic way. Indeed, by remembering the associated sensations that can be felt during execution, such as muscle contraction/relaxation, body heat, pain, as well as tactile information; one can activate and reinforce similar neural networks to when actually executing the movement [6, 7]. Kinesthetic Motor Imagery (KMI) is therefore a relevant complementary tool for athletes.

However, KMI’s physiological manifestations cannot directly be perceived hence providing no feedback and ob-

jectivity. Indeed, unlike physical practice where athletes can adapt execution according to the output or their body’s proprioceptive feedbacks; athletes cannot directly detect brain activity modulations that occur when doing KMI and adapt their strategy. This can have detrimental consequences on athletes’ motivation to diligently practice KMI as feedback is necessary to learn [8].

Because KMI is associated with an event-related desynchronisation (ERD) of sensorimotor rhythms (SMRs, 12-15 Hz) [9] it is possible to use Brain-Computer Interfaces (BCIs) and provide athletes with a real-time feedback on their brain modulations during KMI. Athletes can then visualise the employed strategy’s efficiency and optimise it if needed. Moreover, three recent reviews testify that BCI training improves both the ability to self-regulate brain activity and sport performance [10–12]. Many KMI-BCI protocols reward maximum SMR-ERD [13]. This suggests we consider that growing expertise will be associated with a higher desynchronisation of neurons in the sensorimotor cortices [14]. Indeed, some related fMRI and MEG findings show greater brain activations in high ability imagers [1, 15] or even in expert athletes in comparison to novices [16]. Nonetheless, some results have suggested the existence of a neural efficiency in experts [17–19]. According to this hypothesis, experts happen to have a reduced modulation of neural activity in comparison to novices [20–22], which can be attributed to a more efficient resource distribution. This efficiency would take form of reinforced temporal and spatial stability during MI tasks [16, 20, 23]. Therefore, rewarding a maximum SMR-ERD might not be the optimal solution.

The aim of our work was to investigate the neural correlates of expertise, in sport expertise and perceived KMI expertise, thereby providing elements to contribute to the debate on what neurophysiological markers should be targeted during KMI-BCI training procedures. Our main hypothesis was that experts’ SMR-ERDs would differ from those of novices, in particular when doing KMI of a mastered task. Thus, we planned an experimental design with “Expertise” (2 modalities: basketball-experts, novices; between groups) and “Task” (2 modalities: free-throw, box-reaching; within groups) as fac-

tors. KMI ability and MI frequency of use were assessed with self-reported measures based on validated questionnaires, allowing us to observe potential differences between groups and if so, add them as co-variables in the analysis of the main hypothesis.

## MATERIALS AND METHODS

### *Participants:*

17 basketball players (M age = 20.6 years old, SD = 2.4 years; 9 women and 8 men) and 16 non-basketball players (M age = 22.7 years old, SD = 3.8 years; 8 women and 8 men) were recruited for a two-hour session. According to Edinburgh Handedness Inventory [24], 28 were right handed (M = 88.68%) and 4 were left handed (M = -77.38%). Basketball players were considered as the expert group (Exp-Gp) as it was composed of competitors from District D1 to National Ligue level whereas non-basketball players were included in the novice group (Nov-Gp) as they attested never to have taken proper basketball lessons. Novices also attested that they did not have a particular expertise in any other sport, instrument playing and video games. This inclusion criterion was to prevent them from being experts in KMI as these activities can require using sensory mental models as well. All volunteers were healthy, declared having no sensory or motor deficits and had a medium to good vision. They were also naive regarding neurofeedback. After being informed of the research aims, conditions and financial compensation, all participants gave their informed written consent. This research was approved by the French Protection of Persons' Committee (national number 2022-A00626-37).

### *Experimental design:*

Participants were seated in front of a 27-inch computer screen and started off with two questionnaires. A modified version of the Imagery Use Questionnaire (IUQ) [25] was used to determine at which frequency participants used MI in their daily life. It consisted of items such as "To what extent do you use MI in your training/activities?" that required an answer using a 7 point Likert scale going from "Never" to "Always". The MI frequency use score was calculated with 12 items. The Motor Imagery Questionnaire-Third Version in French (MIQ-3f) [26] was also completed to assess general KMI ability. Participants were asked to execute a task (knee flexion, bust flexion, vertical jump or horizontal arm adduction), imagine it (using visual or KMI) and rate the vividness of the representation on a 7-point Likert scale. General KMI ability score was obtained by summing the 4 items relative to this MI method. A general explanatory video was then shown to give all necessary instructions regarding EEG, KMI and the protocol. The experiment (See Fig. 1) consisted of 2 blocks, one for each task to imagine, composed of a 2min resting state recording, a 3D stick avatar video, 4 runs of 10 KMI trials, where each run lasted approximately 2min30s, and a general KMI ability assessment. A single run consisted of

a 30s resting state period, followed by 10 KMI trials of 10s, separated by 1 to 3s rest periods and 2s of baseline. Therefore, following instructions, a 2min baseline was recorded during which a white cross was displayed on a black screen. Participants had to fixate its center while "letting their thoughts wonder". A video then presented a 3D stick avatar executing the task to imagine in the next steps. The task could either be a basketball free throw (FreeThrow) or a box reaching action (Reaching) depending on the randomised order of conditions. The latter consisted in moving a cardboard box from a knee height shelf up to a second shelf located high enough to require from participants to be on the tip of their toes. As a familiarisation phase, participants had to execute the task and progressively reduce amplitude until ending up in a sitting position while doing KMI only. Instructions were to do KMI of the task once during the 10s trial but it could be repeated a second time if participants still had a few remaining seconds. Participants would let the experimenter know when ready and all four runs would then be recorded for Block 1, with short rest periods between them. At the end of Block 1, participants could rest and Block 2 would start as soon as participants felt ready.

### *EEG recordings and pre-processing:*

EEG was recorded with a 32 channel (FP1, FPz, FP2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, A1, T7, C3, Cz, C4, T8, A2, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, POz, O1, Oz, O2, CPz, AFz, 10–20 system) ANT Neuro eego<sup>TM</sup>sports gel headset and two amplifiers, eego<sup>TM</sup>sports or eego<sup>TM</sup>rt. Data was referenced to CPz, grounded to AFz and targeted channel impedance was set at 5kOhm. EEG signals were recorded via OpenVibe [27] and pre-processed with Matlab/EEGLAB [28] and Fieldtrip toolbox [29]. Offline pre-processing started with applying a 1 to 40 Hz band-pass filter and down sampling the initial data to 250 Hz. Files were then merged to end up with one file per condition per participant. At this stage, for each file, a list of bad channels was made with the EEGLAB Clean Rawdata plugin. A bad channel was considered so if i) it was flat for more than 5s, ii) its high frequency noise standard deviation was above 4 and/or iii) its correlation value with nearby channels was higher than 0.8. However, following most recent recommendations [30] bad channels were kept and removed only after Independent Component Analysis (ICA) step. Following bad channel listing, epochs could then be determined as starting 2.5s before the cross on screen appeared and ending 0.5s after it had disappeared. ICA was then applied to the data using the EEGLAB runica algorithm and components were manually rejected according to signs of artifacted activity, caused for instance by blinking, movement or other sources of noise. Per participant, between 0 and 11 components were excluded out of 32 (M = 5.94). Finally, we removed the channels from the previously saved lists, interpolated them and re-referenced the data to average. Following pre-processing, Fieldtrip toolbox was used for time-frequency decomposition using Morlet wavelets (8-35 Hz with 1 Hz steps). Wavelet cycles were

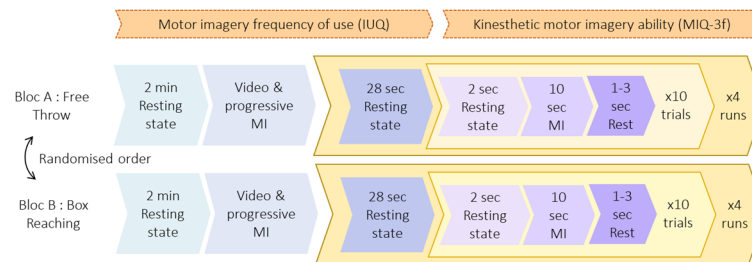


Figure 1: Session's protocol, divided in two blocs, one for each task. A bloc started with a 2min resting state EEG recording. Then, a 3D stick avatar video showed the task to imagine (FreeThrow or Reaching). Participants were asked to execute the task and progressively switch towards KMI. A run could then start with a 28s resting state recording immediately followed by 10 trials. One trial consisted of 2s of resting state, 10s of KMI and 1 to 3s of rest. A bloc was complete when 4 runs were recorded.

increased by 0.1 at each frequency, starting from a width of 3 to 5.7 to ensure a balance between sufficient temporal resolution at lower frequencies and frequency resolution at higher frequencies. As our focus was on SMRs, data from 12 to 15 Hz was then extracted before being normalised. To do so, we measured the relative change from the averaged 10 periods (1s pre-trial) of all trials of a run (1.5–2.5s of the total epoch). The last step consisted in rejecting outliers above or under  $\text{Median} \pm 3 * \text{Median Average Deviation}$  [31].

#### Analysis:

For the behavioural data, two t Tests were used to compare groups' IUQ MI frequency use and MIQ-3f KMI general ability scores. In order to investigate the SMR-ERD power evolution associated to an expertise level and its potential link with general KMI ability, a two-way ANCOVA for repeated measures was performed. Group (Exp-Gp, Nov-Gp) and Task (FreeThrow, Reaching) were used as independent variables, SMR-ERD power as a dependent variable and kinesthetic MIQ-3f score as a co-variable. Effect sizes are reported with a partial eta squared ( $\eta^2 p$ ) for the ANCOVA and with Cohen's d for t Tests. Statistical analyses were computed using Jamovi v.2.4.11.0 [32], a software that implements R statistical language [33].

## RESULTS

Because of a technical issue, two participants had to be excluded from the analyses. Therefore, both groups were composed of 16 participants. We also had one Exp-Gp participant with 10 missing trials out of 40 and another one with 2 missing trials. Considering the low proportion, they were included anyway.

#### MI frequency use:

A Shapiro-Wilk test revealed that no violation of the assumption of normality was made for MI frequency use scores ( $W = 0.969$ ,  $p = 0.464$ ). Thus, we performed a parametric t-test (See Fig. 2) that showed no significant difference between groups concerning the MI frequency of use [ $t(1,16) = 1.18$ ,  $p = 0.246$ ,  $d = 0.418$ ; Exp-Gp ( $M = 3.11/14$ ); Nov-Gp ( $M = 2.55/14$ )].

#### General KMI ability:

For general KMI ability, Shapiro-Wilk test confirmed data was normally distributed ( $W = 0.969$ ,  $p = 0.476$ ). A t Test (See Fig. 2) showed significant difference between groups ( $t(1,16) = 2.09$ ,  $p = 0.045$ ,  $d = 0.739$ ), with Exp-Gp score ( $M = 20.9$ ) being significantly higher than Nov-Gp score ( $M = 17.7$ ).

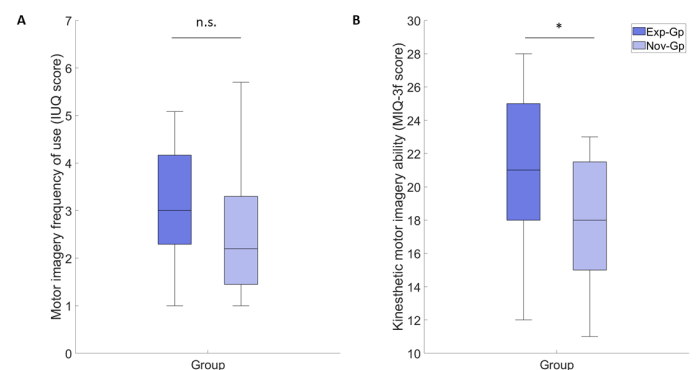


Figure 2: Box plots representing: A. The mean MI frequency of use score as a function of the group (Exp-Gp vs. Nov-Gp) B. The mean general KMI ability score as a function of the group (Exp-Gp vs. Nov-Gp)

#### EEG:

Finally, ANCOVA analyses (See Fig. 3) revealed a main effect of the group [ $F(1,29) = 8.45$ ,  $p = 0.007$ ,  $\eta^2 p = 0.226$ ]; Exp-Gp SMR-ERD change ( $M = -10.48\%$ ); Nov-Gp SMR-ERD change ( $M = 8.45\%$ )] as well as a tendency towards a main effect of KMI ability [ $F(1,29) = 3.03$ ,  $p = 0.092$ ,  $\eta^2 p = 0.095$ ]. However they revealed no main effect of the task [ $F(1,29) = 2.076$ ,  $p = 0.160$ ,  $\eta^2 p = 0.067$ ]; FreeThrow ( $M = -0.215$ ); Reaching ( $M = -1.816$ )] nor any interaction for Group x Task [ $F(1,29) = 0.002$ ,  $p = 0.964$ ,  $\eta^2 p = 0.000$ ]; Exp-Gp Task difference ( $M = 1.44$ ); Nov-Gp Task difference ( $M = 1.75$ )] or Task x KMI ability ( $F(1,29) = 1.857$ ,  $p = 0.183$ ,  $\eta^2 p = 0.060$ ).

## DISCUSSION

The aim of this work was to contribute to the neural efficiency debate by investigating the neurophysiological correlates of expertise during KMI. Our interest was ori-

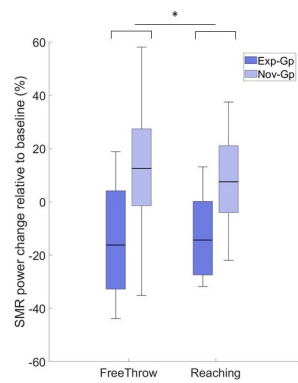


Figure 3: Box plot representing the mean SMR-ERD power as a function of the group (Exp-Gp vs Nov-Gp) and task (FreeThrow vs Reaching)

ented towards knowing if SMR-ERDs evolved with expertise and if this evolution was specific to KMI of mastered tasks. To do so, we compared experts and novices' EEG activities during a task mastered by experts only and a task that both groups mastered. For our experimental design, we chose to observe EEG activities of basketball players experts and basketball novices. The task mastered only by experts was a free throw and the non-specific task was a box reaching action.

Questionnaire results showed that while expert group did not report practicing MI more often than novice group, they self-evaluated their MI abilities higher than novices. In addition, expert group showed significantly stronger SMR-ERDs than novices during KMI whatever the task. As a matter of fact, on average, novices showed an increase of SMR power during KMI as compared to pre-trial baseline although a decrease was expected. Analyses revealed a weak effect of self-reported KMI abilities on that group effect, thus suggesting that the different neurophysiological correlates of KMI are mainly explained by expertise.

Observations of MI frequency use go against well-established findings that suggest MI use is positively linked to athletes' expertise level [25, 34]. The lack of MI use in experts could be due to the fact that half of the participants were competitors at a departmental or regional level. Although their training experience was consequent, these basketball players might not be used to engage in MI as much as higher level amateur or professional players. Indeed, Cumming and Hall, 2002 [34], showed that national athletes perceived imagery to be more relevant to improving their performance and competing effectively than recreational athletes. Therefore, a future inclusion of a third group of high expertise basketball players will allow us to see if MI practice increases with the competitive level. Moreover, to answer IUQ items, novices were asked to evaluate use of MI for all types of motor actions encountered in everyday life and activities (creative activities, skill learning...). On the other hand, basketball players were only asked about their practice of MI to enhance their basketball performance. This potentially could have induced a bias and could have artificially diminished the

discrepancy between groups. Finally, the obtained score covers visual and KMI practice. Future analysis will consider items separately as the other assessed factors of this study focus on KMI only. Indeed, experts could be more familiar with KMI in general, but also in particular during mastered tasks.

Although MI frequency of use was not significantly different between groups, experts happened to have a significantly higher general KMI ability than novices. We can conclude that although practice makes perfect, a high general KMI ability does not seem to be exclusively achieved by having a quantitative MI practice. Hence, basketball expertise seems to allow athletes to develop their general KMI ability through other processes than repetition. Experts could have better MI abilities because of higher sensory mental models, either because they are used to allocating important levels of attention to kinesthetic components during execution and/or have a better ability to memorise and restore them during MI. An important limitation however persists as subjective ease of KMI use may not correlate with quality of KMI.

EEG analyses revealed that experts reached stronger SMR-ERDs than novices, whatever the task. As a matter of fact, it seems that novices increased their SMR power during KMI as compared to baseline on average, while experts decreased it. These results are aligned with many MI-BCI protocols choices to reward a greater ERD. The choice of the baseline, being the second before an MI trial, allowed to counterbalance the signal's non-stationarity. However, we suspect that this portion of recording reflects a pre-KMI state rather than a proper resting state [35]. Kornhuber and Deecke, 1965 [36], refer to this phase as the "readiness potential" and suggest that a surface negative cortical potential happens around 1s prior to movement. Additionally, it is possible that novices initiated KMI too early, which could be explained by a difficulty to voluntarily start and stop MI in an imposed timing. In which case, maximum power decrease would have happened during baseline and would have then be followed by the expected SMR-Event-Related Synchronisation (ERS) [9] during the trial, explaining the positive SMR-ERD power change in novices. It is important to have in mind that SMR-ERD precise modulation patterns during MI are still unknown. Indeed, observations of EEG signal during MI have shown very high variability between individuals [37] but also according to the number of task repetitions [38]. In a previous paper [38], authors suggested that doing MI of a short task once does not result in the same EEG patterns than continuously repeating the MI task during 4 sec. Results showed ERD and ERS components overlap in time when performing MI continuously, meaning ERD could be less detectable and more varied. Our current analysis uses the mean power values of the 10s trials. However, if trial dynamics were to be different than one single ERD per trial, this should be considered. Furthermore, experts could be able to maintain their SMR-ERD through a longer period of time than novices. Like performing MI continu-

ously, ERD and ERS would overlap in novices and explain the SMR power increase compared to baseline. In a near future, we plan on using time-frequency analysis to observe modulations through time within and throughout trials to investigate those dynamics and their relationship with expertise. ANCOVA also provided non-significant results for the Task effect. Our initial choice was to compare two bio-mechanically close tasks. Indeed, choosing two tasks that would have had a different level of complexity as well would not have allowed us to conclude on the basketball specificity's role. Moreover, choosing tasks that mostly implied upper body segments but still involved lower body was a way of assuring us that SMR-ERD wouldn't differ strictly because of the spatially different motor and sensory representations in the Primary Motor and Sensorimotor Cortices [39]. Again, carrying time-frequency analysis throughout trials will be interesting to see if final trials reveal a more important difference between tasks. Furthermore, we found a lack of significant interaction for Task x Group indicating that SMR-ERD difference between tasks was proportionally similar for both groups. This result rejects our hypothesis that SMR-ERD is different between groups, particularly in a mastered task (FreeThrow task for experts). A possible explanation would be that experts benefit from a transfer of competences. The existence of this process has been greatly documented [40] and could be applied to KMI. Indeed, experts could have a facility to do KMI in FreeThrow task that would transfer to Reaching task, illustrated by a negative SMR-ERD change. Whereas novice group would have difficulty to produce lower SMR-ERD whatever the task. Such an interpretation was verified with KMI ability, our ANCOVA's co-variable. Although this factor was not significant, it did not cancel the group effect. We can therefore conclude that there is an influence on SMR-ERD but that globally, this difference is mainly explained by expertise level.

## CONCLUSION

For decades, we have associated an SMR-ERD closely followed by an ERS to MI [41], reflecting the activity of the sensorimotor cortex. We questioned ourselves on the evolution that SMR-ERD could have with expertise and what patterns we should be rewarding when using KMI-BCI to get users to enhance their performance. Currently, different theories exist. The first hypothesis, historically based, stipulates that with expertise, ability to process information increases. This translates into a sensorimotor region activation and an increased recruitment as well as an excitability of cortical neurons [9]. The second one, the neural efficiency hypothesis, goes against it as it suggests that expertise comes with a better cortical and energetic efficiency [19, 21, 22]. This would translate in a decreased activation of pertinent regions. Finally a third hypothesis, suggests that a combination of both these theories could exist [42]. Indeed, in the first stages of learning, we should reward a maximum SMR-ERD and once

expertise level increases, other neurophysiological markers should be identified to reflect the optimisation of resources. Our results suggest that experts have a lower decrease of SMR-ERD compared to novices during KMI of a free throw and reaching action. Future inclusion of a higher expertise group will however be needed to provide more material concerning the mixed hypothesis.

## REFERENCES

- [1] Guillot A, Collet C, Nguyen VA, Malouin F, Richards C, Doyon J. Functional neuroanatomical networks associated with expertise in motor imagery. *Neuroimage*. 2008;41(4):1471–1483.
- [2] Wright CJ, Smith D. The effect of pettiple imagery on strength performance. *International journal of sport and exercise psychology*. 2009;7(1):18–31.
- [3] Reiser M, Büsch D, Munzert J. Strength gains by motor imagery with different ratios of physical to mental practice. *Frontiers in Psychology*. 2011;2:194.
- [4] Laaksonen MS, Ainegren M, Lisspers J. Evidence of improved shooting precision in biathlon after 10 weeks of combined relaxation and specific shooting training. *Cognitive behaviour therapy*. 2011;40(4):237–250.
- [5] Paul M, Ganesan S, Sandhu J, Simon J. Effect of sensory motor rhythm neurofeedback on psychophysiological, electro-encephalographic measures and performance of archery players. *Ibnosina Journal of Medicine and Biomedical Sciences*. 2012;4(02):32–39.
- [6] Neuper C, Wörtz M, Pfurtscheller G. Erd/ers patterns reflecting sensorimotor activation and deactivation. *Progress in brain research*. 2006;159:211–222.
- [7] Hardwick RM, Caspers S, Eickhoff SB, Swinnen SP. Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience & Biobehavioral Reviews*. 2018;94:31–44.
- [8] Abbas ZA, North JS. Good-vs. poor-trial feedback in motor learning: The role of self-efficacy and intrinsic motivation across levels of task difficulty. *Learning and instruction*. 2018;55:105–112.
- [9] Pfurtscheller G. Spatiotemporal erd/ers patterns during voluntary movement and motor imagery. Elsevier, 2000, 196–198.
- [10] Mirifar A, Beckmann J, Ehrlenspiel F. Neurofeedback as supplementary training for optimizing athletes' performance: A systematic review with implications for future research. *Neuroscience & Biobehavioral Reviews*. 2017;75:419–432.
- [11] Xiang MQ, Hou XH, Liao BG, Liao JW, Hu M. The effect of neurofeedback training for sport performance in athletes: A meta-analysis. *Psychology of Sport and Exercise*. 2018;36:114–122.
- [12] Gong A, Gu F, Nan W, Qu Y, Jiang C, Fu Y. A review of neurofeedback training for improving sport performance from the perspective of user experience. *Frontiers in Neuroscience*. 2021;15:638369.
- [13] Horowitz AJ, Guger C, Korostenskaja M. What external variables affect sensorimotor rhythm brain-

computer interface (smr-bci) performance? HCA Healthcare Journal of Medicine. 2021;2(3):143.

[14] Ono T, Kimura A, Ushiba J. Daily training with realistic visual feedback improves reproducibility of event-related desynchronization following hand motor imagery. *Clinical Neurophysiology*. 2013;124(9):1779–1786.

[15] Meulen M Van der, Allali G, Rieger SW, Assal F, Vuilleumier P. The influence of individual motor imagery ability on cerebral recruitment during gait imagery. *Human brain mapping*. 2014;35(2):455–470.

[16] Kraeutner SN, McWhinney SR, Solomon JP, Dithurbide L, Boe SG. Experience modulates motor imagery-based brain activity. *European Journal of Neuroscience*. 2018;47(10):1221–1229.

[17] Haier RJ *et al.* Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence*. 1988;12(2):199–217.

[18] Del Percio C *et al.* Is there a “neural efficiency” in athletes? a high-resolution eeg study. *Neuroimage*. 2008;42(4):1544–1553.

[19] Li L, Smith DM. Neural efficiency in athletes: A systematic review. *Frontiers in Behavioral Neuroscience*. 2021;15:698555.

[20] Del Percio C *et al.* “neural efficiency” of athletes’ brain for upright standing: A high-resolution eeg study. *Brain research bulletin*. 2009;79(3-4):193–200.

[21] Zabielska-Mendyk E, Francuz P, Jaśkiewicz M, Augustynowicz P. The effects of motor expertise on sensorimotor rhythm desynchronization during execution and imagery of sequential movements. *Neuroscience*. 2018;384:101–110.

[22] Wriessnegger SC, Unterhauser K, Bauernfeind G. Limb preference and skill level dependence during the imagery of a whole-body movement: A functional near infrared spectroscopy study. *Frontiers in Human Neuroscience*. 2022;16:900834.

[23] Milton J, Solodkin A, Hluštík P, Small SL. The mind of expert motor performance is cool and focused. *Neuroimage*. 2007;35(2):804–813.

[24] Oldfield RC. The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*. 1971;9(1):97–113.

[25] Hall CR, Rodgers WM, Barr KA. The use of imagery by athletes in selected sports. *The Sport Psychologist*. 1990;4(1):1–10.

[26] Robin N, Coudeville GR, Guillot A, Toussaint L. French translation and validation of the movement imagery questionnaire-third version (miq-3f). *Movement & Sport Sciences*. 2020;(2):23–31.

[27] Renard Y *et al.* Openvibe: An open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. *Presence*. 2010;19(1):35–53.

[28] Delorme A, Makeig S. Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*. 2004;134(1):9–21.

[29] Oostenveld R, Fries P, Maris E, Schoffelen JM. Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational intelligence and neuroscience*. 2011;2011:1–9.

[30] Kim H, Luo J, Chu S, Cannard C, Hoffmann S, Miyakoshi M. Ica’s bug: How ghost ics emerge from effective rank deficiency caused by eeg electrode interpolation and incorrect re-referencing. *Frontiers in Signal Processing*. 2023;3:1064138.

[31] Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*. 2013;49(4):764–766.

[32] Jamovi. The jamovi project: Sydney, Australia. [Online]. Available: <https://www.R-project.org>.

[33] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. [Online]. Available: <https://www.R-project.org>.

[34] Cumming J, Hall C. Deliberate imagery practice: The development of imagery skills in competitive athletes. *Journal of Sports Sciences*. 2002;20(2):137–145.

[35] Rimbart S, Trocellier D, Lotte F. Impact of the baseline temporal selection on the ERD/ERS analysis for Motor Imagery-based BCI. In: EMBC 2023 - 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Sydney, Australia, Jul. 2023.

[36] Kornhuber HH, Deecke L. Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente potentiale. *Pflüger’s Archiv für die gesamte Physiologie des Menschen und der Tiere*. 1965;284:1–17.

[37] Wriessnegger SC, Müller-Putz GR, Brunner C, Sburlea AI. Inter-and intra-individual variability in brain oscillations during sports motor imagery. *Frontiers in human neuroscience*. 2020;14:576241.

[38] Rimbart S, Lindig-León C, Fedotenkova M, Bougrain L. Modulation of beta power in eeg during discrete and continuous motor imageries. In: 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER). 2017, 333–336.

[39] Penfield W, Boldrey E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*. 1937;60(4):389–443.

[40] Oppici L, Panchuk D. Specific and general transfer of perceptual-motor skills and learning between sports: A systematic review. *Psychology of Sport and Exercise*. 2022;59:102118.

[41] Pfurtscheller G, Lopes da Silva F. *Handbook of Electroencephalography and Clinical Neurophysiology – Event-related desynchronization*. Elsevier, Amsterdam, Netherlands (1999).

[42] Zhuang P, Toro C, Grafman J, Manganotti P, Leocani L, Hallett M. Event-related desynchronization (erd) in the alpha frequency during development of implicit and explicit learning. *Electroencephalography and clinical neurophysiology*. 1997;102(4):374–381.

# BAYESIAN MODEL OF INDIVIDUAL LEARNING TO CONTROL A MOTOR IMAGERY BCI

C. Annicchiarico<sup>1,2,3</sup>, F. Lotte<sup>2</sup>, J. Mattout<sup>1,3</sup>

<sup>1</sup> Lyon Neuroscience Research Center, CRNL, INSERM U1028, CNRS UMR5292, Computation, Cognition and Neurophysiology Team, Lyon, France

<sup>2</sup> Inria Center at the University of Bordeaux / LaBRI, France

<sup>3</sup> Université Claude Bernard Lyon 1, Lyon, France

E-mail: [come.annicchiarico@inserm.fr](mailto:come.annicchiarico@inserm.fr)

**ABSTRACT:** The cognitive mechanisms underlying subjects' self-regulation in Brain-Computer Interface (BCI) and neurofeedback (NF) training remain poorly understood. Yet, a mechanistic computational model of each individual learning trajectory is required to improve the reliability of BCI applications. The few existing attempts mostly rely on model-free (reinforcement learning) approaches. Hence, they cannot capture the strategy developed by each subject and neither finely predict their learning curve. In this study, we propose an alternative, model-based approach rooted in cognitive skill learning within the Active Inference framework. We show how BCI training may be framed as an inference problem under high uncertainties. We illustrate the proposed approach on a previously published synthetic Motor Imagery ERD laterality training. We show how simple changes in model parameters allow us to qualitatively match experimental results and account for various subject. In the near future, this approach may provide a powerful computational to model individual skill learning and thus optimize and finely characterize BCI training.

## INTRODUCTION

Motor Imagery is one of the most employed non-invasive BCI paradigm due to its potential in stroke rehabilitation and motor control. Event-related desynchronization (ERD) in motor cortices is associated with motor task execution, observation or mental imagery. It is a key biomarker to pick up to interface the brain with an assistive (e.g. neuroprosthetics) or a rehabilitation (e.g. neurofeedback) device. Studies focusing on MI training have demonstrated notable positive outcomes, including enhanced hand dexterity [1] and post-stroke improvements [2], [3]. These interventions capitalize on the overlapping neural pathways between mental imagery and motor execution. Particularly in the context of hemispheric ischemic stroke, some studies have attempted to address motor control deficits [4] by using neurofeedback training to strengthen MI laterality, with some success [5].

Despite those results, the core neuropsychological mechanisms behind subject self-regulation are still

poorly understood. Some theoretical approaches [6], [7] have proposed unifying frameworks to describe such processes during BCI or NF training. Among those processes, the nature of subject learning has been the main focus of academic debate [8]. Two views mostly prevail and are in relative opposition. Proponents of operant conditioning reflect a model-free (reinforcement learning) view on how subjects learn during BCI training [9]. A different view that also assume that subjects learn from trial and error, supporters of cognitive skill learning [10], [11], [12] suggest that subject actively build an interaction model of the BCI system in order to reliably interact with it. According to this second view, users learn a skill ("interacting with the BCI") in order to control the interface despite the high levels of uncertainty of the paradigm. This form of learning, more akin to "model-based" reinforcement learning (RL), provides more satisfactory explanations for phenomena such as transfer learning and the effect of metacognition on regulation [7]. The true nature of subject experience during BCI training probably stands between these two views on adaptation, with initial interactions generally driven by RL and progressively building a more complete model-based representation of the system.

The general lack of understanding of the self-regulation mechanisms at play during successful and failed training procedures has prompted the scientific community towards the development of models of subject learning in order to explain and hopefully predict the outcome of BCI training given a particular subject, experimental design, etc. These models have built on the above-described R.L. perspective to leverage difficult credit assignment problems as when learning individual neuron activations under high uncertainty [13], [14]. We argue that in order to model the cognitive dynamics of training and account for its metacognitive and transfer learning dimensions, an explicit modelling of the subject's representations is needed. To our knowledge, such an approach to BCI has barely been tackled. In this work, we show how the Active Inference framework [15] may be leveraged to provide an adequate theoretical and computational ground for developing this modelling strategy. To illustrate our modelling approach, we consider a rich and original study that has implemented a

information. Importantly, BCI training can be nicely framed with such an POMDP.

Table 1 provides the description of the above model component and parameters in the context of BCI training. Given this formulation, the subject’s Free Energy can be minimized in three ways: through perception (inference on hidden states), action (transition between hidden states) and learning (updating model parameters). Under this premise, agents may pick actions in order to reduce their (expected) free energy on the basis of anticipated future observations, in a way that optimize a trade-off between information seeking (exploration) and reward maximization (exploitation). In this paper, agents plan their future actions by comparing all the plausible action trajectories within a specific temporal horizon [20]. Finally, learning occurs at a slower pace at which subjects update their model parameters. In the discrete state space leveraged by Active Inference, these model parameters are categorical distributions equipped with conjugate Dirichlet priors. Learning occurs through *counting* co-occurrences between state posteriors and observations (likelihood **a**), or transitions between states following a given action (transition **b**), akin an evidence accumulation process [15]. In essence, Active Inference describes the evolution of subject’s beliefs ( $x, \pi$ ) and representations (**a, b, c, d, e**) depending on environmental parameters (**A, B, D**). This translates directly to BCI training where we may cast the feedback provided to the subject as the observations, and the mental states targeted by the training procedure (attention, hand motor imagery level, ...) as the true hidden states. The subject tries to reach high levels of positive feedback by learning an accurate representation of the BCI system (**a, b, d**).

Table 1: Correspondence between Active Inference graph parameters and BCI training elements

Active Inference parameter	BCI training element
<b>A</b>	Emission rule: relation between feedback and subject's true mental state
<b>B, D</b>	Transition rule: effect of mental action onto mental states
<b>a</b>	Subject's belief about the feedback (affected by instructions, experience...)
<b>b, d</b>	Subject's belief about its mental strategies and the effect of its mental actions (idem)
<b>c, e</b>	Subject preferences (towards positive feedback) and habits
<b><math>\hat{s}, s</math></b>	True and belief about mental states, respectively
<b>o</b>	Observations (feedback)
<b><math>\pi, u</math></b>	Subject's mental policy and possible actions

The diagram illustrates a partially observable Markov decision process (POMDP) model, showing the interaction between a **Subject model** and an **Environment (process)**.

**Subject model (Internal State and Actions):**

- Policy and mental actions:** The subject model maintains a belief state  $\pi$  (represented by a blue circle). It receives a current belief  $b_t$  and a previous action  $a_{t-1}$  to produce a new action  $a_t$  (represented by a blue circle).
- Belief about mental state:** The belief state  $\pi$  is updated based on the previous belief  $b_{t-1}$  and the current action  $a_t$  to produce the next belief  $b_{t+1}$  (represented by a grey circle).

**Environment (process):**

- True environment rules (learned):** The environment maintains a hidden state  $s_t$  (represented by a grey circle). It receives a current state  $s_{t-1}$  and a current action  $a_t$  to produce the next state  $s_{t+1}$  (represented by a grey circle).
- Observations (feedback / interoceptive signals):** The environment produces an observation  $o_t$  (represented by a blue circle) based on the current hidden state  $s_t$  and the current action  $a_t$ .
- True mental states (hidden):** The hidden state  $s_t$  is updated based on the previous hidden state  $s_{t-1}$  and the current action  $a_t$  to produce the next hidden state  $s_{t+1}$  (represented by a grey circle).

**Interactions:**

- The subject model's action  $a_t$  is sent to the environment's hidden state  $s_t$ .
- The environment's observation  $o_t$  is sent back to the subject model's belief state  $\pi$ .

**Legend:**

- Subject representation (learned):**
  - $a$ : Likelihood model
  - $b, d$ : Transition model
- True environment rules (learned):**
  - $A$ : Emission rule
  - $B, D$ : Transition rule

*A Motor Imagery Neurofeedback training task:* To illustrate our modeling approach, we consider a simplified version of the task implemented by Perronet, Lioi et al. [16], [17]. In their first experiment, (N=10)

subjects were instructed to perform kinesthetic right hand motor imagery and to “find their own strategy” in order to control a feedback gauge across 3 x 10 blocks. Each block comprised a 20s rest and a 20s task block. The task was multimodal as both fMRI and EEG data were recorded and the feedback was based on either EEG alone, fMRI alone or both signals (two feedback gauges simultaneously). Importantly, the gauge levels were always based on a measure of lateral asymmetry between left and right motor cortex activities (For EEG: an asymmetry index computed on the normalized difference in  $\mu$  (8-12 Hz) band power between C3 and C4, updated every 250 ms; for fMRI: a laterality index as described in [21], updated every 2 s).

This study is quite unusual, namely because of the two neuroimaging modalities employed. However, it offers an appealing example to model, for at least two reasons. First, the well-defined laterality biomarker permits fairly simple assumptions regarding the subject's self-regulatory process. Second, data availability [17] allows for broad model calibration. In what follows, we propose a computational model of this protocol and provide general predictions regarding long-term training outcomes.

**Modeling Motor Imagery laterality training:** The Motor Imagery neurofeedback loop is formalized as a high uncertainty self-regulation task. The agent trains over  $N_{trials}$ , each trial being composed of an arbitrary 40 rest and 40 MI timesteps (each timestep corresponding to 2 EEG feedback update for the experimental task). During MI, the agent is given a feedback based on its hidden states and attempts to reach highly rewarding outcomes. No feedback is provided during rest. During the whole training, the agent activity was defined by two hidden states based on electrophysiology : the left and right ERD levels.

$$\begin{cases} ERD_L(t) = \hat{i}(t) \cos(\hat{\alpha}(t)) + \epsilon \\ ERD_R(t) = \hat{i}(t) \sin(\hat{\alpha}(t)) + \epsilon \end{cases} \quad \text{Eq.2}$$

Where the radius  $\hat{i}(t) \in [0; 1]$  captures the global ERD strength and the angle  $\hat{\alpha}(t) \in [0; \frac{\pi}{2}]$  its lateralization or orientation.  $\epsilon$  is a baseline level accounting for spontaneous desynchronizations outside of MI, which we set to 0.01 (weak baseline level).

Agents have no direct observation of these two physiological states that reflect cortical motor excitability and could be associated with mental states such as motor preparation and sensorimotor expectation. In this framework, agents entertain a belief or prior over these states ( $i; \alpha$ ) and use the feedback provided (see *Emissions*) to update this belief.

We further adopt a discretized, POMDP compatible formulation of our model, considering that  $\hat{i}, \hat{\alpha}$  (process states) and  $i, \alpha$  (model states) can each span a finite set of  $N_s$  possible states. For the sake of simplicity, the simulations were conducted with  $N_s(i) = N_s(\alpha) = 4$  {0: null, 1: low, 2: medium, 3: high ERD strength} and  $N_s(\hat{\alpha}) = N_s(\alpha) = 5$  {L: left (0), CL: center-left ( $\frac{\pi}{8}$ ), C: center ( $\frac{\pi}{4}$ ), CR: center-right ( $\frac{3\pi}{8}$ ), R: right ( $\frac{\pi}{2}$ ) ERD orientation}. Note that discrepancies between the model

and the process in terms of state space dimensions could be accounted for and their effect on training simulated within this framework.

**Emissions:** Agents receive outcomes  $o_t$  based on their true state  $\hat{s}_t = (\hat{i}(t), \hat{\alpha}(t))$ . This feedback modality (denoted as *AsI*) is based on the laterality of the ERD. It is computed using an asymmetry index between the left and right ERDs, for  $\hat{i}(t) > 0$  : (Eq.3)

$$\hat{o}_t = AsI(\hat{\alpha}) = \frac{ERD_L(t) - ERD_R(t)}{ERD_L(t) + ERD_R(t)} \in [-1, 1]$$

To account for the noise in the biomarker and feature extraction process, the categorical emission matrix **A** encodes the emission rule of the BCI pipeline as a discretized gaussian distribution  $Cat(N(\hat{o}_t; \sigma_{proc}))$  with  $N_{AsI} = 5$  possible feedback values.

During the experimental task [16], the strength of left ERD was continuously monitored but was not provided as a feedback signal. We mimic this observation channel with a second emission modality (referred to as *L-ERD*) based on the simulated left ERD level. Similarly, these outcomes are not observed by the synthetic subject during training. They are used to compare physiological measurements to model predictions and broadly estimate which parameter values best matched the study results (see *Results*). Just like with the *AsI* modality, the L-ERD observations are noisy (same noise parameter  $\sigma_{proc}$ ) and discretized so as to take one out of 5 possible values.

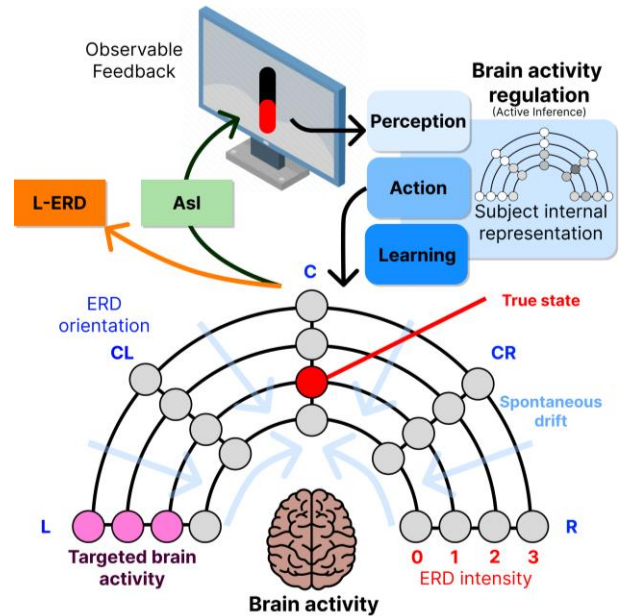


Figure 2: The modeled Motor Imagery intensity / orientation training. The agent internal representation drives the brain activity based on the feedback provided.

**Transitions:** At each timestep  $t$ , the subject's true states evolve depending on previous state value  $\hat{s}_{t-1}$  and mental actions  $u_{t-1}$ . During the actual task, agents could potentially explore and use a large number of mental strategies (attentional / sensorial exercises, relaxation efforts, etc.), among which only a limited amount would prove “effective” and allow the subject to control their mental state with a probability  $p_{effect} = 0.99$ . Since

mental actions are poorly understood, we assumed a synthetic topological state space that satisfies the three as following constraints:

- *Continuity*: for a specific state factor ( $i / \alpha$ ), the mental states could only move from value  $k$  to adjacent (or same) values  $\{k-1, k, k+1\}$ .
- *Invariability*: for a specific state factor, the effect of “effective” actions was independent from the occupied state.
- *Resting states*: to reflect the natural tendency of Motor imagery intensity to return to a resting state, non “effective” actions pulled the mental state of the subject towards this resting state with probability  $p_{decay} = 0.1$ . The resting states were  $\hat{t} = 0$  for MI intensity and  $\hat{a} = C$  (center) for MI orientation.

For each state factor ( $i / \alpha$ ),  $N_{up} = N_{down} = 1$  action were “effective” and allowed the subject to control their mental states.  $N_{neutral} = 10$  actions were non “effective” and resulted in spontaneous drift towards the resting state.

*Subject priors*: the agents entertained representational priors about the BCI loop before starting the training. This included belief about the feedback modality (also called the ‘likelihood model’ **a**) and beliefs about the effect of their mental actions (**b**).

We assumed biased agents. We model them as having the expectation that high levels of motor imagery intensity would lead to higher feedback levels. This is in fact misleading but fits with the initial instructions they actually received in this experiment: “to perform Motor Imagery”. This assumption is supported with further arguments in the discussion. The agents’ model of the feedback was initiated using the Dirichlet conjugate prior for the categorical likelihood **a**:

$$\mathbf{a}_0 = c_a \mathbf{1} + s_a \text{Cat}(N(i, \text{AsI}(\alpha); \sigma_{model})) \quad \text{Eq.4}$$

Where  $c_a$  and  $s_a$  are the initial concentration and confidence parameters, which we set to 1 and 100, respectively. This means that subjects were very confident that the feedback actually reflects (albeit with some noise) their mental imagery level. *AsI* is the asymmetry index previously formulated and  $\sigma_{model}$  is a noise term encoding subject’s prior confidence in the feedback modality. It was set to 0.5.

Finally, subject prior beliefs about their mental actions were set as the combination of three terms: a prior concentration parameter  $c_b$  indicating how much new evidence is needed for the subjects to change their prior beliefs, a ‘stickiness’ parameter  $s_b$  that encodes subject’s belief about actions not affecting their mental state, and an initial mental action confidence vector  $b_{pre}$  that encodes previous knowledge about the effect of their mental actions. Importantly,  $b_{pre}$  is a vector with one value for each state factor ( $i / \alpha$ ). The initial mental action model of the agents was thus, for each state factor:

$$\mathbf{b}_0 = c_b \mathbf{1} + s_b \mathbf{Id} + b_{pre} \mathbf{B} \quad \text{Eq.5}$$

With  $\mathbf{Id}$  the identity matrix. Simulations were conducted with  $c_b = 1.0$  and  $s_b = 1.0$  (i.e. subjects were opened to new evidence regarding their mental strategies). Of course, subjects started the training with relatively low

values of  $b_{pre}$ , as high values of the parameter would render the training useless (this would mean the subject was already knowing how to perform the task optimally).

*Goals & simulations*: Using this simple model of self-regulation, our goal was to predict training outcome depending on the individual priors of each subject. Therefore, several families of agents were instantiated with various initial mental imagery familiarity levels. We demonstrate how these priors affect the way subjects learn how to perform the task and the evolution of the overall quality of their mental imagery models. To that end, we conducted simulations of agents performing Active Inference using the parametrized graph parameters **a0**, **b0**, **A**, **B**. The process parameters used in these simulations are  $\sigma_{process}$ ,  $b_{pre}(i)$  and  $b_{pre}(\alpha)$ .

All simulations in this paper were conducted using *active pyinfer*, a freely available Python package for running sophisticated inference schemes. The code used in these simulations is freely available at: [https://github.com/Erresthor/ActivPynference\\_Public/blob/main/paper\\_scripts/paper\\_grazBCI/simulations.ipynb](https://github.com/Erresthor/ActivPynference_Public/blob/main/paper_scripts/paper_grazBCI/simulations.ipynb).

## RESULTS

*Agents already familiar with MI*: Figure 3 illustrates the outcome of 10 simulated agents performing 10 trials each, starting with informed action priors  $b_{pre}(i) = 1$  and  $b_{pre}(\alpha) = 1$ . These subjects thus started the training with high mental imagery control skills, rendering the training unnecessary. The feedback provided was noisy, but informative ( $\sigma_{process} = 1.5$ ). The average simulated mental states (true ERD intensity and orientation) are shown as well as the provided feedback (green). These can be compared to the corresponding performances of neurofeedback subjects from [16] shown below for a few subjects (Figure 3.A). The quite large mismatch between the empirical and simulated time series suggest that subjects entertained less precise action priors. Interestingly, the agents quickly learned to maintain a weak ERD strength while correctly lateralizing their ERDs, leading to less effortful, more optimal behavior.

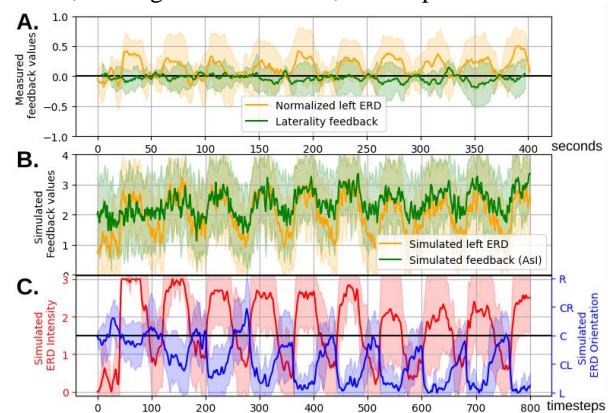


Figure 3: First empirical trials from [16] (A.) compared with the simulated laterality feedback and left ERD (B.) and motor imagery states (C.) from 10 simulated agents with high initial motor imagery control.

*Agents initially unable to perform MI lateralization:* Another class of agents was instantiated who were initially unable to produce lateralized motor imagery. They had no priors on how to control the orientation of their ERD ( $b_{pre}(\alpha) = 0.0$ ), but had some poor priors on how to control their intensity ( $b_{pre}(i) = 0.1$ ). They thus had to fully rely on the feedback to learn these transitions. To facilitate their training, a fairly reliable biomarker was assumed ( $\sigma_{proc} = 0.5$ ). The training results are shown in Figure 4. Overall, agents managed to reliably produce a lateralized ERD, although after quite a long training.

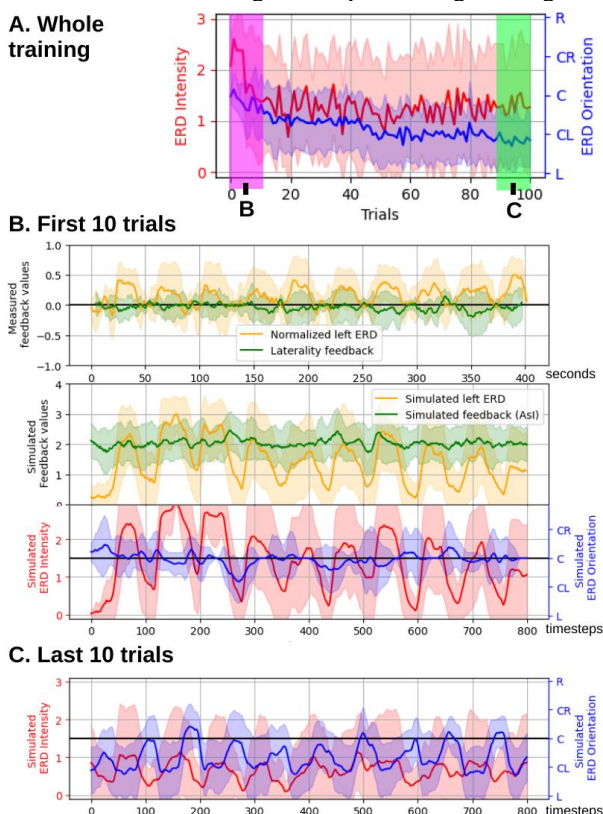


Figure 4: Agents with no prior knowledge of ERD lateralization performed 100 simulated neurofeedback trials. We show their average ERD strength (red) and orientation (blue) across the training (A) and at specific points of the training (B, C). The similarity between the simulated initial MI levels and the empirical observations (B) suggests that this set of parameters better matches the data than the over-optimistic previous simulations.

*Agents with mixed prior abilities:* Finally, 21 x 21 group of 10 agents with intermediate MI lateralization priors were simulated. Each group had a different pair of parameter values  $\{b_{pre}(i), b_{pre}(\alpha)\}$ , set between 0 and 2. This reflected individual differences in subjects starting BCI training with different Motor Imagery prior experience. The feedback provided was very noisy ( $\sigma_{proc} = 1.5$ ). Figure 5 shows the evolution of average Motor Imagery performance in each group of subjects, at the start of training and at the end. Our simulations reveal counter-intuitive training effects, such as poor training

results from subjects initially well versed in their ability to lateralize their ERDs but lacking the ability to reliably perform an ERD (e.g. subjects who misinterpret Motor Imagery by performing right hand visual instead of kinesthetic motor imagery). Conversely, subjects who were very good at performing mental imagery but lacked control over their MI laterality tended to benefit from training and managed to learn how to direct their attention, despite the noisy feedback.

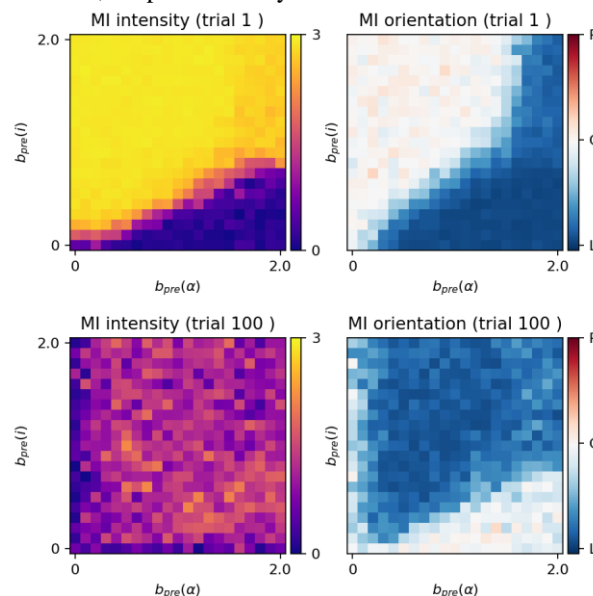


Figure 5: Simulated motor imagery performance before (top) and after (bottom) neurofeedback depending on initial experience  $b_{pre}(i)$  (y-axis) and  $b_{pre}(\alpha)$  (x-axis).

## DISCUSSION

The reported simulations provided an account of Motor Imagery training using Neurofeedback for various groups of subjects parametrized mostly by their past experience with Motor Imagery: (i) subjects familiar with motor imagery and who had good initial priors, (ii) subjects with poor initial ability to lateralize their ERD and had to learn from scratch, and (iii) intermediate subjects who started with mixed priors about MI laterality and strength, but had to finetune them in order to perform the task efficiently.

Simulations showcased very different training curves and general subject classes that would more or less benefit from the training depending on their initial situation. They illustrated the crucial role of subject's prior skills (i.e. previous experience), expectations about the feedback, training and beliefs following task instructions. Subjects starting training with uninformed priors performed poorly. This was in part due to the sparse feedback modality (low temporal resolution / low dimensionality) which made learning from scratch a very tricky task. This suggests that reducing the amount of targeted mental dimensions may be instrumental to guarantee successful training [6]. The lackluster ability of the subjects when they had to build a model of interaction from scratch also suggests that more basic

learning mechanisms such as classical (model free) Reinforcement Learning may play a significant role in the initial phase of the training, with a more complex representational learning taking over later on [13]. The proposed framework is very general and flexible enough to capture a large variety of experimental paradigms. For instance, the multidimensional feedback based learning implemented in [16] may be modelled by agents learning simultaneously several sensory mappings of the same internal dynamical state.

## CONCLUSION

This paper presents a computational account of neurofeedback/BCI Motor Imagery training using the Active Inference framework. Preliminary simulations reveal that the Active Inference framework has great potential to provide an account of individual self-regulation dynamics. Future work will consist in fitting alternative instantiations of such models to actual data in order to disentangle between learning profiles and identify individual traits for BCI learning curves and empirically observed neurophysiological dynamics.

## REFERENCES

- [1] Y. Ota *et al.*, “Motor Imagery Training With Neurofeedback From the Frontal Pole Facilitated Sensorimotor Cortical Activity and Improved Hand Dexterity,” *Front. Neurosci.*, vol. 14, p. 34, 2020,
- [2] A. Kruse, Z. Suica, J. Taeymans, and C. Schuster-Amft, “Effect of brain-computer interface training based on non-invasive electroencephalography using motor imagery on functional recovery after stroke - a systematic review and meta-analysis,” *BMC Neurol.*, vol. 20, no. 1, p. 385, 2020,
- [3] M. Mihara *et al.*, “Near-infrared Spectroscopy-mediated Neurofeedback Enhances Efficacy of Motor Imagery-based Training in Poststroke Victims: A Pilot Study,” *Stroke*, vol. 44, no. 4, pp. 1091–1098, 2013,
- [4] J. Yan, X. Guo, Z. Jin, J. Sun, L. Shen, and S. Tong, “Cognitive Alterations in Motor Imagery Process after Left Hemispheric Ischemic Stroke,” *PLoS ONE*, vol. 7, no. 8, p. e42922, 2012,
- [5] S. Boe, A. Gionfriddo, S. Kraeutner, A. Tremblay, G. Little, and T. Bardouille, “Laterality of brain activity during motor imagery is modulated by the provision of source level neurofeedback,” *NeuroImage*, vol. 101, pp. 159–167, 2014,
- [6] A. Gaume, A. Vialatte, A. Mora-Sánchez, C. Ramdani, and F. B. Vialatte, “A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback,” *Neurosci. Biobehav. Rev.*, vol. 68, pp. 891–910, 2016,
- [7] U. Strehl, “What learning theories can teach us in designing neurofeedback treatments,” *Front. Hum. Neurosci.*, vol. 8, 2014,
- [8] L. H. Sherlin *et al.*, “Neurofeedback and Basic Learning Theory: Implications for Research and Practice,” *J. Neurother.*, vol. 15, no. 4, pp. 292–304, 2011,
- [9] N. Lubianiker, C. Paret, P. Dayan, and T. Hendler, “Neurofeedback through the lens of reinforcement learning,” *Trends Neurosci.*, vol. 45, no. 8, pp. 579–593, 2022,
- [10] A. V. P. Veilahti, L. Kovarskis, and B. U. Cowley, “Neurofeedback Learning Is Skill Acquisition but Does Not Guarantee Treatment Benefit: Continuous-Time Analysis of Learning-Curves From a Clinical Trial for ADHD,” *Front. Hum. Neurosci.*, vol. 15, p. 668780, 2021,
- [11] F. Lotte, F. Larrue, and C. Mühl, “Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design,” *Front. Hum. Neurosci.*, vol. 7, 2013,
- [12] D. J. McFarland and J. R. Wolpaw, “Brain-computer interface use is a skill that user and system acquire together,” *PLOS Biol.*, vol. 16, no. 7, p. e2006719, 2018,
- [13] E. J. Davelaar, “Mechanisms of Neurofeedback: A Computation-theoretic Approach,” *Neuroscience*, vol. 378, pp. 175–188, 2018,
- [14] E. F. Oblak, J. A. Lewis-Peacock, and J. S. Sulzer, “Self-regulation strategy, feedback timing and hemodynamic properties modulate learning in a simulated fMRI neurofeedback environment,” *PLOS Comput. Biol.*, vol. 13, no. 7, p. e1005681, 2017,
- [15] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, J. O’Doherty, and G. Pezzulo, “Active inference and learning,” *Neurosci. Biobehav. Rev.*, vol. 68, pp. 862–879, 2016,
- [16] L. Perronnet *et al.*, “Unimodal Versus Bimodal EEG-fMRI Neurofeedback of a Motor Imagery Task,” *Front. Hum. Neurosci.*, vol. 11, p. 193, 2017,
- [17] G. Lioi *et al.*, “Simultaneous MRI-EEG during a motor imagery neurofeedback task: an open access brain imaging dataset for multi-modal data integration,” *Neuroscience*, preprint, 2019.
- [18] K. Friston, “The free-energy principle: a unified brain theory?,” *Nat. Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, 2010,
- [19] K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny, “Variational free energy and the Laplace approximation,” *NeuroImage*, vol. 34, no. 1, pp. 220–234, 2007,
- [20] K. Friston, L. Da Costa, D. Hafner, C. Hesp, and T. Parr, “Sophisticated Inference,” 2020,
- [21] M. Chiew, S. M. LaConte, and S. J. Graham, “Investigation of fMRI neurofeedback of differential primary motor cortex activity using kinesthetic motor imagery,” *NeuroImage*, vol. 61, no. 1, pp. 21–31, 2012,

# USING TRANSFORMER NETWORKS FOR STREAMING SPEECH SYNTHESIS FROM INTRACRANIAL EEG

Joaquín Amigó-Vega<sup>1,2 †</sup>, Maxime Verwoert<sup>2</sup>, Maarten C. Ottenhoff<sup>2</sup>, Pieter L. Kubben<sup>2</sup>,  
Christian Herff<sup>2</sup>

<sup>1</sup>Gran Sasso Science Institute, Computer Science Department, L'Aquila, Italy

<sup>2</sup>Neural Interfacing Lab, Department of Neurosurgery, Mental Health and Neuroscience Research  
Institute, Maastricht University, Maastricht, the Netherlands

† joaquin.amigo@gssi.it

**ABSTRACT:** Speech Neuroprostheses have the potential to enable users to communicate without the need for overt muscle movement. Several recent approaches have demonstrated the feasibility of decoding textual and acoustic representations of speech from invasively measured neural activity. However, most approaches decode or synthesize speech after several seconds or complete utterances. While this provides tremendous communicative ability to patients, it lacks the full expressive power of natural conversations. Ideally, a speech neuroprosthesis would synthesize speech without a noticeable delay.

Here, we present a real-time speech decoding pipeline that generates speech output in a streaming fashion, i.e., with delays of less than 40 ms. Intracranial EEG data is measured, processed, decoded, and synthesized into an audio waveform using our fast and modular framework. Notably, we employ a Transformer architecture for the decoding step from neural features to a spectral representation of speech.

## INTRODUCTION

Speech plays an important role in human interaction, serving as a primary means of conveying thoughts and emotions. It is integral to the fabric of our social existence and personal identity. However, various conditions, such as Amyotrophic Lateral Sclerosis (ALS) and locked-in syndrome, can impair one's ability to speak, significantly impacting the quality of life. These diseases may leave cognitive functions intact, while debilitating the muscular activity required for speech production.

Speech Brain-Computer Interfaces (BCIs), also called speech neuroprostheses, are a groundbreaking technology designed to help people in need. By harnessing neural signals through invasive or noninvasive methods, these BCIs decode speech-associated brain activity. This process involves extracting and translating neural patterns related to speech formation into actionable outputs, thereby enabling communication or device control.

The ultimate goal of a speech BCI is to facilitate seamless, naturalistic conversation, akin to normal speech. Achieving this requires real-time processing of neural

signals, a technical challenge that remains at the forefront of current research. Despite ongoing advancements, many existing speech BCI systems rely on offline evaluations, where signal analysis and method validation occur after data collection [1–5]. While recent studies have made strides toward closed-loop systems capable of generating textual representations [6–8] or synthesized sentences [9], these technologies typically operate with delays, processing complete sentences or phrases [10] before producing output. For completely natural communication, the patient needs to produce speech output immediately to ensure natural flow, e.g. to interrupt the conversational partner or to modulate their own speech.

This paper introduces a novel real-time streaming synthesis pipeline for speech BCIs, distinguished by its low latency and modular framework. Developed in Python and based on the framework Timeflux [11], our pipeline processes and decodes neural data into a speech waveform with less than 40 ms of delay. Longer delays have been found to severely impair speech production [12]. Notably, our system employs a transformer encoder to translate sequences of neural data into speech spectral sequences. The attention mechanism [13] in transformers is particularly well suited for learning the temporal dynamics in the neural and speech data and has successfully been used on offline data before [14].

To validate our streaming speech BCI, we conducted simulated online studies using a previously recorded dataset of intracranial EEG during speech production [15].

## MATERIALS AND METHODS

### *Participants:*

Our closed-loop experiments are conducted with voluntary participants implanted with sEEG electrodes as part of the clinical therapy for their pharmaco-resistant epilepsy. All participants gave written informed consent before joining the study, and the electrode locations were purely determined based on clinical necessity. All participants were Dutch native speakers and had normal speech, hearing, and language functions.

For this simulated online evaluation, we employ our pre-

viously published open-access Single Word Production Dutch-iBIDS (SWPD) dataset [15], consisting of 10 participants speaking 100 words each.

#### *Data recording:*

Patients were implanted with platinum-iridium sEEG electrode shafts (Microdeep intracerebral electrodes; Dixi Medical, Beçanson, France) with a diameter of 0.8 mm, a contact length of 2 mm and an inter-contact distance of 1.5 mm with each shaft containing between 5 and 18 contacts. Neural data was recorded using two or more Micromed SD LTM amplifier(s) (Micromed S.p.A., Treviso, Italy) with 64 channels each. Electrode contacts were referenced to a common white matter contact. Data was recorded at either 1024 Hz or 2048 Hz.

#### *Simulated Online Experiment:*

To assess the real-time capabilities of our closed-loop speech decoding pipeline, we conducted simulated online experiments using the SWPD dataset [15]. The dataset's recording environment mirrors the anticipated operational scenario for our pipeline, making it an ideal choice for our evaluation process. As part of the assessment, we divided the data from each participant into training and testing sets, allocating 75% for model training and the remaining 25% for testing. After training the pipeline with the designated data, we streamed the testing dataset through LabStreamingLayer (LSL), emulating the amplifier characteristics used in our real setup. This approach was designed to closely replicate the dynamics of real neural signal acquisition and processing, thereby providing a realistic approximation of how the pipeline would perform in live application scenarios.

#### *Closed-loop pipeline:*

##### *Pipeline Design and Requirements*

In the initial phase of constructing our speech Brain-Computer Interface (BCI) pipeline, we focused on identifying key requirements to ensure its effectiveness for real-time communication. Among our primary objectives were ensuring *real-time decoding*, *rapid model training*, and a high degree of *modularity and configurability*. Real-time decoding is crucial as the pipeline must process neural signal samples swiftly to minimize latency, thereby enabling near-instantaneous speech synthesis. Given the constraints of on-site training, it was imperative that the machine learning models employed could be trained quickly to avoid reducing valuable data collection time with participants. Additionally, to facilitate rapid experimentation and adaptation of new approaches, the system architecture needed to be both modular and easily configurable.

##### *Framework and Technology Selection*

We used Python and the framework Timeflux [11] for building the pipeline. Timeflux facilitates the creation of applications as directed acyclic graphs (DAGs), where processing nodes are interconnected through YAML syntax, enabling efficient data flow and simultaneous processing. For communication between graphs we used ZeroMQ, an asynchronous messaging library, ensuring robust data exchange without interrupting the execution.

##### *Pipeline Architecture*

The pipeline involves two main stages: *Initialization* and *Real-time decoding*. Each stage consists of a series of graphs and nodes executed concurrently, optimizing data processing speed.

**Initialization Stage:** This stage prepares the system for the online decoding. It uses the open-loop recorded experiment data, which includes synchronized neural signals, audio, and markers, to extract the relevant parameters and train the machine learning models.

First, we segregate the data into distinct datasets labeled “neural” and “audio” and adjust their format, length, and type.

Irrelevant channels, such as clinical markers and heart-rate, are eliminated from the “neural” dataset, and the power line noise and its first harmonic are filtered out, using causal IIR bandstop-filters. Afterwards, the signal is extracted in a broadband high-frequency range (70–170 Hz) and windowed, subsequently calculating the log power for each window. At the same time, another graph extracts the “audio” features by decimating the audio signal to lower its sampling rate, then using a sliding window with the same window size and frameshift as the neural data to extract a mel-scaled spectrogram, aligned to the neural features.

The size of the window used for the audio and neural data can be different. However, the shift needs to always be the same, allowing seamless alignment between the features.

Both feature sets are aligned and scaled before being used to train the Machine Learning model. When the training is completed, we save the model parameters and additional helpful metadata for the decoding stage.

**Real-Time Decoding Stage:** This stage, presented in Fig. 1, is responsible for the on-the-fly decoding of neural signals into audible speech. It encompasses system initialization, data intake, feature extraction, neural decoding, audio reconstruction, and finally, data management and preservation.

Initially, all nodes remain inactive and await the initialization parameters saved during the first stage. After reading the parameters, they are broadcast to all the pipeline's nodes while the LSL flow of incoming “neural” data is paused. The pause lasts 10 s and ensures that all the nodes are ready to promptly process the data once the flow resumes.

After the flow of LSL-streamed “neural” data is resumed, the features are extracted. The process is similar to the one described in the *Initialization Stage* where the data is filtered, windowed, and log power is extracted.

The features are scaled and fed into the Transformer model to obtain audible acoustic representations. We use 1.3 s of features to produce a representation of that same size, however, only the last 34.69 ms are passed to the following synthesis stage.

Subsequently, the Griffin-Lim algorithm transforms these spectral representations into an audio waveform that is

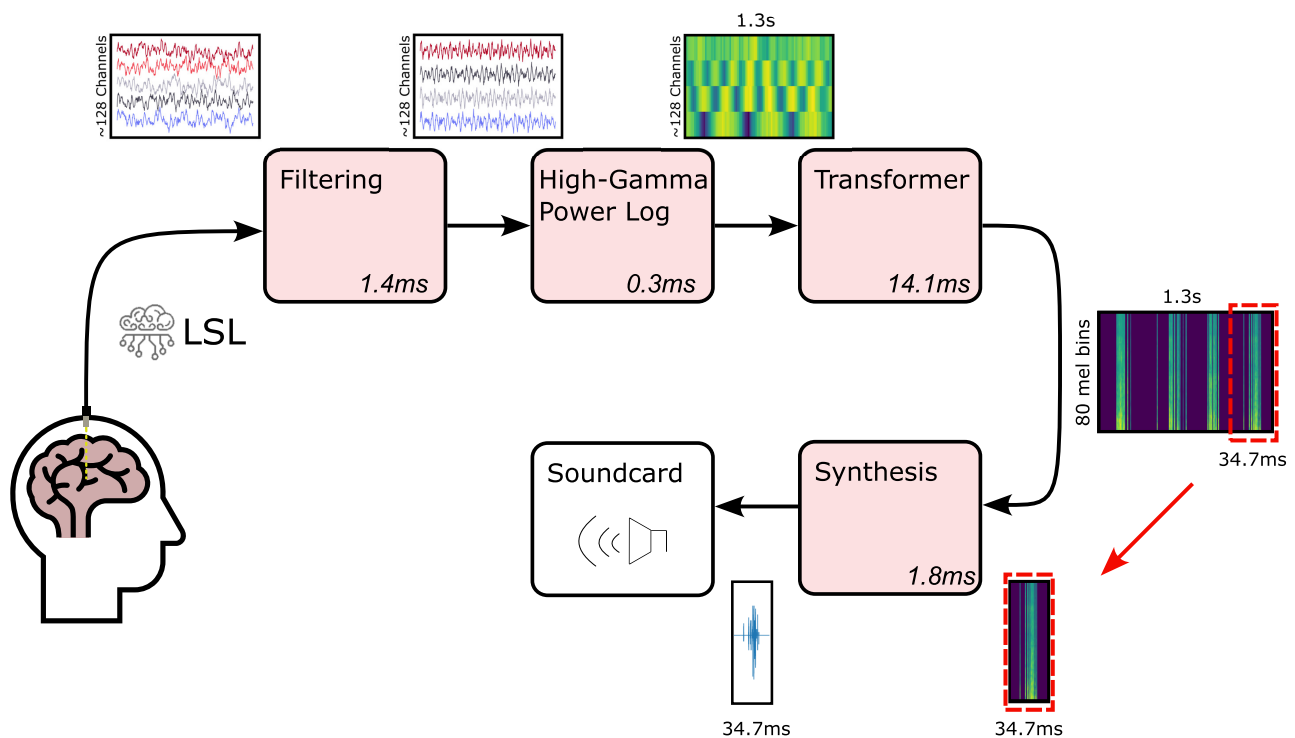


Figure 1: Illustration of the process of converting neural signals into audible speech. The system initializes by setting up nodes and pausing data intake. Neural data is then resumed and processed—filtered, feature-extracted, and transformed into a 1.3-second audio representation by a Transformer model, with only the last 34.7 ms used for sound synthesis via the Griffin-Lim algorithm. The generated waveform is sent to a sound card for playback. Concurrently, a data management system ensures the integrity and continuity of the operation without data loss.

send to the sound card for immediate auditory feedback. We include an efficient data management paradigm that prevents data loss and does not disrupt the pipeline operation or hinder its speed. The paradigm includes a resource-friendly data-saving routine executed concurrently to maintain operational speed and integrity.

#### Transformer Architecture:

As mentioned, the pipeline can be easily configured to extract different neural and audio features and use different Machine Learning models. Here, we present initial results with a real-time-ready transformer architecture.

A Transformer model is an advanced neural network architecture that excels in processing data sequences using self-attention mechanisms [13]. These mechanisms permit the efficient extraction of hidden context and relationships within data. Transformers are highly efficient, scalable, and flexible, making them superior for tasks requiring a deep understanding of complex relationships. This is why they have become the foundation for many state-of-the-art solutions in natural language processing and beyond. Transformers have also been used in decoding speech from offline data successfully [14].

In our context of having a small amount of time-series data with limited time to train the model, it is challenging to use a Transformer because they typically require large amounts of data and significant computational resources to effectively learn the complex patterns and relationships in time-series. Their architecture, designed

for capturing long-range dependencies, struggles to generalize from small datasets without overfitting and may not achieve optimal performance within a short training time-frame.

Despite these hurdles, we used a Transformer model to reconstruct auditory data from neural signals. Using only the self-attention mechanism and the encoder block, the model focuses on efficiently extracting and analyzing temporal features [16]. This approach reduces computational demands and training time, while still capturing complex patterns with less risk of overfitting. Focusing on prediction rather than sequence generation aligns the model's strengths directly with the requirements of time-series analysis, making it better suited for our tasks.

A challenge in real-time decoding with a sequential model lies in balancing the need to analyze significant temporal contexts to accurately decode complex patterns against the constraints of immediate processing. Recently published BCI works address this challenge by recording a large enough sequence of neural data and then producing the mapping to reproducible audio or text [10, 17].

Processing extensive historical data introduces latency for real-time applications like audio synthesis from neural signals, which conflicts with our real-time requirements. Our proposed solution, which exploits an idea presented by Shigemi *et al.* [18] involves using a predefined large-enough context size for analysis but synthesizing only the latest segment of the sequence. This approach allows the

Transformer to leverage enough historical data for accurate predictions, while maintaining the ability to produce outputs in real-time. It effectively addresses the challenge of adapting sequence-to-sequence mapping for real-time decoding, ensuring accuracy and immediacy in applications such as closed-loop neural interfaces.

Fig. 2 presents our conceived model architecture with the most relevant parameters. Our model first maps all the  $F$  channels of the input sequence into a 125-dimensional space through a linear transformation. Then, we use six standard encoder layers [13], each containing two main components: a Multi-Head Attention and a Feed Forward neural network, followed by an Add & Norm step to facilitate layer normalization. The Multi-Head Attention mechanism has five attention heads, and the Feed Forward neural network has a dimensionality of 2048 on the inner layer. The output of the last encoder layer undergoes another linear transformation to match the desired output dimension. A dropout rate of 0.25 is applied throughout the network to prevent overfitting. We use Mean Squared Error (MSE) as the loss criterion, and the learning rate is set to a modest  $5e-4$ , which balances the speed of convergence with the stability of the learning process.

## RESULTS

Given the constrained interaction duration with participants, it was critical to minimize model training times. The pipeline averaged approximately 133.11 s for model training, with a standard deviation of 12.00 s. This duration aligns well with our experimental requirements, offering a balanced compromise between training efficiency and subsequent decoding performance.

Processing latency per sample was another critical metric. Notably, each decoding operation by the transformer yields a 1.27-second audio window, from which only the latest 34.69 ms are utilized for audio reconstruction. To ensure near-real-time functionality, processing for each sample must therefore be completed in under 34.69 ms. Our performance results indicate an average processing time of approximately 17.62 ms per sample (standard deviation 1.01 ms), significantly below the 34.69 ms threshold. This efficiency meets our near-real-time criteria and provides flexibility for exploring other, more complex decoding approaches or even switching to higher-quality vocoders, such as HIFIGan [19] or VocGAN [20].

The qualitative aspect of our results involves the reconstruction of speech from neural signals. Fig. 3 aggregates the correlation outcomes across all participants, with data points indicating the correlation coefficient between the spectrograms of the original recorded and the synthesized audio for each individual, providing a visual representation of the decoding accuracy and variability among participants. Correlation are stable across the entire frequency spectrum of the mel-scale (Fig. 3 b), but vary dramatically between participants. Best results exceed average correlation coefficients of 0.66 (sub-06, Fig. 3

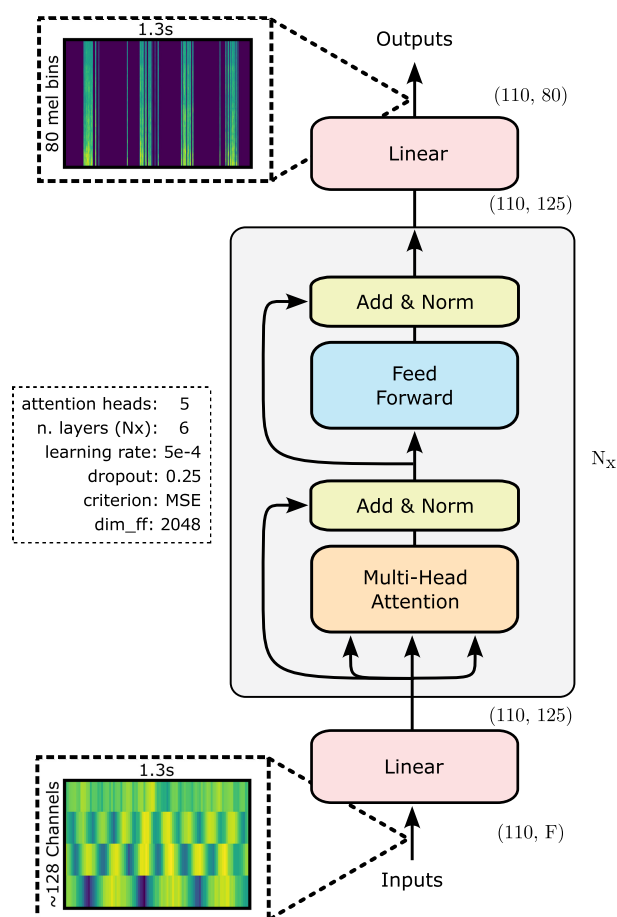


Figure 2: Transformer Model Architecture and Parameters. The model inputs are linearly transformed from  $F$  channels to a 125-dimensional space, followed by six encoder layers. Each layer consists of a Multi-Head Attention with five heads and a Feed Forward network with an inner-layer dimension of 2048, followed by an Add & Norm step. The final encoder output is linearly transformed to the desired output size. The model employs a dropout of 0.25, uses MSE as the loss function, and has a learning rate of  $5e-4$ .

a).

## DISCUSSION

The presented results, particularly concerning processing speeds and model training efficiency, precisely align with our pipeline's rapid training and real-time decoding objectives. This achievement highlights our pipeline's effectiveness in enabling real-time communication for individuals with speech impairments and its proficiency in decoding speech from new, unseen words. This latter capability underscores the system's robust generalization, a critical feature for practical Brain-Computer Interface (BCI) applications where pre-defining a comprehensive vocabulary is impractical.

Variations in decoding results across participants likely mirror the differential placement of sEEG electrodes. This suggests that proximity to speech-related brain areas might significantly influence both neural signal decoding quality and model training success. Notably, these

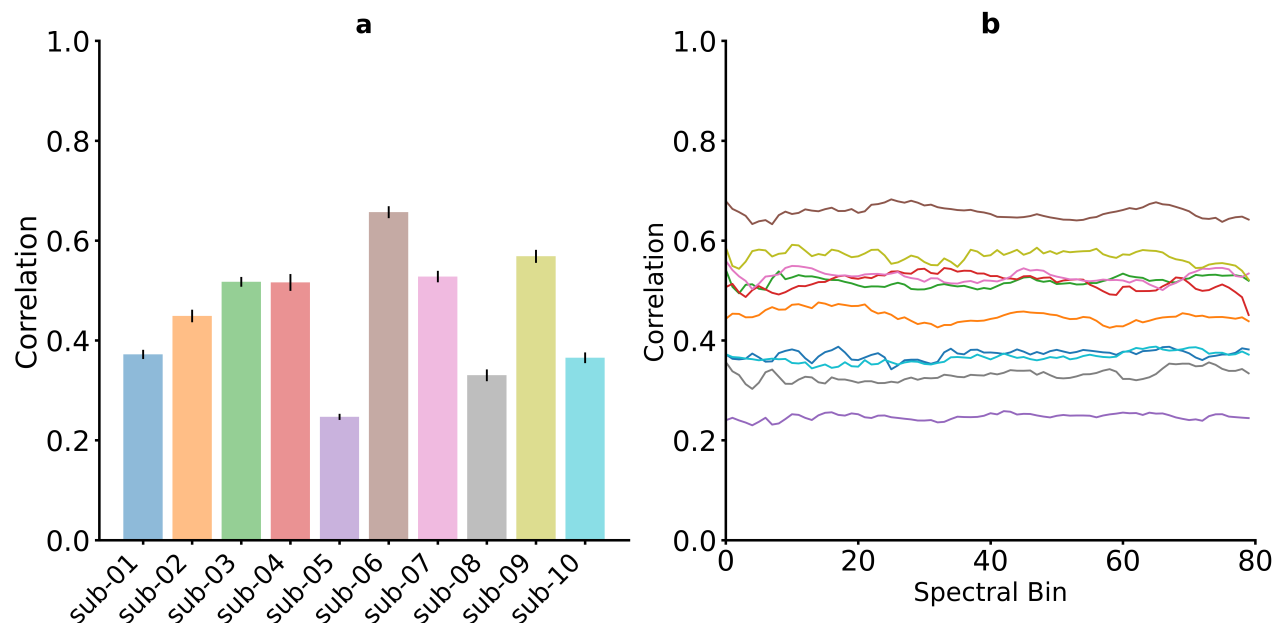


Figure 3: Correlation between original and reconstructed spectrograms. **a)** Mean correlation coefficients across all spectral bins for each participant, with error bars indicating the standard deviation. **b)** Mean correlation coefficients for each spectral bin.

achievements come from training on relatively limited data, approximately 225 s per participant. This is in stark contrast to the vast datasets employed in training the current state-of-the-art speech BCIs, which often utilize data ranging from dozens of minutes to several hours [6, 7, 10, 21], highlighting the efficiency and potential of our approach even with constrained datasets.

Currently, real-time reconstructed speech results are not intelligible, leaving room for further improvements in decoding approach and vocoder.

While the current results stem from simulations using the SWPD dataset, delineating the pipeline’s capability for real-time speech decoding from neural signals, presenting online results falls outside this paper’s scope. Nevertheless, addressing this gap is a priority in our ongoing research.

The promising outcomes achieved with the Transformer model open exciting future research directions, such as refining model architectures and devising new strategies to minimize further decoding latency. The 17.62 ms extra in processing time also permits using a more complex synthesizer that better reconstructs the audible speech from the spectrogram.

## CONCLUSION

This paper introduced a closed-loop speech decoding pipeline designed for real-time operation. Our system is distinctively characterized by its low latency and modular framework, facilitating seamless, near-instantaneous communication. Utilizing Python and the Timeflux framework, we developed a modular pipeline that allows for swift prototyping and testing, catering to the dynamic needs of BCI research.

We demonstrated the feasibility of real-time streaming speech synthesis from neural signals through rigorous offline validations using aligned neural and audio recordings from the SWPD dataset. Our pipeline employs a Transformer model optimized for time-series data, achieving fast decoding speed and reasonable results. Despite the challenges associated with limited data availability and the constraints of working within clinical settings, our system managed to train models efficiently, with an average model training time of approximately 133.11 s and a decoding processing time of about 17.62 ms per sample, well below the threshold required for real-time functionality [12]. Notably, our streaming approach could allow for natural conversation, as sound is produced almost immediately, as opposed to other approaches which produce chunks of audio corresponding to whole sentences.

The qualitative results further underscore the efficacy of our pipeline in reconstructing audible speech. The reconstructed spectrograms and the correlation coefficients across participants highlight the potential of our technology to provide a voice for those who have lost their natural ability to speak due to neurological conditions.

Our work showcases an improvement in speech BCIs and opens new avenues for research and development toward more intuitive and accessible communication solutions. Future work will focus on enhancing the decoding results, reducing latency further, and expanding the system’s adaptability. By continuing to refine and validate our pipeline, we aim to bring this technology closer to widespread clinical application, offering hope for improved quality of life for individuals with severe speech impairments.

## REFERENCES

- [1] C. Herff *et al.*, “Brain-to-text: Decoding spoken phrases from phone representations in the brain,” *Frontiers in neuroscience*, vol. 9, p. 217, 2015.
- [2] M. Angrick *et al.*, “Speech synthesis from ecog using densely connected 3d convolutional neural networks,” *Journal of neural engineering*, vol. 16, no. 3, p. 036 019, 2019.
- [3] J. Berezutskaya, Z. V. Freudenburg, M. J. Vansteensel, E. J. Aarnoutse, N. F. Ramsey, and M. A. van Gerven, “Direct speech reconstruction from sensorimotor brain activity with optimized deep learning models,” *Journal of Neural Engineering*, vol. 20, no. 5, p. 056 010, 2023.
- [4] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, “Speech synthesis from neural decoding of spoken sentences,” *Nature*, vol. 568, no. 7753, pp. 493–498, 2019.
- [5] J. G. Makin, D. A. Moses, and E. F. Chang, “Machine translation of cortical activity to text with an encoder–decoder framework,” *Nature neuroscience*, vol. 23, no. 4, pp. 575–582, 2020.
- [6] F. R. Willett *et al.*, “A high-performance speech neuroprosthesis,” *Nature*, vol. 620, no. 7976, pp. 1031–1036, 2023.
- [7] S. L. Metzger *et al.*, “Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis,” *Nature Communications*, vol. 13, no. 1, p. 6510, 2022.
- [8] D. A. Moses *et al.*, “Neuroprosthesis for decoding speech in a paralyzed person with anarthria,” *New England Journal of Medicine*, vol. 385, no. 3, pp. 217–227, 2021.
- [9] S. L. Metzger *et al.*, “A high-performance neuroprosthesis for speech decoding and avatar control,” *Nature*, vol. 620, no. 7976, pp. 1037–1046, 2023.
- [10] M. Angrick *et al.*, “Online speech synthesis using a chronically implanted brain-computer interface in an individual with als,” *medRxiv*, 2023.
- [11] P. Clisson, R. Bertrand-Lalo, M. Congedo, G. Victor-Thomas, and J. Chatel-Goldman, “Time-flux: An open-source framework for the acquisition and near real-time processing of signal streams,” in *BCI 2019-8th International Brain-Computer Interface Conference*, 2019.
- [12] A. Stuart, J. Kalinowski, M. P. Rastatter, and K. Lynch, “Effect of delayed auditory feedback on normal speakers at two speech rates,” *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2237–2241, 2002.
- [13] A. Vaswani *et al.*, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [14] J. Kohler *et al.*, “Synthesizing speech from intracranial depth electrodes using an encoder-decoder framework,” *Neurons, Behavior, Data analysis, and Theory*, 2022.
- [15] M. Verwoert *et al.*, “Dataset of speech production in intracranial electroencephalography,” *Scientific data*, vol. 9, no. 1, p. 434, 2022.
- [16] Y.-E. Lee and S.-H. Lee, “Eeg-transformer: Self-attention from transformer architecture for decoding eeg of imagined speech,” in *2022 10th International winter conference on brain-computer interface (BCI)*, IEEE, 2022, pp. 1–4.
- [17] F. R. Willett, D. T. Avansino, L. R. Hochberg, J. M. Henderson, and K. V. Shenoy, “High-performance brain-to-text communication via handwriting,” *Nature*, vol. 593, no. 7858, pp. 249–254, May 2021. (visited on 03/13/2024).
- [18] K. Shigemi *et al.*, “Synthesizing speech from ecog with a combination of transformer-based encoder and neural vocoder,” in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023, pp. 1–5.
- [19] J. Kong, J. Kim, and J. Bae, “Hifi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis,” *Advances in neural information processing systems*, vol. 33, pp. 17 022–17 033, 2020.
- [20] J. Yang, J. Lee, Y. Kim, H. Cho, and I. Kim, “Vocgan: A high-fidelity real-time vocoder with a hierarchically-nested adversarial network,” *arXiv preprint arXiv:2007.15256*, 2020.
- [21] N. S. Card *et al.*, “An accurate and rapidly calibrating speech neuroprosthesis,” *Neurology*, Preprint, Dec. 2023. (visited on 12/27/2023).

# NEUROPHONE: REAL-TIME BRAIN-MOBILE PHONE INTERFACE

Norhan Abdelhafez<sup>1</sup>, Manal Tantawy<sup>1</sup>, Abdelrahman Sayed<sup>1</sup>, Nora Ekramy<sup>1</sup>, Mohammed Nour-Eldin<sup>1</sup>

<sup>1</sup> Scientific Computing Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

E-mail: norhan.abdelhafez3@gmail.com

**ABSTRACT:** The extensively studied P300 component of the human event-related potential in cognitive neuroscience has significant applications, including constructing BCI systems for individuals with motor disabilities. However, accurately and efficiently identifying the P300 component in EEG data poses challenges due to the low signal-to-noise ratio and biological diversity among subjects. To address this, cutting-edge deep learning architectures were developed and employed. Initially, digital signal processing techniques were applied, followed by training and evaluation of DL models like Chrononet, EEGNet, DCRNN, CNNs, and RNNs. Results revealed that our lightweight CNN model, combined with K-fold cross-validation and weighted class, achieved the highest average classification accuracy of 98% surpassing other models for subject-dependent P300 classification. This high-performing CNN model facilitated the creation of NeuroPhone, a communication application grounded in the core principles of BCI systems.

## INTRODUCTION

Electroencephalography (EEG) has opened a window into the human brain, allowing us to visualize its electrical activity and delve into the hidden language of its neurons. From its widespread medical applications to its growing presence in research and consumer domains, EEG offers a powerful tool for understanding and interacting with the mind. At the heart of this interaction lies the fascinating world of brainwaves, different patterns reflecting different states of consciousness. From the high-frequency beta waves associated with focused attention to the slow delta waves accompanying deep sleep, each frequency serves as a neural biomarker for specific cognitive states [1].

Among these brainwaves, the P300 event-related potential (ERP) holds a special position. This distinctive positive spike, peaking roughly 300 milliseconds after a specific stimulus, reveals much about our cognitive processes. Researchers have extensively studied the P300, recognizing its crucial role in attention, memory, decision-making, and information processing [2]. Its potential, however, extends beyond research labs, paving the way for revolutionary technology called Brain-Computer Interfaces (BCIs).

BCIs offer a direct communication channel between the

brain and external devices, bypassing traditional input methods. By harnessing the power of P300 and other EEG signals, BCIs empower individuals to control computers, prosthetic limbs, and even communicate through their thoughts [3]. Yet, despite the immense promise of BCIs, their path to widespread adoption is met with two key challenges: achieving robust and accurate P300 detection and overcoming the computational limitations of existing BCI systems. Current models often struggle to extract the subtle P300 signal from the inherent noise of EEG data, and their demanding computational requirements prevent seamless integration with mobile devices, a crucial step towards accessibility for a wider population.

In this study, we present the development of NeuroPhone, an efficient BCI in the form of a communication application designed to break down these barriers. NeuroPhone leverages the P300 peak, enabling individuals with motor disabilities to control their smartphones and engage in digital communication solely through their visual attention. By employing cutting-edge deep learning techniques tailored for mobile device processing power, NeuroPhone aims to overcome the previous limitations of accuracy and accessibility. This paper delves into the development of NeuroPhone including the implemented digital signal processing (DSP) techniques, deep learning architectures used, and specifics of the development process of NeuroPhone's application software.

## METHODS

We aimed to develop a comprehensive and computationally efficient classification model based on the detection of P300 event-related potential (ERP). We first went on exploring and evaluating different architectures, such as ChronoNet, EEGNet, DCRNN, and others, to determine their effectiveness and performance in a subject-dependent task. Then, we introduced a lightweight convolutional neural network (CNN) architecture that excels in capturing unique ERP features, leading to superior classification accuracy compared to existing state-of-the-art architectures.

In the subsequent sections, we first present details about the datasets we utilized in our work, including the online EPFL BCI Group dataset [4] and the data we collected offline using Emotiv EPOC headset. After that, we

provide a comprehensive overview of the models we investigated. Then we present the architecture of our CNN model. Additionally, we outline the training methodology we employed, highlighting the steps and techniques utilized to optimize and fine-tune the models for optimal performance. And finally, we explain the details of NeuroPhone's application software, and the technology used.

**Datasets:** The EPFL BCI group dataset, which was employed in our research, played a crucial role in evaluating the performance of various models. This dataset was specifically curated by the Brain-Computer Interface (BCI) group at École Polytechnique Fédérale de Lausanne (EPFL) and is widely recognized in the field. The dataset consists of meticulously recorded electroencephalogram (EEG) signals, making it a valuable resource for investigating brain-computer interfaces. The dataset has a population of five disabled and four able-bodied subjects. Subjects were facing a laptop screen on which six images were displayed. The images were selected according to an application scenario in which users can control electrical appliances via a BCI system. The EEG was recorded at 2048 Hz sampling rate from 32 electrodes placed at the standard positions of the 10–20 international system. Each subject recorded 4 sessions and each session had 6 runs. For a single run, the images were flashed in random sequences, one image at a time. Each flash of an image lasted for 100 ms and during the following 300 ms none of the images was flashed, i.e. the interstimulus interval was 400 ms, see (Fig. 1).

For our collected dataset, we followed the same recording paradigm as EFPL dataset. We used the famous Emotiv EPOC headset with 14 channels placed at the standard positions of the 10–20. We recorded the EEG signal from a single male subject. The subject was faced by NeuroPhone's application screen which displayed 6 images (icons), each represented a certain functionality that allows the user to communicate with others, see (Fig. 2). More details are provided at the application subsection. The subject recorded 6 sessions; each session had a duration of 90 seconds with a sampling rate of 256 Hz.

**Preprocessing:** The data underwent several preprocessing steps to ensure optimal analysis. The re-referencing step involved utilizing the average signal from the two mastoid electrodes for re-referencing purposes. To obtain a desired signal range of 1 to 12 Hz, a band-pass Butterworth filter of order 3 was applied to filter the signal [5]. Subsequently, the signal was down sampled by 64 Hz to reduce computational load. Then, data was segmented such that each segment was corresponding to an event. A duration of 1 second was taken after each stimulus event and given that the duration of the flashing event was 400 ms, there was a 600 ms overlap. On a single segment, z-score normalization was implemented to normalize the signal. To handle extreme values, the signal underwent a Winsorizing process, where the 10th and 90th percentiles were calculated for samples from each electrode. Any

amplitude values falling below the 10th percentile or above the 90th percentile was substituted with the respective 10th or 90th percentile value [4]. These preprocessing steps collectively aimed to optimize the data for deep learning models' training. The input signal shape for the models was  $(32 \times 32)$  where the first dimension is the number of time samples, and the second dimension is the number of channels. And in our collected data input shape was  $(32 \times 14)$  because EMOTIV dataset contained only 14 channels.

**Deep Learning architectures:** In our quest for the optimal deep learning architecture for subject-dependent P300 classification, we explored a diverse range of models, each offering its own set of advantages and limitations. Chrononet and EEGNet, specifically designed for EEG analysis, leverage convolutional layers to efficiently capture the temporal characteristics of the P300 component, making them well-suited for this task [6][7]. However, their deep architectures can be computationally expensive to train and might require a substantial amount of data for optimal performance. DCRNNs, combining the strengths of CNNs and RNNs, excel at capturing both the spatial and temporal information crucial for P300 detection [8]. Despite their effectiveness, DCRNNs can be more complex to design and train effectively, requiring careful hyperparameter tuning to unlock their full potential. Finally, standard RNNs, while adept at learning sequential data like EEG signals, can suffer from vanishing gradients, hindering their ability to learn from long sequences.

**Our CNN model architecture:** The proposed CNN architecture is designed to extract salient features from 2D EEG signals for robust P300 component classification. The model comprises two convolutional layers with 32 and 64 filters (3x3 kernel size), applying learned filters to extract spatial patterns relevant to P300 detection. ReLU activation introduces non-linearity. A max-pooling layer (2x2 pool size) down-samples feature maps, reducing dimensionality and promoting spatial invariance. A flatten layer prepares the extracted features for classification by two fully connected layers (128 neurons with ReLU activation, and 1 neuron with sigmoid activation), see (Fig. 3).

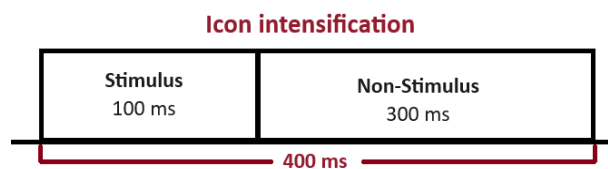


Figure 1: The interstimulus interval of each flashing event.

**Training Methodology:** To ensure robust model performance and mitigate the effects of overfitting, a K-fold cross-validation strategy was employed during training. Specifically, a 5-fold cross-validation approach was implemented (K=5). This technique divides the dataset into five partitions (folds) while preserving the proportion of P300 and non-P300 examples in each fold. Iteratively, one fold is designated as the testing set while

the remaining folds are used for training. Model evaluation metrics are computed on the held-out testing set after each training iteration.

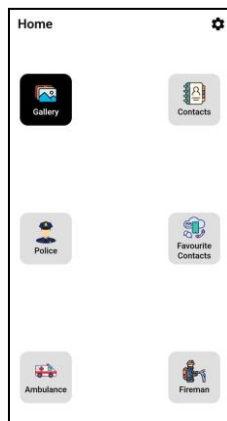


Figure 2: NeuroPhone's Mobile Application interface which contains 6 flashing images (icons).

To address potential class imbalance within the EEG dataset, where the number of P300 events (positive class) is significantly lower compared to non-P300 events (negative class), class weights were computed and incorporated into the training process. This approach assigns higher weights to the minority class (P300 events) during training. The specific weights are calculated based on the class frequencies within the training data. By assigning higher weights, the model is effectively forced to pay closer attention to the less frequent P300 examples, leading to a more balanced learning process and improved classification performance for the minority class.

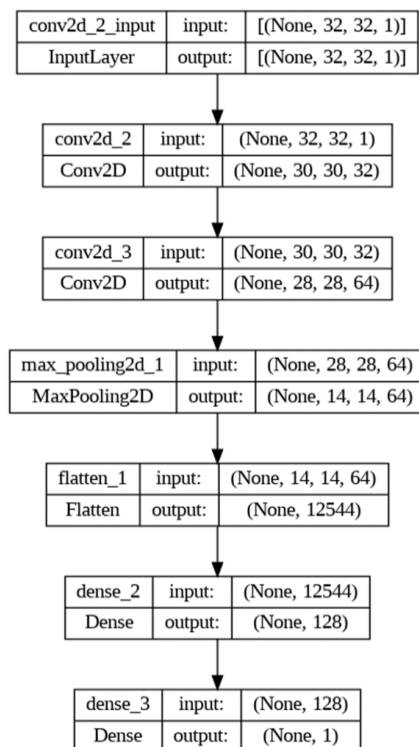


Figure 3: Architectural configuration of our CNN model.

The deep learning models were compiled with the Adam optimizer, a common choice for its adaptive learning rate capabilities, and binary cross-entropy loss, suitable for binary classification. Accuracy served as the primary evaluation metric.

Finally, average loss and accuracy scores across all folds were calculated to provide a comprehensive assessment of model performance under the K-fold cross-validation procedure.

*Application:* The purpose of our application is to help disabled people use their mobile phones and perform some important functions through it using only their visual attention. Our application is designed in a way that visually stimulates the user to choose the icon they desire. Each icon represents a functionality that enables them to control their smartphone. They are: Gallery, Contacts, Police, Favorite Contacts, Ambulance, and Fireman. The whole set of icons would flash in random order. The user would focus their attention on any icon they want to choose, and after some repetitions of flashing the whole set of icons, the DL model would detect the P300 peak that synchronized with the timing of the desired icon's flash, and thus, would fire the start of the execution of that icon's functionality.

We used Emotiv EPOC X 14 which consists of 14 EEG channels and 2 reference channels. The electrodes are located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 according to the International 10-20 system, see (Fig. 4). The headset was connected to its software EMOTIV-PRO on the laptop then we start streaming data from Lab Streaming Layer (LSL) option in the application. The LSL feature allows efficient, two-way communication between EmotivPRO and other third-party software and devices.

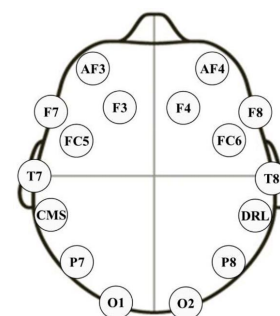


Figure 4: The location of EMOTIV electrodes.

We have created an API with Flask Library in Python language. The main functionality of this API is to synchronize between the EEG signal coming from EMOTIV and the flash timing coming from Neurophone's mobile application.

The flask API receives raw EEG signal samples from EMOTIV in addition to timestamp of each sample, so the first step it performs is to segment the raw EEG signal and preprocess it. Each segment is 1000 ms long which corresponds to 256 signal samples because EMOTIV's sampling rate is 256 HZ. After preprocessing, the segment's length would be 32 samples because of the

down sampling step in preprocessing. We also perform bandpass Butterworth filter between 1 and 12 HZ and Z-score normalization. After preprocessing, each segment is passed to the DL model to determine whether it contains P300 or not. The DL model we utilized in the API was our CNN model that we presented its architecture earlier in (Fig. 3). We chose CNN because it outperformed the others in the offline evaluation. Along with the received EEG signal samples and timestamp of each sample, Flask API receives each icon's flash timing from Neurophone's flutter mobile application.

We used Flutter to create the mobile application. The application was running on a Galaxy M31 Phone with an Octa-core Exynos 9611 (10nm) Processor and Android 12 operating system. The Application would continuously send each icon's flash time and icon's index to Flask API. The second step performed by the API is synchronization. If the EEG segment that was synchronized with an icon flash time contained a P300 peak, the API would send a firing response back to the flutter application to start the execution of that icon's functionality, see (Fig. 5).

## RESULTS

In this section, we present the experimental results of our study focusing on the classification of P300 in EEG signals in subject-dependent task. To achieve this objective, we trained a variety of models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Dynamic Convolutional Recurrent Neural Networks (DCRNN), EEGNet, and ChronoNet. Those models are evaluated on the EPFL BCI group dataset and our collected dataset. In Tab. 1, we demonstrate the results on the EPFL dataset. We provide the average k-fold accuracy and F1-score (across the 5 folds).

Among accuracy results, CNN achieved the highest average classification accuracy on EPFL data. We also explored their performance on our collected dataset. In Tab. 2, we demonstrate the results of our single male subject data. The demonstrated results on our collected are offline results, meaning that the collection and evaluation were performed offline and then the best performing model in the offline evaluation was utilized in the real-time scenario (it was the CNN model in this case).

Table 2: Results on our collected data

Model	Accuracy	F1-score
EEGNET	0.94	0.86
DRCNN	0.92	0.81
RNN	0.97	0.89
CNN	<b>0.98</b>	<b>0.95</b>

The results on our in-house data also demonstrated that CNN model outperforms the others by achieving a 98% average classification accuracy, highlighting the eligibility of CNN to be utilized in real-time.

## DISCUSSION

We could notice from results in Tab. 1 that subject 8 achieved the highest average classification accuracy across most of the DL models. It may be attributed to the fact that the subject was highly motivated during the experiments as stated by the authors who collected the data [4]. Even though EPFL contained 9 subjects, they excluded the data of the fifth subject due to the difficulty of communication with him. We could also notice that CNN achieved the highest average classification accuracy and F1 score across all the subjects compared to the other models.

We acknowledge that the in-house data were small, and a larger dataset is required for the results to be generalizable. The use of parameters such as information transfer rate (ITR) is an essential metric to be utilized for an improved evaluation of the BCI system.

To compare our results with other studies implementing P300-based BCI systems, we find that Eric Sellers and Emanuel Donchin [9] achieved an average classification accuracy of 72% for ALS patients and 85% for abled subjects. We could see that Hubert Cecotti and Axel Gräser [10] achieved a classification accuracy of 95.5% using a CNN model. To compare our results to other studies that utilized the same dataset we used (EPFL dataset), we find that the authors in [11] achieved an average classification accuracy of 95.68% for the healthy subjects and 94.69% for disabled patients through their CNN model that uses 2-D EEG scalogram images. We also see Shojaedini et al. [12] reached a classification accuracy of 95.34% using a CNN model with a new adaptation method for hyperparameters. Our methods achieved a higher classification accuracy for P300 detection among the studies that used the same dataset.

While NeuroPhone provides a robust BCI system to detect P300 and control the mobile application, there are some limitations of the system. The first one is the inevitable time delay between the mobile application and the API. This delay is attributed to the quality of the connection between the phone and the API. We haven't accurately measured the delay time, but it was believed to be around a few seconds. The second limitation is the number of repetitions, in real-time, the user requires around 3 to 4 repetitions to select the desired icon. A single repetition is the flashing of all the icons, and it lasts for 2.4 seconds, so 3 repetitions would be around 8 seconds.

## CONCLUSION

This research demonstrates the power of deep learning for EEG analysis and brain-computer interface development. Our lightweight CNN model, combined with K-fold cross-validation and class weighting, achieved superior P300 classification accuracy compared to other architectures. This enabled the successful creation of the NeuroPhone application. Future research will explore transfer learning for improved model generalization across subjects and investigate hybrid

Table 1: Results on EPFL BCI group dataset

	Sub 1		Sub 2		Sub 3		Sub 4		Sub 6		Sub 7		Sub 8		Sub 9	
Model	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
EEGNET	0.89	0.70	0.84	0.64	0.93	0.80	0.90	0.74	0.89	0.71	0.90	0.77	<b>0.95</b>	0.87	0.86	0.68
ChronoNet	0.93	0.83	0.90	0.75	<b>0.94</b>	0.84	0.89	0.79	0.93	0.83	0.93	0.85	0.92	0.82	0.91	0.79
DRCNN	0.86	0.65	0.84	0.61	0.89	0.73	0.90	0.72	0.91	0.75	0.90	0.75	<b>0.93</b>	0.81	0.86	0.64
RNN	0.94	0.83	0.94	0.83	0.96	0.88	0.94	0.83	0.95	0.86	0.96	0.88	<b>0.97</b>	0.92	0.95	0.84
CNN	0.99	0.96	0.98	0.96	0.99	0.98	0.98	0.96	0.98	0.96	0.99	0.97	<b>0.99</b>	0.97	0.99	0.97

deep learning and signal processing approaches for further enhancements.

## ACKNOWLEDGMENT

The authors acknowledge the support of the Information Technology Industry Development Agency (ITIDA) for their funding of this research.

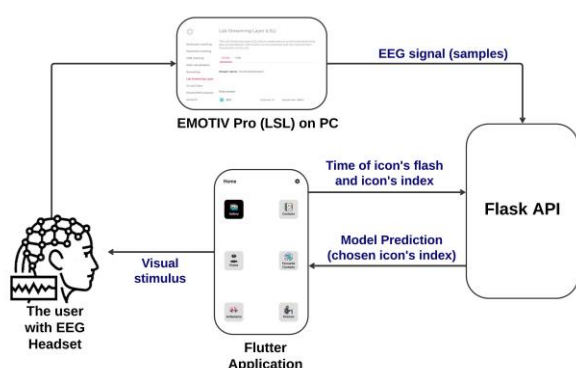


Figure 5: Using NeuroPhones' Flutter application interface, the user gets visually stimulated by the flashing, and at the same time, the EEG signal is transmitted to Flask API, the API synchronizes between the EEG signal and the flash timing and sends back a response to Flutter App to execute the desired icon's functionality.

## REFERENCES

- [1] Kumar JS, Bhuvaneshwari P. Analysis of Electroencephalography (EEG) Signals and Its Categorization—A Study. *Procedia Eng.* 2012;38:2525-2536.
- [2] Zhong R, Li M, Chen Q, Li J, Li G, Lin W. The P300 Event-Related Potential Component and Cognitive Impairment in Epilepsy: A Systematic Review and Meta-analysis. *Front Neurol.* 2019;10:943.
- [3] Rashid M, Sulaiman N, P P Abdul Majeed A, Musa RM, Ab Nasir AF, Bari BS, Khatun S. Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. *Front Neurobot.* 2020;14:25.
- [4] Hoffmann U, Vesin JM, Ebrahimi T, Diserens K. An efficient P300-based brain-computer interface for disabled subjects. *J Neurosci Methods.* 2008;167(1):115-25.
- [5] Shojaedini SV, Morabbi S, Keyvanpour M. A New Method for Detecting P300 Signals by Using Deep Learning: Hyperparameter Tuning in High-

Dimensional Space by Minimizing Nonconvex Error Function. *J Med Signals Sens.* 2018;8(4):205-214.

- [6] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, & Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J of Neural Eng.* 2018;15(5):056013.
- [7] Roy S, Kiral-Kornek I, Harrer S. ChronoNet: A Deep Recurrent Neural Network for Abnormal EEG Identification. *ArXiv abs/1802.00308* (2018): 4.
- [8] Li Q, Liu Y, Shang Y, Zhang Q, Yan F. Deep Sparse Autoencoder and Recursive Neural Network for EEG Emotion Recognition. *Entropy (Basel).* 2022;24(9):1187.
- [9] Sellers EW, Donchin E. A P300-based brain-computer interface: initial tests by ALS patients. *Clin Neurophysiol.* 2006;117(3):538-48.
- [10] Cecotti H, Gräser A. Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE Trans Pattern Anal Mach Intell.* 2011;33(3):433-45.
- [11] Singh SA, Meitei TG, Devi ND, Majumder S. A deep neural network approach for P300 detection-based BCI using single-channel EEG scalogram images. *Phys Eng Sci Med.* 2021;44(4):1221-1230.
- [12] Shojaedini SV, Morabbi S, Keyvanpour M. A New Method for Detecting P300 Signals by Using Deep Learning: Hyperparameter Tuning in High-Dimensional Space by Minimizing Nonconvex Error Function. *J Med Signals Sens.* 2018;8(4):205-214.

# NOVEL MATERIALS FOR BRAIN COMPUTER INTERFACES: PERSPECTIVES AND ASPECTS OF COMBINATION OF A MAGNETOELECTRIC STIMULATOR AND A GRAPHENE MICROTRANSISTOR ARRAY RECORDING SYSTEM

S. Matsoukis<sup>1,2</sup>, J. Scharinger<sup>2</sup>, J. Covelo<sup>3</sup>, N. Cancino-Fuentes<sup>3</sup>, M. V. Sanchez-Vives<sup>3</sup>, G. Edlinger<sup>1</sup>, C. Guger<sup>1</sup>

1. g.tec medical engineering GmbH, Austria

2. Institute for Computational Perception, Johannes Kepler University, Linz, Austria

3. Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

E-mail: matsoukis@gtec.at

**ABSTRACT:** In this paper, we explore the innovative combination of magneto-electric nanoparticles (MENPs) and graphene solution-gated field-effect transistors (gSGFETs) to advance brain-computer interfaces (BCIs). ME materials, known for their wireless and minimally invasive brain stimulation capabilities, are combined with gSGFETs, known for their high-resolution neural recording. Our research explores the potential benefits of this hybrid approach, including reduced artifacts, enhanced spatial resolution, and improved detection of subthreshold phenomena and DC potentials. A hardware and software setup is proposed and possible data analysis methods that will assist in the further development of the system are reviewed. This combined technology offers a promising direction for advanced BCIs and represents a significant advance in neural engineering.

## INTRODUCTION

In the rapidly evolving field of neural engineering, significant advances have been made to improve brain-computer interfaces (BCIs) through innovative materials and technologies. One notable development is magnetoelectric (ME) materials, which are known for their unique ability to convert magnetic fields into electric fields and vice versa through mechanical coupling between magnetostrictive and piezoelectric components [1]. The application of ME materials in brain stimulation is a promising option for low magnetic field stimulation in the range of a few mT. They can be fabricated in various sizes, from milliliters to nanometers, thus facilitating minimally invasive procedures. At the nanoscale, MENPs can even be administered intravenously, providing a wireless, less invasive approach to stimulating deep brain regions [2-6]. While previous in vitro and in vivo research on MENPs for brain stimulation has shown promising results for their modulatory effects on brain activity, these studies have primarily used calcium imaging to measure neuronal activation [3-5]. This technique has the advantage of simultaneously visualizing the activity of

large populations of neurons while not requiring the use of implants, thus reducing invasiveness. Another key advantage, particularly relevant in magnetic field stimulation, is the absence of interference from induced voltages in the measurements. Calcium imaging, however, has significant limitations. It detects changes in calcium ion concentrations within neurons, which indicate neuronal activity. When neurons fire, these calcium ions flow into the cells and are detected by fluorescent calcium indicators. However, the kinetics of these indicators and related physiological processes limit the temporal resolution of the method. In addition, calcium imaging cannot detect subthreshold changes in activity [7].

Therefore, this method is not suitable for a complete study of stimulation-related phenomena that require good temporal resolution, such as evoked potentials or entrainment at a specific stimulation frequency.

Electrophysiological methods, on the other hand, can directly record electrical potentials with high temporal resolution, making them well suited for use with neural stimulation techniques. However, they offer limited spatial resolution, and induced voltages become a significant issue when stimulation involves magnetic fields [8].

To address these challenges and fully exploit the capabilities of MENPs stimulation, we propose the combination with a graphene microtransistor array recording system. The graphene-based active sensors, specifically graphene solution-gated field-effect transistors (gSGFETs) are notable for their flexibility, biocompatibility, high carrier mobility, chemical stability, and mechanical conformability [9]. Recent advancements in gSGFETs have demonstrated their effectiveness for broadband recordings and their potential for spatially resolved mapping, making them ideal for exploring various neural activities, including Infra-slow oscillations [10-12].

The combination of these two systems promises to provide the high spatial resolution and minimal interference typical of imaging methods, along with the high temporal resolution and subthreshold phenomenon

detection capabilities of full-band electrophysiology methods. We also propose a detailed hardware and software setup and suggest the exploration of specific data analysis methods that will further the development of the system. By combining the wireless, minimally invasive stimulation capabilities of ME materials with the high-resolution recording capabilities of gSGFETs, this research aims to create a more comprehensive and effective approach to studying and modulating brain activity, making a significant contribution to neurological research and potential treatments for neurological disorders.

## MAGNETOELECTRIC MATERIALS FOR NEURAL STIMULATION

The ability of ME materials to convert magnetic fields into electric fields with significant high performance has led to increasing research on the application of ME materials in neural stimulation. ME materials consist of a magnetostrictive component that deforms under a magnetic field and a piezoelectric component that converts this deformation into voltage. These components can be configured in several forms: two bonded linear thin films are common at the submillimeter scale, while a magnetostrictive core with a piezoelectric shell is typical for MENPs. Depending on their size, ME materials can be either implanted or injected—submillimeter devices are usually implanted, whereas nanoscale devices like MENPs are injected. Injection can be performed stereotactically in the targeted brain region or even intravenously, depending on the MENPs' nanodiameter, which influences their ability to cross the blood-brain barrier. In this work, we focus on MENPs (Figure 1) due to their minimal invasiveness, which is significant for biomedical applications. However, the proposed hardware and software system can also be integrated with submillimeter ME devices, which offer a comparatively lower level of invasiveness than other solutions.

Stimulation using low-intensity magnetic fields and MENPs in deep brain regions could provide a wireless and less invasive alternative to traditional deep brain stimulation (DBS) [2]. Furthermore, MENPs could enhance precision and depth in targeting neural activity compared to techniques like transcranial magnetic stimulation (TMS), transcranial alternating current stimulation (tACS), and transcranial direct current stimulation (tDCS) [2-6]

Significant experiments with ME materials have shown their ability to modulate brain activity, using nanoparticles for targeted stimulation in *in vitro* studies and demonstrating neuromodulation feasibility *in vivo*. These studies suggest ME materials could offer new treatments for neurological disorders [3-5]. However, they often rely on indirect methods like calcium imaging to assess neural responses, highlighting a gap in direct neural activity measurement through electrophysiology.

This gap suggests a need for integrating ME materials with advanced systems like graphene electrode arrays for a deeper, more accurate understanding of neural dynamics and stimulation effects.

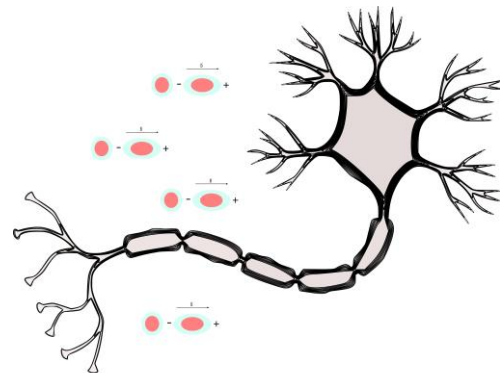


Figure 1: Magneto-electric nanoparticles deform when a magnetic field is applied, creating electrical dipoles that can potentially modulate neuronal activity.

## SOLUTION-GATED GRAPHENE FIELD-EFFECT TRANSISTORS FOR NEURAL RECORDINGS

Electrolyte-gated transistors have emerged as a promising technology in the field of active sensors. The unique properties of graphene, such as its high electrical conductivity, flexibility, and biocompatibility, make gSGFETs especially suitable for interfacing with biological systems. GSGFETs offer several advantages over traditional neural recording systems. Their high carrier mobility allows for rapid response to neural signals, enhancing the temporal resolution of recordings. Additionally, the thin and flexible nature of graphene enables gSGFETs to conform to neural tissues, reducing mechanical mismatches and improving signal fidelity. Compared to conventional metal-oxide-semiconductor field-effect transistors (MOSFETs), gSGFETs exhibit lower noise levels, which is crucial for detecting subtle neural activities. [9-10].

Several studies have demonstrated the efficacy of gSGFETs in neural recording applications. For instance, a landmark study [9] illustrated how gSGFETs could be used to record electrophysiological signals from cardiac cells with higher clarity than traditional methods. Another study [11] successfully employed gSGFETs in recording complex neural networks, showcasing their potential in understanding neural dynamics and disorders. Furthermore, experiments have shown that gSGFETs are capable of operating in harsh biochemical environments, maintaining their stability and functionality over extended periods, a crucial factor for long-term neural monitoring [12]. gSGFETs have been successful in capturing a wide range of frequencies, including those below 0.1Hz, known as infra-slow oscillations, with performance similar to glass micropipettes. They also provide the capability for detailed spatial imaging. [10].

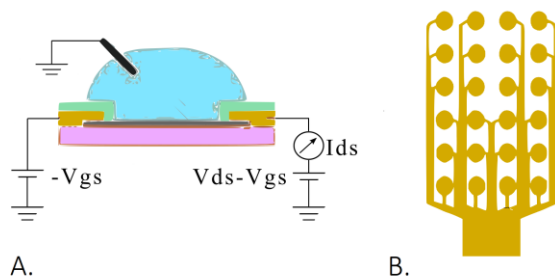


Figure 2: gSGFET visualization with applied gate-source and drain-source polarizations. Drain-source current  $I_{ds}$  is measured at drain B. Multielectrode array structure visualization with positions for multiple gSGFETs.

## ADVANTAGES OF THE COMBINATION OF THE TECHNOLOGIES

The use of gSGFETs to record the outcomes of magnetic stimulation with MENPs offers several benefits, demonstrating a promising match between these two technologies and marking a significant advancement in neurotechnology.

### A. Reduced Induced Artifacts from Magnetic Fields and Less Interactions

A significant advantage of using MENPs for stimulation is their reliance on low-intensity magnetic fields [18], which opens up possibilities for employing various waveform types beyond those used in traditional transcranial magnetic stimulation (TMS). In TMS, short pulses of high intensity, up to 1.5 T, typically saturate recording systems, which are only able to record when stimulation is not occurring. This causes manageable problems in recordings since the duration of the pulse is only in the range of microseconds. In contrast in the case of MENPs, using lower magnetic fields of some mT allows for the exploration of alternative waveforms, such as sinusoidal signals. However, even at these reduced intensities, saturation issues in recording systems can occur, necessitating measures to mitigate these artifacts.

Graphene material exhibits low magnetic susceptibility, making gSGFETs an effective solution for recording in environments with magnetic interference. The stability of graphene's electrical properties in magnetic fields, as demonstrated by Harrysson Rodrigues et al. [19], further underscores the suitability of gSGFETs for such uses. Additionally, a study by Zhao et al. [20] showed that graphene fiber electrodes used for deep brain stimulation in conjunction with fMRI effectively mapped activation patterns without being disrupted by the MRI scanner's magnetic field. This confirms that graphene electrodes can function reliably in magnetic environments, making them ideal for simultaneous stimulation and recording applications.

Therefore, combining MENPs and gSGFETs has several beneficial effects, such as reducing artifacts in

electrophysiological recordings and enabling the examination of more waveforms and frequencies of stimulation. Additionally, the minimal magnetic force interaction between the MENPs and gSGFETs prevents the alteration of MENP distribution that could occur with metallic electrodes.

### B. Enhanced Spatial Resolution

High resolution in neural recording and stimulation is essential to precisely target and record from specific neuronal populations or individual neurons. This level of precision is critical for unraveling the complex mechanisms of neural communication, synaptic dynamics, and network functionality. The integration of MENPs with gSGFETs, could greatly enhance this capability, allowing researchers to probe neural circuits with the necessary spatial fidelity. MENPs offer targeted brain stimulation with high spatial precision [18], while graphene electrodes provide the flexibility and high-resolution recording critical for capturing neural activity [13-16]. Furthermore, the inherent non-uniformity of the magnetic field in magnetic stimulation applications, coupled with the potentially non-uniform distribution of the applied nanoparticles, underscores the need for high spatial resolution in recordings. This is critical for accurate interpretation of neural responses to magnetic stimulation and for precise delivery of therapeutic interventions. Together, MENPs and gSGFETs promise to advance our understanding of brain function by facilitating precise modulation and detailed observation of neuronal activity.

### C. Recording DC potentials and Infra-slow oscillations

The incorporation of graphene electrodes, known for their sensitivity in recording direct current (DC) potentials and Infra-slow oscillations [13-16], in conjunction with a stimulation device is critical. Recent literature suggests that such neural activity may occur during or after transcranial direct current stimulation (tDCS) [16]. Given the shared principles between magnetic and electrical stimulation, it is plausible to expect similar results in experiments involving stimulation with magnetoelectric nanoparticle systems (MENPS).

Exploring the relationship between subthreshold stimulation and potential changes in direct current (DC) potentials or infra-slow neural activity presents an intriguing research opportunity. Stimulation with magnetoelectric (ME) materials often involves magnetic fields in the range of tens of milliteslas for brain stimulation [2-6]. This intensity is generally considered to be below the threshold required to produce noticeable effects. Therefore, a study investigating whether such subthreshold stimulation levels can lead to changes in DC potentials or infra-slow activity could significantly enhance our understanding of neural responses to ME materials.

This investigation is feasible, especially in experiments with MENPS. Typically, only magnetic field is used as a negative control in these studies. Using a plain magnetic field as a baseline allows for a clearer distinction between the unique effects of magnetoelectric materials and the inherent activity of the brain. This methodological approach may provide valuable insights into the subtleties of brain responses to subthreshold magnetic stimulation.

#### D. Minimal invasiveness, broadband recordings and biocompatibility

MENPs enable miniature, wireless neural stimulation devices combined with the flexibility of graphene electrodes could lead to minimally invasive neural interfaces. The wide frequency response range of graphene complements the ability of ME materials to operate over various frequencies, allowing versatile neural modulation and recording. Finally, gSGFETs and MENPs are both known for their biocompatibility, making them well-suited for use in neural interfaces, with minimal risk of biological rejection or adverse reactions.

In conclusion combining magnetoelectric stimulators with graphene electrode recording systems unlocks several benefits. This innovative approach holds great promise for future developments in neural technology, including advanced brain-computer interfaces and sophisticated neurological research tools.

### A PROPOSED SETUP FOR ELECTROPHYSIOLOGY EXPERIMENTS

#### A. HARDWARE DESCRIPTION

The proposed setup consists of a ME stimulator and a gSGFET multichannel recording system. The ME stimulator is similar with the one presented in [21], it is designed for *in vitro* and *in vivo* experiments, features a two-channel power capability using a Class-D audio amplifier, each channel delivering up to 100W. This design enables the operation of two experimental setups simultaneously. The stimulator can generate magnetic fields up to 20 mT RMS, adjustable based on selected protocols and frequencies. A microcontroller collects temperature, current and other important measurements. A capacitance board allows the capacitance to be adjusted to meet specific frequency requirements. The system includes a specially designed circular coil, ensuring effective magnetic field generation across various experimental conditions, including those in electrophysiology chambers.

The recording system is an innovative 64/128-channel amplification system for neural signal processing, incorporating graphene-based transistors and has been tested in several experiments [16-17], [22]. The system includes key components such as gSGFETs, a breakout board, a preprocessing device, an amplifier, and a

software interface.

A critical element, the preprocessing device, is responsible for converting and preamplifying analog signals and setting bias voltages for the gSGFETs. Its architecture includes a mainboard, several analog modules, a digital board, and an accumulator board. The mainboard's primary role is to process signals from electrodes through analog modules, converting them from current to voltage, and segregating them into DC and AC components, as detailed in the circuitry described by C. Hébert et al [8]. This design allows versatility in using different electrode arrays and includes functionalities for electrode characterization.

Each analog module on the mainboard can process eight channels, with the capability to adjust to various electrode configurations. Furthermore, these modules are equipped with bypass switches to facilitate electrode characterization. The digital board features a Bluetooth module-microcontroller for wireless adjustments of voltage settings and managing bypass switches.

Post-preprocessing, the signal gets divided into 64 AC and 64 DC channels. The amplification stage employs the g.RAPHENE device, an advanced version of the g.Hiamp with 128 input channels, 64 for AC and 64 for DC amplification. The g.Hiamp [23], a high-performance biosignal amplifier developed by g.tec medical engineering, is noted for its excellent signal resolution and sensitivity. Its application in ultra-high-density electroencephalography has been pivotal, as indicated by recent data from g.Hiamp recordings.

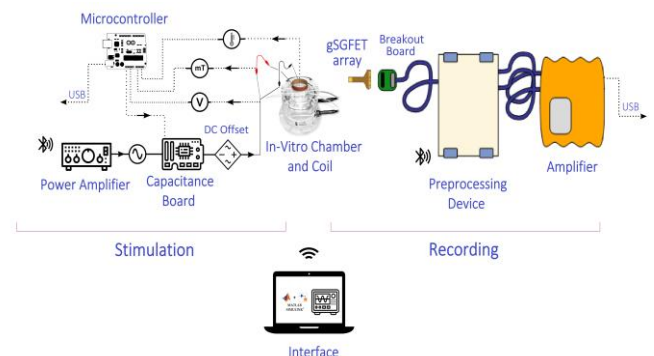


Figure 3 Hardware diagram of the proposed system

#### B. SOFTWARE DESCRIPTION

A common software interface controls the two systems developed in MATLAB/Simulink.

The user interface of the ME stimulator communicates seamlessly with the amplifier via Bluetooth, facilitating easy control even through mobile apps. The software supports various waveform options, enhancing the adaptability of the system to diverse research requirements. Its ability to simulate the magnetic/electric field effects and estimate coil temperature rise based on

selected protocols is crucial for evaluating stimulation parameters and ensuring safety.

For the part of the graphene device the interface supports intricate tasks like electrode characterization and recording execution. It includes features for configuring Bluetooth communication, setting amplifier parameters, and specifying voltage sweeps. The system provides real-time visualizations of both AC and DC signal components, allowing for detailed analysis of neural activities

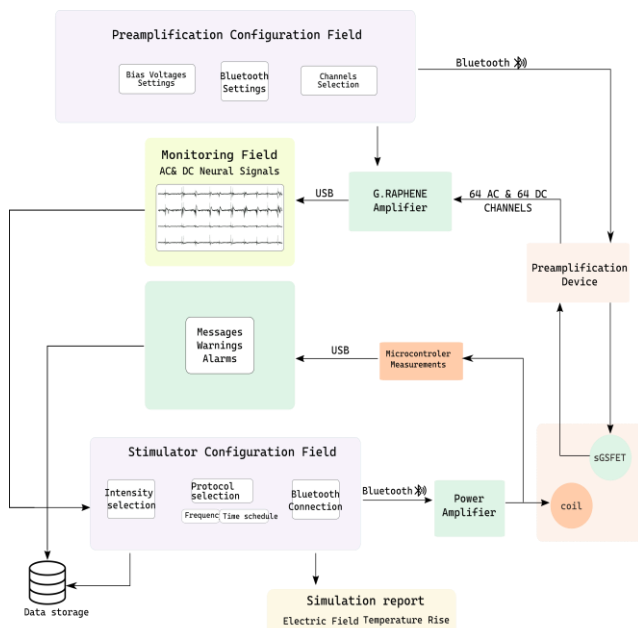


Figure 4 Software diagram of the proposed system.

In summary, the combined magnetoelectric stimulator and graphene electrode recording systems present a sophisticated, highly adaptable, and user-friendly platform. This integrated approach enhances the capability for precise neural stimulation and recording, opening new avenues for advanced neuroscience research and potential therapeutic applications.

## DATA ANALYSIS METHODS

In the direction of further development and evolution of a system integrating gSGFETs with MENP-based magnetic stimulation, especially in local field potential (LFP) recordings, a thorough research approach could focus on some data analysis methods.

Time-frequency analysis is a technique in which LFP signals are decomposed into their component frequencies over time using methods such as wavelet transforms or short-time Fourier transforms. It's useful for identifying changes in power over different frequency bands, including Infra-slow oscillations. Research such as that of Unakafova and Gail [24] provides a practical guide for neuroscientists in selecting open-source toolboxes for spike and LFP data analysis, including those with time-frequency analysis functionality. Spike-field coherence

analysis examines the relationship between neuronal spiking activity (UP states) and LFPs. It can show how changes in UP and DOWN states correlate with fluctuations in slower frequency bands, including DC shifts. Henningson and Illes [25] proposed a model to study subthreshold fluctuations.

Phase-amplitude coupling (PAC) analysis examines the interaction between the phase of lower frequency oscillations, such as infraslow or theta bands, and the amplitude of higher frequency activity. This approach is particularly relevant for studying the interplay between Infra-slow oscillations and fast neuronal dynamics. An important application of PAC analysis is demonstrated in the study by Hiroaki Hashimoto et al [26], who found that PAC between infra-slow and high-frequency activity can effectively discriminate between preictal and interictal states in epilepsy, underscoring its potential as a useful biomarker. Cross-frequency coupling (CFC) examines the relationship between different frequency ranges in neural signals [27]. It can be used to study how infraslow oscillations influence, or are influenced by, other frequency bands in LFP data. Finally, machine learning approaches [28], such as neural networks or support vector machines, can be trained to classify and predict patterns in LFP data, taking into account both fast neural fluctuations and slower DC shifts. infra-slow activity or DC potentials following stimulation.

## DISCUSSION

The current work explores the combination of neural stimulation with MENPs and recording with gSGFETs as a novel technique for future brain-computer interfaces. Significant benefits such as reduced artifacts, enhanced spatial resolution, and detection of subthreshold phenomena are highlighted. The proposed hardware and software setup is designed to accommodate a range of experimental conditions with ease of use, while comprehensive data analysis methods, including time-frequency analysis and machine learning approaches, enable detailed interpretation of the intricate neural signals recorded.

Overall, this integrated approach represents a significant step forward in neural engineering, promising advances in neurological research and potential therapeutic applications.

## ACKNOWLEDGMENT

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101130650 (META-BRAIN).

## REFERENCES

- [1] H. Palneedi, V. Annapureddy, S. Priya, and J. Ryu, "Status and Perspectives of Multiferroic

- Magnetoelectric Composite Materials and Applications,” *Actuators*, vol. 5, no. 1, p. 9, Mar. 2016.
- [2] S. Kopyl, R. Surmenev, M. Surmeneva, Y. Fetisov, and A. Kholkin, “Magnetoelectric effect: principles and applications in biology and medicine– a review,” *Mater. Today Bio*, vol. 12, p. 100149, Sep. 2021.
  - [3] A. Singer et al., “Magnetoelectric Materials for Miniature, Wireless Neural Stimulation at Therapeutic Frequencies,” *Neuron*, vol. 107, no. 4, pp. 631–643.e5, Aug. 2020.
  - [4] K. L. Kozielski et al., “Nonresonant powering of injectable nanoelectrodes enables wireless deep brain stimulation in freely moving mice,” *Sci. Adv.*, vol. 7, no. 3, p. eabc4189, Jan. 2021.
  - [5] T. Nguyen et al., “In Vivo Wireless Brain Stimulation via Non-invasive and Targeted Delivery of Magnetoelectric Nanoparticles,” *Neurotherapeutics*, vol. 18, no. 3, pp. 2091–2106, Jul. 2021.
  - [6] P. Wang et al., “Colossal Magnetoelectric Effect in Core–Shell Magnetoelectric Nanoparticles,” *Nano Lett.*, vol. 20, no. 8, pp. 5765–5772.
  - [7] M. Z. Lin and M. J. Schnitzer, “Genetically encoded indicators of neuronal activity,” *Nature Neuroscience*, vol. 19, no. 9, pp. 1142–1153, Sep. 2016, doi: <https://doi.org/10.1038/nn.4359>.
  - [8] Z. Wei, B.-J. Lin, T.-W. Chen, K. Daie, K. Svoboda, and S. Druckmann, “A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology,” *PLOS Comput. Biol.*, vol. 16, no. 9, p. e1008198, Sep. 2020.
  - [9] C. Hébert et al., “Flexible Graphene Solution-Gated Field-Effect Transistors: Efficient Transducers for Micro-Electrocorticography,” *Advanced Functional Materials*, vol. 28, no. 12, p. 1703976, Nov. 2017.
  - [10] E. Masvidal-Codina et al., “High-resolution mapping of infraslow cortical brain activity enabled by graphene microtransistors,” *Nature Materials*, vol. 18, no. 3, pp. 280–288, Dec. 2018.
  - [11] A. Camassa et al., “Chronic full-band recordings with graphene microtransistors as neural interfaces for discrimination of brain states,” *Nanoscale Horizons*, Feb. 2024.
  - [12] M. Lee et al., “Graphene-electrode array for brain map remodeling of the cortical surface,” *NPG Asia Materials*, vol. 13, no. 1, pp. 1–10, Oct. 2021.
  - [13] A. Mitra et al., “Spontaneous infra-slow brain activity has unique spatiotemporal dynamics and laminar structure,” *Neuron*, vol. 98, no. 2, Apr. 2018.
  - [14] Sandro Lecci et al., “Coordinated infraslow neural and cardiac oscillations mark fragility and offline periods in mammalian sleep,” *Science Advances*, vol. 3, no. 2, Feb. 2017.
  - [15] R. Garcia-Cortadella et al., “Graphene active sensor arrays for long-term and wireless mapping of wide frequency band epicortical brain activity,” *Nat Commun*, vol. 12, no. 1, p. 211, Jan. 2021.
  - [16] N. Cancino-Fuentes et al., “Recording physiological and pathological cortical activity and exogenous electric fields using graphene microtransistor arrays in vitro,” *Nanoscale*, vol. 16, no. 2, pp. 664–677, 2024.
  - [17] E. Masvidal-Codina et al., “Characterization of optogenetically-induced cortical spreading depression in awake mice using graphene microtransistor arrays,” *Journal of Neural Engineering*, vol. 18, no. 5, p. 055002, Apr. 2021.
  - [18] S. Fiochi, E. Chiaramello, A. Marrella, M. Bonato, M. Parazzini, and P. Ravazzani, “Modelling of magnetoelectric nanoparticles for non-invasive brain stimulation: a computational study,” *Journal of Neural Engineering*, vol. 19, no. 4, June 2022.
  - [19] S. Zhao, G. Li, C. Tong, W. Chen, P. Wang, J.-k. Dai, X. Fu, Z. Xu, X. Liu, L. Lu, Z. Liang, and X. Duan, “Full activation pattern mapping by simultaneous deep brain stimulation and fMRI with graphene fiber electrodes,” *Nature Communications*, vol. 11, no. 1, May 2020.
  - [20] I. Harrysson Rodrigues, A. Generalov, M. Soikkeli, A. Murros, S. Arpiainen, and A. Vorobiev, “Geometrical magnetoresistance effect and mobility in graphene field-effect transistors,” *Applied Physics Letters*, vol. 118, no. 8, Feb. 2022.
  - [21] S. Matsoukis et al., “Development of a magnetoelectric neural stimulator for in vitro experiments,” 2023 IEEE Biomedical Circuits and Systems Conference (BioCAS), Toronto, ON, Canada, 2023.
  - [22] A. Bonaccini Calia et al., “Full-bandwidth electrophysiology of seizures and epileptiform activity enabled by flexible graphene microtransistor depth neural probes,” *Nature Nanotechnology*, vol. 17, no. 3, pp. 301–309, Dec. 2021.
  - [23] “g.HIamp 256-Channel Biosignal Amplifier,” g.tec medical engineering GmbH. <https://www.gtec.at/product/ghiamp/> (accessed Mar. 12, 2024)
  - [24] V. A. Unakafova and A. Gail, “Comparing Open-Source Toolboxes for Processing and Analysis of Spike and Local Field Potentials Data,” *Frontiers in Neuroinformatics*, vol. 13, Jul. 2019.
  - [25] M. Henningson and S. Illes, “Analysis and Modeling of Subthreshold Neural Multi-Electrode Array Data by Statistical Field Theory,” *Frontiers in Computational Neuroscience*, 2017.
  - [26] H. Hashimoto et al., “Phase-amplitude coupling between infraslow and high-frequency activities well discriminates between the preictal and interictal states,” *Scientific Reports*, vol. 11, no. 1, Aug. 2021.
  - [27] J. Schumacher, R. Haslinger, and G. Pipa, “Statistical modeling approach for detecting generalized synchronization,” *Physical Review E*, vol. 85, no. 5, May 2012.
  - [28] M. I. Fabietti, M. Mahmud, A. Lotfi, et al., “Early Detection of Alzheimer’s Disease From Cortical and Hippocampal Local Field Potentials Using an Ensembled Machine Learning Model,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023

# COMPARING FINGERS AND GESTURES FOR BCI CONTROL USING AN OPTIMIZED CLASSICAL MACHINE LEARNING DECODER

D. Keller<sup>1</sup>, M. J. Vansteensel<sup>1</sup>, S. Mehrkanoon<sup>2</sup>, M. P. Branco<sup>1</sup>

<sup>1</sup>UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, Utrecht, The Netherlands

<sup>2</sup>Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

E-mail: m.pedrosobranco@umcutrecht.nl

**ABSTRACT:** Severe impairment of the central motor network can result in loss of motor function, clinically recognized as Locked-in Syndrome. Advances in Brain-Computer Interfaces offer a promising avenue for partially restoring compromised communicative abilities by decoding different types of hand movements from the sensorimotor cortex. In this study, we collected ECoG recordings from 8 epilepsy patients and compared the decodability of individual finger flexion and hand gestures with the resting state, as a proxy for a one-dimensional brain-click. The results show that all individual finger flexion and hand gestures are equally decodable across multiple models and subjects (>98.0%). In particular, hand movements, involving index finger flexion, emerged as promising candidates for brain-clicks. When decoding among multiple hand movements, finger flexion appears to outperform hand gestures (96.2% and 92.5% respectively) and exhibit greater robustness against misclassification errors when all hand movements are included. These findings highlight that optimized classical machine learning models with feature engineering are viable decoder designs for communication-assistive systems.

## INTRODUCTION

Dysfunction of the neuromotor system may precipitate transient or, in severe cases, enduring global loss of motor control. Global dysfunction may be referred to as Locked-In Syndrome (LIS) [1], often characterized by quadriplegia and aphonia. In recent decades, efforts to replace dysfunctional motor control have seen pioneering developments in Brain-Computer Interfaces (BCIs) [2]. BCIs extract information directly from cortical activity to control mechanical or digital effectors without relying on neuromuscular activation, essentially bypassing the muscular output. Restoration of effector control can serve several purposes, ranging from object manipulation [3–5], locomotion and mobility [6] and speech production [7–9]. However, for individuals with severe impairment, the restoration of the communicative agency has been identified as one of the most urgent needs [10]. A simple approach towards communication BCI is automatic letter selection on a digital keyboard [11]. An attractive signal recording modality for communication BCIs is electrocorticography (ECoG) due to its high spa-

tiotemporal precision, good signal-to-noise ratio, and reliable signal stability over extended periods [11, 12]. Several studies have demonstrated that hand movement recognition from ECoG recordings can be performed with high accuracy. Consequently, a variety of hand movements have been explored for this purpose, including but not limited to finger flexion [13–17], reaching and grasping [3–5, 18, 19], and wrist flexion and extension [15], more complex hand gestures [15, 20–22] and handwriting [23]. To identify a reliable motor signature for a unidimensional BCI control signal (i.e., 'brain-click') many studies have examined different types of hand movements in isolation [13, 14, 20–22], but few have compared different hand movements against each other within a unified framework.

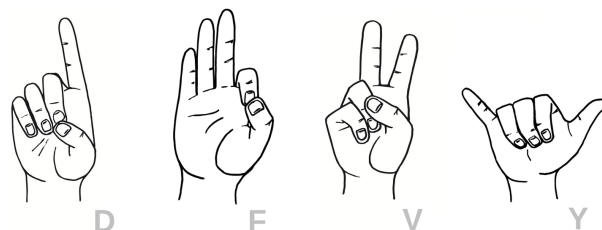


Figure 1: The four gestures executed by subject S1 - S5.

The goal of this work is to contribute to a deeper understanding of the decodability of individual finger flexion and hand gestures against the resting state. Specifically, we explored which hand strategy is the most promising for a reliable brain-click and which is more transferable across subjects. In addition, we aim to extend our analysis to the prospect of multidimensional control with four and eight degrees of freedom (DoF) to investigate which type of hand movement intrinsically yields a better within and across-category discriminability. To overcome the notorious data sparsity in this domain, we employed an optimized feature selection decoder with different classification models and assessed which (offline) machine learning approach yields the best performance on individual hand movements (2-DoF), within hand movement types (4-DoF), and within all hand movements (8-DoF).

## MATERIALS AND METHODS

**Data description:** This study is based on two ECoG data sets, consisting of finger flexion, including the index finger, little finger, and thumb, and hand gestures associated with the American Sign Language letters D, F, V, and Y [21, 24] (see Figure 1).

The data were collected from 8 subjects ( $N_G = 5$ ,  $N_F = 6$ ) in an epilepsy monitoring unit of the University Medical Center Utrecht (see Table 1). These subjects had 32, 64 or 128 high-density subdural ECoG electrodes with an inter-electrode distance of 3 or 4mm and an exposed diameter of 1 to 1.3mm (AdTech, Racine, USA; or PMT Corporation, Chanhassen, MN, USA) implemented over the hand-knob region of the sensorimotor cortex. The ECoG data were recorded using a 128-channel Micromed LTM system (subjects 1 - 5; Treviso, Italy; 22 bits, hardware bandpass filter 0.15–134.4Hz; sampling frequency 512Hz) and a Blackrock system (subjects 6 - 8; Microsystems LLC, Salt Lake City, USA, digital bandpass filter 0.3 - 500Hz; sampling frequency 2000Hz). Data were converted to the BIDS standard format [25].

Table 1: Subject Details

Sub- ject	Task	Trials (C / T)	Age	Sex	Hand	Hand- edness	Hemi- sphere	Grid (incl.)
S1	G F	37 / 74 90 / 181	19	F	Right	Right	Left	4x8 (32)
S2	G	68 / 138	45	F	Left	Left	Right	8x8 (59)
S3	G	34 / 69	29	M	Right	Right	Left	4x8 (29)
S4	G F	32 / 67 90 / 181	19	M	Right	Right	Left	4x8 (31)
S5	G F	34 / 69 88 / 177	42	M	Right	Right	Left	4x8 (32)
S6	F	89 / 179	30	F	Left	Right	Right	16x8 (123)
S7	F	85 / 171	20	F	Right	Right	Left	8x8 (64)
S8	F	84 / 169	36	F	Right	Right	Left	16x8 (128)

*Note.* Trials are presented as the ratio of hand movement trials per condition (C) out of all trials (T; including the rest trials). In Grid, (incl.) indicates the number of channels included. Abbreviations: Gesture, G; Finger, F; Male, M; Female, F.

**Experimental Design:** Subjects were instructed to initiate movements based on visual cues that were presented in a randomized, event-driven design. For the gestures, the subjects imitated the depicted gesture after stimulus onset and maintained the posture until the end of the trial before returning to a resting position. Each subject performed 10 trials with an intertrial interval of 4.4s and a run duration of 6.7m. Rest trials were implicitly calculated from a small time interval before the onset of the next movement. For finger flexion, the subjects performed two finger flexions immediately after cue onset and then returned to a resting position afterward. In contrast to the gestures, each movement was interleaved with an explicit resting trial. The design consisted of 30 trials with an intertrial interval of 7s and a run duration of 8.2m. In both experiments, each subject performed the tasks with the hand contralateral to the grid location, and subject 2 performed the task twice. In addition, a data

glove (5DT, Irvine CA, USA, 20 ms sampling time) was used during both experiments to record motor activity.

**Preprocessing:** Data preprocessing included the removal of bad trials and channels (identified by [21, 24] based on data glove data and raw signal inspection), followed by common average referencing, notch, and band-pass filtering (56 Hz - 130 Hz) to remove artifacts. Finger flexion data, sampled at 2000 Hz, were downsampled to 512 Hz for consistency across subjects. The data were then subsequently aligned with movement onset markers obtained from data glove recordings and segmented accordingly. For decoding individual hand movements and within hand movement types a segmentation window of  $W_F = [-0.5, 1.5s]$  and  $W_G = [-0.5s, 2.5s]$  was used. For decoding all hand movements, the two 4-DoF settings for subject 5 were combined with a common segmentation window of  $W_{FG} = [-0.5s, 2s]$ . In all three settings  $t = 0$  represents the motion-aligned stimulus presentation.

Features were extracted using a continuous Morlet wavelet transformation, which produced spectral power features for the high-frequency band (60 Hz - 126 Hz) in 2 Hz frequency bins. To reduce the feature space the power was averaged and the time dimension was decimated to  $T_G = 154$  and  $T_F = 102$  time points per channel for the fingers and gestures, respectively. The resulting feature vectors were used for subsequent model training. Preprocessing was conducted in Python (v3.9) using the MNE library (v1.16).

**Decoder:** The architecture of the decoder, depicted in Figure 2, revolves around an optimized data-driven feature engineering approach for conventional classical machine learning classifiers. The decoder encompasses four modules: Normalization, Incremental Feature Selection (IFS), Feature Reduction (FR), and Classification.

The initial step of the decoder normalises the spectral power of the spatio-temporal feature vector ( $N$ ) via a Box-Cox transformation [26], followed by mean centering and unit variance scaling to ensure data normality and variance stabilisation. Feature selection employs Variance Thresholding (VT), Mutual Information Criterion (MIC) [27, 28], and Recursive Feature Elimination (RFE). RFE iteratively removes a set of features ( $S = 1e-3 * N$ ) corresponding to the least important coefficients, and akin to MIC, retains a subset of the best  $K$  temporal features across all channels. Each selection method generates a binary mask indicating the retained features. The individual modules are applied incrementally, with the specific combination and its parameter configuration being delegated to a Bayesian optimization algorithm, which avoids manual tuning and efficiently navigates through the high-dimensional parameter space. In particular, this approach aims to balance the advantages of filter and wrapper methods [29] to remove noisy and redundant features while prioritising discriminative ones. The Feature Reduction (FR) method can be applied in isolation or in conjunction with IFS processing. In the classification phase, four algorithms were selected to compete with each other: Logistic Regression (LR), Linear Discrim-

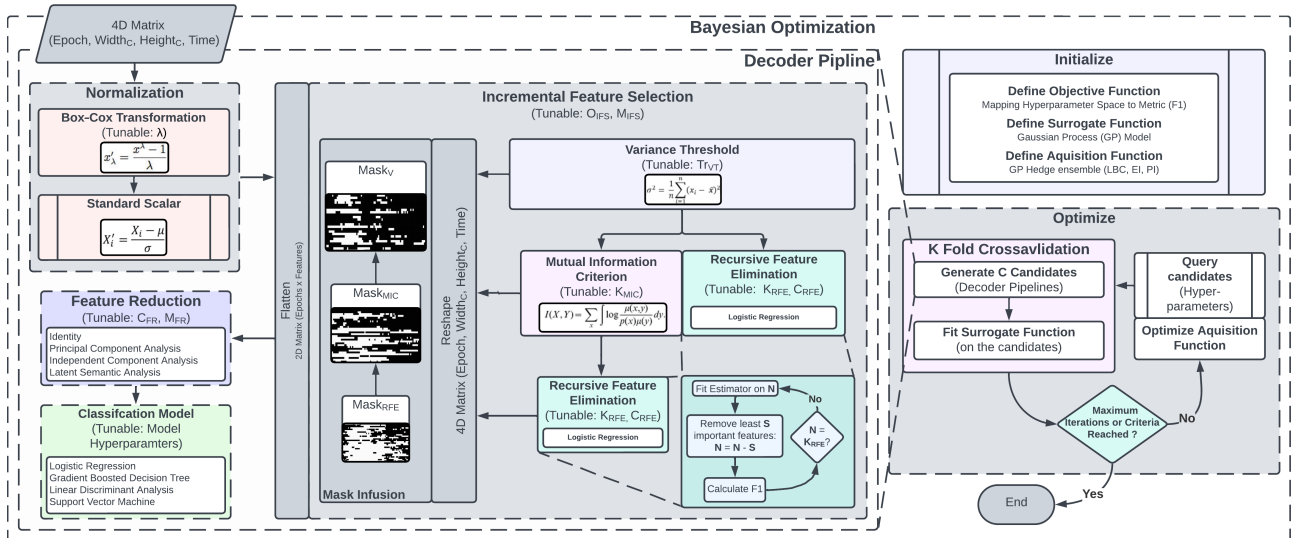


Figure 2: General decoder architecture. The pipeline is composed of four modules: Normalization, Incremental Feature Selection (IFS), Feature Reduction (FR), and a Classification Model. A Box-Cox (BC) transformation and a Standard Scalar normalize the signal. The incremental selection procedure can recruit different combinations of Variance Thresholding (VT), Mutual Information Criterion (MIC), and Recursive Feature Elimination (RFE) to select the relevant features from the spatio-temporal representation. RFE iteratively removes the set of  $S$  least important features from the pool  $N$  until it reaches  $K_{RFE}$  features. After feature selection, the feature space can be further compressed with various Feature Reduction (FR) methods. The resulting vector is forwarded to one of four classifiers. A Bayesian optimization algorithm orchestrates the order and method of feature selection ( $O_{IFS}$ ,  $M_{IFS}$ ), other pipeline configuration ( $\lambda_{BC}$ ,  $\mu$ ,  $\sigma$ ,  $Tr_{VT}$ ,  $K_{MIC}$ ,  $K_{RFE}$ ,  $C_{RFE}$ ,  $C_{FR}$ ), and various model-specific hyperparameters. Dotted lines represent meta-routing processes; single and double-lined boxes represent dynamic and predefined processes, respectively. Abbreviations: LBC, Lower Confidence Bound; EL, Negative Expected Improvement; PI Negative Probability Improvement

inant Analysis (LDA), Boosted Decision Trees (BDT), and Support Vector Machines (SVM). A Majority Class Predictor, which predicts the most frequent hand movement, was evaluated on the data to establish a 'chance' baseline. Performance is assessed using the F1 scoring metric, adjusted for label imbalance with the inversely weighted class distribution.

For this 'black-box' optimization problem, the Bayesian algorithm [30] approximates an expensive non-smooth objective function by inference, essentially guiding the search process based on prior results. To find an optimal decoder candidate in the large parameter space, a Gaussian Process model [31] with a hedging portfolio strategy [32] is used, where hedging probabilistically choose the best acquisition function from three candidates: Lower Confidence Bound, Negative Expected Improvement or Negative Probability Improvement. The search is restricted to a maximum of 256 candidates, exploring a hyperparameter space, ranging from 9 (LDA) to 20 (BDT) configurations for different algorithms, of which up to 60% are conditional hyperparameters; the number-of-components hyperparameter for the FR step was shared among all three methods. Model performance is evaluated using stratified 10-fold crossvalidation, with the best candidate further assessed through leave-one-out crossvalidation. The decoder pipeline adheres to the scikit-learn architecture, ensuring compatibility with the scikit-learn library and its derivatives. The implementation is in Python 3.9, using scikit-learn (v1.4.0) and xgboost (v2.0.3).

**Statistical Analysis:** The analysis relies on a Friedman ANOVA to identify a general effect and Dunn's test with Benjamini-Hochberg's false discoveries rate correction the post hoc analysis and pairwise comparison. The statistical analysis was performed in Python 3.9, using scipy (v1.13.0) and scikit-posthocs (v0.9.0).

## RESULTS

**Individual Hand Movements:** In the context of individual hand movements, all classification models exhibited a high F1 performance (averaged across subjects), exceeding 98.0% for each finger flexion and gesture (details summarised in Table 2 and Figure 3 A). Notably, the Index finger (99.59%), Gesture V (99.11%), and Gesture F (98.09%) were the most promising candidates for brain-click BCI control. Interestingly, they share a commonality in index finger flexion. Within the hand gestures, no gesture significantly outperformed the others (Friedman ANOVA and Dunn's test; ns). Similarly, within the finger flexions, the overall effect was significant,  $\chi^2(3) = 15.32$ ,  $p = 4.71e^{-4}$ , but no finger flexion was significantly different from the others (Dunn's test). Moreover, decoding performance remained remarkably stable for each subject across all hand movements and models, with consistent trends in variability observed for each subject (i.e., subject 1 consistently had the lowest and highest scores for the fingers and the gestures, respectively).

**Hand Movements Types:** For the hand movement types, a different trend emerged. For 4-DoF classifica-

Table 2: Mean performance across all subjects of the Optimize Feature Selection Decoder for Individual Movements (2-DoF), Within Types (4 DoF) and Within all Hand Movements (8-DoF).

Models	Fingers (vs. rest)						Gestures (vs. rest)								Multi-DoF (incl. rest)					
	Index		Little		Thumb		Gesture D		Gesture F		Gesture V		Gesture Y		4-Finger		4-Gesture*		8-Hand**	
	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1
Chance	75.3	64.6	75.5	64.9	74.9	64.2	83.3	75.7	79.8	70.8	79.4	70.3	79.8	70.9	50.3	33.7	56.6	40.9	50.4	33.8
LR	98.2	98.3	98.3	98.4	95.9	96.0	97.4	97.5	98.2	98.2	98.7	98.7	96.6	96.8	95.7	95.7	91.3	91.5	87.4	87.7
BDT	99.6	<b>99.6</b>	98.1	98.1	98.1	<b>98.0</b>	97.6	97.7	98.7	98.7	98.4	98.5	98.2	98.2	95.8	95.8	92.0	92.0	88.6	88.7
LDA	99.3	99.3	98.7	<b>98.8</b>	97.5	97.6	98.1	<b>98.1</b>	98.7	98.7	98.0	98.0	96.4	96.6	96.1	96.1	91.1	91.1	93.1	<b>92.9</b>
SVM	99.5	99.4	98.8	<b>98.8</b>	97.7	97.6	97.9	98.0	98.8	<b>98.89</b>	99.1	<b>99.1</b>	98.4	<b>98.4</b>	96.7	<b>96.6</b>	92.5	<b>92.5</b>	92.7	92.6

Note. Values are in %. \*The 4-DoF gesture decoding includes Gesture F, Y, and V. \*\* The 8-DoF decoding of all hand movements was only obtained from subject 5. Abbreviations: Accuracy, A; Versus, vs; Inclusive, incl.

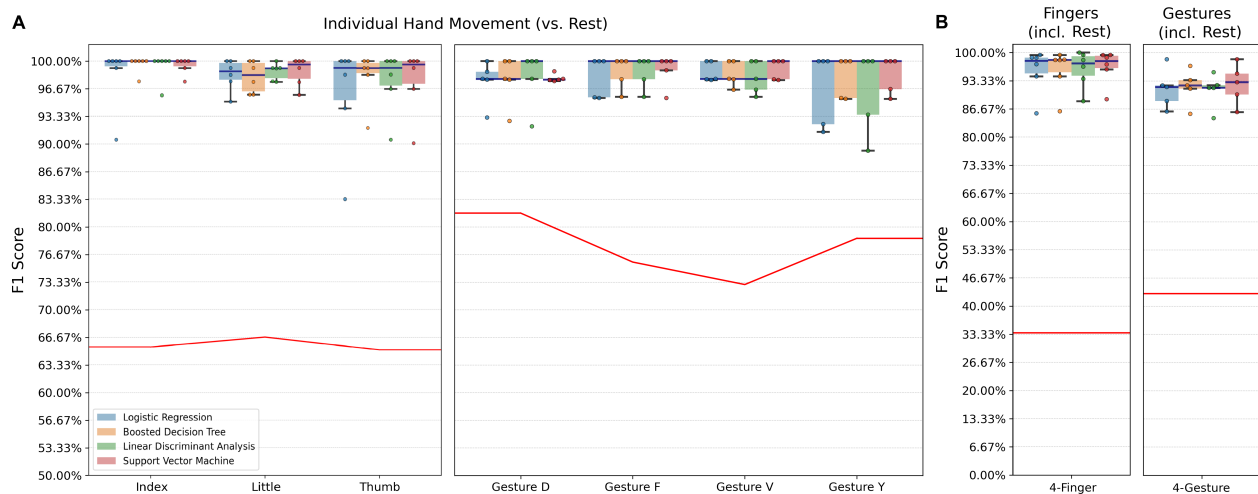


Figure 3: Box plots for different machine learning models of the decoder for (A) all individual fingers and (B) the three multi-DoF comparisons. Each point represents one subject, with the upper and lower error bars representing an interquartile range of 25 and 75, respectively, and where omitted when performance for one or more subjects exceeded this range. The 4-DoF gesture decoding includes Gesture F, Y, and V. The red line represents the highest chance level among all subjects.

tion, finger flexion (95.7% - 96.6%) outperformed the hand gestures (91.1% - 92.5%; average across subjects), which could be statistically verified (Dunn's test),  $p = 0.031$ ; we excluded the worst decodable gesture (Gesture D) to ensure a similar task complexity. An extension to 8-DoF classification preserves a high F1 score for Subject 5 (88.7% - 92.9%), with remarkably minimal confusion between gestures and fingers. On visual inspection, fingers exhibit more confusion with the resting state, while gestures are more often confused among themselves (as depicted in Figure 4).

**Classification Models:** When training classification models within an optimised feature engineering framework, no model emerges as significantly superior to the others (Friedman ANOVA;),  $\chi^2(4) = 60.48$ ,  $p = 2.3e^{-12}$ , (Dunn's test; ns), although, all perform significantly above chance level (Dunn's Test),  $p < 1.28e^{-8}$ . In general, Boosted Decision Trees (BDT) and Support Vector Machines (SVM) demonstrate the highest classification performance across all conditions, except for Little Finger and Gesture D.

## DISCUSSION

The current work demonstrates that optimised spatio-temporal feature engineering of finger flexion and hand gestures, recorded from high-density ECoG, enables reliable decoding for one-dimensional brain-click, 4- and 8-DoF decoding tasks, even with very small data volumes. Notably, within each category, no single hand movement emerged as superior decodable. However, upon qualitative inspection, the index finger and Gestures V and F, all sharing index finger flexion, appeared as the most promising candidates. Moreover, the fingers exhibited a better performance in more complex 4- and 8-DoF decoding; finger flexion may possess more discriminative properties for multi-DoF tasks. Finally, no classical machine learning model outperformed the others, but BDT and SVM may have a small advantage.

Hand gestures entail a more complex interplay of motor components than simple finger flexion, including wrist and finger flexion, and lateral extension, among others. However, our results revealed that individual finger flexion alone yields a near perfect neuroelectrical signature.

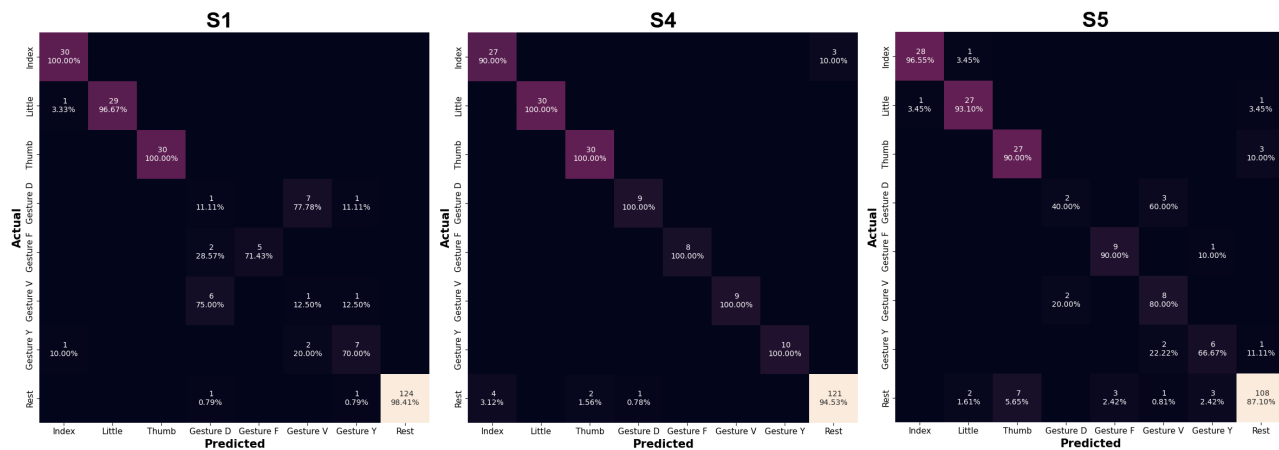


Figure 4: Confusion matrices of the LR model for subjects 1, 4, and 5. In each confusion matrix, the horizontal axis represents the predicted hand movement (or rest), and the vertical axis represents the ground truth hand movement (or rest). Henceforth, the diagonal elements represent the total correct values predicted per hand movement. The colour coding represents the proportion of absolute counts for each hand movement ranging from black (0%) to white (100%).

Consequently, the addition of supplementary motor components may not increase decodability but rather confront the decoder with a motor signature that inherently has a higher variability in its signal. As we evaluated the decoder on small data volumes, an effect of additional components on decoder performance may emerge with larger sample sizes. Extending this rationale, gestures might possess a more intricate spatial and temporal pattern, recruiting various neuronal populations from a larger effector field in the sensorimotor cortex. The absence of advanced feature transformation techniques in our classical machine learning approach might hinder the decoder from exploiting the full potential of the gesture's electrical signature. In addition, differences in sample size, with fewer than 30 trials for fingers and only 10 trials for gestures, may impact the comparison, while the overall small sample size may not provide sufficient power to detect a potential statistical difference.

In line with the evident discernible difference between the 4-DoF types, misclassification errors were more pronounced for gestures, with frequent confusion among different fingers and the resting state. Notably, substantial confusion between hand movements and the resting state suggests a potential contamination of the rest periods with unintended movement. This may be attributed to the design of the experimental design of the gestures, which lacked separate explicit rest trials. Incorporating a threshold based on data glove recordings could be beneficial. However, defining true rest may not be practically feasible for real BCI applications, as it necessitates extensive subject training to suppress such activity [11, 12, 19], and a more naturalistic approach would be to build a decoder that can successfully discriminate between meaningful and non-task-related sporadic motor activity. Furthermore, some features along the temporal dimension may not reflect actual motor activity, but 'resting' activity, especially before the movement onset and towards the end of the segmentation window.

The proposed decoder design surpassed prior approaches evaluated on gesture data for four out of five subjects [20, 21], albeit a direct comparison is difficult due to differences in task complexity. For communication assistance systems, the proposed decoder design can offer a viable alternative to deep learning approaches for one-dimensional brain-click tasks [19] and even larger DoF applications [14, 17], where data acquisition is challenging. Although trained offline, the decoder can process individual segments of preprocessed data in as little as 2 - 10ms. Consequently, although theoretically deployable in an online setting, regular offline retraining on new data is necessary to address concept drifts for ensuring long-term stability, in particular for individuals with neurodegenerative diseases. Importantly, the experiments involved movement execution by epileptic individuals, further validation in attempted movement is imperative to extend applicability to online BCI settings for individuals with LIS.

## CONCLUSION

Electrocorticography data provides a high-resolution spatiotemporal feature representation, forming a suitable foundation to tailor an optimised classical machine learning decoder with automatic feature engineering to the large feature space. We demonstrate that within this framework, both finger flexion and hand gestures enable reliable decoding across multiple subjects, and when extended to a multiple degrees of freedom, maintain high discriminability between hand movements. For click-based letter selection in communication-assistive BCI systems, the index finger flexion emerges as an optimal candidate. Moreover, all tested models consistently exhibit high classification performance across multiple subjects - a comparable performance to deep learning approaches.

## REFERENCES

- [1] Plum F, Posner JB. The diagnosis of stupor and coma. Oxford University Press, USA (2000).
- [2] Thakor NV. Translating the brain-machine interface. *Science translational medicine*. 2013;5(210):210ps17–210ps17.
- [3] Hochberg LR *et al.* Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*. 2012;485(7398):372–375.
- [4] Collinger JL *et al.* High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*. 2013;381(9866):557–564.
- [5] Bouton CE *et al.* Restoring cortical control of functional movement in a human with quadriplegia. *Nature*. 2016;533(7602):247–250.
- [6] Lorach H *et al.* Walking naturally after spinal cord injury using a brain–spine interface. *Nature*. 2023;1–8.
- [7] Moses DA *et al.* Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine*. 2021;385(3):217–227.
- [8] Willett FR *et al.* A high-performance speech neuroprosthesis. *Nature*. 2023;620(7976):1031–1036.
- [9] Metzger SL *et al.* A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*. 2023;620(7976):1037–1046.
- [10] Kageyama Y *et al.* Nationwide survey of 780 Japanese patients with amyotrophic lateral sclerosis: Their status and expectations from brain–machine interfaces. *Journal of Neurology*. 2020;267:2932–2940.
- [11] Vansteensel MJ *et al.* Fully implanted brain–computer interface in a locked-in patient with ALS. *New England Journal of Medicine*. 2016;375(21):2060–2066.
- [12] Pels EG *et al.* Stability of a chronic implanted brain–computer interface in late-stage amyotrophic lateral sclerosis. *Clinical Neurophysiology*. 2019;130(10):1798–1803.
- [13] Xie T, Zhang D, Wu Z, Chen L, Zhu X. Classifying multiple types of hand motions using electrocorticography during intraoperative awake craniotomy and seizure monitoring processes—case studies. *Frontiers in neuroscience*. 2015;9:353.
- [14] Xie Z, Schwartz O, Prasad A. Decoding of finger trajectory from ecog using deep learning. *Journal of neural engineering*. 2018;15(3):036009.
- [15] Thomas TM *et al.* Simultaneous classification of bilateral hand gestures using bilateral microelectrode recordings in a tetraplegic patient. *MedRxiv*. 2020:2020–06.
- [16] Pradeepkumar J, Anandakumar M, Kugathasan V, Lalitharatne TD, De Silva AC, Kappel SL. Decoding of hand gestures from electrocorticography with lstm based deep neural network. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC). 2021, 420–423.
- [17] Yao L, Zhu B, Shoran M. Fast and accurate decoding of finger movements from ecog through riemannian features and modern machine learning techniques. *Journal of Neural Engineering*. 2022;19(1):016037.
- [18] Jiang T *et al.* Characterization and decoding the spatial patterns of hand extension/flexion using high-density ecog. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2017;25(4):370–379.
- [19] Crone N *et al.* A click-based electrocorticographic brain-computer interface enables long-term high-performance switch-scan spelling. *Research Square*. 2023.
- [20] Bleichner MG, Freudenburg ZV, Jansma JM, Aarnoutse EJ, Vansteensel MJ, Ramsey NF. Give me a sign: Decoding four complex hand gestures based on high-density ecog. *Brain Structure and Function*. 2016;221:203–216.
- [21] Branco MP, Freudenburg ZV, Aarnoutse EJ, Bleichner MG, Vansteensel MJ, Ramsey NF. Decoding hand gestures from primary somatosensory cortex using high-density ecog. *NeuroImage*. 2017;147:130–142.
- [22] Li Y *et al.* Gesture decoding using ecog signals from human sensorimotor cortex: A pilot study. *Behavioural neurology*. 2017;2017.
- [23] Xu M, Zhou W, Shen X, Wang Y, Mo L, Qiu J. Swin-tcnet: Swin-based temporal-channel cascade network for motor imagery EEG signal recognition. *Biomedical Signal Processing and Control*. 2023;85:104885.
- [24] Siero JC, Hermes D, Hoogduin H, Luijten PR, Ramsey NF, Petridou N. Bold matches neuronal activity at the mm scale: A combined 7 t fmri and ecog study in human sensorimotor cortex. *Neuroimage*. 2014;101:177–184.
- [25] Pernet CR *et al.* Eeg-bids, an extension to the brain imaging data structure for electroencephalography. *Scientific data*. 2019;6(1):103.
- [26] Box GE, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1964;26(2):211–243.
- [27] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Physical review E*. 2004;69(6):066138.
- [28] Ross BC. Mutual information between discrete and continuous data sets. *PloS one*. 2014;9(2):e87357.
- [29] Tsamardinos I, Aliferis CF. Towards principled feature selection: Relevancy, filters and wrappers. In: *International Workshop on Artificial Intelligence and Statistics*. 2003, 300–307.
- [30] Mockus J, Mockus J. The bayesian approach to local optimization. Springer (1989).
- [31] Rasmussen CE, Williams CK, *et al.* Gaussian processes for machine learning. Springer (2006).
- [32] Hoffman M, Brochu E, De Freitas N, *et al.* Portfolio allocation for bayesian optimization. In: *UAI*. 2011, 327–336.

# REFERENCING SCHEMES AND THEIR EFFECT ON OSCILLATIONS AND BROADBAND POWER SPECTRAL SHIFTS IN STEREOELECTROENCEPHALOGRAPHY

Alexander P. Rockhill<sup>1</sup>, Michael A. Jensen<sup>2</sup>, Nicole C. Swann<sup>3</sup>, Ahmed M. Raslan<sup>2</sup>, Dora Hermes<sup>1</sup>, Kai J. Miller<sup>2</sup>

<sup>1</sup> Department of Neurological Surgery, Oregon Health & Science University, Portland, OR, United States

<sup>2</sup> Department of Neurosurgery, Mayo Clinic, Rochester, MN, United States

<sup>3</sup> Department of Human Physiology, University of Oregon, Eugene, OR, United States

E-mail: miller.kai@mayo.edu

**ABSTRACT:** Choosing a referencing scheme for stereoelectroencephalography (SEEG) is complicated by the varying depth of contact locations and, consequently, the different tissue that is being recorded from. In order to better understand how changes in electrophysiology related to movement are affected by the choice of reference, we examined how 16 different referencing schemes effected alpha (8 - 13 Hz) and beta (13 - 30 Hz) oscillations and high-frequency broadband (HFB) power (65 - 115 Hz). We found the choice of referencing scheme has more complicated effects than previously described and recommend using different referencing schemes as a methodological tool to optimize brain-computer interface (BCI) performance.

## INTRODUCTION

Stereoelectroencephalography (SEEG) measures electric potential as a differential between two voltage measurements, a reference and a recording, like all methods of measuring electric potential. SEEG recording contacts are implanted so that each contact is typically in a different layer of gray matter, in white matter or in a subcortical structure whereas electrocorticography (ECoG) contacts are placed above the dura usually mostly covering a small number of gyri. This causes SEEG to have relatively large variations in the statistical properties of the signals because of different brain-to-electrode impedance from the relative lipid contact in the tissue, compartmentalization of intracellular & extracellular solutes and formation of a fluid sheath around the SEEG shaft as well as sampling brain areas with more diverse cytoarchitecture and functional specialization.

In order to study the effects of referencing on SEEG recordings, we focused on well-replicated results; spatially focal high-frequency broadband (HFB) power increases in primary motor cortex, beta (13 - 30 Hz) power decreases across much of primary motor cortex and alpha (8 - 13 Hz) power changes (increases and decreases) during movement [1] [2] [3]. HFB changes are theorized

to facilitate action selection and gate action selection respectively [4]. HFB has been shown to be closely correlated with increases in firing rates in single units [5] and beta oscillations have been shown to be modeled accurately as traveling waves [6]. Alpha power changes related to movement-related in primary motor cortex are not as well understood but likely relate to the mirror neuron system [7]. Although the function of this rhythm is not as well described, it is observed reliably. Thus, the electrophysiological characteristics of these signals are relatively well understood, so we chose them to compare the effect of different referencing schemes. Previous work has explored how a subselection of these referencing schemes effect HFB and alpha (8 - 12 Hz) movement-related signal and found that more local rereferencing methods, like bipolar and Laplacian rereferencing, increased decodability of oscillations whereas more global rereferencing, like common average referencing, increased decodability of HFB [8]. In this study, we explored why this tradeoff occurs and how including other rereferencing strategies can give us a fuller picture in order to inform reference selection for brain-computer interface (BCI) applications.

## MATERIALS AND METHODS

*Ethics Statement:* This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board (IRB) of Mayo Clinic under IRB number 15-006530, which also authorized sharing of the data. Each patient/representative voluntarily provided independent written informed consent to participate in this study as specifically described in the IRB review (with the consent form independently approved by the IRB).

*Patients:* Thirteen patients (6 females, 11 - 20 years of age) from Mayo Clinic were included in this study. These data are publicly available from a previous publication [9]. These patients underwent placement of 10 - 20 SEEG electrode leads to characterize epileptogenic brain areas

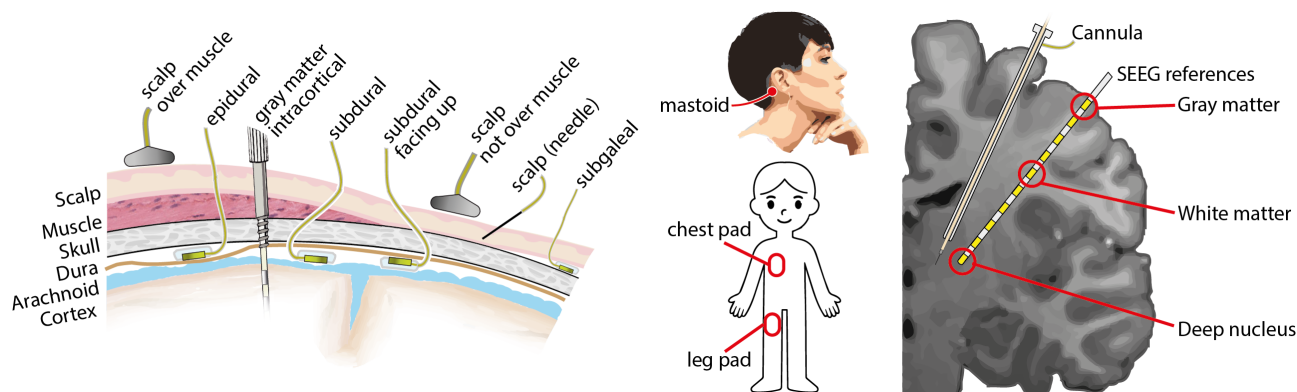


Figure 1: Native referencing options during SEEG recordings. Generally, the aim is for the reference to have the same environmental electrical noise so that it is removed by the differential amplifier. If the reference contains signal, such as from the brain or muscles, this signal will be present in all recording channels, so an electrically inactive location is ideal.

for treatment of drug-resistant partial epilepsy.

**Task:** Patients were visually cued to move their hand, tongue or foot for three seconds alternated with the same period of rest as in [9]. The task was administered using BCI-2000 [10].

**SEEG Recording:** The data was recorded with a g.HiAmp amplifier (gTec, Schiedlberg, Austria). Recordings were sampled at 1,200 Hz including electromyography (EMG) measured from the forearm flexors/extensors (hand), base of chin (tongue) and anterior tibialis (foot).

**Referencing:** SEEG data was recorded with a native reference as close to the recordings as possible to eliminate common noise. Of the common choices for referencing online recordings shown in Fig. 1, we chose a low-amplitude white matter contact since it has the same electrical environment as the other recording contacts, does not have signal from muscle activation and has low-amplitude neural signal. After the signal was recorded, a new reference signal was generated in 16 different ways:

1. Native: the original white-matter reference
2. Average: an average of all the channels (Fig. 2a)
3. Lead average: an average of all channels on a single lead (Fig. 2a)
4. Headbox average: an average of all contacts being amplified by the same headbox (Fig. 2a)
5. Bipolar: the next neighboring contact (Fig. 2a)
6. White matter: the average of all contacts predominantly located in white matter (Fig. 2b)
7. Laplacian: the average of two neighboring contacts or the single neighbor for the ends of the electrode
8. Position: the weighted average of two contacts on either side or as many as there are (Fig. 2c)
9. Distance: the average of all the contacts weighted by the distance to the recording contact (Fig. 2d)

10. Low PSD: an average of the 50% of contacts with power spectral density (PSD) below 45 Hz most frequency below the mean PSD (Fig. 2e)
11. Low RMS: an average of the 50% of contacts with the lowest root mean square (RMS) amplitude (Fig. 2f)
12. Low PSD per lead: same as low PSD but per lead
13. Low RMS per lead: same as low RMS but per lead
14. Low PSD per headbox: same as low PSD but per headbox
15. Low RMS per headbox: same as low RMS but per headbox
16. PCA: the first (largest) principal component using all channels as observations and samples over time as features (Fig. 2g)

**Locality analysis:** In order to determine how much each re-referencing scheme caused correlation between channels, or spread the signal, Pearson's  $r$  was computed on the time-series data. We report the median Pearson's  $r$  correlation across samples aggregated by epoch, including both movement and rest periods, from -500 ms to 1500 ms relative to the start of the period.

**Movement-rest analysis:** The time-series data was converted into power spectral density using Welch's method with 1024 points per segment and a 75% window overlap between segments using Hann windowing. An activation  $r^2$  metric was computed as in [9] using the following equation:

$$r^2 = \frac{(\bar{m} - \bar{r})^2}{|\bar{m} - \bar{r}| \sigma_{m \cup r}^2} \frac{N_m N_r}{N_{m \cup r}^2}$$

where  $\bar{m}$  is the mean PSD for movement,  $\bar{r}$  is the mean PSD for rest,  $\sigma_{m \cup r}^2$  is the variance of both movement and rest PSDs and  $N_m$ ,  $N_r$  and  $N_{m \cup r}$  are the number of movement epochs, rest epochs and the combined sum, respectively. This signed metric quantifies the difference between the mean of the movement PSDs across epochs

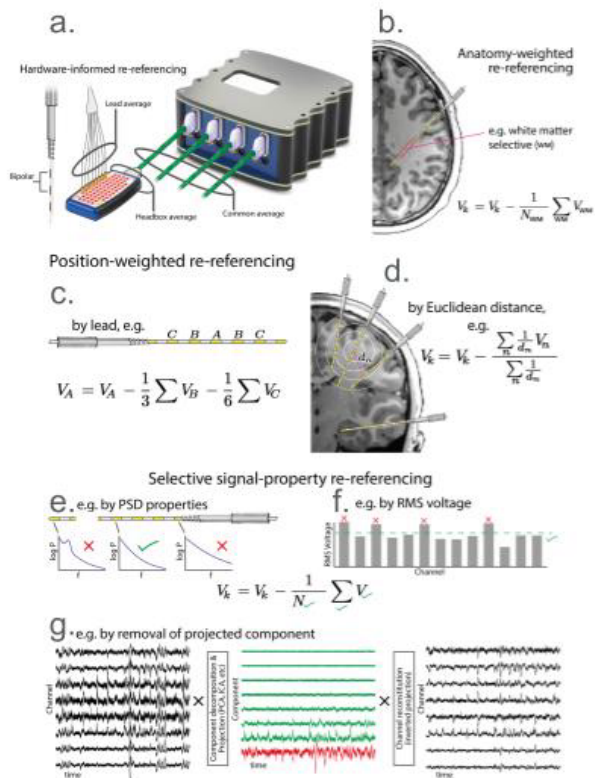


Figure 2: Many of the re-referencing options for after the recording has been digitized. Average and bipolar referencing tend to be used more commonly but there are many other options.

compared to mean rest PSDs across epochs relative to the variances of PSDs across epochs of movement and rest separately. Power spectral density was averaged in 5 Hz bins. The best  $r^2$  in the alpha, beta and HFB ranges for were then found separately for each patient and effector (hand, tongue and foot). Time-series data were also filtered with a fourth order Butterworth 2 Hz on either side of the peak oscillation frequency for beta and from 65 to 115 Hz in bandwidths of 10 Hz for HFB to visualize the time-course of changes observed in the PSDs.

## RESULTS

SEEG contacts that were modulated with movement generally had the pattern of an increase or decrease in alpha power, decreased beta power and increased HFB during movement, as shown in Fig. 3. Power decreased in some recording contacts in every patient in the beta range (Fig. 3, gray box), where blue indicates a decrease in power during movement relative to rest. Similarly, power increased in the HFB range (Fig. 3, black box) for some recording contacts in every patient, where the yellow color indicates an increase in power during movement relative to rest. The number of contacts with beta power decreases was more than the number of contacts with HFB increases. Alpha power changed for at least one contact in every patient as well (Fig. 3, white box)

but was less consistent in direction. The patterns of spectral differences were similar between average and bipolar references, and this was generally the case for all re-reference schemes.

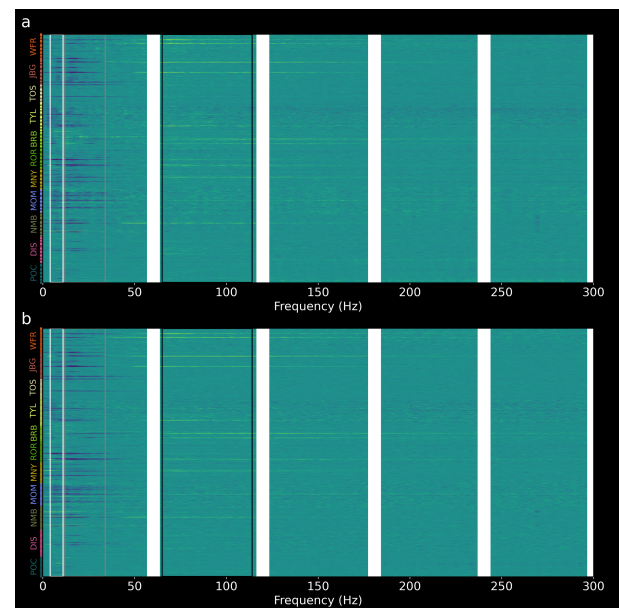


Figure 3:  $r^2$  activation maps of the signed ratio of the difference between the mean PSD across epochs of movement and rest relative to the variance unique to movement and rest for average (a) and bipolar (b) references. In both references, there are contacts with beta power decreases during movement and high-frequency broadband power increases during movement but only at electrode contacts positioned at brain areas that are modulated with hand movement. The all the recording contacts for all patients are shown on the y-axis with each electrode shaft striped using alternating dark-light colors. The frequency of the PSD is shown on the x-axis, with boxes around alpha in white, beta in gray and HFB in black.

We found that more local re-referencing methods, such as bipolar and Laplacian, had lower correlations between channels while more global re-referencing methods, such as using an average reference, had greater correlations. Here, we are referring to referencing schemes that include more recording contacts in the signal used as a reference for a given contact as more global and referencing schemes that include fewer contacts as more local. The greater amount of yellow off the diagonal in each of the plots in Fig. 4a is summarized as an average in Fig. 4b. Laplacian and bipolar referencing schemes had the least correlation between recording contacts and that correlation was near the diagonal suggesting that neighboring contacts on the same electrode shaft contributed most to that correlation. Average referencing had an intermediate amount of correlation between channels which was spread out less near the diagonal than Laplacian and bipolar correlations, suggesting that the signal spread was less local to an electrode shaft. Native referencing had by far the most correlation between recording contacts, including relatively large correlations between contacts on different electrode shafts, suggesting that signal from the

chosen white matter reference contact was largely present in all of the recording contacts.

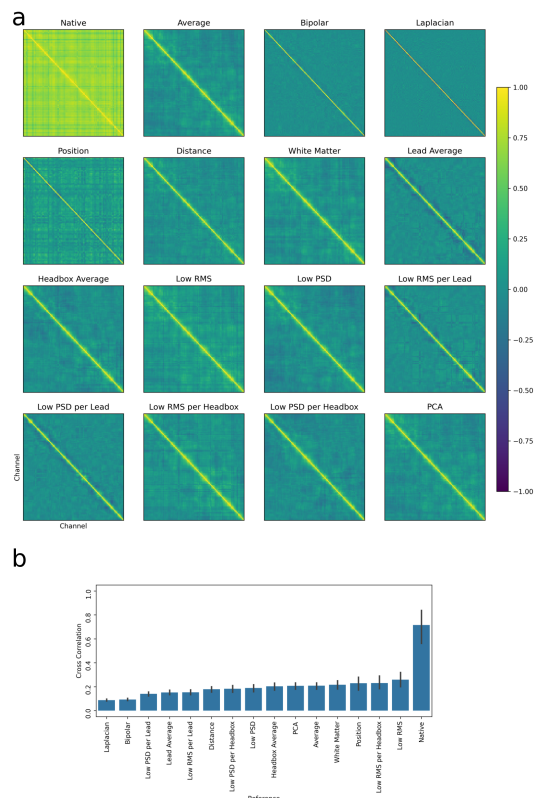


Figure 4: a. Correlation plots between channels computed pairwise are shown for each referencing scheme. The autocorrelation on the diagonal is always one. Off the diagonal, correlation implies that recording contacts are detecting the same source so minimizing this increases the specificity of the location of the source of the signal when interpreting the data. b. The average off-diagonal correlation across patients.

Next, we quantified how referencing schemes effected  $r^2$  values in the alpha, beta and HFB ranges across the patients and for the three different effectors (hand, tongue and foot) as shown in Fig. 5. The recording contact with the most movement-related modulation was determined using the maximal absolute value  $r^2$  in the frequency range of each spectral feature. The mean of all these values across references was then subtracted from each  $r^2$  to get  $\delta r^2$ . We found movement related changes in the oscillatory frequencies, alpha and beta, were better detected by more local referencing schemes (Fig. 5a and b). More global referencing schemes, on the other hand, better differentiated movement-related changes in HFB (Fig. 5c). However, was considerable variation across patients and effectors; bipolar or Laplacian referencing was not the best choice for detecting oscillations in many cases (different patients and effector combinations) and average referencing was not the best choice for detecting HFB in many cases.

We looked into the effect of the referencing scheme on beta and HFB power using data from an example patient (Fig. 6). As shown in Fig. 6a, the HFB increase that

peaked immediately after movement onset was maximal at the fourth most superficial contact. The neighboring contacts also had a peak in HFB, causing the bipolar reference to diminish this HFB increase after movement. In Fig. 6b, beta oscillations were observed to constructively interfere at the contact with the greatest movement-related changes before the movement onset and destructively interfere during the movement. The result was that bipolar referencing caused the difference between beta power before movement (during the rest period) to have a larger difference compared to beta during the movement. In Fig. 6c, the amount of time shown is reduced to show few enough cycles that the phase can be seen. The opposite phase in the recording contact most modulated by movement compared to the neighboring contact, shows the constructive interference increasing beta power during rest for bipolar referencing.

## DISCUSSION

For each patient and effector, there was a relatively local source of HFB activity during movement and a relatively diffuse distribution of alpha and beta oscillations during rest as previously described [2]. This was generally best captured using bipolar or Laplacian referencing for oscillations and average referencing for HFB. This was because, when the recording contact was positioned in line with the source of the activity, oscillations constructively interfered, increasing the signal-to-noise ratio. HFB, on the other hand, does not interfere like oscillations do and so was better captured by an average reference scheme. There was considerable variability depending on position of the recording contact and the tissue types of each contact and its neighbors. These differences cause different referencing schemes to be preferred and dispreferred in a pattern that is unique to each case, suggesting that referencing is a tool that can be used to optimize the detection of signal in the data by compensating for some of these effects, especially in applications such as BCI where interpretation of the signal is secondary to performance.

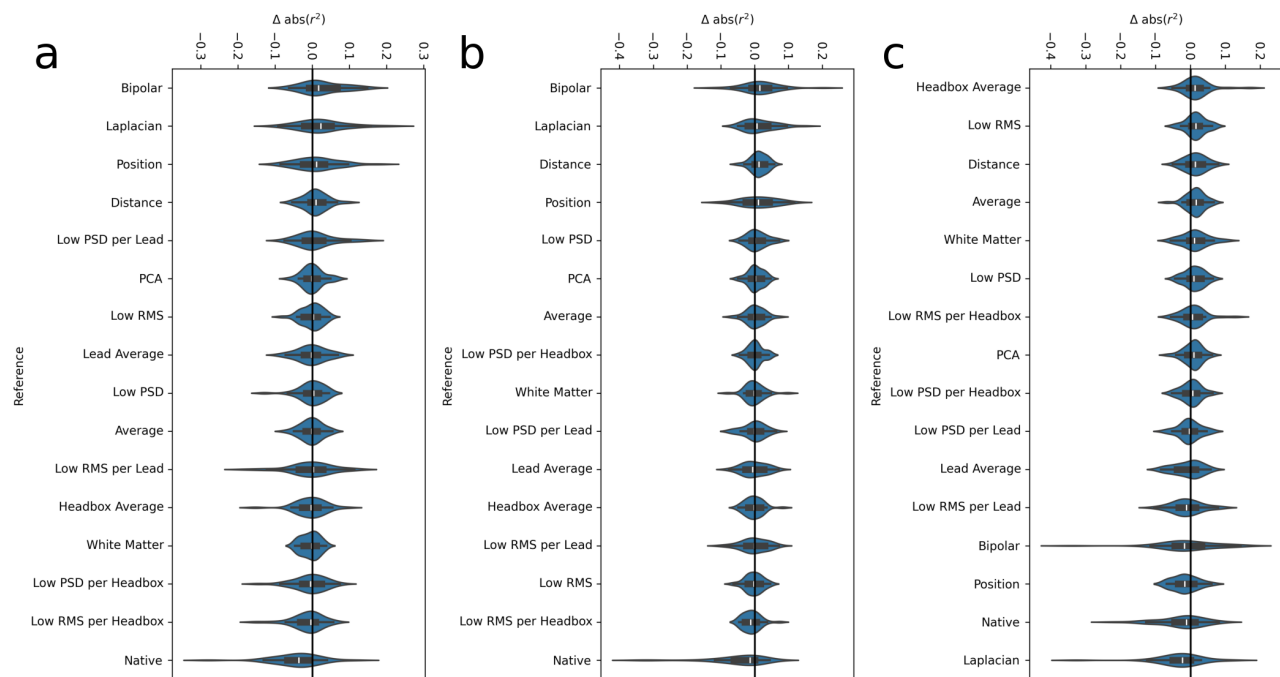


Figure 5: The change in  $r^2$  metric for different referencing schemes are shown for alpha (a), beta (b) and HFB (c). More local referencing strategies such as bipolar and Laplacian cause larger differences between the rest and movement conditions, as measured by  $r^2$  for oscillations (alpha and beta). More global referencing schemes, such average referencing, cause larger differences for HFB. However, there is such great variability between patients and effectors such that this approach is not optimal in many cases.

## CONCLUSION

There is considerable variability in the  $r^2$  activation metric between patients and effectors compared to the variability between referencing strategies. We interpret this to be due to effects of the position and orientation of the electrode recording contacts relative to neural sources of activity which depend on the trajectory of the electrode shaft and the micro-organization of the brain. In general, using Laplacian or bipolar referencing to study oscillations and average referencing to study HFB yields better signal-to-noise, however, this is not the case much of the time so referencing strategy should be used as a tool to optimize data interpretation.

## REFERENCES

- [1] Miller KJ *et al.* Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*. 2007;27(9):2424–2432.
- [2] Miller KJ, Honey CJ, Hermes D, Rao RP, denNijs M, Ojemann JG. Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *NeuroImage*. 2014;85:711–720.
- [3] Rockhill A *et al.* Intracranial electrode location and analysis in MNE-python. *Journal of Open Source Software*. 2022;7(70):3897.
- [4] Miller KJ *et al.* Human motor cortical activity is selectively phase-entrained on underlying rhythms. *PLoS Computational Biology*. 2012;8(9):e1002655.
- [5] Manning JR, Jacobs J, Fried I, Kahana MJ. Broad-band shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *Journal of Neuroscience*. 2009;29(43):13613–13620.
- [6] Stolk A *et al.* Electrocorticographic dissociation of alpha and beta rhythmic activity in the human sensorimotor system. *eLife*. 2019;8:e48065.
- [7] Hobson HM, Bishop DV. Mu suppression - a good measure of the human mirror neuron system? *Cortex*. 2016;82:290–310.
- [8] Li G *et al.* Optimal referencing for stereo-electroencephalographic (SEEG) recordings. *NeuroImage*. 2018;183:327–335.
- [9] Jensen MA *et al.* A motor association area in the depths of the central sulcus. *Nature Neuroscience*. 2023;26(7):1165–1169.
- [10] Schalk G, McFarland D, Hinterberger T, Birbaumer N, Wolpaw J. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*. 2004;51(6):1034–1043.

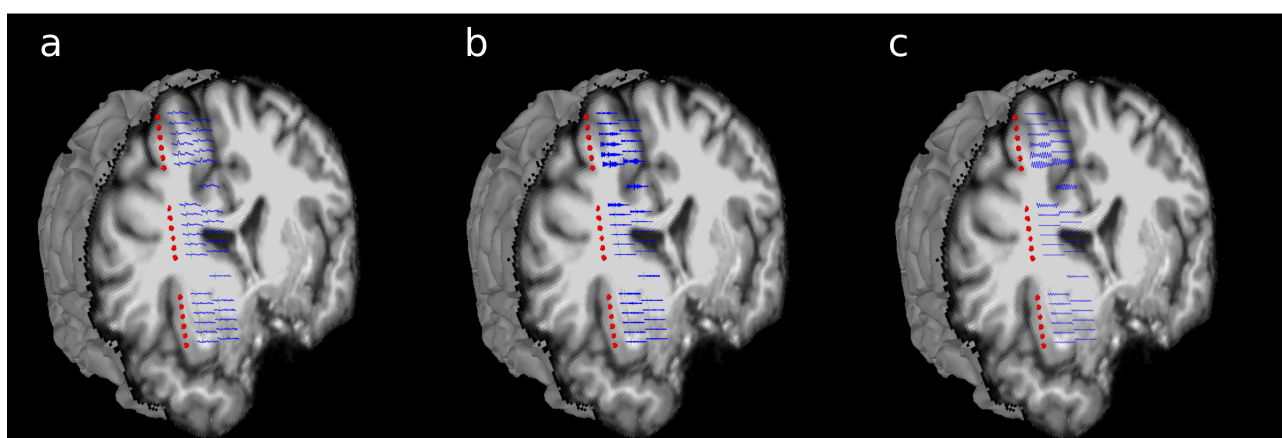


Figure 6: a. HFB activity around movement onset (green vertical line) for an example recording. Activity using an average reference is shown to the right of the recording contacts (red spheres). HFB with a bipolar reference is shown to the right of that, interleaved between the two contacts being referenced. The HFB signal decreases as distance increases from the spatial location with maximal modulation during movement. Since the HFB modulation in the movement-related contacts that are shown only contributes a small amount to the average reference, the reference signal has low HFB. The difference between a maximally active contact and the reference is greater than the difference between that maximally active contact and its neighbor which is also has HFB activation. b. Time-series data filtered in the beta range is shown for average reference to the right of contacts and bipolar reference to the right of that. When oscillations are aligned in phase, they constructively interfere as in the pre-movement period for the 5<sup>th</sup> and 6<sup>th</sup> most superficial contacts. Whereas, when the phases are opposite, as in the next most superficial bipolar pair, they destructively interfere. This is shown with fewer points in time in (c) in order to see this effect.

# FUNCTIONAL REPRESENTATION OF SOMATOSENSORY, VISUAL, AND REINFORCEMENT PROCESSING ON THE CANINE BRAIN SURFACE

*An early feasibility chronic use demonstration: BrainInterchange-BCI2000 ecosystem*

Frederik Lampert<sup>1</sup>, Filip Mivalt<sup>2,3</sup>, Inyong Kim<sup>2</sup>, Nuri F. Ince<sup>1,4</sup>, Jiwon Kim<sup>2</sup>, Jhan L. Okkabaz<sup>4</sup>,  
Max A. Van den Boom<sup>4</sup>, Vaclav Kremen<sup>2</sup>, Rushna Ali<sup>1</sup>, Volker A. Coenen<sup>5</sup>, Gerwin Schalk<sup>1,6</sup>,  
Peter Brunner<sup>7</sup>, Gregory A. Worrell<sup>2,4</sup>, Kai J. Miller<sup>1,4</sup>

<sup>1</sup>Department of Neurosurgery, Mayo Clinic, MN, USA

<sup>2</sup>Department of Neurology, Mayo Clinic, MN, USA

<sup>3</sup>Department of Biomedical Engineering, Brno University of Technology, Brno, Czechia

<sup>4</sup>Department of Biomedical Engineering, Mayo Clinic, MN, USA

<sup>5</sup>Department of Stereotactic and Functional Neurosurgery, Freiburg University, Germany

<sup>6</sup>Chen Frontier Lab, Tianqiao and Chrissy Chen Institute, Shanghai, China

<sup>7</sup>Department of Neurosurgery, Washington University School of Medicine, St Louis, MO, USA

E-mail: lampert.frederik@mayo.edu; miller.kai@mayo.edu

## ABSTRACT

Implantable brain-computer interface (BCI) systems, promising for neurological disorder treatment, often encounter high technical barriers. Our fully-implanted CorTec BrainInterchange-BCI2000 ecosystem, aimed for widespread open-source adoption, demonstrates functionality through a year-post-implant canine study, using a brain surface electrocorticography (ECoG) construct. Broadband power-spectrum increases have been shown to track neural population activity in humans, and we find that they reveal distinct functional representation for processing of visual, somatosensory, and auditory reinforcement stimuli in the canine (captured at 65-150Hz). Canine visual and somatosensory rhythms resemble human alpha and beta rhythms but at different frequencies: a ~15Hz visual rhythm in the occipital analog (marginal gyrus) suppresses with light exposure, and a ~24Hz somatosensory rhythm diminishes upon petting. These findings indicate a unique canine neurophysiology and confirm the BCI2000-BrainInterchange ecosystem's robustness a year after the implantation. This ecosystem holds promise for developing open-source BCI devices to assist patients with neurological conditions.

## INTRODUCTION

The technology behind practical brain computer interfacing has an exclusivity problem arising from a number of factors. Hardware that can record sufficient channels for the coding brain activity requires a large infrastructure to build, maintain, and troubleshoot. Engineers with the rare expertise needed to manage these devices typically professionally reside far from the clinical setting that patients with neurological disorders present to for therapy. The software skills and signal processing know-how needed

to translate signals measured from the brain into closed-loop commands for external applications or internal recursive stimulation of the brain are similarly exclusive. For this and other clinical reasons, we have been developing a combined hardware-software open source ecosystem with the CorTec BrainInterchange (BIC) device [2] and the BCI2000 software environment [3] that will serve as a general purpose platform that can be easily applied for clinicians focused on a specific patient need.

Our development of this ecosystem has begun with canine (*canis familiaris*) implants. For chronic studies, dogs offer a viable model for cognitive and neurodegenerative studies, attributable to their trainability, cooperative nature, and neurophysiological similarities to humans [4, 5]. We are able to test the device on a daily basis without any restraint of the dog and upon the completion of the study the dogs can be adopted into homes. In this demonstration, a series of basic sensory input tasks are performed approximately a year after the time of implant. While the implanted neurophysiology of the canine brain is relatively unexplored, we expect that many aspects of well-described phenomena in the human brain will generalize.

Broadband changes in the brain surface electrical potential have been shown to be a robust correlate of local neural activity and an effective control signal in brain computer interfaces [6, 7]. However, these changes are often obscured at low frequencies by coincident oscillations, and so broadband changes must often be captured at frequencies above ~50-60Hz, setting a performance threshold that BCI devices must rise to. In this manuscript, we demonstrate that the BrainInterchange-BCI2000 ecosystem has this capability and, in the process, uncover functional representation of somatosensory, visual, and social reinforcement in the canine brain.

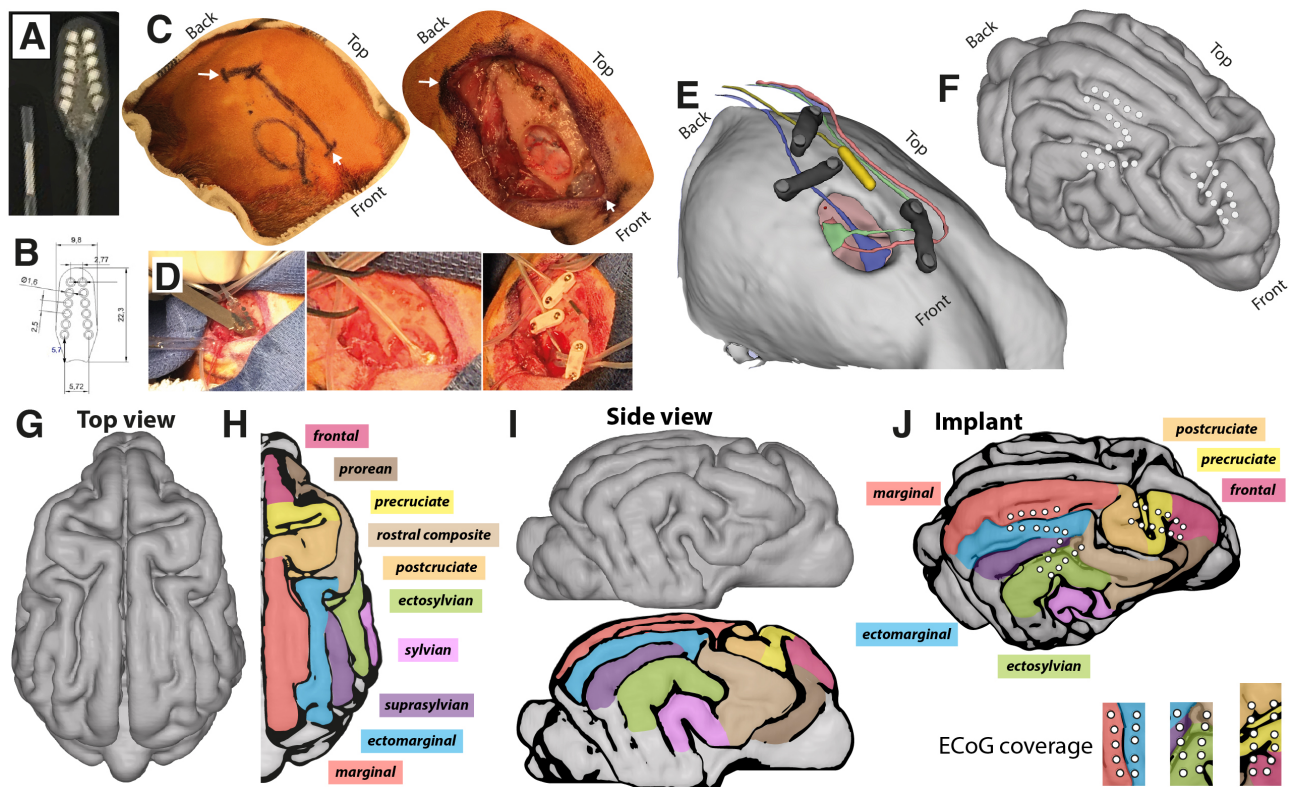


Figure 1: **Right hemisphere implant and anatomic segmentation** (A) Electrode grid and ground electrode. (B) Schematic with dimensions of grid (in mm). (C) Planned incision and craniotomy on scalp (left) with opening and craniotomy, showing epidural space (right). White arrows correspond to same location on left and right. (D) Insertion and anchoring of electrodes. (E) Rendering of skull and craniotomy with electrodes in situ, ground electrode in yellow. (F) Brain rendering showing three grids in situ, extracted from pre-implant MRI and post-implant CT. (G) Top view of the canine cortex. (H) Color-map representation of canine right hemisphere gyral anatomy [1]. (I) Sagittal view of canine right hemisphere and color-map representation of its gyral anatomy. (J) Schematic placement of electrode grids over the canine cortex.

## MATERIALS AND METHODS

**Surgical implant:** A 2-year-old female beagle, “Belka”, was implanted with the 32-channel sensing-and-stimulation Cortec BrainInterchange device as previously described [2], according to a public operative protocol [8]. Three arrays (32 ECoG channels) were implanted epidurally over the right hemispheric convexity with the FDA-approved AirRay electrodes [9] (Fig. 1).

**Anatomic co-registration:** A pre-implantation 3T MRI and a post-implantation CT were obtained. The brain was manually segmented from the MRI using 3D Slicer [10], and the CT was co-registered with electrodes aligned to the anatomy using the CTMR package as previously described [11], which was also used for subsequent plotting. Anatomic segmentation of the brain surface was determined manually with reference to the *Stereotactic Cortical Atlas of the Domestic Canine Brain* [1] (Fig. 1). The 3 grids were localized to the 1) *frontal*, *precruciate* and *postcruciate* gyrus; 2) *ectosylvian* gyrus extending to the border of the *suprasylvian* & *ectomarginal* gyrus and the *rostral composite*; and 3) flanking the *marginal* and *ectomarginal* gyri.

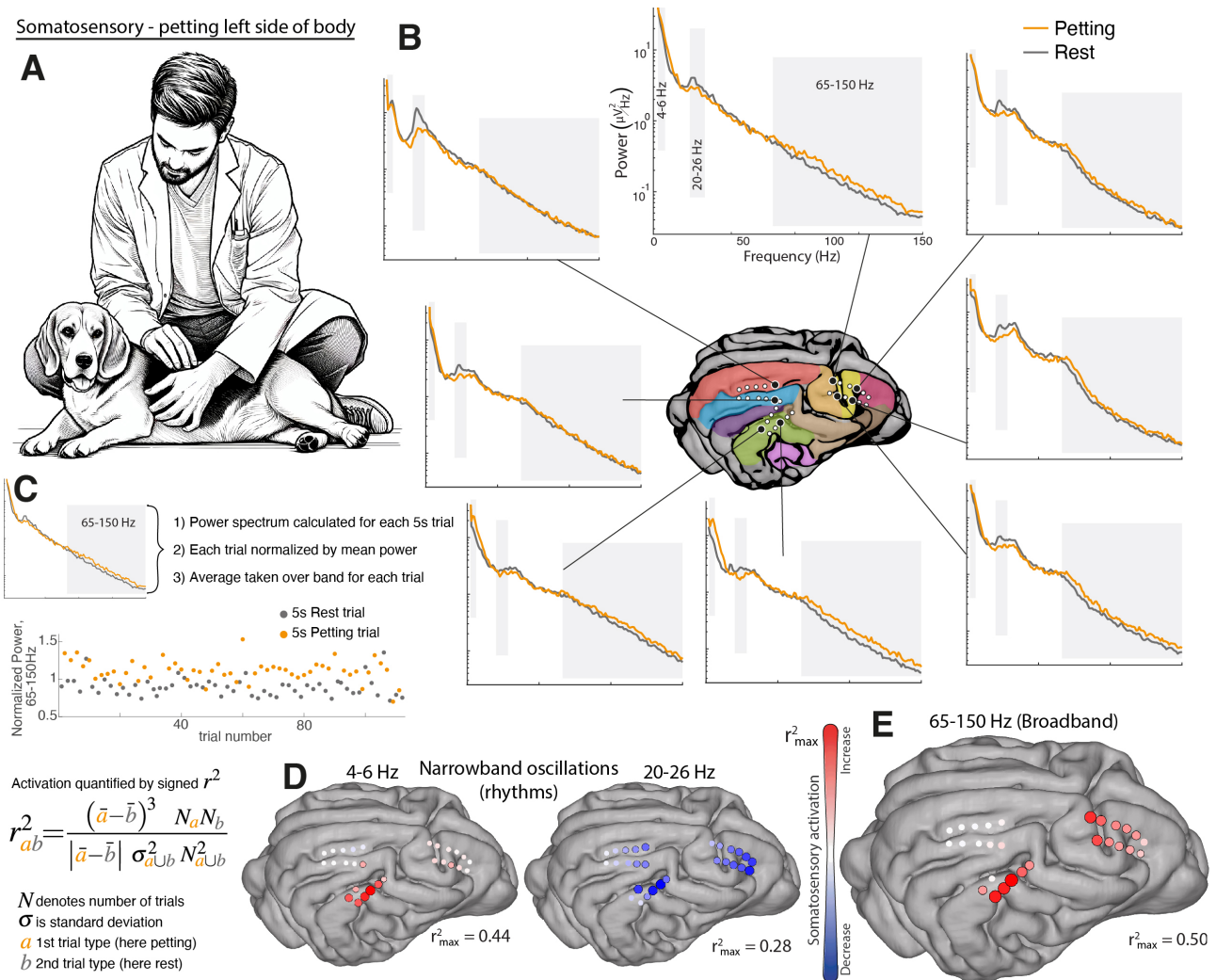
**Canine tasks:** Three types of tasks were performed - somatosensory, visual, and reinforcement (auditory input with face touching). For each task, the four best runs

(as determined by behavior, prior to data analysis) were selected for further analysis. Each run consisted of 15 repetitions of active & inactive task blocks.

**Somatosensory** - The dog was positioned unrestrained within the examination room with the room lights on, next to the examiner. 3-second blocks of tactile stimulation (petting along the left side of the dog, encompassing the whiskers, front and hind limbs, and torso) were interleaved with 5s blocks of rest (Fig.2A).

**Visual** - Examination took place in a closed room with all external sources of light blocked. The dog was placed on a leash in the middle of the room, allowing her to move within the reach of the leash. 5-second “lights OFF” blocks with the room lights turned off were interleaved with 5s “lights ON” blocks, where the rooms lights were turned on. A laptop was in the room, with monitor light exposed, cuing the examiner to turn the room lights on and off.

**Reinforcement with auditory encouragement and & face touching** - The dog was positioned unrestrained in the examination room next to the examiner with the room lights on and calming classical music playing at a low volume. 5-second reinforcement blocks where the examiner provided verbal reinforcement (“Good girl Belka!”) & gently touched the left side of the face were interleaved with rest periods.



**Figure 2: Somatosensory activation** (A) Cartoon representation of the somatosensory stimulation task, where 3s blocks of tactile stimulation (petting left whiskers, front and hind limbs, and torso) were interleaved with 5s blocks of rest. (B) Power spectral densities (PSDs) of petting and rest blocks. Gray shaded regions highlight 4-6Hz, 20-26Hz and 65-150Hz frequency ranges. (C) For analysis, PSDs for each task block were normalized by the average PSD over the whole experiment, and averaged normalized power was quantified for each frequency range. Task-associated changes were quantified using a signed  $r^2$  metric (which can range from -1 to 1). (D) Scaled activation maps (by signed  $r^2$ ) shows domains of change for the low frequency brain rhythms. Black circle outline indicates task-induced significant change at  $p < 0.05$  (after Bonferroni correction, unpaired t-test in mean normalized power in band for each trial). (E) Maps of local brain activity, reflected by broadband spectral change, is captured at (65-150Hz), showing engagement of the pre- and post-cruciate, ectosylvian, and rostral composite gyri.

**Electrophysiological measurements:** Data were measured using BCI2000 general-purpose software [12], which provides a graphical user interface for data acquisition, online processing for closed-loop application (though not used in this study), and stimulus presentation. Data were sampled at 1000 Hz, with an amplification gain of 57.5 dB / 1  $\mu$ V, and initially referenced to the channel 1 (Fig. 3). Missing samples due to the packet loss were replaced with the first valid sample preceding the packet loss. This is a default setting for handling missing data packages in Cortec's BIC device (see the work of Ayyoubi *et al.* [13] for further packet loss discussion).

**Signal processing:** Data were examined by raw trace as well as relative signal power to identify bad channels, which were discarded prior to common-average re-referencing of the data. Power spectral densities

(PSDs) up to 150Hz were calculated for each task block using Welch's averaged periodogram method [14] with 1s Hann windowing [15] and 50% window overlap. Several data blocks were rejected (across all channels) due to significant transient artifact (4/6/0 blocks for somatosensory/visual/reinforcement). Individual block PSDs were normalized by the mean power at each frequency (mean calculated over each full task). Signed  $r^2$  cross-correlations comparing task conditions (Fig. 2C) were calculated across all the channels at each frequency, and plotted on featuremaps (Fig. 3) to characterize the spatial and frequency-specific structure of neurophysiological changes associated with each task. Based on the visual examination of the raw PSDs and the featuremaps, three low-frequency narrow-band ranges (4-6Hz, 14-16Hz [visual only], & 20-26 Hz) were cho-

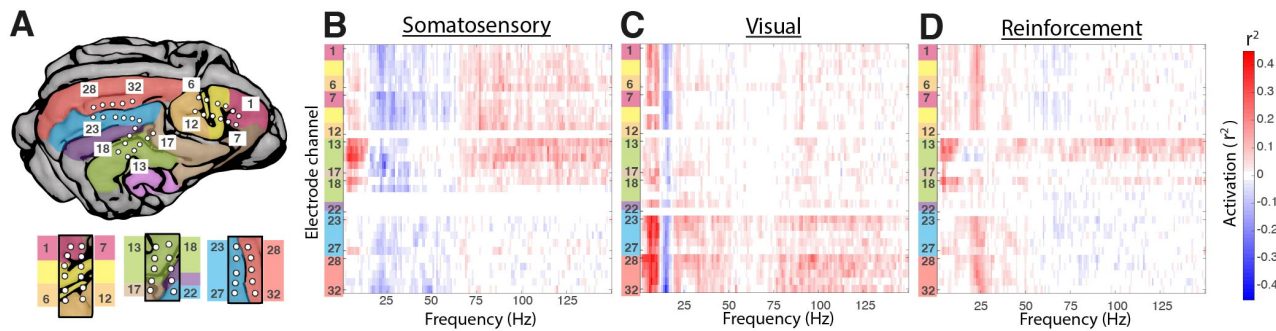


Figure 3: **Task featuremaps** (A) Segmentation of the canine right hemisphere gyral anatomy and electrode placement locations, with channel numbering that is referenced in (B-D). (B-D)  $r^2$  feature-maps for the somatosensory, visual, and reinforcement tasks.

sen for statistical analysis. Separately, a high-frequency broadband range was chosen *a priori* at 65-150Hz to capture the 1/f structure that has been shown in humans to be a correlate of local population activity [6, 16]. Averaged power (after normalizing by the mean PSD) was calculated across each frequency range for each block. Blocks of each type within each task were then compared with one another using a signed  $r^2$  metric and an unpaired (2-sample) t-test (mean normalized power in band for trials of petting vs. rest, lights on vs. lights off, and reinforcement vs. rest). Maps of  $r^2$  values were projected onto the rendered brain to show task-associated brain activity (Figs. 2,4), with channels that reached threshold significance ( $p < 0.05$  after Bonferroni correction), were marked with a black outline.

**Ethics statement:** This research is conducted under Mayo Clinic IACUC protocol A00001713-16-R19. We maintain our canines in an IACUC-approved environment. In addition, according to State of Minnesota statute 135A.191, the canines will be made available for adoption at the conclusion of research. In the event of serious illness or decline, the animals may be humanely euthanized by the veterinary team according to an IACUC mandated protocol. The canine subject, Belka, is a 3 year old female (implanted at 2 years old). She is housed in a communal environment, and receives daily social interactions with veterinary staff as well as open time with other canines. The intent of this animal research is to test and develop a platform for novel human therapeutics.

## RESULTS

**Somatosensory** - Somatosensory petting stimulation contralateral to the side of implantation showed robust activation (as revealed by broadband spectral power increase) over the pre- and post-cruciate gyri, as well as the ectosylvian / rostral composite gyri (Figs.2–4). A narrowband rhythm (oscillation) that was observed with a peak of  $\sim 24$ Hz (20-26Hz) decreased in power over these same regions during petting.

**Visual** - Visual task stimulation showed a significant activation (as revealed by broadband spectral power increase) over the marginal and ectomarginal gyri (Fig. 3C, 4A). There was a very prominent oscillation, peaked at

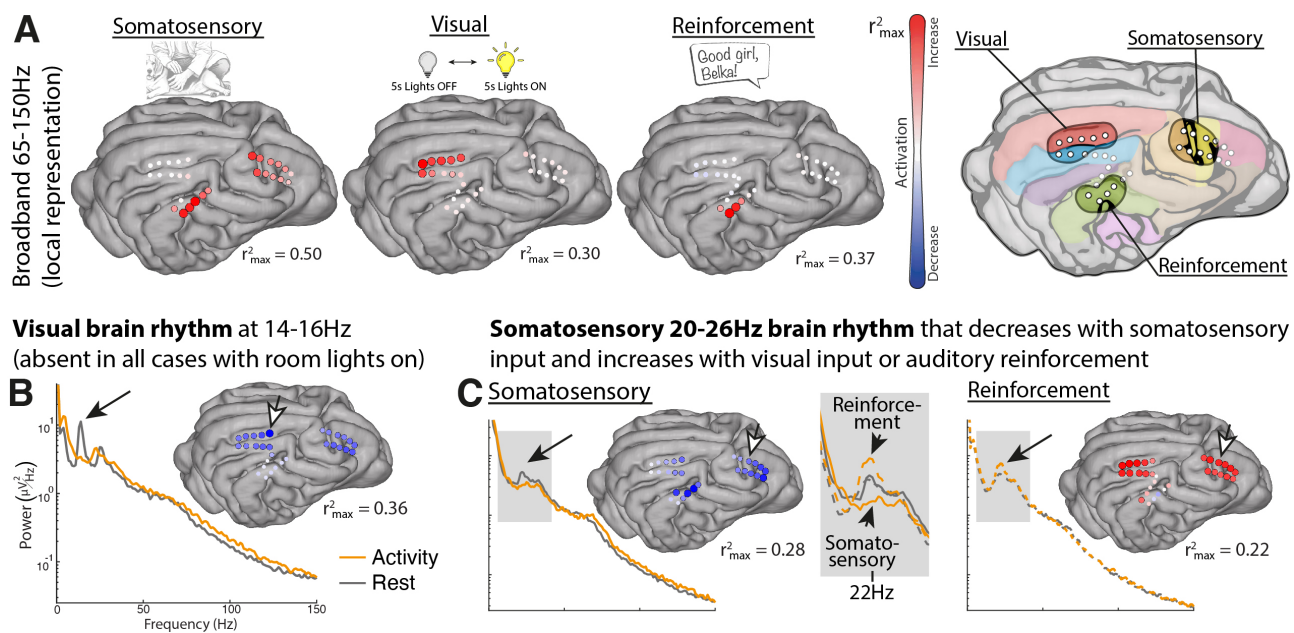
$\sim 15$ Hz, that emerged over most of the sampled sites during the lights off period, and diminished when the room lights were turned on (Fig. 3C, 4B).

**Reinforcement with auditory encouragement and & face touching** - During reinforcement with coincident verbal praising and stroking of the left cheek, we found robust activation (broadband spectral increase) of the anterior measurement sites of the ectosylvian gyrus. Additionally, there was an increase in the rhythm centered around  $\sim 24$ Hz, present over the pre- and post-cruciate, marginal, and ectomarginal gyri (Fig. 3D, 4).

**Comparison across tasks** Comparison of local representation, as reflected by 65-150Hz broadband spectral change, shows clear functional representation in the canine brain, with visual representation in the marginal gyrus, somatosensory representation in the pre- and post-cruciate gyri, and reinforcement (both tactile and auditory) in the ectosylvian gyrus. Notably, there was also: **1)** An oscillation/rhythm peaked at  $\sim 5$ Hz (4-6Hz) that increased in power at all measured sites when lights were turned on in the visual task (with  $r_{max}^2 = 0.43$ ), and selectively increased in the ectosylvian sites during the somatosensory ( $r_{max}^2 = 0.44$ ) and reinforcement ( $r_{max}^2 = 0.28$ ) tasks. **2)** A  $\sim 15$ Hz rhythm was only seen with the lights off blocks of the visual task and was not seen during any other tasks. **3)** Although not modulated in the visual task, an oscillation peaked at  $\sim 24$ Hz was present over much of the brain surface - it was selectively depressed during somatosensory input and augmented during reinforcement input (Figs. 3A&C, 4C).

## DISCUSSION

These simple sensory tasks demonstrate for the first time that, as in humans [6, 16], local neural activity can be captured by broadband spectral change from the surface of the canine brain (here captured at 65-150Hz, Fig. 4). In agreement with emerging canine fMRI studies [17], clear functional representation for somatosensory processing was found most robustly surrounding the cruciate sulcus, which is the canine homolog of the human central sulcus, and also the ectosylvian & rostral composite gyri. The ectosylvian gyrus has previously been implicated in canine somatosensory function using peripheral-stimulation evoked brain potentials (SSEPs) [18]. Com-



**Figure 4: Graphical summary of the obtained results.** (A) The functional representation of the canine brain, revealed by broadband increase in the power spectrum (65-150 Hz), shows distinct representation for each modality. (B) A clear  $\sim 15$ Hz brain rhythm emerges in the dark that is suppressed when the lights are turned on ( $r^2_{\max}=0.36$ ), and is not seen in either of the other 2 tasks. (C) Interestingly, there is an oscillation with peak at  $\sim 24$ Hz was present over much of the brain surface that is selectively depressed during somatosensory input and augmented during reinforcement input. Note that rest condition PSDs are approximately equal at this frequency range for both tasks (solid and dashed black lines in middle gray square), but that petting/reinforcement selectively suppresses/augments it. White arrows in (B)&(C) show sites where PSDs are from.

parison of lights-on to lights-off brain activity localized visual processing to the marginal gyrus, which agrees with fMRI localization [19], and the removal of which has been shown to blind dogs [20]. Combined light tactile and auditory reinforcement selectively activated only at sites in the ectosylvian gyrus and nowhere else, loosely agreeing with fMRI measurement [21].

Alongside broadband power spectral changes, we also observed a number of prominent oscillatory rhythms below 50Hz (Figs. 2-4). For example, a canine analog of the human occipital “alpha” rhythm emerged when the dog was in the dark, with a peak at  $\sim 15$ Hz that is most prominent in the marginal gyrus, and is slightly higher in frequency than the 8-13Hz range reported by *Lopes da Silva, et. al.* [22]. Oddly, there was an increase in power at an  $\sim 5$ Hz oscillation where *Kujala et al.* found a visually-induced decrease from the scalp [23]. The  $\sim 24$ Hz rhythm that is selectively suppressed with somatosensory stimulus and augmented during reinforcement (Fig. 4C) appears to be a novel observation.

Importantly, human ECoG studies have shown that behaviorally-induced oscillatory brain rhythms are generally not functionally specific in the same way that the broadband changes are [6, 7, 24]. For this reason, it is important that chronically implanted hardware can be able to capture the broadband spectral changes. However, two factors make this technically non-trivial: 1) large power changes in oscillations/rhythms obscure smaller amplitude broadband power changes below  $\sim 50$ Hz; 2) the broadband phenomena falls off in power at higher frequencies as  $P \sim 1/f^4$  [16]. Implanted

hardware must therefore have a sufficiently low noise floor such that behaviorally-associated broadband power increases can be resolved at higher frequencies. These canine behavioral results demonstrate, empirically, that the BrainInterchange-BCI2000 ecosystem accomplishes this, even after the device has been implanted for a year. The BrainInterchange-BCI2000 ecosystem is undergoing continuous development and optimization [2]. All software, data, resources, & protocols for this initiative are fully open-source, with a vast documentation to teach the community how to use it without having extensive technical expertise. When fully developed, this ecosystem may enable clinical teams to create personalized BCI therapies tailored specifically to the needs of their patient population.

## CONCLUSION

Our work demonstrates the utility of the BrainInterchange-BCI2000 platform in the setting of ECoG recordings during three sensory experiments almost a year after implantation. Captured local neural activity, revealed by broadband spectral changes, show robust sensorimotor, visual, and reinforcement functional organization in the brain of a canine. This represents an important step towards the development of an open source platform for clinical use, capable of closed-loop stimulation and applicable in personalized BCI therapies for patients suffering from neurological disorders.

## ACKNOWLEDGEMENTS

We are grateful to be able to work with this noble animal Belka, and for the care and attention provided by the Veterinary staff at Mayo Clinic. Belka will be available for adoption at the conclusion of the research in accordance with the Minnesota state Beagle Freedom Bill. This work was supported by the NIH U01-NS128612 (KJM, GAW, PB). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Our founder played no role in data collection and analysis, study design, decision to publish, or manuscript preparation.

## REFERENCES

- [1] Johnson PJ *et al.* Stereotactic cortical atlas of the domestic canine brain. *Scientific Reports*. 2020;10(1).
- [2] Schalk G *et al.* Toward a fully implantable ecosystem for adaptive neuromodulation in humans: Preliminary experience with the cortec braininterchange device in a canine model. *Frontiers in Neuroscience*. 2022;16:932782.
- [3] Schalk G, Mellinger J. A practical guide to brain-computer interfacing with BCI2000: General-purpose software for brain-computer interface research, data acquisition, stimulus presentation, and brain monitoring. Springer Science & Business Media (2010).
- [4] Bunford N, Andics A, Kis A, Miklósi Ádám, Gácsi M. Canis familiaris as a model for non-invasive comparative neuroscience. *Trends in Neurosciences*. 2017;40(7):438–452.
- [5] Story BD *et al.* Canine models of inherited musculoskeletal and neurodegenerative diseases. *Frontiers in Veterinary Science*. 2020;7.
- [6] Miller KJ, Honey CJ, Hermes D, Rao RP, Ojemann JG, *et al.* Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *Neuroimage*. 2014;85:711–720.
- [7] Miller KJ, Schalk G, Fetz EE, Den Nijs M, Ojemann JG, Rao RP. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*. 2010;107(9):4430–4435.
- [8] Bci2000.org. [Accessed 14-03-2024]. [Online]. Available: \url{https://bci2000.org/downloads/doc/Canine\_Cortec\_Surgical\_Protocol.pdf}.
- [9] Gierthmuehlen M *et al.* Mapping of sheep sensory cortex with a novel microelectrocorticography grid. *Journal of Comparative Neurology*. 2014;522(16):3590–3608.
- [10] 3d slicer image computing platform. <https://www.slicer.org>. [Accessed 14-03-2024].
- [11] Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF. Automated electrocorticographic electrode localization on individually rendered brain surfaces. *Journal of neuroscience methods*. 2010;185(2):293–298.
- [12] Schalk G, McFarland D, Hinterberger T, Birbaumer N, Wolpaw J. BCI2000: A general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*. 2004;51(6):1034–1043.
- [13] Ayyoubi AH *et al.* Benchmarking signal quality and spatiotemporal distribution of interictal spikes in prolonged human ieeg recordings using cortec wireless brain interchange. *Scientific Reports*. 2024;14(1).
- [14] Welch P. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*. 1967;15(2):70–73.
- [15] Harris F. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*. 1978;66(1):51–83.
- [16] Miller KJ, Sorensen LB, Ojemann JG, Den Nijs M. Power-law scaling in the brain surface electric potential. *PLoS computational biology*. 2009;5(12):e1000609.
- [17] Guran CNA *et al.* Functional mapping of the somatosensory cortex using noninvasive fmri and touch in awake dogs. *bioRxiv*. 2023:2023–12.
- [18] Vanderzant CW, Schott RJ, Natale JE, Pondo CA, D'Alecy LG. Somatosensory evoked potentials of the dog: Recording techniques and normal values. *Journal of Neuroscience Methods*. 1989;27(3):253–263.
- [19] Boch M, Karl S, Sladky R, Huber L, Lamm C, Wagner IC. Tailored haemodynamic response function increases detection power of fmri in awake dogs (canis familiaris). *Neuroimage*. 2021;224:117414.
- [20] Mirka A, Brookhart JM. Role of primary visual cortex in canine postural control. *Journal of Neurophysiology*. 1981;46(5):987–1003.
- [21] Andics A, Gácsi M, Faragó T, Kis A, Miklósi Ádám. Voice-sensitive regions in the dog and human brain are revealed by comparative fmri. *Current Biology*. 2014;24(5):574–578.
- [22] Lopes da Silva F, van Lierop T, Schrijer C, Storm van Leeuwen W. Organization of thalamic and cortical alpha rhythms: Spectra and coherences. *Electroencephalography and Clinical Neurophysiology*. 1973;35(6):627–639.
- [23] Kujala MV *et al.* Reactivity of dogs' brain oscillations to visual stimuli measured with non-invasive electroencephalography. *PLOS ONE*. 2013;8:1–8.
- [24] Miller KJ *et al.* Human motor cortical activity is selectively phase-entrained on underlying rhythms. *PLoS Computational Biology*. 2012;8(9):e1002655.

# MOVEMENT ASSOCIATED INCREASE IN THALAMIC BROADBAND SPECTRAL POWER IS A POTENTIAL FEATURE FOR BCI CONTROL

Bryan T. Klassen<sup>1</sup>, Matthew R. Baker<sup>2</sup>, Gabriela Ojeda Valencia<sup>2</sup>, Kai J. Miller<sup>2</sup>

<sup>1</sup> Department of Neurology, Mayo Clinic, Rochester, MN, USA

<sup>2</sup> Department of Neurosurgery, Mayo Clinic, Rochester, MN, USA

E-mail: [klassen.bryan@mayo.edu](mailto:klassen.bryan@mayo.edu)

## ABSTRACT

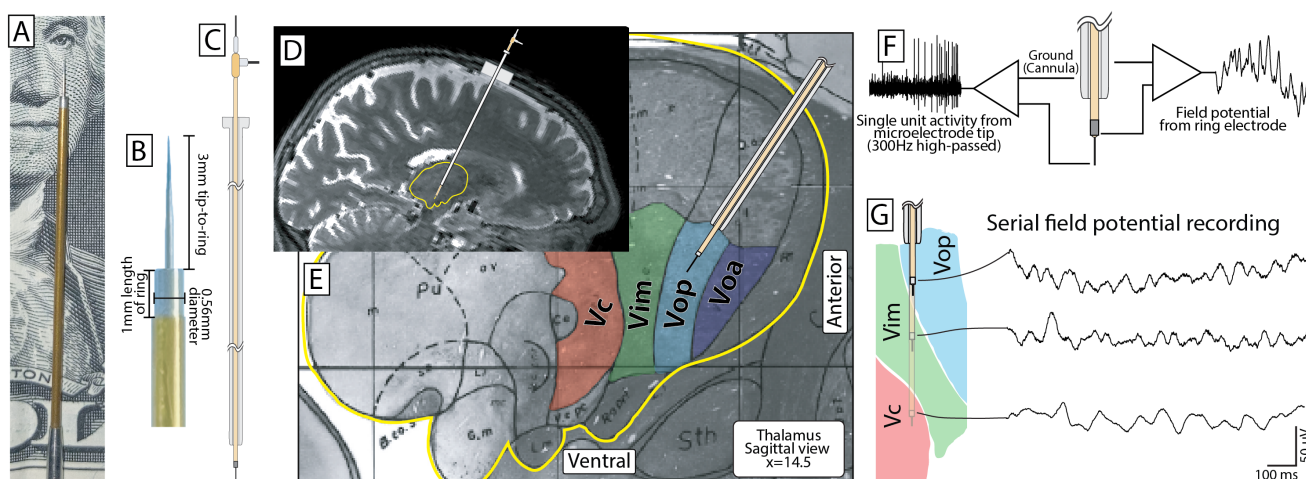
Signals within the subcortical brain regions may be useful as control signals for brain computer interfaces (BCI). In this study we show, using a simple hand movement task, that focal increases in broadband spectral power, which are commonly used to control cortically-based BCI interfaces, may also be observed in the ventralis intermedius (VIM) thalamic nucleus, a key relay for the cerebellar outputs to the motor cortex that help to regulate voluntary movement.

## INTRODUCTION

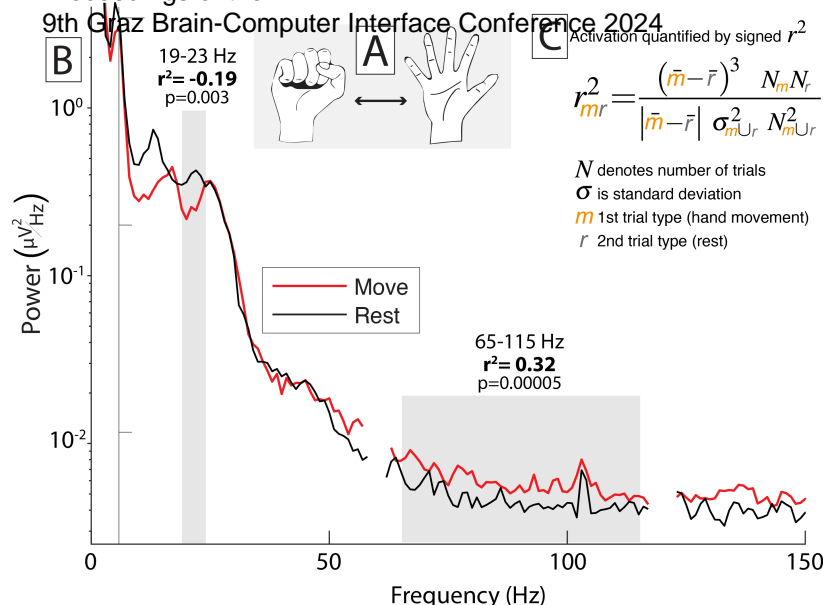
The cerebral cortex has been the predominant source for control signals in brain-computer interface (BCI) technologies to date, largely due to its direct involvement in motor control and sensory processing activities[1]. While this approach has yielded functional BCIs, tapping into the rich potential of subcortical structures such as the thalamus and other deep brain nuclei may confer distinct advantages. These regions, which play crucial roles in the modulation and relay of motor and sensory information,

offer a different set of neural signals that could enhance the practicality, fidelity, and functionality of BCIs.

Among these deep nuclei, the motor thalamus is an intricate structure with functional topography parcellated by the motor homunculus and overlapping representations from basal ganglia and cerebellar nuclei. Such a configuration suggests its potential as a robust source of control signals for brain-computer interfaces. Within the thalamus, the motor homunculus representation of the entire body can be accessed within a spatial span of mere millimeters. Like cortical activity, thalamic movement related oscillatory activity in the beta frequency range (13-30 Hz) is suppressed during movement[2]. However, the broad spatial distribution of this desynchronization within the motor thalamus suggests it may be too diffuse to resolve somatotopic movement, constraining the scope of a BCI driven by this activity. Increases in broadband spectral power during movement are observed in motor cortex, with a distribution much more focal than for low frequency desynchronization[3]. Whether focal increases in



**Figure 1: Experimental setup.** (A) Recordings were obtained using a clinical DBS electrode both (B) with a high-impedance tip and with a low-impedance ring having surface area of 0.56 mm<sup>2</sup>. The electrode was passed through a cannula (C) that had been inserted into the dorsal thalamus, as shown on a sagittal MRI slice (D), aligned along a trajectory targeting the VIM thalamic nucleus (E). (F) Local field potentials were recorded from the ring electrode and single unit activity was recorded from the microelectrode tip; the shaft of the cannula was referenced in all cases. (G) This study analyzes the field potentials at multiple evenly spaced sites along the trajectory recorded serially as the electrode was passed to target.



**Figure 2: Movement associated spectral shifts.** At each recording site, subjects performed a simple motor task alternating between (A) opening/closing of the hand and rest. (B) Averaged power spectral densities (PSD) from one recording site for Subject 1 are shown for movement and rest epochs. In this case, power within low frequency (<30 Hz) oscillations decreased during movement while broadband spectral power (65-115 Hz) increased. PSD for a subject were normalized to the global mean across all trials, and a signed  $r^2$  cross-correlation value was calculated from the mean PSD for movement versus rest trials, as shown in equation (C). The  $r^2$  for both broadband change and change in a subject-specific 5 Hz wide low frequency band were computed.

broadband spectral power can be observed in this region has not been well studied.

The posterior region of motor thalamus, the ventral intermedius (VIM) nucleus, relays motor signals between cerebellum and motor cortex and is a key therapeutic target for deep brain stimulation treating tremor and other movement disorders[3]. This provides an opportunity to study the electrophysiology of motor thalamus during movement. In the present study, we recorded from the VIM nucleus in awake patients undergoing deep brain stimulation surgery for tremor and hypothesized that movement associated increases in broadband power would be observed.

## MATERIALS AND METHODS

**Subjects:** Fourteen subjects undergoing deep brain stimulation electrode placement into the bilateral ventralis intermedius (Vim) nucleus of the thalamus for treatment of disabling tremor consented to participate in a research protocol during the awake surgery. The study and consent procedures were approved by Mayo Clinic's internal review board (IRB no. 19-009878).

**Recordings:** Serial thalamic recordings during task performance were obtained at multiple evenly-spaced recording sites (2 or 3 mm apart) as a microelectrode/macroelectrode (AlphaOmega Sonus STR-009080-00) was advanced towards the inferior border of the thalamus (Fig. 1). In 4 subjects, simultaneous recordings were obtained from two electrodes arranged parallel to each other along the anterior-posterior plane. Data were recorded to an

AlphaOmega Neuromega system, referenced to the shaft of the electrode cannula and using a sampling rate of 44 kHz. Surface EMG was recorded using pairs of bipolar-referenced Ag/AgCl electrodes placed 2 cm apart overlying the forearm muscles for finger flexion/extension.

**Motor Task:** At each recording site, subjects were verbally and visually cued for two alternating conditions: 1) rest, 2) continuous opening/closing movements of the dominant hand. Individual task epochs were 5 seconds in duration. The sequence was repeated 20 times per site. Compliance with the task was assessed in real-time by monitoring of EMG activity.

**Power Spectral Density and Cross Correlations:** All analyses were performed in Matlab. Epochs of rest versus movement were manually segmented via visual inspection of the rectified EMG; ambiguous epochs were

rejected.

Averaged power spectral densities (PSD) for individual movement or rest epochs were calculated from 1 to 300 Hz, with 1 Hz frequency resolution, using Welch's method of overlapping periodograms with a 1 second Hann window and 0.5 second overlap to attenuate edge effects. Averaged PSDs were then normalized to the global mean across all trials.

At each recording site, we calculated signed  $r^2$  cross-correlation values ( $r^2$ ) by comparing the mean PSD between movement and rest trials (Fig. 2). As a proxy for broadband activity,  $r^2$  was calculated for the 65-115 Hz band (avoiding line noise at 60 and 120 Hz). Because the center frequency for movement-related oscillations in the lower (<30 Hz) frequency bands varied between subjects, we systematically determined for each subject which 5 consecutive frequency bins, within the range of 8-30 Hz, was able to maximally discriminate between movement and rest as follows: 1)  $r^2$  values for each 1 Hz bin within the 8-30 Hz range were averaged across all recording sites in a subject and 2) the 5 contiguous bins with the highest sum of  $r^2$  values was selected.

**Plotting Power Changes on Subject-Specific MRI:** The site of each recording relative to the final lead position was known, allowing for MRI coordinates for each recording site to be computed using offsets to the lead artifact as seen on a postoperative CT scan co-registered to a preoperative T1-weighted MRI using mutual information in SPM12. The MRI was resliced in plane with recording sites and served as the background on which data was plotted[5].

Separate plots were prepared for movement associated power changes in the low frequency band and in broadband. We set significance at 0.95, uncorrected. Recording sites with a significant  $r^2$  were plotted in red (movement associated increase) or blue (movement associated decrease), with the sizes scaled to the maximum  $r^2$  value for that subject. Sites without significant power change were plotted with a white circle of fixed diameter.

## RESULTS

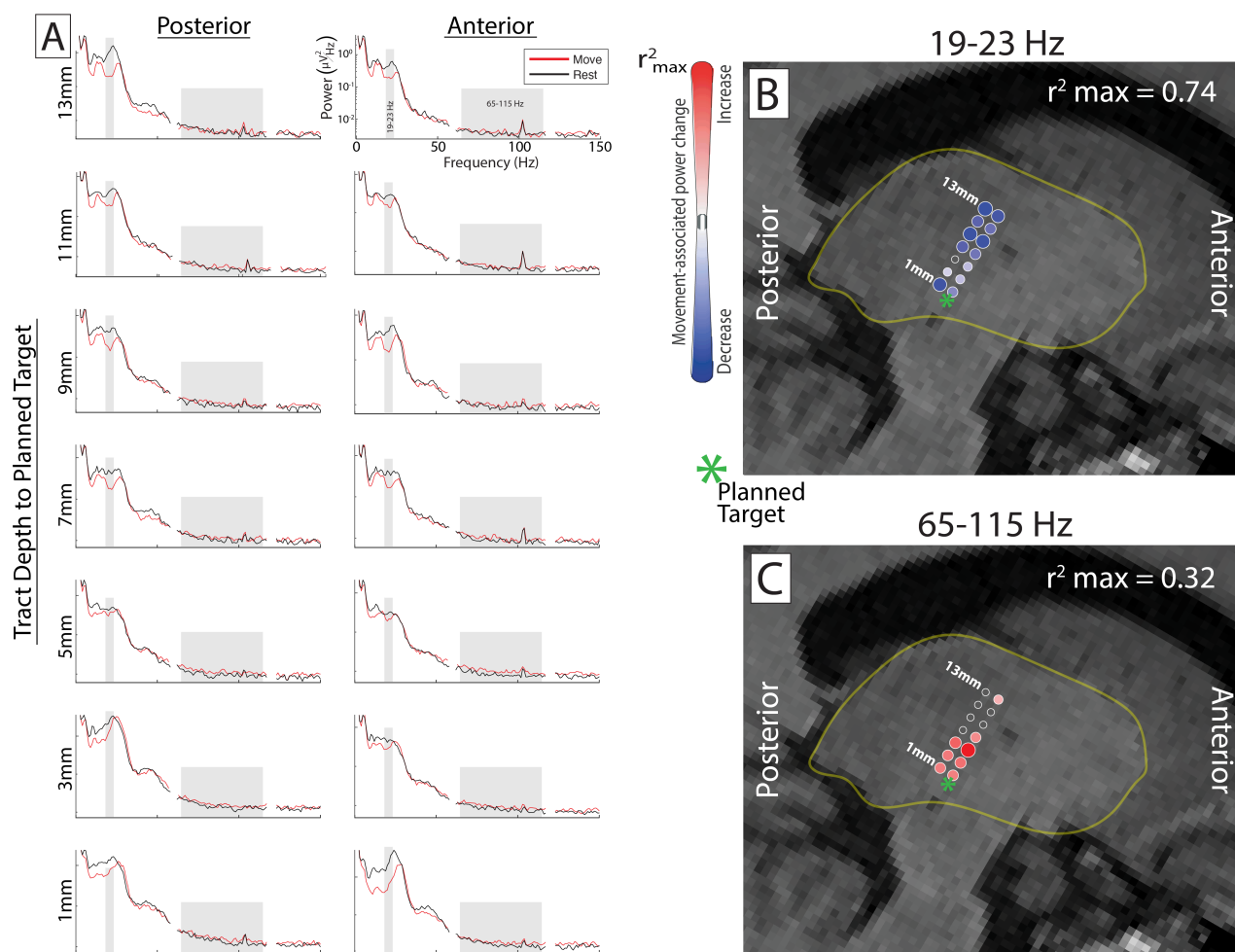
**Broadband Spectral Changes:** Example power spectral densities (PSD) for movement and rest epochs from one recording site for Subject 1 are shown in figure 2B. Figure 3 shows the spectra at each recording site for this single subject (A) and  $r^2$  maps for the low frequency band (B) and broadband (C). A significant increase in broadband power during movement was observed in at least one recording site for 8 of the 14 subjects ( $r^2_{\max}$  range = 0.11-0.44) (Fig 4C). Sites showing significant broadband power increase were most often found at the inferior recording sites, in the dorsal thalamus near the

predicted region of the VIM nucleus.

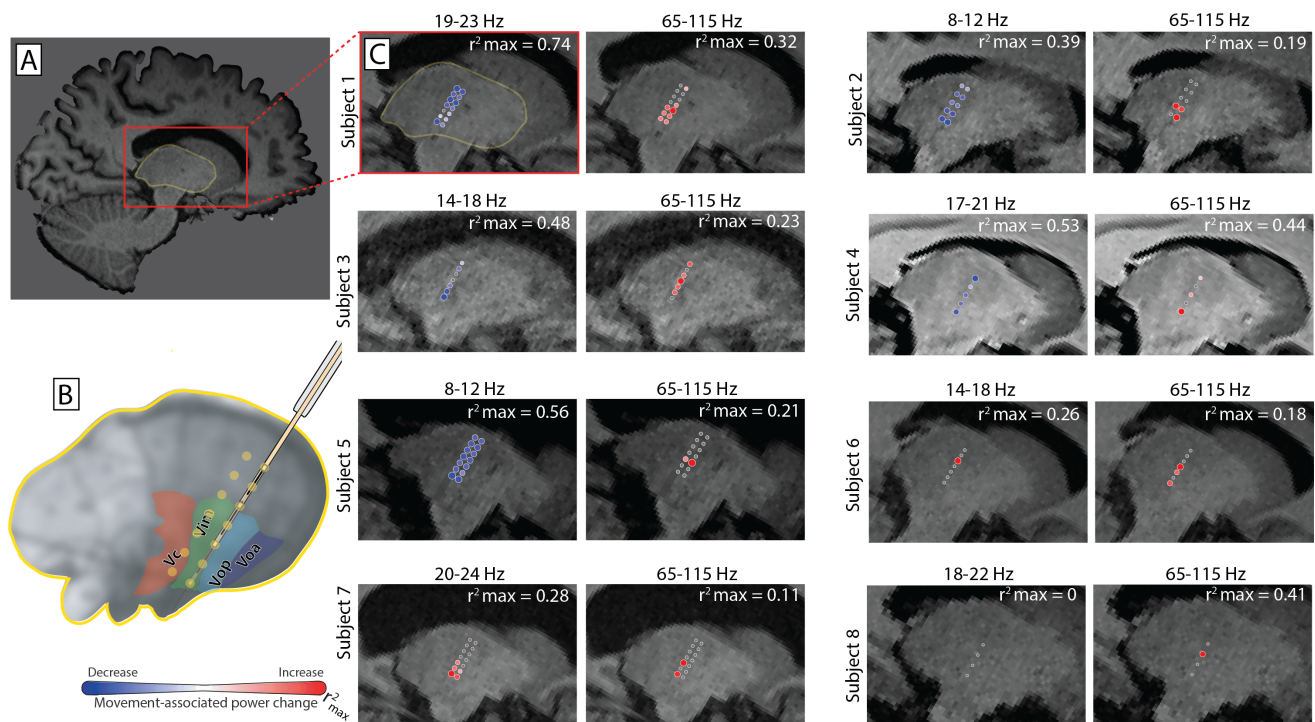
**Low Frequency Oscillations:** We found that the low frequency band showing the greatest power change with movement varied between subjects, from as low as 8-12 Hz, up to 20-24 Hz. Significant decreases in power within the subject-specific low frequency bands were present in at least one recording site for 11 of 14 total subject (data not shown), and for 5 of the 8 subjects with increase in broadband (Fig. 4c). In two cases (subjects 6 and 7) there was increased power within the band, and in one (case 8) there was no significant change with movement. Sites showing significant a power decrease in the low frequency band were widely distributed through the sampled region of dorsal and ventral thalamus.

## DISCUSSION

We show that a focal movement associated increase in broadband spectral power may be observed in the VIM thalamus during voluntary movement. In our subjects, increases in broadband power tended to be more spatially



**Figure 3: Spectra and  $r^2$  plots for a single subject.** (A) Average power spectral densities during movement (red) and rest (black) shown for each of 14 recording sites sampled in subject 1. The low (19-23 Hz) and broadband (65-115 Hz) frequency ranges are highlighted with gray boxes. The  $r^2$  values at each site are shown for (B) the 19-23 Hz band and (C) the 65-115 Hz band as dots plotted on sagittal T1 MRI slice. A red dot indicates an increase in power with movement, and a blue dot indicates a decrease. The size of each dot represents the absolute value of  $r^2$  scaled to the maximum  $r^2$  value for the frequency band (as shown in upper right-hand corner of each plot). In this subject, the sites were arranged along two parallel tracts 2mm apart in the anterior/posterior plane with serial recordings taken every 2 mm during electrode decent to the planned target (green asterisk).



**Figure 4: Subjects with significant increase in broadband spectral power during movement.** (A) A sagittal T1 MRI slice from subject 1, showing the thalamus (yellow outline) and the cropped/enlarged region used for  $r^2$  plots (red box). (B) Cartoon thalamus showing the ventral sensorimotor nuclei including the VIM target. Approximate recording locations are shown relative to these nuclei for the case of two parallel tracts (anterior/posterior) and for seven recording sites spaced 2mm apart. (C) All subjects with significant movement associated broadband increase. The  $r^2$  at each recording site is plotted on a sagittal T1 MRI for both the subject specific low frequency band and broadband. Red dots indicate and increase in spectral power with movement, blue a decrease. The size of each dot represents the absolute value of  $r^2$ , scaled to the maximum  $r^2$  value for that band.

discrete than were the desynchronizations of low frequency oscillations, as has been seen previously in motor cortex.

However, we did not observe broadband power increases in nearly half of the subjects. This may be because the frequency range of interest approaches the noise floor of our recording paradigm. Also, the operating room has many idiosyncratic sources of electrical noise which vary from case-to-case and even through the course of a single surgical case.

In most patients, a diffuse desynchronization of low frequency oscillations was seen during movement. Qualitative review of the spectra shows that there is often more than one oscillation in this range, and that in some cases the center frequency of the most prominent oscillation shifted with movement. Therefore, our limiting analysis to the 5 Hz band yielding the highest  $r^2$  value was an overly simplistic approach.

It is possible that the serial nature of recordings allowed subtle changes in behavioral state of the patient (attention, etc.) to confound results. This may be one explanation for the patchy spatial distribution of responses in a few subjects. Future studies will focus on simultaneous recordings at multiple sites during simple hand movements to allow for a more consistent behavioral state under which power differences across the recording sites can be more directly compared.

Upcoming work will include simple hand/tongue/foot

movement tasks and radially segmented recording electrodes to attempt to resolve somatopic representation using the broadband power shifts we report in this study.

## CONCLUSION

We find that focal increases in thalamic broadband spectral power are detected in a majority of subjects during a simple hand movement task. Further studies are needed to determine if this may be a signal robust enough to serve as an alternative control for BCI applications.

## ACKNOWLEDGEMENTS

We are grateful to the patients who volunteered their time to participate in this research. This work was supported by the state of Minnesota through a Minnesota Partnership Grant for Biotechnology and Medical Genomics MNP#21.42 (KJM/BTK), NIH U01-NS128612 (KJM), the Brain and Behavior Research Foundation (KJM), and the Foundation for OCD Research (KJM). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Our funders played no role in data collection and analysis, study design, decision to publish, or manuscript preparation.

## REFERENCES

- [1] Gallego JA, Makin TR, McDougale SD. Going beyond primary motor cortex to improve brain-computer interfaces. *Trends Neuroscience*. 2022;45(3):176-183.
- [2] Guehl D *et al.* Usefulness of thalamic beta activity for closed-loop therapy in essential tremor. *Scientific Reports*. 2023;13:22332.
- [3] Miller KJ *et al.* Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*. 2007;27(9):2424-2432.
- [4] Benabid AL *et al.* Long-term suppression of tremor by chronic stimulation of the ventral intermediate thalamic nucleus. *Lancet*. 1991;337:403-406.
- [5] Richner TJ, Klassen BT, Miller KJ. An in-plane mirror-symmetric visualization tool for deep brain stimulation electrodes. In: 2020 42<sup>nd</sup> Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020, 1112-1115.

# DYNAMIC SUPPRESSION OF THE CORTEX THROUGH SYNCHRONISATION DURING BRAIN COMPUTER INTERFACING

F.E. Permezel<sup>1</sup>, M.A. Jensen<sup>2</sup>, D. Hermes<sup>1,3</sup>,  
K.J. Miller<sup>1,4</sup>

<sup>1</sup>Department of Neurology, Mayo Clinic, MN, USA

<sup>2</sup>Department of Neurosurgery, Mayo Clinic, MN, USA

<sup>3</sup>Department of Biomedical Engineering, Mayo Clinic, MN, USA

E-mail: Permezel.fiona@mayo.edu

## ABSTRACT:

This study investigates the neural dynamics of motor imagery and brain-computer interface (BCI) feedback through electrocorticography (ECoG). It focuses on how 12-20Hz rhythm entrainment with broadband power indicates cortical synchronization and suppression. The research examines 12-20Hz rhythm entrainment across rest and active phases in a motor task and BCI imagery feedback task. Using speech-associated broadband power increases in a speech motor area, a patient controlled a BCI system with word repetition imagery. The study examined broadband power shifts between rest and active task conditions, revealing increased power shifts in the speech motor area during the BCI imagery task as well as a unique activation in the dorsal motor area. Notably, it found increased broadband power to 12-20Hz rhythm coupling, indicating suppression of cortical activity, in the dorsal motor cortex during the BCI imagery feedback task's rest phase compared to the motor task rest phase, which may be suggestive of "cognitive control" over cortical suppression.

## INTRODUCTION:

Recent advancements in brain-computer interface (BCI) technology have underscored the potential of motor imagery in enhancing motor skill acquisition and rehabilitation, particularly in individuals afflicted with neurological conditions [1-6]. The intricate neural mechanisms underlying these processes, including the activation of neocortical areas similar to those engaged during actual motor movements, have been illuminated through a variety of neuroimaging techniques, albeit with ongoing debates regarding the role of primary motor cortex in motor imagery [6-15]. Notwithstanding, the precision of electrocorticography (ECoG) in mapping somatotopic functions and capturing high-frequency cortical dynamics [16-22] has significantly advanced our understanding of these neural phenomena.

In parallel, the exploration of the beta rhythm in somatomotor regions has elucidated its inverse relationship with sensory processing and motor

production, alongside its modulation in response to movement and motor imagery [18, 23-32]. These dynamics are often assessed using Phase Locking Value (PLV), a method well-documented by Lachaux et al. [47] for measuring the synchrony in brain signals, particularly in cortical-subcortical circuits which play a potential role in organizing somatomotor functions. This rhythm's involvement in cortical-subcortical circuits, particularly in relation to local neuronal activity and its potential role in organizing somatomotor function, has begun to be quantified, revealing a complex interplay of rhythm phase and broadband signal amplitude [33-39]. Miller [40] used ECoG electrodes in humans to study the phase-coupling of neuronal activity with the beta rhythm during finger movements. This study uncovered that broadband neuronal activity is phase-coupled with the beta rhythm, with this phase-coupling decreasing in amplitude during movement. The ECoG recording channels of greatest phase-coupling change with movement however, were not the same as those with the greatest beta activity change with movement. This study suggested that assessment of rhythm phase and broadband amplitude coupling may offer insights into the interplay of distinct brain rhythms and broadband amplitude during different activities. Building on these insights, our study leverages data from a participant engaged in motor and BCI feedback during motor imagery conditions collected previously by the Ojemann [41] group. We aim to dissect the nuanced shifts at specific electrode sites, with comparison between the motor vocalization and BCI imagery feedback conditions in i) broadband amplitude in response to action, ii) amplitude shifts of various rhythm frequencies in response to action, and iii) to delineate the changes with action and rest in the coupling between 12-20Hz rhythm phase and broadband amplitude. By examining these dynamics our work seeks to illuminate the underlying neural mechanisms of motor control and a BCI imagery feedback task, offering novel insights into their application in rehabilitative and BCI technologies.

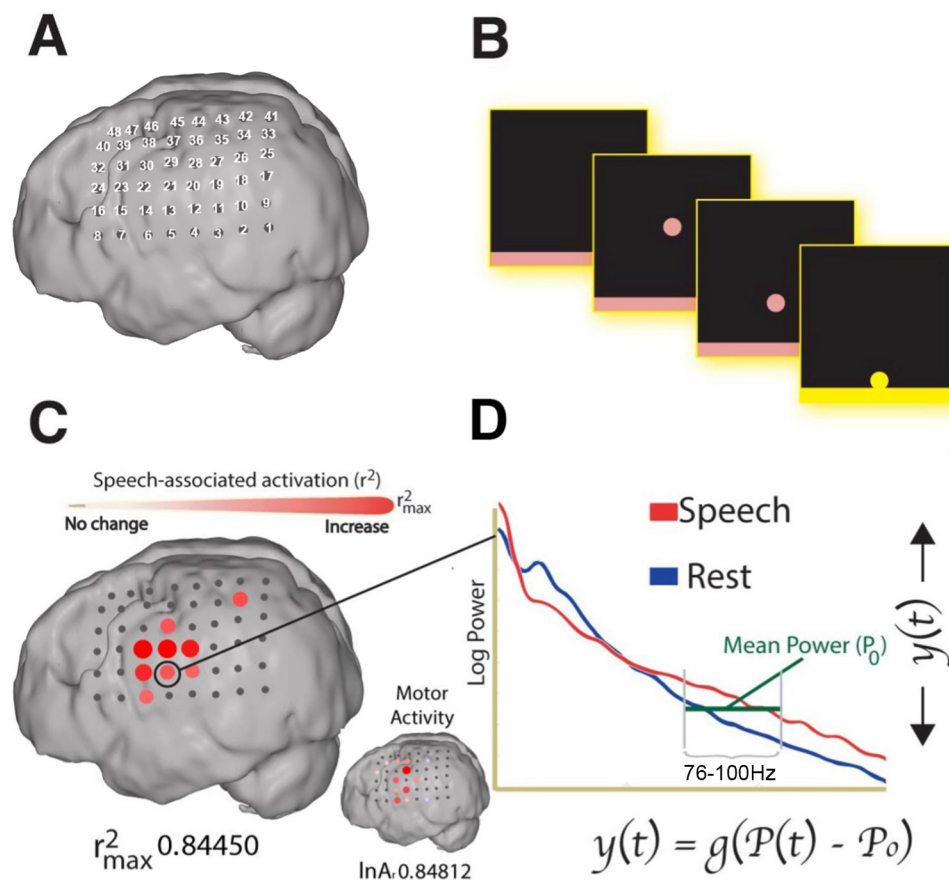


Figure 1: **BCI Imagery Feedback Task and Relation to Broadband Shift.** (A) Map of ECoG grid showing electrode positions. (B) Visual feedback of this broadband power change which causes moving of a dot on a screen in one direction. Reduction this broadband power increase will cause movement of the dot in the other direction. In the BCI imagery feedback experiment the patient is visually cued to try to move the dot in one direction or the other. (C) Dots represent significant ( $p < 0.05$  FDR corrected) broadband 76-100Hz power increase during BCI imagery feedback task active condition (motor imagery of saying „move“) compared to a rest condition. The smaller brain image shows the distribution of the same for the motor vocalization experiment. (D) The power spectra demonstrating the broadband increase with BCI imagery feedback compared to rest.

## MATERIALS AND METHODS

**Participant:** The study included an epilepsy patient undergoing craniotomy for seizure localization, with informed consent under U. Washington IRB (#12193) approval. Data was downloaded from the ECoG Library database [42].

**Stimulation:** Electro-cortical mapping identified motor/speech cortices using 5–10 mA pulses.

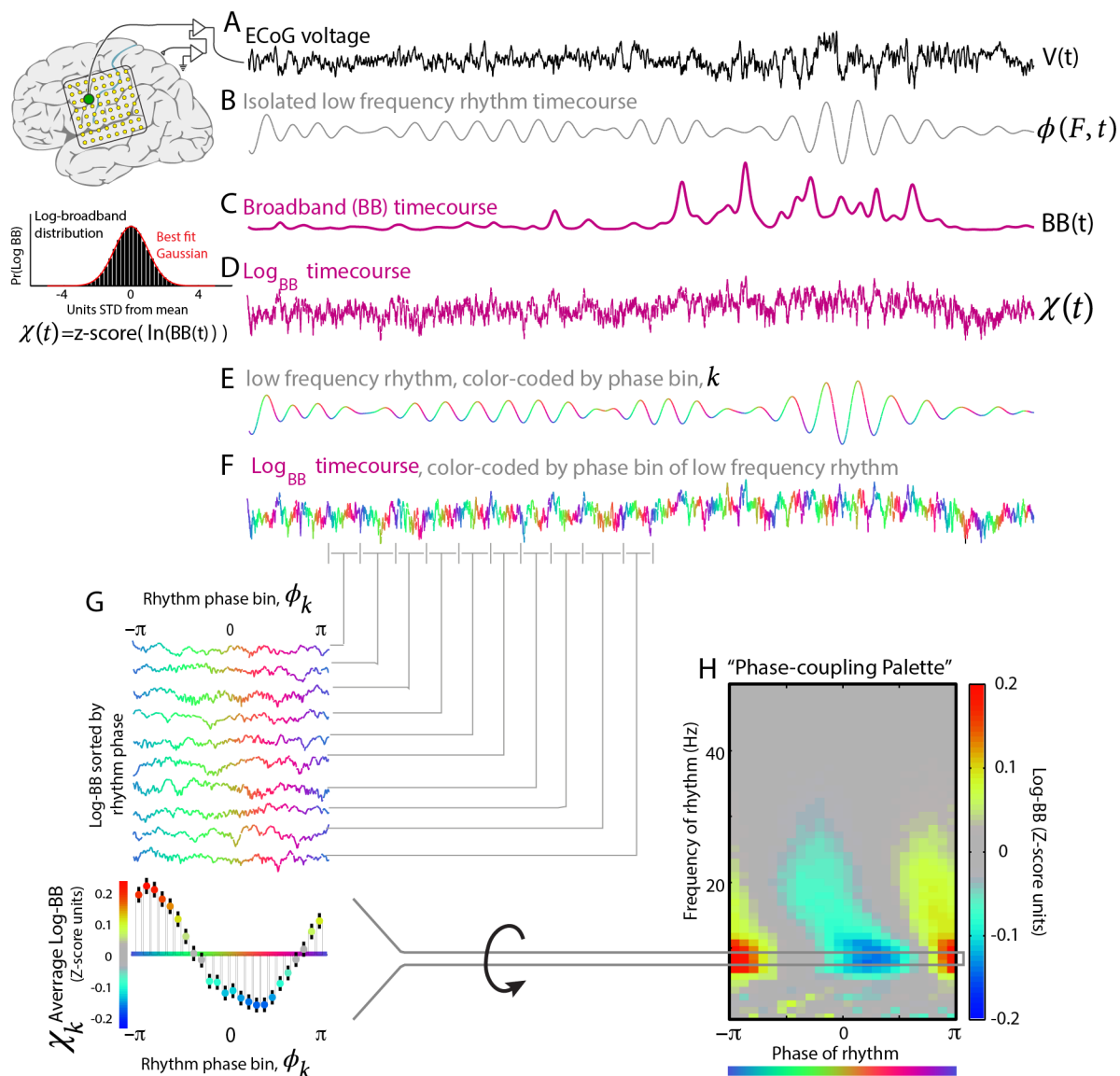
**Recordings:** Signals from subdural platinum ECoG electrode arrays, with a total of 48 electrodes (Fig. 1.), were recorded at 1000Hz and processed for analysis.

**Tasks:** The patient performed the motor task of saying the word „move“, with multiple trials interspersed with rest periods. The patient then performed a BCI feedback task, where motor imagery of saying the word „move“ was utilized to move a cursor to a target, via ECoG recording detecting broadband power increase at an

electrode chosen by the greatest broadband power increase in the motor task. The patient then is instructed to move the cursor away from the target (Fig. 1), which is actuated by reduction of this broadband increase at the specified electrode. It should be noted that the „rest“ condition in the feedback task was an „active rest“ task as the patient was deliberately attempting to move the cursor away from the „active vocalization imagery“ task.

**Electrode Localization and Brain Mapping:** Locations were estimated from x-rays and mapped to visualize activity distribution.

**Data Analysis:** A high-frequency broadband (HFB) ranging from 76–100 Hz was selected for the analysis. In brief this particular interval was chosen because it lies within this broad increase, avoids 60 Hz contamination, and matches the 25 Hz width of the low-frequency band (LFB), with further discussion on this



Figure

**2: Relation between broadband power and phase of ECoG rhythms.** (A) ECoG potential is measured from the cortex (green dot). (B) Example low frequency rhythm obtained by convolving ECoG with a simple Morlet wavelet. (C) Fluctuations in broadband power are also extracted from ECoG potential. (D) Log values of the time-dependent broadband have a normal distribution, so the z-scored log broadband timeseries is obtained. (E) The rhythm timeseries from B is shown color coded for instantaneous phase (relative to positivity peak of the potential). (F) The timeseries of the z-scored log of the broadband, color coded by the coincident phase of the low frequency rhythm. (G) The log-broadband signal is aligned with the phase of the low frequency rhythm (as color-coded). The average of the log-broadband amplitude is obtained for phase bins. Error bars denote 3 times the standard error of the mean (3\*SEM) for each phase bin. This can be appreciated in one dimension as a row in the phase coupling palette (in H). (H) The full "Phase coupling palette" obtained by repeating the process detailed in E-G at each frequency from 1-50 Hz., showing modulation of broadband power with the full range of frequencies.

HFB frequency range selection outlined in prior studies [18]. In the patient's tests, which included verbalizing the word "move" and actual leg movement, only the verbal command was used for the feedback task. Therefore, comparisons were made between the periods of saying "move" and the rest periods immediately after, to account for the task-specific beta rebound following movement as noted by Pfurtscheller [31]. Cursor speed in the feedback task was updated every 40 ms, based on the power calculated from 79-95Hz and

electrode numbers 13 and 22 over the preceding 280 ms.

1. Power Spectral Analysis: Power spectral density (PSD) of the signal was calculated for each epoch surrounding movement or imagery events.

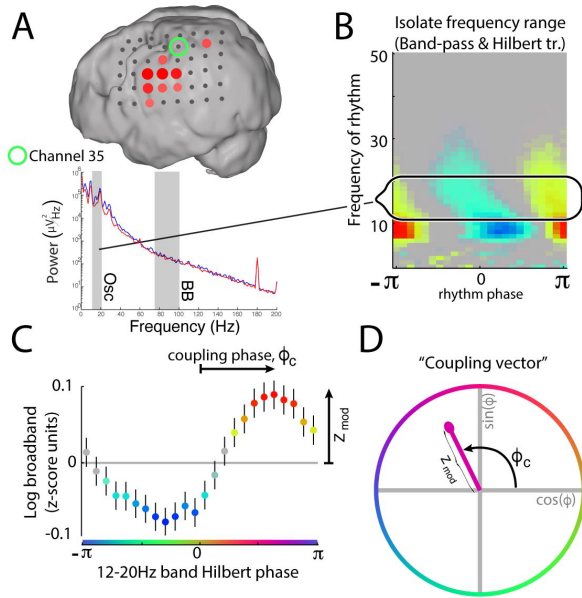
$$(P(f, q) = \frac{1}{T} \left| \sum_{t=-T/2}^{T/2} V(t_q + t) \cdot H(t) \cdot e^{i2\pi f t} \right|^2)$$

Here ( $T$ ) is the epoch duration, ( $t$ ) is event duration, ( $f$ ) is frequency, ( $q$ ) is the event, ( $V(t)$ ) is the electrode potential, ( $t_q$ ) are event markers and ( $H(t)$ ) is the Hann window function used to transform the time-domain data into frequency-domain representations, encompassing all frequencies during both rest and action phases.

2. High-Frequency Band Extraction (76-100 Hz): The averaged PSD for each movement or rest trial was normalized to the global mean across all trials. The power in the 76-100 Hz frequency range was extracted from the normalized PSD for each channel to focus on the high-frequency band (HFB) activity (Fig. 3A).

3. Phase-Amplitude Coupling (PAC) Analysis: Utilizing the Hilbert transform, the phase of the 12-20Hz rhythm and the amplitude of the HFB were used to generate phase-amplitude coupling metrics (Fig. 2):

$$(\tilde{V}(F, t) = r(F, t)e^{i\phi(F, t)})$$



**Figure 3: Broadband and Phase Amplitude Coupling Palette at One Electrode.** (A) Demonstration of Channel 35 power spectra with regards to the broadband power change during BCI imagery feedback task active condition (motor imagery of saying „move“) compared to a rest condition. (B) Phase amplitude coupling palette at Channel 35 with the 12-20Hz frequency range circled. (C & D) Phase of rhythm depicted by color and angle on the complex plane.

Where ( $\tilde{V}(F, t)$ ) is the complex PAC vector, for frequency range ( $F$ ) = 12-20Hz at time ( $t$ ). The “analytic amplitude” of the range ( $F$ ) at time ( $t$ ) is ( $r(F, t)$ ) and the “phase” is ( $\phi(F, t)$ ). Generation of metrics included constructing ‘palettes’ (Fig. 3) for each channel for each task to examine the relationship between phase and amplitude across frequencies. The

coupling vector magnitude ( $Z_{mod}$ ) and its preferred phase ( $w_c$ ) are calculated as:

$$(Z_{mod}e^{iw_c} = \frac{1}{2K} \sum_k x_k e^{iw_k})$$

where ( $x_k$ ) represents the average log-broadband amplitude in the  $k$ -th phase bin, ( $w_k$ ) is the central phase of the  $k$ -th bin, and ( $K$ ) is the total number of bins.

For assessing the distribution of phase-amplitude coupling values across trials (Fig. 4), the projected distribution is used:

$$(Z_{mod,q}(n) = Z_{mod}(n) \cdot \cos(w(n) - w_q))$$

where ( $n$ ) is the trial number, ( $w(n)$ ) is the preferred phase for the  $n$ th trial, and ( $w_q$ ) is the preferred phase of the mean coupling vector for trials of type  $q$ , and ( $N$ ) is this number of trials. The mean coupling vector is computed as:

$$(Z_{mod,q}e^{iw_q} = \frac{1}{N} \sum_n Z_{mod}(n)e^{iw(n)})$$

This method allows for the evaluation of the significance of phase-amplitude coupling by assessing the distribution of ( $Z_{mod,q}(n)$ ) values.

4. Broadband and Rhythmic Amplitude Analysis: For each channel during rest and action phases, the broadband amplitude, the phase-amplitude coupling amplitude, and the amplitudes for rhythmic activities in the specified frequency band (12-20Hz) were calculated.

5. Statistical testing: Phase and amplitude of coupling vectors were calculated per trial for rest and active conditions for both BCI imagery feedback and motor tasks. For all conditions, each trial’s coupling vector was then projected to the mean phase angle across all trials.

Signed  $r^2$  cross-correlation values at each channel were produced for each experimental condition (active vs rest) for each of broadband amplitude, phase-amplitude coupling amplitude and rhythmic amplitude.

A two-sample t-test was performed for each of broadband amplitude, phase-amplitude coupling amplitude and rhythmic amplitude and a significance  $p$  value calculated.

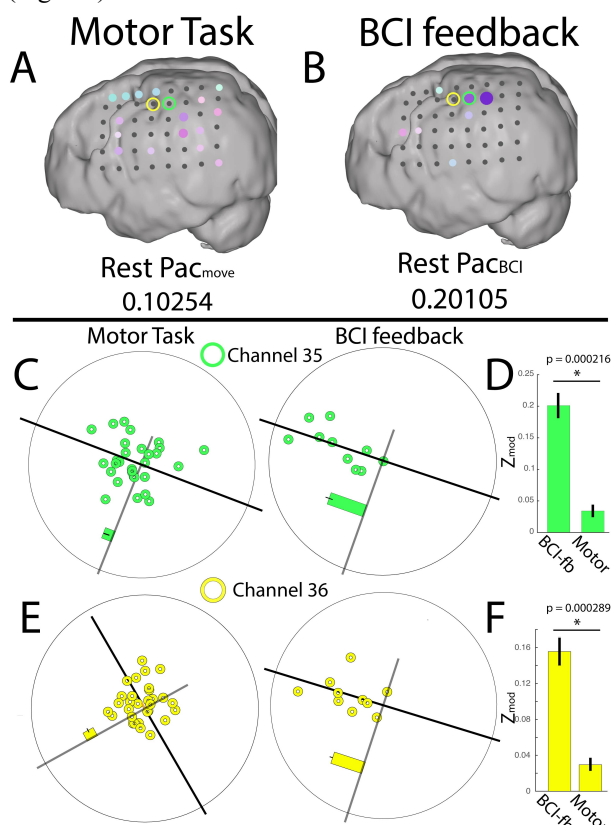
Channels with significant ( $p < 0.05$ )  $r^2$  values are demonstrated by colored circles on a 3D rendering of the patient’s brain with the diameter scaled to the magnitude of these  $r^2$  values.

Power spectra of action vs rest conditions were plotted for motor vocalization and imagery feedback tasks.

Phase-amplitude coupling amplitudes ( $Z_{mod}$ ) for motor and feedback rest conditions were highlighted on a 3D rendering of the patient’s own brain with again only channels with  $Z_{mod}$  with significant  $p$  values  $< 0.05$  with FDR correction highlighted with channel dot size proportional to  $Z_{mod}$  values. Dot color was related to the phase coupling angle  $\phi_c$ .

## RESULTS

Comparison of broadband 76-100Hz power increase between the motor vocalization task vs. rest and the BCI imagery feedback task vs. rest (significant R values,  $p < 0.05$  after FDR correction) revealed a similar area of broadband increase, although the area with broadband increase extended over more channels in the speech area in the imagery feedback task (Fig. 3A). There was also a distinctly separate increase in one channel (37) in the dorsal motor area, only in the imagery feedback task (Fig. 3A).



**Figure 4: Phase and Amplitude of the Coupling Vectors in the Rest Conditions.** (A) Shows the significant ( $p < 0.05$  FDR corrected) values for the amplitude of the coupling vector ( $Z_{mod}$ ) of 12-20Hz to broadband power in the rest condition of the motor vocalization task. The size of the dot indicates the  $Z_{mod}$  value, scaled to  $Z_{mod} = 0.2$ . The color of the dot indicates the degree of phase. (B) The same as for A except for depicting the rest condition of the BCI imagery feedback task. (C) The phase (angle from center) and amplitude (distance from center) of coupling vectors for each trial (each dot) of the rest condition in the motor task and BCI feedback task at Channel 35. The black line parallel to the bar indicates the mean phase angle across trials. (D) The significant difference from a two-sample t-test (Bonferroni corrected  $p < 0.001$ ) between the mean  $Z_{mod}$  (coupling amplitude) over trials between the BCI imagery feedback task and the vocalization motor task at Channel 35. (E) The same as (C) except for Channel 36. (F) the same as (D) except for Channel 36.

There was significant 12-20Hz rhythm decrease (significant R values,  $p < 0.05$  after FDR correction) around the face area of the primary motor and premotor cortex in the motor vocalization task, however there was no significant 12-20Hz shift in the imagery feedback task.

There was no significant increase or decrease in 12-20Hz to broadband coupling vector amplitude for task vs. rest for either the motor vocalization or imagery feedback task.

Examination of the amplitude and phase of the 12-20Hz to broadband coupling vectors for the rest condition revealed significantly increased coupling ( $p < 0.001$  Bonferroni adjusted for  $p < 0.05$ ) in channels 35 and 36 in the imagery feedback task at rest when compared to the motor vocalization task at rest. Channels 35 and 36 (Fig. 4) are in the dorsal motor area, distinctly separate from the speech area, and from the motor face area that demonstrated the 12-20Hz decrease with motor vocalization.

## DISCUSSION

The finding of greater broadband 76-100Hz power increase over the speech area for task vs rest in the imagery feedback task compared to the motor vocalization task vs rest (Fig. 3A) is consistent with the results in the Ojemann group 2010 [41] study and other scientific investigations [43]. The heightened broadband power during BCI imagery feedback, relative to motor action, might stem from the initial selection of channel (13) for imagery feedback, based on its significant broadband power change during motor action. However, in our study, channel (13) was not the primary site of increased broadband power in the imagery feedback condition; instead, adjacent channels exhibited more substantial broadband increases compared to the motor vocalization task. This broader increase in broadband power during the BCI imagery feedback task, versus a motor task, suggests that the feedback mechanism itself might induce larger broadband shifts, a concept supported by Ushiba [44] and Neuper [45].

The single channel (37) of broadband power increase in the dorsal motor area (Fig. 3A) is difficult to explain during the motor vocalization imagery task and could be related to a motor association area activation [46].

The Ojemann group [41] found higher R values for low frequency (8-32Hz) power decrease in their imagery feedback task than their motor task. The current study found significant R values for 12-20Hz decrease with task only in the motor vocalization task and not the imagery feedback task. The reason for this difference in findings may be the different range of low-range frequencies tested. It may be that 20-32Hz is an important rhythm range for task related modulation in a vocalization imagery feedback task in the face motor area. In light of variations observed in correlation

between low-frequency power and task performance across studies, it is crucial to consider how different approaches to power estimation and normalization might influence findings. The methods employed in this study, detailed in the Methods section, follow established protocols but could still yield results that differ from those obtained using alternative analytical techniques. These differences might stem from how power spectra are normalized or from the spectral components that are emphasized or attenuated through different processing steps.

Miller in 2012 [40] demonstrated a broadband amplitude entrainment with 12-20Hz rhythm during rest periods (fixation) in the peri-central cortex. During finger movement this phase-entrainment was diminished or eliminated, suggesting the beta rhythm in the peri-central cortex to be a gating mechanism of motor function. The results of the present study did not demonstrate diminished entrainment of the broadband amplitude with 12-20Hz rhythm during motor vocalization or the active imagery phase of the feedback task. This may be related to the primary sites of motor activation in this study being outside of the dorsal peri-central cortex, as the task was motor vocalization and imaginary vocalization. There was however significantly higher broadband amplitude entrainment with a 12-20Hz rhythm in dorsal motor area channels (35 & 36) in the imagery feedback rest condition, when compared to the motor vocalization condition (Fig. 4). This relative dorsal motor cortex suppression during the “active relaxation” phase of the imagery feedback task may be evidence of unconscious suppression of motor cortex when attempting to moving a cursor in the opposite direction to that of the active imagery vocalization task. This suggests a degree of “cognitive control” over suppression of the cortex during the reinforced “rest” state of the BCI imager feedback experiment.

These results suggest a promising direction for future research in the BCI field, particularly in the use of phase-amplitude coupling (PAC) to assess deliberate cortical suppression during 'active rest' states. This approach may offer a novel method to electrophysiologically identify cognitive intentions, enhancing the responsiveness and adaptability of BCI systems. Further exploration into how variations in PAC correlate with specific cognitive tasks could lead to more intuitive interfaces that better align with user intent, potentially improving outcomes in therapeutic and rehabilitative applications.

## REFERENCES

- [1] Murphy, S.M., *Imagery interventions in sport*. Med Sci Sports Exerc, 1994. **26**(4): p. 486-94.
- [2] Dijkerman, H.C., et al., *Does motor imagery training improve hand function in chronic stroke patients? A pilot study*. Clin Rehabil, 2004. **18**(5): p. 538-49.
- [3] Page, S.J., P. Levine, and A. Leonard, *Mental*

- practice in chronic stroke: results of a randomized, placebo-controlled trial*. Stroke, 2007. **38**(4): p. 1293-7.
- [4] Alkadhi, H., et al., *What disconnection tells about motor imagery: evidence from paraplegic patients*. Cereb Cortex, 2005. **15**(2): p. 131-40.
- [5] Sharma, N., V.M. Pomeroy, and J.C. Baron, *Motor imagery: a backdoor to the motor system after stroke?* Stroke, 2006. **37**(7): p. 1941-52.
- [6] Hochberg, L.R., et al., *Neuronal ensemble control of prosthetic devices by a human with tetraplegia*. Nature, 2006. **442**(7099): p. 164-71.
- [7] Naito, E., et al., *Internally simulated movement sensations during motor imagery activate cortical motor areas and the cerebellum*. Journal of Neuroscience, 2002. **22**(9): p. 3683.
- [8] Jeannerod, M. and V. Frak, *Mental imaging of motor activity in humans*. Curr Opin Neurobiol, 1999. **9**(6): p. 735-9.
- [9] de Lange, F.P., K. Roelofs, and I. Toni, *Motor imagery: A window into the mechanisms and alterations of the motor system*. Cortex, 2008. **44**(5): p. 494-506.
- [10] Porro, C.A., et al., *Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study*. J Neurosci, 1996. **16**(23): p. 7688-98.
- [11] Roth, M., et al., *Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study*. Neuroreport, 1996. **7**(7): p. 1280-4.
- [12] Schnitzler, A., et al., *Involvement of primary motor cortex in motor imagery: a neuromagnetic study*. Neuroimage, 1997. **6**(3): p. 201-8.
- [13] McFarland, D.J., et al., *Mu and beta rhythm topographies during motor imagery and actual movements*. Brain Topography, 2000. **12**(3): p. 177-186.
- [14] Guillot, A., et al., *Brain activity during visual versus kinesthetic imagery: An fMRI study*. Human Brain Mapping, 2009. **30**: p. 2157.
- [15] Hermes, D., et al., *Functional MRI-based identification of brain areas involved in motor imagery for implantable brain-computer interfaces*. Journal of Neural Engineering, 2011. **8**(2): p. 025007.
- [16] Crone, N.E., et al., *Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band*. Brain, 1998. **121** ( Pt 12): p. 2301-15.
- [17] Ball, T., et al., *Differential representation of arm movement direction*. Journal of Neural Engineering, 2009. **6**: p. 016006.
- [18] Miller, K.J., et al., *Spectral changes in cortical surface potentials during motor movement*. J Neurosci, 2007. **27**(9): p. 2424-32.
- [19] Aoki, F., et al., *Increased gamma-range activity in human sensorimotor cortex during performance of visuomotor tasks*. Clin Neurophysiol, 1999. **110**(3): p. 524-37.

- [20] Kubánek, J., et al., *Decoding flexion of individual fingers using electrocorticographic signals in humans*. Journal of Neural Engineering, 2009. **6**: p. 066001.
- [21] Miller, K.J., et al., *Decoupling the Cortical Power Spectrum Reveals Real-Time Representation of Individual Finger Movements in Humans*. Journal of Neuroscience, 2009. **29**(10): p. 3132.
- [22] Manning, J.R., et al., *Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans*. Journal of Neuroscience, 2009. **29**(43): p. 13613–13620.
- [23] Jasper, H.H., *Electrical activity of the brain*. Annual Review of Physiology, 1941. **3**(1): p. 377-398.
- [24] Miller, F.R., G.W. Stavraky, and G.A. Wootton, *Effects of eserine, acetylcholine and atropine on the electrocorticogram*. Journal of Neurophysiology, 1940. **3**(2): p. 131.
- [25] Bartley, S.H. and P. Heinbecker, *The response of the sensorimotor cortex to stimulation of a peripheral nerve*. American Journal of Physiology--Legacy Content, 1937. **121**(1): p. 21.
- [26] Penfield, W., *Mechanisms of voluntary movement*. Brain, 1954. **77**(1): p. 1.
- [27] Jasper, H. and W. Penfield, *Electrocorticograms in man: effect of voluntary movement upon the electrical activity of the precentral gyrus*. European Archives of Psychiatry and Clinical Neuroscience, 1949. **183**(1): p. 163-174.
- [28] Bates, J.A., *Electrical activity of the cortex accompanying movement*. J Physiol, 1951. **113**(2-3): p. 240-57.
- [29] Jasper, H.H. and H.L. Andrews, *Brain potentials and voluntary muscle activity in man*. Journal of Neurophysiology, 1938. **1**(2): p. 87.
- [30] Crone, N.E., et al., *Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization*. Brain, 1998. **121** ( Pt 12): p. 2271-99.
- [31] Pfurtscheller, G., *Event-Related Desynchronization (ERD) and Event Related Synchronization (ERS)*. Electroencephalography: Basic Principles, Clinical Applications and Related Fields, ed. E. Niedermeyer and F. Lopes da Silva. 1999, Baltimore: Williams and Wilkins. 958-967.
- [32] Miller, K.J., et al., *Cortical activity during motor execution, motor imagery, and imagery-based online feedback*. Proceedings of the National Academy of Sciences, 2010. **108**(9): p. 4430-4435.
- [33] Murthy, V.N. and E.E. Fetz, *Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys*. Proc Natl Acad Sci U S A, 1992. **89**(12): p. 5670-4.
- [34] Murthy, V.N. and E.E. Fetz, *Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys*. J Neurophysiol, 1996. **76**(6): p. 3968-82.
- [35] Reimer, J. and N.G. Hatsopoulos, *Periodicity and evoked responses in motor cortex*. J Neurosci, 2010. **30**(34): p. 11506-15.
- [36] Manning, J.R., et al., *Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans*. Journal of Neuroscience, 2009. **29**(43): p. 13613.
- [37] Miller, K.J., *Broadband Spectral Change: Evidence for a Macroscale Correlate of Population Firing Rate?* Journal of Neuroscience, 2010. **30**(19): p. 6477.
- [38] Miller, K.J., et al., *Power-law scaling in the brain surface electric potential*. PLoS computational biology, 2009. **5**(12): p. e1000609.
- [39] Ray, S. and J.H. Maunsell, *Different origins of gamma rhythm and high-gamma activity in macaque visual cortex*. PLoS Biol, 2011. **9**(4): p. e1000610.
- [40] Miller, K.J., et al., *Human motor cortical activity is selectively phase-entrained on underlying rhythms*. PLoS Comput Biol, 2012. **8**(9): p. e1002655.
- [41] Miller, K.J., et al., *Cortical activity during motor execution, motor imagery, and imagery-based online feedback*. Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4430-5.
- [42] Miller, K.J., *A library of human electrocorticographic data and analyses*. Nature Human Behaviour, 2019. **3**(11): p. 1225-1235.
- [43] Wolpaw, J.R., J.D.R. Millán, and N.F. Ramsey, *Brain-computer interfaces: Definitions and principles*. Handb Clin Neurol, 2020. **168**: p. 15-23.
- [44] Ushiba, J., et al. *Feeling of Bodily Congruence to Visual Stimuli Improves Motor Imagery Based Brain-Computer Interface Control*. in *Converging Clinical and Engineering Research on Neurorehabilitation II*. 2017. Cham: Springer International Publishing.
- [45] Neuper, C., et al., *Motor imagery and action observation: Modulation of sensorimotor brain rhythms during mental control of a brain-computer interface*. Clinical Neurophysiology, 2009. **120**(2): p. 239-247.
- [46] Jensen, M.A., et al., *A motor association area in the depths of the central sulcus*. Nature Neuroscience, 2023. **26**(7): p. 1165-1169.
- [47] Lachaux, J. P., et al., *Measuring phase synchrony in brain signals*. Human brain mapping, 1999. **8**(4), 194–208.

# WAVELET PACKET DECOMPOSITION TO EXTRACT FREQUENCY FEATURES FROM SPEECH IMAGERY

A. Tates<sup>1</sup>, A. Matran-Fernandez<sup>1</sup>, S. Halder<sup>1</sup>, I. Daly<sup>1</sup>

<sup>1</sup>Brain-Computer Interfaces and Neural Engineering Lab, School of Computer Science and  
Electronic Engineering, University of Essex, Colchester, UK

E-mail: at18157@essex.ac.uk

**ABSTRACT:** Speech Imagery (SI) is considered an intuitive paradigm for Brain-Computer Interface designs in particular for communication applications. In this work, we use Electroencephalography (EEG) for offline SI decoding. We recorded covert speech from 17 participants. We tested two types of wavelet decomposition techniques. Specifically, we considered coefficients from 6 decomposition levels with Discrete Wavelet Transform (DWT) and multiple 2 Hz spaced packets with Wavelet Packet Decomposition (WPD), we computed different statistical features from such coefficients to form vector inputs for our binary-class classification approach. We approached the issue of feature/sample gap by using the Maximum Relevance and Minimum Redundancy (MRMR) feature selector algorithm to select the most informative features. We achieved a mean accuracy of  $76.6\% \pm 16$  and demonstrated the potential of WPD to extract narrow-band features, and how its refined representation outperforms DWT in SI decoding.

## INTRODUCTION

Speech Imagery (SI) has become an attractive paradigm due to its intuitiveness [1, 2]. The Brain-Computer Interface (BCI) user is prompted to covertly say or repeat a speech unit (e.g., a letter, word, or phrase). With accurate classification of such tasks, a user can convey different messages or commands, e.g., to change an application state. One potential application of SI-based BCIs is as an assistive technology to restore communication for people who have lost the ability to speak. Researchers have approached SI-based BCI designs using different speech units as vowels [3, 4], syllables [5, 6] or words [7, 8] and were able to achieve higher than chance decoding accuracies suggesting the potential use of this paradigm.

To classify the speech unit from recorded EEG signal, informative features need to be extracted, EEG dynamics are known for their non-stationarity therefore a need for techniques that capture time and frequency domain information [9].

The widely known Fast Fourier Transform (FFT) has been applied to extract SI frequency information, Bajestani et.al (2022) [10] used FFT coefficients to classify between tasks with higher than-chance accuracy. Modified forms of FFT have also shown promising results when

extracting SI features, the Discrete Gabor transform was applied by Jahangiri et.al (2018) [6] where the coefficients helped identify the relevance of the gamma band ( $> 60$  Hz). Mel Frequency Cepstral Coefficients initially used for audio decomposition were used as EEG features, and showed classifiable properties between SI tasks [11, 12]. These FFT-based methods represent well-frequency information but omit time domain features which may also be important for SI decoding.

Wavelet Decomposition is a method proven useful in extracting both, time and frequency domain features [13, 14], in particular, Discrete Wavelet Transform (DWT) has been used as a feature extraction technique in SI approaches [15, 16]. DWT decomposes the signal with a transformation analogous to high and low-pass filtering. However, it may not be optimal for accessing specific frequency ranges as the obtained decomposition levels are derived from the low-pass filtered version of the scaled signal [17]. Additionally, Wavelet Packet Decomposition (WPD) performs a more detailed representation as the decomposition levels derive from the low and high-pass filtered versions of the signal resulting in a representation with more frequency ranges to access [18].

The issue of participant-dependent frequency variability is known in the area of EEG decoding, as the prominent frequencies elicited from imagery tasks tend to change for individuals, the appropriate selection of frequency information would lead to better classification results [19], thus we investigated the use of WPD to find participant-specific frequency ranges and compare its performance with features from fixed frequency ranges from DWT.

Due to the relatively small number of SI samples in comparison with the large number of features obtained from the wavelet decomposition levels, a dimensionality reduction step is needed to select a reduced number of features for optimal performance of a machine learning classifier. We investigated the capabilities of the Maximum Relevance Minimum Redundancy (MRMR) feature selection algorithm as it has proven useful for selecting informative features from large feature sets [20].

We have chosen the two phonetically distinct monosyllabic words 'left' and 'right' for our SI experiments and emphasized the participants to focus on the inner pronunciation rather than its meaning.

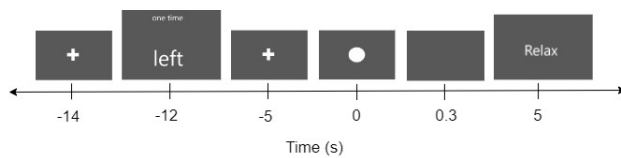


Figure 1: Timeline of the experimental protocol.

## MATERIALS AND METHODS

**Participants:** Seventeen right-handed able-bodied participants (9 female) between the ages of 20 and 35 ( $\mu = 25.65$ ,  $\sigma = 8.3$ ) were recruited from the student population of the University of Essex. Participants received a compensation voucher worth £10 (GBP) for their time. All volunteers read, understood and signed the consent form based on the recommendations of the Ethical Committee of the University of Essex in January 2023 (Reference Number ETH2223-0220).

**EEG Instrumentation:** EEG was recorded using a 64-channel Biosemi Active-Two system. Electrode placement was done via the international 10-20 system, plus one electrode close to the pterion after each eyebrow for electrooculography (EOG) and one electrode behind each ear on the mastoids for electromyography (EMG) recording. Data was recorded at a sampling rate of 2048 Hz unaffected by hardware cut-off.

**Experimental Protocol:** Participants were seated in a comfortable chair facing a 52-inch screen. A graphical user interface developed with PsychToolbox 9.0 [21] in Matlab R2022 was used to display the prompts over a plain grey screen. We used a stimulus masking approach where we first showed the imagery prompt and then had a visual cue presented as a circle in the middle of the screen that remained for 300 ms, having a flash-like effect. See Figure 1 for the timeline of the experiment.

Participants were asked to perform the speech imagery of the words 'left' and 'right' as soon as they saw the cue stimulus.

We first presented a fixation cross for 2 seconds followed by the imagery prompt for 6 seconds and a time-variant (1.5–2 s) fixation cross before the cue. We cued our participants with the described flash stimulus and proceeded to leave a plain screen for 5 seconds until the 'relax' prompt was shown.

**Signal Analysis:** Raw EEG data were first downsampled to 1024 Hz from the original 2048 Hz, we then applied a notch filter (zero-phase, Hamming window FIR) at a cutoff frequency of 50 Hz and its harmonics at 100 Hz and 150 Hz to reduce the power line noise. We divided the data into regularly spaced epochs from  $t_1 = -2$  s to  $t_2 = 8$  s with respect to stimulus onset ( $t = 0$ ), 25 trials per class were initially recorded. Channels were visually inspected and rejected when they looked overly noisy with respect to their neighbours. Epochs were visually inspected to reject those with bad movement artifacts. Between 4 to 7 epochs and 6 to 10 channels were dropped for each participant. Common Average Referencing (CAR) was then applied after to improve the

signal-to-noise ratio.

In order to remove EOG and EMG artifacts, the signal components were estimated using Independent Component Analysis (ICA) with the Picard algorithm [22] to select and discard components encompassing evident eye blinks, lateral ocular movements or muscular artifacts based on their spatial or temporal locations and frequency distributions. Between one and four components were removed for each participant. The remaining components were used to reconstruct the data.

**Feature Extraction:** We used a 1.2-second-long post-stimulus window and tested the signal decomposition algorithms WPD, and DWT. We used Daubechies wavelet (db4) as the mother wavelet as it has been widely used for EEG approaches [8, 23].

DWT is known for its ability to represent time and frequency information [14, 24], it assumes that a signal is a linear combination of a particular set of wavelet functions, and these functions are scaled and shifted versions of a mother wavelet [17] WPD is a more refined version of wavelet decomposition which solves the scaling limitation of DWT as the decomposition happens on both detail and approximation coefficients at each level generating a larger frequency space, thus for the 6th WPD decomposition level, 64 packets of coefficients would be obtained[18].

To decompose the signal we first applied a low-pass filter (zero-phase, Hamming window FIR) at 128 Hz cut-off. For DWT we considered 6 levels of detail coefficients, D1(64–128 Hz), D2(32–64 Hz), D3(16–32 Hz), D4 (8–16 Hz), D5 (4–8 Hz), D6 (2–4 Hz) and one of approximation coefficients A6 (0–2 Hz). For WPD, we considered the 2 Hz step packets at the 6th decomposition level. These packets encompassed frequencies from 4–30 Hz and 70–128 Hz. We selected Alpha and Beta as previous approaches found informative features in such bands [10, 11, 25, 26] and also accounted for frequencies higher than 70 to explore the gamma band, also known to be relevant in SI-related activity [6, 27]. We did not consider frequencies between 30–70 Hz for WPD to narrow the number of options to select from therefore reducing computation cost.

We computed the next statistical and wavelet features from each level/packet: mean value, standard deviation, root mean square, slope, kurtosis, energy, entropy, mean absolute difference, negative turnings, positive turnings and wave centroid.

To find the most informative features we tested the classification performance of each statistical feature from each level/packet on a one-to-one basis, then combined the features with the top 3 classification accuracies to check for performance improvement.

**Feature Selection:** Each feature-level/packet combination formed a feature vector of shape  $\text{channels} \times 1$ , as the average number of epochs per class was  $21 \pm 4$ , we aimed for an ideal features vector shape of  $10 \times 1$ . To reduce the vector dimensionality we used the Minimum Redundancy Maximum Relevance (MRMR) method on every

run of our cross-validation procedure. MRMR aims to maximize the relevance of features to the target variable while minimizing the redundancy among selected features [20], it uses a relevance score based on mutual information and a redundancy score based on Pearson correlation.

**Classification:** For each participant, we had an average of 21 trials  $\pm 4$  per class. We evaluated the 2-class classification performance of our model with the median accuracy from a 6-fold cross-validation. We repeated the cross-validation 15 times, with a different seed at the time, and used the median score of each repetition to better estimate the model's performance.

Linear discriminant analysis has been widely used in BCIs. As large dimensionalities and overfitting are common problems in BCI, regularized LDA has been found to be useful for small training sample settings, we used the shrunken version of LDA [28], which adds a penalty term to the loss function, using the scikit-learn library [29] and the 'auto' shrinkage parameter that finds an optimal value based on the lowest error.

## RESULTS

To get the most informative frequency ranges from the SI-EEG data, we recursively tested 11 statistical features computed from WPD packet and DWT level coefficients, we found the best-performing setting for each decomposition modality and participant based on the classification accuracy. We then reported the obtained accuracies and compared the results as seen in Figure 2. The use of features from multiple narrowed frequency intervals with WPD achieved 13% higher accuracy than the limited levels of decomposition from DWT, with ( $p < 0.01$ ) from a two-sample test. WPD scores are above the 99% confidence interval, computed based on the trial number per class [30], marked for the black horizontal lines, while most of DWT results lay below this interval.

The MRMR algorithm for feature selection was shown to be useful in reducing dimensionality while retaining informative features. We have counted the occurrence of selection of each WPD packet over cross-validation folds and present them as a channel-feature heatmap in Figure 3. We observe that some relevant channels involve locations that may be reflecting speech processing-related areas as left-central channels C1, C5, frontal-temporal channel FT7 or temporal channel T7 one of the most selected by our feature selection process. However, informative features spread across different regions, as with Fp2 chosen along different frequencies or P8 highly relevant on the frequency band (26–28 Hz). Even if neural dynamics are considered to be produced by left hemisphere dominant processes [31–33], SI-relevant features from EEG appeared to spread around different regions depending on individuals.

Similar to frequency domain features, relevant information seems to be spread along all the tested bands above 10 Hz with particular highlights on bands at 26–28 Hz,

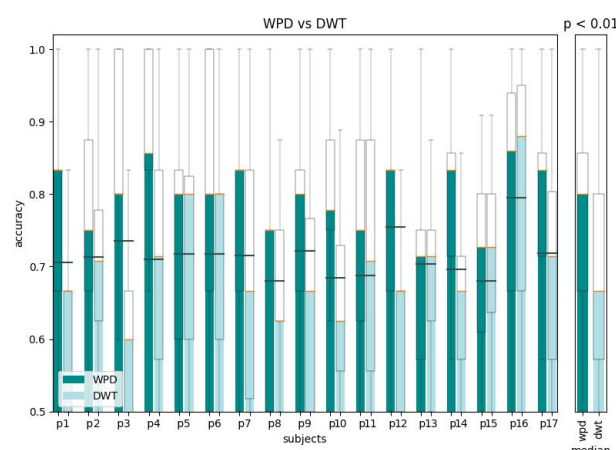


Figure 2: Comparison of obtained accuracies between the DWT and WPD decomposition methods across repeated 6-fold cross-validation, black horizontal lines represent the 99% confidence interval.

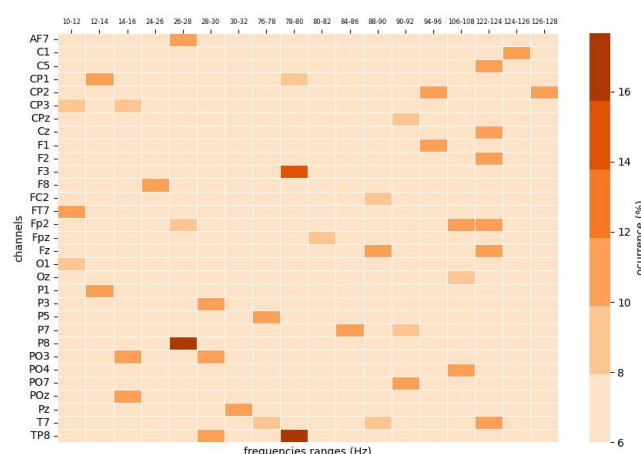


Figure 3: Heatmap of occurrences of channels vs features from the MRMR algorithm on WPD features.

78–80 Hz and 122–124 Hz.

During our analysis, we checked for the statistical features with discriminative properties between the SI classes on DWT and WPD coefficients, we counted the number of times that each feature gave a higher-than-chance result, Figure 4 shows the occurrence of significant results from each feature, where the slope of the coefficients, appeared as the most discriminative property from these wavelet representations.

## DISCUSSION

Research into the Speech Imagery paradigm is gaining traction, different experiments and designs prove that SI can be classified from EEG signals [3, 6, 7].

The Motor Imagery (MI) paradigm, whose event-related desynchronization (synchronization) is well known to have a predominant range of frequencies (Alpha and Beta) and location in the central Motor Cortex [34, 35]. In contrast, the SI-related potentials are not fully under-

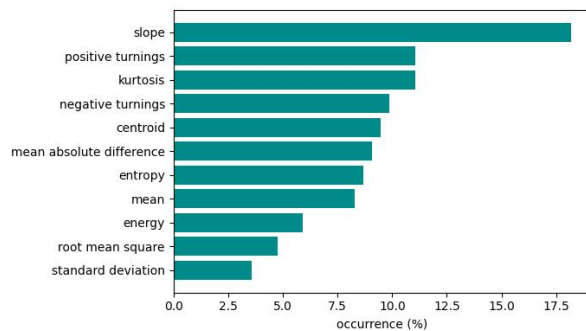


Figure 4: Occurrences of statistical features obtained from both DWT and WPD.

stood [36].

Speech Imagery involves more complex processing for the brain than MI [37], as a Language process, the brain regions known to be active during speech processing may be active during SI, the literature suggests that SI activity has a left hemisphere dominant processing, that involves different brain regions. Some commonly mentioned regions are the temporal-parietal junction that has been related to a memory and semantic decoding step, the frontal-temporal regions possibly handling syllabification and premotor and motor regions for the activity related to the somatosensory SI experiences [33, 38, 39].

Out of the most relevant features by the MRMR selector, we find channels located on regions that may be influenced by areas known to be active during language production, FT7 around Broca's area, C1 in the Motor cortex, T7 and C5 around the superior temporal region, P5 and P3 in the temporal parietal junction [33, 40]. However, the encountered relevant features are not restricted to these areas and are spread around different regions, as features from channels P8, TP8 or Fp2.

Studies of SI with Electroencephalography (EEG) and EEG have found that this imagery paradigm involves broad-frequency dynamics and highlights the important contribution from the gamma band ( $> 60$  Hz) [6, 38, 41]. Our results suggest that many informative features come from the narrow frequency ranges between 26–28 Hz, 78–80 Hz, or 122–124 Hz. It can also be noticed that relevant features appeared to be chosen nearly continuously in the Gamma range between 76–108 Hz but no features were significantly chosen between 96–106 Hz. We have tested WPD frequencies laying on Alpha, Beta and high Gamma bands and found relevant information is spread along different frequencies. Therefore we suggest that future SI analysis should consider a broad spectrum of the frequency domain.

The issue of participant-dependant frequency variability in SI from EEG data was demonstrated in our comparison between the two wavelet decomposition strategies. The general decomposition levels extracted with DWT in most of the cases did not lead to significant classification performance, however selecting participant-specific narrow frequency bands with the use of WPD significantly improved the classification accuracy, as shown in Figure

2. To test a decoding pipeline with multiple WPD configurations can be computationally expensive due to the total amount of available packets. In this work, we have pointed out some frequency ranges which combinations could be the starting testing point for future SI-related work.

## REFERENCES

- [1] Wang L, Liu X, Liang Z, Yang Z, Hu X. Analysis and classification of hybrid bci based on motor imagery and speech imagery. *Measurement*. 2019;147:106842.
- [2] Wang L, Zhang X, Zhang Y. Extending motor imagery by speech imagery for brain-computer interface. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013, 7056–7059.
- [3] DaSalla CS, Kambara H, Sato M, Koike Y. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*. 2009;22(9):1334–1339.
- [4] Chengaiyan S, Anandan K. Effect of functional and effective brain connectivity in identifying vowels from articulation imagery procedures. *Cognitive Processing*. 2022;23(4):593–618.
- [5] D’Zmura M, Deng S, Lappas T, Thorpe S, Srinivasan R. Toward eeg sensing of imagined speech. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2009, 40–48.
- [6] Jahangiri A, Sepulveda F. The relative contribution of high-gamma linguistic processing stages of word production, and motor imagery of articulation in class separability of covert speech tasks in eeg data. *Journal of Medical Systems*. 2018;43(2).
- [7] Kaongoen N, Choi J, Jo S. Speech-imagery-based brain-computer interface system using ear-eeg. *Journal of Neural Engineering*. 2021;18(1):016023.
- [8] Torres-García AA, Reyes-García CA, Villaseñor-Pineda L, García-Aguilar G. Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification. *Expert Systems with Applications*. 2016;59:1–12.
- [9] Lotte F *et al*. A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update. *Journal of Neural Engineering*. 2018;15(3):031005.
- [10] Asghari Bejestani MR, Mohammad Khani GR, Nafisi VR, Darakeh F. Eeg-based multiword imagined speech classification for persian words. *BioMed Research International*. 2022;2022:1–20.
- [11] Riaz A, Akhtar S, Iftikhar S, Khan AA, Salman A. Inter comparison of classification techniques for vowel speech imagery using eeg sensors. In: *The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014)*. 2014, 712–717.
- [12] Cooney C, Folli R, Coyle D. Mel frequency cepstral coefficients enhance imagined speech decoding accuracy from eeg. In: 2018 29th Irish Signals and Systems Conference (ISSC). 2018, 1–7.
- [13] Adeli H, Zhou Z, Dadmehr N. Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods*. 2003;123(1):69–87.
- [14] SUBASI A. Eeg signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*. 2007;32(4):1084–1093.
- [15] Mahapatra NC, Bhuyan P. Multiclass classification of imagined speech vowels and words of electroencephalography signals using deep learning. *Advances in Human-Computer Interaction*. 2022;2022:1–10.
- [16] Sree RA, Kavitha A. Vowel classification from imagined speech using sub-band eeg frequencies and deep belief networks. In: 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN). 2017, 1–4.
- [17] Mallat S. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1989;11(7):674–693.
- [18] Xue JZ, Zhang H, Zheng CX, Yan XG. Wavelet packet transform for feature extraction of eeg during mental tasks. In: *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*. 2003, 360–363 Vol.1.
- [19] Ang KK, Chin ZY, Zhang H, Guan C. Filter bank common spatial pattern (fbcsp) in brain-computer interface. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). 2008, 2390–2397.
- [20] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005;27(8):1226–1238.
- [21] DH B. The psychophysics toolbox. *Spatial vision*. 1997;10(4).
- [22] Ablin P, Cardoso JF, Gramfort A. Faster independent component analysis by preconditioning with hessian approximations. *IEEE Transactions on Signal Processing*. 2018;66(15):4040–4049.
- [23] Agarwal P, Kumar S. Electroencephalography based imagined alphabets classification using spatial and time-domain features. *International Journal of Imaging Systems and Technology*. 2021;32(1):111–122.
- [24] Gramfort A *et al*. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*. 2013;7(267):1–13.
- [25] Idrees BM, Farooq O. Vowel classification using wavelet decomposition during speech imagery. In: 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN). 2016, 636–640.
- [26] Biswas S, Sinha R. Wavelet filterbank-based EEG rhythm-specific spatial features for covert speech classification. *IET Signal Processing*. 2021;16(1):92–105.
- [27] Nitta T *et al*. Linguistic representation of vowels in speech imagery eeg. *Frontiers in Human Neuroscience*. 2023;17.
- [28] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of erp components — a tutorial. *NeuroImage*. 2011;56(2):814–825.
- [29] Pedregosa F *et al*. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825–2830.

- [30] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random? a closer look on bci results. *International Journal of Bioelectromagnetism*. 2008;10(1):52–55.
- [31] Si X, Li S, Xiang S, Yu J, Ming D. Imagined speech increases the hemodynamic response and functional connectivity of the dorsal motor cortex. *Journal of Neural Engineering*. 2021;18(5):056048.
- [32] Stephan F, Saalbach H, Rossi S. The brain differentially prepares inner and overt speech production: Electrophysiological and vascular evidence. *Brain Sciences*. 2020;10(3):148.
- [33] Tian X, Zarate JM, Poeppel D. Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*. 2016;77:1–12.
- [34] Korostenskaja M *et al.* Characterization of cortical motor function and imagery-related cortical activity: Potential application for prehabilitation. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2017, 3014–3019.
- [35] Jahangiri A, Chau JM, Achancaray DR, Sepulveda F. Covert speech vs. motor imagery: A comparative study of class separability in identical environments. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018, 2020–2023.
- [36] Kim HJ, Lee MH, Lee M. A bci based smart home system combined with event-related potentials and speech imagery task. In: 2020 8th International Winter Conference on Brain-Computer Interface (BCI). 2020, 1–6.
- [37] Kraft E, Gulyás B, Pöppel E. Neural correlates of thinking. In: *On Thinking*. Springer Berlin Heidelberg, 2009, 3–11.
- [38] Martin S *et al.* Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuroengineering*. 2014;7.
- [39] Hickok G. The dual stream model of speech and language processing. In: *Aphasia*. Elsevier, 2022, 57–69.
- [40] Broca's region. Oxford University Press New York (2006).
- [41] Martin S *et al.* Word pair classification during imagined speech using direct brain recordings. *Scientific Reports*. 2016;6(1).

# EEG CORRELATES OF ERROR-RELATED ACTIVITY DURING BALLISTIC COMPUTER MOUSE MOVEMENTS

Idorenyin Amaunam<sup>1</sup>, Mushfika Sultana<sup>1</sup>, Borja Rodriguez-Herreros<sup>2</sup>, Tej Tadi<sup>3</sup>, Robert Leeb<sup>3</sup>,  
Serafeim Perdakis<sup>1</sup>

<sup>1</sup>Brain-Computer Interface and Neural Engineering Laboratory, School of Computer Science and  
Electronic Engineering, University of Essex, United Kingdom

<sup>2</sup>Service des Troubles du Spectre de l'Autisme et apparentés, Département de psychiatrie, Lausanne  
University Hospital (CHUV), Lausanne, Switzerland

<sup>3</sup>Mindmaze SA, Lausanne, Switzerland  
E-mail: ia21792@essex.ac.uk

**ABSTRACT:** It has been repeatedly shown that processing of perceived errors in the human brain may elicit some type of evoked response in electroencephalography (EEG) collectively termed as Error-Related Potential (ErrP). The study of ErrP signatures offers a potential back door to better understanding how the brain encodes and reacts to errors and a useful tool for poking adaptation and learning, but also has several practical applications in Brain-Computer Interface (BCI) and general Human-Computer Interaction (HCI). The bulk of this literature has focused on so-called “interaction” ErrP, reflecting the response to discrete events occurring during self-paced, casual interaction of a subject with their environment. Here we present a two-case study investigating the existence and characteristics of ErrP EEG correlates in an eye-hand coordination task consisting in “ballistic” computer mouse movements, where the action and reaction time constraints imposed on the subject are extremely tight. We show that clear EEG substrates of error processing can be retrieved for both subjects and bare strong similarities with the interaction ErrP waveforms. The findings of this work suggest the possibility of detecting, in real-time, errors committed during fast-paced interaction, thus potentially enabling automatic ErrP-based error correction in real-world BCI and HCI scenarios.

## INTRODUCTION

Error-Related Potentials (ErrPs) are Event-Related Potential (ERP) waveforms in EEG time-locked to the realization of committed errors [1]. Mainly owing to its utility as an automatic, subconscious means of error correction during BCI [2] or general HCI, the topic of ErrP signatures in various contexts has been extensively studied in the last 20 years [3]. ErrP correlates are relatively slow signals characterized by a negative fronto-central peak around 100 ms after the error onset, and followed by a larger, positive, centro-parietal peak with latency about 500 ms, which has been associated with the subject's awareness of the error [1, 4]. Based on source imaging and localization studies, ErrP correlates are believed

to originate in the anterior cingulate cortex [1, 4].

Besides the inherent interest of cognitive neuroscience in ErrPs as a means to elucidate the brain's error processing mechanisms [1, 5], ErrP detection has attracted a lot of attention due to a wide spectrum of promising applications in BCI and HCI. First and foremost, ErrP recognition offers a seamless avenue for automatic error correction in human-machine interfaces [6, 7] that requires no direct manual intervention by the user [3]. In particular, this is convenient for BCI applications where EEG or other brain signal monitoring is already available to support the main interface control modality [8–14], so that ErrP detection yields no additional burden for the system's apparatus. As ErrPs are a natural physiological “reward”/punishment signal, they have also been used in the context of BCI human-machine co-adaptation [15], especially with respect to reinforcement learning approaches [16–18]. Interestingly, under the same framework ErrPs may also play the role of the main BCI modality [19, 20]. In order to optimize such applications, a lot of studies have been dedicated to the design of machine learning and other techniques for enabling high-accuracy single-trial classification of ErrP correlates, overcoming various engineering challenges [4, 12, 21–23].

The aforementioned prototypical error-related pattern is commonly observed in experimental protocols involving “discrete” errors of the interface that clearly constitute single events: there is an abrupt, profound error onset to which the ErrP signal is time-locked, and the duration of the erroneous action or feedback are relatively short [4, 11, 23, 24], even when such errors are embedded in continuous interaction and unpredictable feedback tasks [6, 12, 14, 25, 26]. Very few investigations have been carried out with regard to gradually unfolding errors [27]. However, despite the similarities of extracted ErrP waveforms in this regular discrete ErrP category, important differences or even complete absence of ErrPs have been denoted depending on the particular task (e.g., interaction, response, or observation [3]) and contextual circumstances [24] of the experimental design.

As previously argued [1], a critical factor in the elicita-

tion and shape of ErrPs may be the time constraints imposed on a human subject when requested to detect errors. Notwithstanding the fact that many of the studies in this literature yielded high mental workload for the user during the error recognition task induced by concurrent observation or interaction tasks, typically, the amount of time available to consolidate the occurrence of an error is ample. Here, we aimed to investigate whether ErrP signatures can be elicited and captured in EEG activity when errors happen while human individuals are engaged in an extremely fast-paced eye-hand coordination activity.

The elegance and efficiency of human movements owes much to our ability to compensate for inaccuracies and environmental perturbations. Performing online corrections relies on real-time monitoring of the hand/body trajectory from the onset of the movement [28]. We devised an experimental protocol with two related but distinct goals: (i) to assess the facilitatory effects of excitatory neuromodulation, specifically, anodal high-density transcranial Direct Current Stimulation (tDCS) [29] of the intraparietal sulcus on visuo-motor coordination, and (ii) to determine the existence and features of ErrP with simultaneous EEG monitoring, while subjects are asked to perform a “double-step” reaching task. Here, we provide preliminary analysis on the second goal of describing potential EEG correlates of error perception in this protocol.

Online eye-hand coordination has often been investigated via double-step reaching tasks, in which the target of a reaching movement suddenly moves to a new location after the movement onset and subjects are asked to adjust the movement trajectory to compensate for the perturbation. In the present study, subjects were required to put the computer mouse pointer into a narrow circular target at the top of the screen starting from a base location at the bottom with “ballistic” movements, as accurately and, very importantly, as fast as possible, with maximum possible velocity and acceleration. We hypothesized that target displacement during ballistic movements may evoke ErrP waveforms and seek to describe them. Our preliminary results with two subjects confirm this hypothesis and suggest that, despite the task’s extreme timing demands, the elicited error-related responses resemble the interaction ErrPs identified in the literature. We further show that trials with target displacement can be identified with a shallow classifier, opening the road for enriching naturalistic interaction with elaborate error detection and correction capabilities even for hurried tasks.

## MATERIALS AND METHODS

### *Participants:*

We report on two subjects, S1 and S2, randomly picked for initial analysis from a larger dataset of 28 healthy right-handed volunteers (11 female; mean age  $24.9 \pm 5.8$  years). All participants were naïve with respect to the experimental procedures and the hypothesis of the study. Participants had normal or corrected-to-normal visual acuity and reported no history of neuropsychiatric

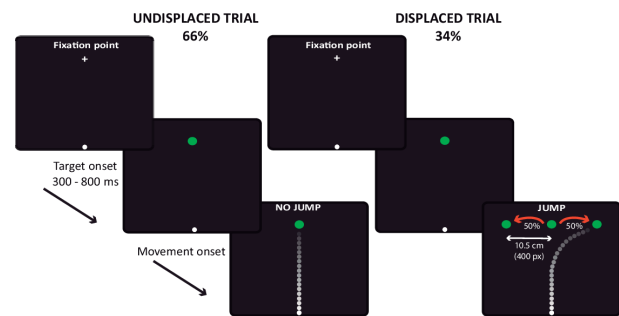


Figure 1: Experimental protocol. Trial timeline and illustration of the protocol’s events and visual elements.

disorders. Prior to their inclusion in the study, participants signed written informed consent. The study was performed according to the declaration of Helsinki and was approved by the local Ethics Committee. All participants were screened for HD-tDCS tolerance [29]. The Edinburgh handedness inventory was administered to assess handedness.

### *Experimental setup and protocol:*

Participants were seated in front of a table positioned 45-50 cm below their eyes. Visual stimuli (see below) were generated using Python v3.6 running on Windows 10 in an Intel Core i7-7700 3.6 GHz computer and displayed on a 24” HP E232 monitor with a refresh rate of 60 Hz and a resolution of 1920x1080 pixels. A high-speed Logitech G G203 prodigy cable mouse, sampling at 250 Hz, was used to track the hand’s 2D spatial position during reaching movements.

EEG data were recorded at 500 Hz sampling rate with a Starstim 8 hybrid EEG/tDCS 8-channel, active, wet, Ag/AgCl electrode system (Neuroelectronics, Barcelona, Spain). We measured positions FCz, FC1, Cz, CP1, CP5, P3, PO3 and PO7 of the 10-20 EEG localization system. This channel montage was designed to satisfy and compromise the needs of both the goals set during the study’s design, including the possibility to focally excite the intraparietal sulcus through tDCS. Enough, strategically placed electrodes covering fronto-central and parietal cortical areas were included to capture ErrP activity. Each subject executed 3 experimental sessions, which consisted of 5 runs each. Each run lasted approximately 6 minutes, and included 100 trials performed with the subject’s dominant hand, for a total of 1500 trials per subject. Subject S1 of this study executed an additional 3 sessions of 3 runs each, for a total of 2400 trials. The experimental task consisted in performing a fast goal-directed reaching movement towards a visual target located in the center of the screen (Fig. 1). Prior to the initiation of the trial, participants were required to move the mouse to the starting position in the bottom center of the screen. After 1000ms with the mouse placed at this starting point, a small white fixation cross was automatically shown as a warning signal in the centre of the screen. The fixation cross preceded the target onset by a variable period (300 or 800ms) to avoid participants from predicting the timing of that onset. Trials with (34%) and without (66%) target dis-

placement were presented in pseudo-random order. In non-displaced trials, the target remained static in the center of the screen. In displaced trials, however, the target showed an unexpected 10.5cm (400 pixels) lateral displacement—disappeared and instantly re-appeared to the left or to the right with a 50-50% chance—, after the initiation of the reaching movement. The movement onset was detected by a 50 mm/s velocity threshold of the mouse movement. The target displacement was timed exactly at the movement onset (maximum 6 ms delay due to technical limitations of the experimental protocol) to assure that participants did not have relevant visual cues about the final position of the target during the initial movement planning.

The total longitudinal distance between the starting point and the target was 30 cm (1200 pixels) in the screen. The ratio between the distance travelled by the mouse in the real world and the cursor in the screen was 2:3, that is, when the mouse moved for example by 20 cm, the cursor advanced by 30 cm in the screen. To discard trials with long reaction times, a warning sound was provided when the velocity threshold required to start the movement was not reached during the 500 ms that followed the target onset. Participants were instructed to hit the target as fast and as accurate as possible performing a ballistic movement, exhibiting maximum velocities and accelerations over a very short period of time. When the target was displaced, participants had to adjust their hand trajectory to succeed in hitting the target in its final location. The target was presented for 1000 ms. At the end of the reaching movement, participants brought their hand back to the starting point, and prepared to start the next trial. Participants were instructed not to move their trunk with respect to the chair and avoid head movements.

#### *Data Analysis and Evaluation:*

EEG data were band-passed with a 3<sup>rd</sup>-order Butterworth filter within [1-20] Hz to remove signal drifts and isolate the spectral range within which ErrP components are known to be found [7]. The final epochs considered for analysis corresponded to the trial segment [-1, 1] s where  $t = 0$  s the movement onset. Subsequently, automatic artifact removal with FORCE [30] and DC removal baselining were applied to each epoch. Finally, we removed all epochs whose maximum filtered amplitude exceeded 100  $\mu$ V. We assessed the statistical significance of the difference in amplitude among trials of a subject with (Error) and without (Correct) displacement through two-sided, unpaired t-tests with  $\alpha = 0.05$  and Bonferroni correction for multiple ( $N = 8000$ , 8 channels  $\times$  1000 time points) comparisons. Two-class (Correct vs Error) classification accuracy is derived with 10-fold cross-validation employing a binary Decision Tree (DT) classifier. DT is selected for a first attempt to classify these novel ErrPs as less vulnerable to overfitting and class-bias than other shallow models (including linear or quadratic discriminant analysis with regularization/shrinkage), taking into account the fact that there are double Correct than Error trials in the dataset. The 100-best, in terms of  $r^2$  fea-

ture fitness, spatio-temporal (i.e., channel/time-point combinations) amplitude features are selected using the fold's training data. Average and standard deviation (across folds) of the total and class-wise classification accuracy are reported per subject.

## RESULTS

We present results on 2306 trials of S1 and 1494 trials of S2 that survived the trial rejection. Fig. 2 and 3 establish beyond doubt both the elicitation of evoked potentials during the performance of this protocol's task, and the fact that pronounced and statistically significant differences exist between Correct (with no target displacement) and Error (with target displacement) trials. The grand average waveforms are consistent across subjects. Correct trials (blue) exhibit a large positive peak around  $t = 500$  ms from movement/displacement onset ( $t = 0$  ms), followed by a very small "refractory" negative peak at  $t = 680$  ms for S1; this is completely absent for S2. Early negative peaks can be also identified around  $t = 260$  ms, especially for S2 and for Error trials. On the other hand, Error trials (red) show a similar shape, which is however delayed: the large positive peak is located around  $t = 600$  ms and the refractory negative peak around  $t = 750$  ms for S1 and slightly before  $t = 1.0$  s for S2. As a result, the average difference Error-Correct (black) demonstrates a first negative peak at  $t = 500$  ms (very consistently for both subjects) ahead of a larger positive peak in the interval  $t = [650, 720]$  ms. The differences Correct versus Error are statistically significant in the period [400 – 800] ms that includes both the negative and the positive peak, for both subjects; there are no significant differences outside this interval (apart from few very short, spurious ones), which further points to the observed effects corresponding to ErrPs.

The grand average difference between Correct and Error trials is very similar to that reported in the literature for interaction ErrP in different contexts [1, 4, 7, 11, 12]. Along the same lines, Fig. 4 shows that the derived waveform, despite being fairly spread for both subjects, is stronger in fronto-central (channels FCz, Cz) and to a lesser extent parietal (PO3) regions, and is not particularly lateralized (i.e., the phenomena fade out for peripheral channels), which also aligns well with the findings of previous studies on ErrPs.

Fig. 5 shows that above-chance classification accuracy can be obtained for S1 and S2 with a DT classifier discriminating Error from Correct trials. However, the classification is biased towards the dominant Correct class, and only half of the ErrPs can be identified correctly.

## DISCUSSION

The results confirm both the generation of evident ErrP EEG correlates in the framework of this protocol, and that these are similar to interaction ErrPs. However, significant differences are also noted. In particular, while

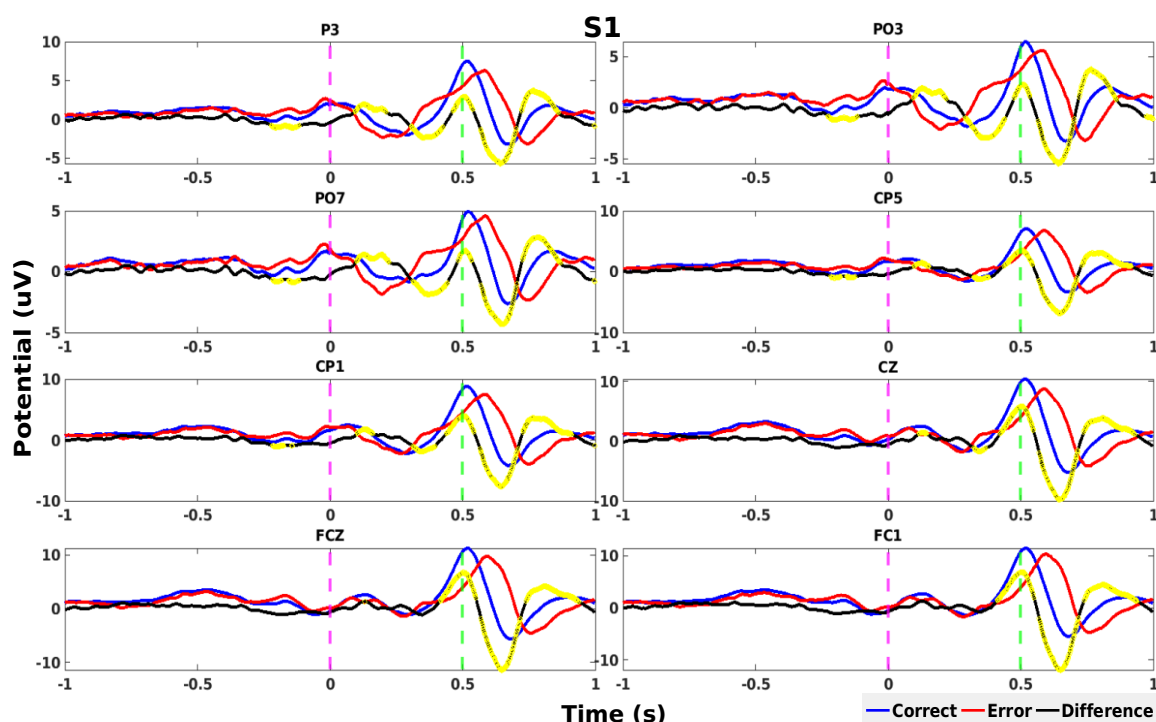


Figure 2: Grand averages of S1 trials. Correct trial average in blue, Error in red. The black line illustrates the difference Error-Correct. Vertical, dashed lines indicate salient time points: magenta for  $t = 0$  ms (basalistic movement onset) and green for  $t = 500$  ms. Yellow circles illustrate statistically significant difference between Error and Correct amplitudes for this time point.

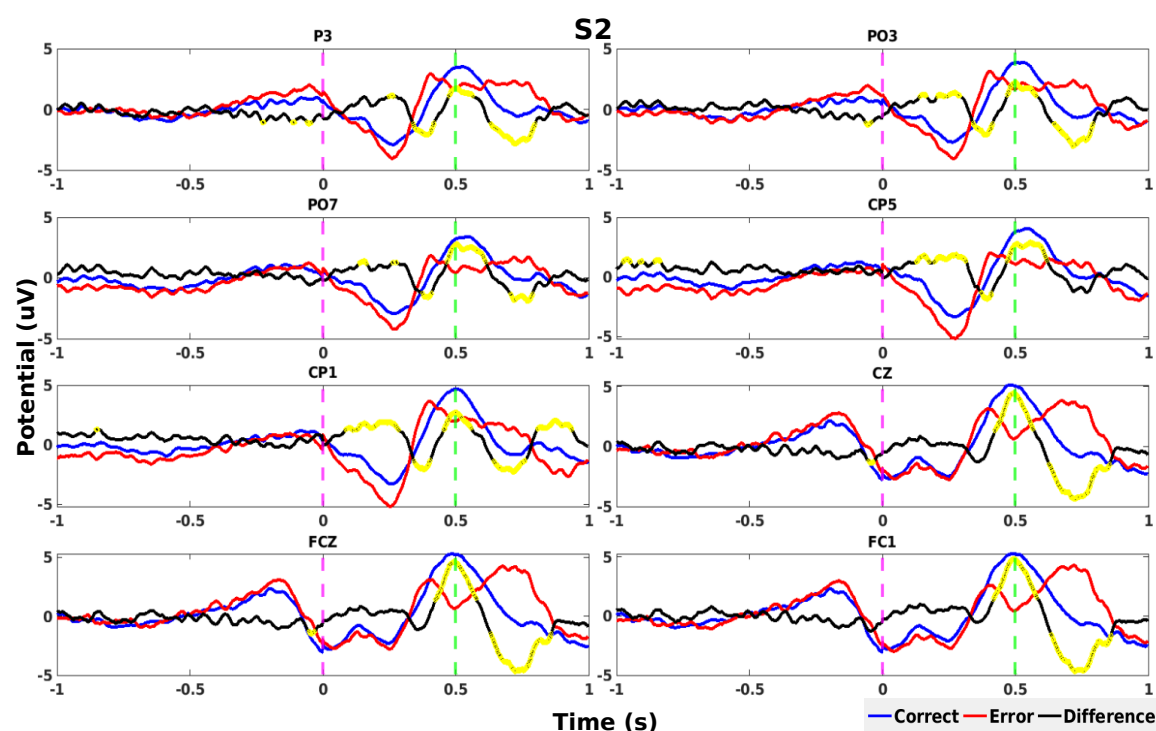


Figure 3: Grand averages of S2 trials. See Fig. 2 caption for details.

the difference Error-Correct (black curve in Fig. 2 and 3) exhibits an anticipated “large positive, then small negative peak” pattern on grand averages [1, 4, 7, 11, 12], compared to the literature, both peaks seem to be delayed by approximately 100-200 ms (Fig. 4). Another,

maybe more peculiar, discrepancy concerns the manifestation of profound, strong peaks also in Correct trials. In relevant works, the Error-Correct curve’s average waveform seems to result from the corresponding modulation in Error trials alone, with Correct trials remaining flat [1,

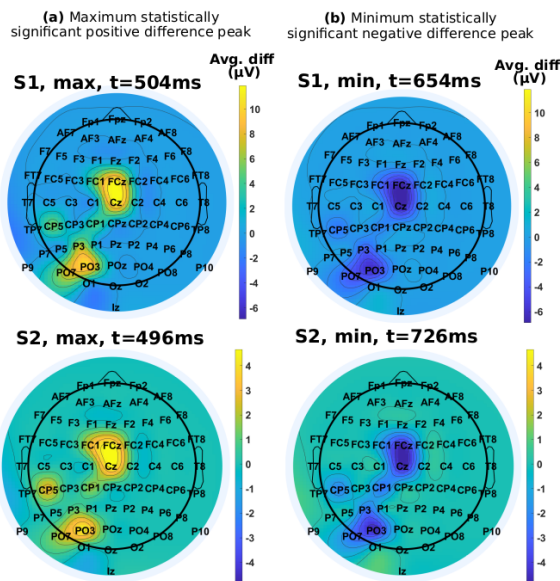


Figure 4: Topographic distribution of the average EEG amplitude difference Error-Correct within the narrow 8-channel layout used, for the maximum statistically significant positive difference peak (left) and the minimum, statistically significant negative difference peak (right) for subjects S1 (top) and S2 (bottom). Text on top of the plots specifies the exact time point where the corresponding peaks shown are located. Bright yellow indicates large positive average difference, deep blue large negative average difference, and green no average difference.

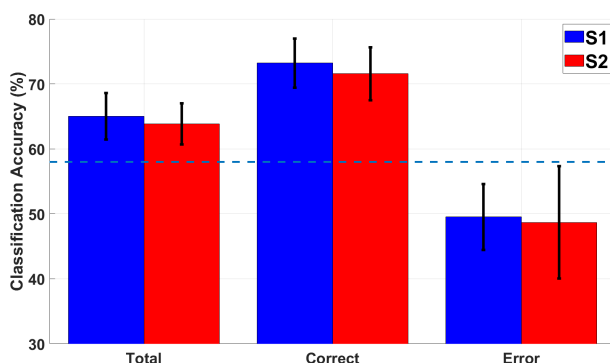


Figure 5: Average and standard deviation (across cross-validation folds) of the Total and class-wise (Correct, Error) classification accuracy for subjects S1 (blue) and S2 (red). The horizontal, dashed, light blue line marks the 58% random classification threshold [31].

12], or only reaching modest peak amplitudes [11], normally distinctly smaller than that of Error trials, an effect referred to as Correct-Response Negativity [32]. On the contrary, here we denote for both subjects a similar signature for both displaced and non-displaced target trials, whose peaks are of comparable amplitude; consequently, the Error-Correct grand average resembles that of regular interaction ErrPs solely due to the fact that ErrPs are delayed with respect to Correct potentials.

One possible explanation is that both Correct and Error patterns may receive contributions by Visually Evoked Potentials generated by the target stimulus appearances, resulting (through some complex and currently unclear

process) in the observed waveform shapes. The delayed Error signature seems to somewhat align with this theory. Of note, the original target stimulus happens at a random, subject-dependent time in the interval  $[-0.5, 0]$  ms before the movement onset ( $t = 0$ ) (the trial was discarded if no movement occurred within 0.5 s after the target projection). The displaced target appearance (only for Error trials) happens at most 6 ms after  $t = 0$ . Hence, the derived signatures cannot be explained on the basis of Visually Evoked Potential (VEP) contributions alone. The high mental workload exerted here could also be implicated with the delayed responses.

Whatever the rationale behind the features of the ErrPs emerging in this protocol, strong peaks in Correct trials render their single-trial detection a particularly difficult task, as shown by the compromised classification accuracy obtained (yet, significantly above the random classification level overall [31]).

## CONCLUSION

Our two-case study corroborates the elicitation of ErrP EEG signals in an extremely hurried eye-hand coordination task comprising “ballistic” computer mouse movements, and substantiates that the corresponding signatures are similar to the interaction ErrPs described in the relevant literature for various protocols involving EEG-monitored error processing. We show that these ErrPs can be recognized in single-trial by means of a shallow classifier, albeit accurate classification seems to be challenging. These preliminary results suggest that it may be possible to detect errors from EEG even under extremely tight time constraints of the underlying task, thus potentially enabling automatic ErrP-based error correction in rushed, real-world BCI and HCI scenarios. Future work entails confirming these results using data from all participants, delving into the mechanisms leading to this kind of novel ErrPs and improving the classification outcome leveraging the large number of subjects and trials available to apply deep and transfer learning techniques.

## REFERENCES

- [1] Chavarriaga R, Sobolewski A, Millán JdR. Errare machinale est: The use of error-related potentials in brain-machine interfaces. *Frontiers in Neuroscience*. 2014;8.
- [2] Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain-computer interfaces for communication and rehabilitation. *Nature Reviews Neurology*. 2016;12(9):513–525.
- [3] Pires G, Castelo-Branco M, Guger C, Cissotto G. Editorial: Error-related potentials: Challenges and applications. *Frontiers in Human Neuroscience*. 2022;16.
- [4] Ferrez PW, Millán JdR. Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE Transactions on Biomedical Engineering*. 2008;55(3):923–929.

- [5] Hoffmann S, Falkenstein M. Predictive information processing in the brain: Errors and response monitoring. *International Journal of Psychophysiology*. 2012;83(2):208–212.
- [6] Milekovic T, Ball T, Schulze-Bonhage A, Aertsen A, Mehring C. Error-related electrocorticographic activity in humans during continuous movements. *Journal of Neural Engineering*. 2012;9(2):026007.
- [7] Spüler M, Niethammer C. Error-related potentials during continuous feedback: using EEG to detect errors of different type and severity. *Frontiers in Human Neuroscience*. 2015;9.
- [8] Spüler M, Bensch M, Kleih S, Rosenstiel W, Bogdan M, Kübler A. Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. *Clinical Neurophysiology*. 2012;123(7):1328–1337.
- [9] Schmidt NM, Blankertz B, Treder MS. Online detection of error-related potentials boosts the performance of mental typewriters. *BMC Neuroscience*. 2012;13(1).
- [10] Cruz A, Pires G, Nunes UJ. Double ErrP Detection for Automatic Error Correction in an ERP-Based BCI Speller. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2018;26(1):26–36.
- [11] Bevilacqua M, Perdakis S, Millán JdR. On error-related potentials during sensorimotor-based brain-computer interface: Explorations with a pseudo-online brain-controlled speller. *IEEE Open Journal of Engineering in Medicine and Biology*. 2019;1:17–22.
- [12] Lopes-Dias C, Sburlea AI, Müller-Putz GR. Online asynchronous decoding of error-related potentials during the continuous control of a robot. *Scientific Reports*. 2019;9(1).
- [13] Iwane F, Iturrate I, Chavarriaga R, Millán JdR. Invariability of EEG error-related potentials during continuous feedback protocols elicited by erroneous actions at predicted or unpredicted states. *Journal of Neural Engineering*. 2021;18(4):046044.
- [14] Iwane F, Sobolewski A, Chavarriaga R, Millán JdR. EEG error-related potentials encode magnitude of errors and individual perceptual thresholds. *iScience*. 2023;26(9):107524.
- [15] Perdakis S, Millán JdR. Brain-machine interfaces: A tale of two learners. *IEEE Systems, Man, and Cybernetics Magazine*. 2020;6(3):12–19.
- [16] Llera A, Gerven M van, Gómez V, Jensen O, Kappen H. On the use of interaction error potentials for adaptive brain computer interfaces. *Neural Networks*. 2011;24(10):1120–1127.
- [17] Kim SK, Kirchner EA, Stefes A, Kirchner F. Intrinsic interactive reinforcement learning – using error-related potentials for real world human-robot interaction. *Scientific Reports*. 2017;7(1).
- [18] Ehrlich SK, Cheng G. Human-agent co-adaptation using error-related potentials. *Journal of Neural Engineering*. 2018;15(6):066014.
- [19] Chavarriaga R, Millán JdR. Learning from EEG error-related potentials in noninvasive brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2010;18(4):381–388.
- [20] Iturrate I, Chavarriaga R, Montesano L, Minguez J, Millán JdR. Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control. *Scientific Reports*. 2015;5(1).
- [21] Millán J, Mouriño J. Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2003;11(2):159–161.
- [22] Iturrate I, Montesano L, Minguez J. Task-dependent signal variations in EEG error-related potentials for brain-computer interfaces. *Journal of Neural Engineering*. 2013;10(2):026024.
- [23] Usama N, Niazi IK, Dremstrup K, Jochumsen M. Detection of error-related potentials in stroke patients from EEG using an artificial neural network. *Sensors (Basel)*. 2021;21(18):6274.
- [24] Si-Mohammed H *et al.* Detecting system errors in virtual reality using EEG through error-related potentials. In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2020*. United States, Mar. 2020, 653–661.
- [25] Kreiling A, Neuper C, Müller-Putz GR. Error potential detection during continuous movement of an artificial arm controlled by brain-computer interface. *Medical and Biological Engineering and Computing*. 2012;50(3):223–230.
- [26] Lopes Dias C, Sburlea AI, Müller-Putz GR. Masked and unmasked error-related potentials during continuous control and feedback. *Journal of Neural Engineering*. 2018;15(3):036031.
- [27] Omedes J, Iturrate I, Minguez J, Montesano L. Analysis and asynchronous detection of gradually unfolding errors during monitoring tasks. *Journal of Neural Engineering*. 2015;12(5):056001.
- [28] Bard C, Turrell Y, Fleury M, Teasdale N, Lamarre Y, Martin O. Deafferentation and pointing with visual double-step perturbations. *Experimental Brain Research*. 1999;125(4):410–416.
- [29] Nitsche MA, Paulus W. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology*. 2000;527(Pt 3):633.
- [30] Daly I, Scherer R, Billinger M, Müller-Putz G. FORCe: Fully online and automated artifact removal for brain-computer interfacing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2014;23(5):725–736.
- [31] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random? A closer look on BCI results. *International Journal of Bioelectromagnetism*. 2008;10(1):52–55.
- [32] Gehring WJ, Liu Y, Orr JM, Carp J. The Error-Related Negativity (ERN/Ne). In: *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press, Dec. 2011.

# FROM CUE-BASED TO SELF-PACED MOVEMENT DETECTION: INFLUENCE OF THE CUE ON TRAINING DATA

Patrick Suwandjjeff<sup>1</sup>, Gernot R. Müller-Putz<sup>1,2</sup>

<sup>1</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria

<sup>2</sup>BioTechMed Graz, Austria

E-mail: gernot.mueller@tugraz.at

**ABSTRACT:** The utilization of a visual cue plays a significant role in enhancing the operational efficiency of brain-computer interface (BCI) systems for individuals with Locked-In Syndrome (LIS). This significance arises from the absence of a reliable method to discern the actual initiation of attempted movements in these patients. First, the decoders for identifying or classifying self-initiated movements need to be trained on cue-based paradigms. However, these cues can elicit neural activity (e.g., visual/auditory evoked potentials, cognitive processing, etc.) that obscures the neural dynamics of movement, thus negatively influencing the performance of the decoder. Therefore, we implemented four novel visual cues with the intention to reduce these effects to a minimum. Our research findings indicate that the effectiveness of classification performance in self-paced EEG recordings when the decoder is trained on cue-based data for movement tasks, is significantly impacted by the design of the cue.

## INTRODUCTION

A brain-computer interface (BCI) is a sophisticated system designed to facilitate communication between the human brain and external devices. This is achieved by capturing and interpreting bioelectrical signals, which are indicative of the user's intentions. These signals can be obtained through non-invasive methods such as electroencephalography (EEG) or invasive techniques like electrocorticography (ECoG). The BCI serves as a bridge, translating the user's cognitive intentions into actionable commands for seamless interaction with external technologies [1],[2]. We aim to make use of four different gestures and movement-related cortical potentials (MRCPs) elicited by these gestures to decode the user's intention. The rationale for incorporating different gestures in this study was to anticipate their potential use in the INTRECOM Project (<https://intrecom.eu/>) where we aim to use them for a four directional control of a speller [3]. The MRCP, a crucial EEG signal tied to voluntary movement preparation and execution, has been extensively studied for its insights into neural processes governing motor planning and control [4], [5]. Comprising components such as the readiness potential (RP) and the movement-related potential, MRCPs offer

a window into the mechanisms of movement-related neural events. In case of paralyzed participants, the movement itself cannot be measured to retrieve the time point of movement onset. In such cases cues seem unavoidable. The challenge arises when the cues essential for the precise timing of a paradigm elicit visual or auditory evoked potentials (V/AEPs) after their stimulus. Such V/AEPs can inadvertently interfere with the analysis of MRCPs, especially in experimental paradigms involving self-initiated neural patterns such as voluntary movements or attempted movements [6], [7], [8] in an asynchronous BCI application in a later stage. Meaning, in the realm of BCIs, asynchronous configurations pose challenges to decoding performance because the classifier is trained on MRCPs influenced by cue-related potentials, which are absent in asynchronous usage. This underscores the need to refine and optimize the visual cue to produce MRCPs minimally influenced for training a classifier that can detect self-paced movement.

Therefore, the primary motivation for this study is to address a fundamental question: can visual cues be designed to exert minimal impact on MRCPs during movement attempts? While past research has compared cue-based and self-paced MRCPs and explored the influence of visual, auditory and vibrotactile cues [9], [10], [11], [12], [13] limited attention has been given to strategies mitigating the impact of visual cues on MRCPs. To fill this gap, we draw inspiration from previous work by Ofner et al. [7], who introduced a gradually appearing visual cue gradually appearing to minimize abrupt changes. Consequently, this gradual adaptation aims to mitigate the interference of cues on EEG signals. Building on this foundation, we propose three new cues designed with the principle of exerting minimal influence on signals associated with movement execution. By rejecting or minimizing cue-induced effects, our goal is to ensure that signals recorded during cue-based activities closely mirror those generated during self-initiated movement attempts.

## MATERIALS AND METHODS

### A. EEG recordings: cue-based and self-paced data

In this research, 22 individuals in good health,

averaging  $26.2 \pm 4.2$  years of age, participated in EEG activity recording. Each participant willingly gave written consent after receiving detailed information from the researchers regarding the study's goals, content, and procedures. Participants were assured the autonomy to cease their involvement at any point without obligation to provide a reason. The experimental protocol obtained approval from the ethics committee at TU Graz before initiation. Additionally, the recorded data for each participant underwent anonymization. Participants engaged in the recording of EEG activity as they executed four distinct hand movements (gestures) using their right hand while seated in front of a computer screen. The EEG signals were recorded using a 64-channel actiCAP system (Brain Products GmbH, Gilching, Germany) at a sampling rate of 500 Hz. EEG signals were collected from all cortical areas through 60 electrodes positioned based on the 10-10 electrode system, while the remaining four electrodes served as electrooculography (EOG) electrodes. EOG electrodes were strategically placed at the outer canthi of both eyes and above and below the left eye to monitor saccades and blinks. The ground electrode was situated at the right mastoid, and the reference electrode was positioned at FCz. To identify actual movement onsets in both cue-based and self-paced sessions, we utilized a motion capture system developed in the institute with a sampling frequency of 30 Hz. A marker was positioned at the nail of the participant's pointer finger. The data output provided spatial information along the x, y, and z axes.

The gestures (Fist, Pincer, Y, Pistol) were chosen based on classification results of previously performed studies [7], [14], [15] in light of the assumption that gestures can differ significantly due to variations in involved joints and rotation thus producing variations in the EEG movement-related dynamics. The experiment consisted of two parts: (i) cue-based data collection applying different cues and gestures. (ii) a self-paced phase where participants were instructed to freely execute corresponding gestures at any time. Data recorded during this phase were used to evaluate the performance of an offline asynchronous decoder, trained on the data of (i).

For the cue-based part participants were instructed to execute four specific hand movements precisely at predetermined start times, signaled by various visual cues. Six successive movements of all four gestures were performed resulting in 5-minute runs, followed by a 30-second rest period. The cue was constant throughout this period, while the presentation order of the gestures was randomized for each trial. This process was repeated over 32 separate runs, each with a randomly shuffled cue, resulting in a cumulative total of 192 trials for each individual gesture (48 trials per cue and per gesture). The core concept underlying these cues is their gradual appearance; they do not appear abruptly but transition smoothly to their initial positions through methods like shrinking, rotating, or fading. A single trial, exemplified on the reference cue [7], has the

following sequence (Fig. 1): the gesture was displayed for 1s, followed by a fixation cross positioned in front of a filled green circle ('ready cue'). After a variable period (2-3s), the green circle would gradually shrink. This phase was designed to function as a preparatory period for participants, serving as a smooth visual transition between cues and minimizing visual cue effects on EEG. The preparation phase was succeeded by the 'go cue,' signaling the initiation of movement execution (3s of execution and holding the end position of the gesture). After that a rest phase with a blank screen was presented for 1.5s. In this study, alongside the reference cue, three novel visual cues were introduced which are based on the same principle (see Fig. 2) the fading cue, the rotation cue, and the star cue.

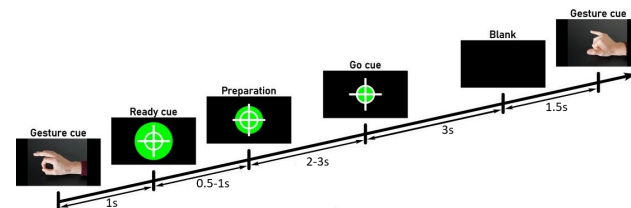


Figure 1: Timings of the different phases during one gesture trial. Starting with the gesture presentation, followed by the ready cue, thereafter the preparation phase, then the go cue and at last the blank (rest) phase.

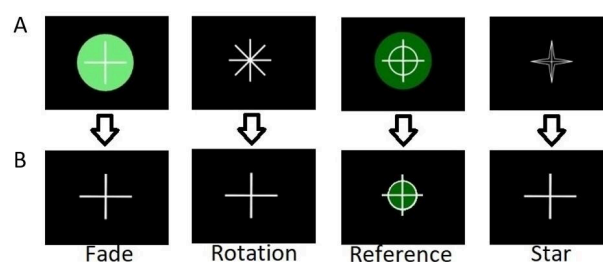


Figure 2: Four cues. A. Initial position/shape of the different cues. B. End position/shape.

In the self-paced part, participants were instructed to execute the same gesture and hold it for 3s (same procedure as in the cue based part) at approximately 10-second intervals over the course of a 5-minute run. The timing of one run can be seen in Fig. 3. This protocol was repeated across a total of 8 runs, leading to 60 trials for each distinct gesture. It is worth mentioning that the amount of movement trials for each subject differ slightly, since the self-paced paradigm instructs the participants to do the movements approximately every 10s, leading to some variation between subjects.

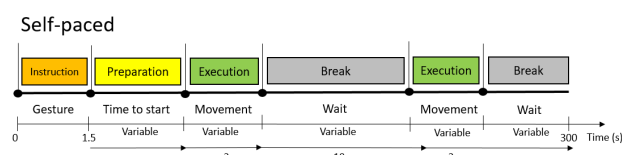


Figure 3: Timing of the self-paced paradigm and different steps during one 5-minute run.

## B. Processing of recordings

The recorded signals underwent offline processing and analysis using MATLAB R2019a and the EEGLAB toolbox ([16]). For the offline analysis, the EEG signals from 60 channels in a standard 10-10 setup underwent preprocessing steps. Initially, a zero-phase band-pass filter, implemented as a third-order Butterworth filter was applied to the signals within the frequency range of 0.3 to 70 Hz. To eliminate power line interference at 50 Hz, a notch filter was employed. Independent component analysis (ICA) was utilized to remove artifacts related to eye and muscle activity. A common average reference (CAR) was then applied. As we concentrate specifically on MRCPs for this study, we bandpass filtered the data in the low-frequency range of (0.5 - 5 Hz) by using a 3th order IIR filter and thereafter we resampled the whole signal at 10 Hz to decrease computational workload. Temporal alignment of all trials occurred with respect to the cue onset, within a window spanning from -2 s to 2 s. After a thorough visual inspection, epochs exceeding the threshold of  $\pm 50$   $\mu$ V were excluded. Kinematic data related to gestures were used to calculate the velocity of participants' hand movements. Movement onset was determined when the hand's velocity surpassed a predefined threshold (which was set to the same velocity value for all participants) between the "go" cue and the cue for the break, ensuring accurate detection while minimizing false positives (FP) for small movements during rest. We had to exclude one participant's dataset due to exceptionally poor signal quality and kinematic tracking.

## C. Training a classifier on the different cues and applying it on self-paced data

In this study, our main goal was to develop an MRCP-based classifier capable of predicting when a gesture occurs regardless of the specific gesture itself. We therefore combined, for each participant, trials from all four gestures into a single class, labeled as 'movement' (4 gestures x 48 trials = 192 trials). Data from the rest condition (independently of the cue type) were utilized as a 'rest' class (4 cues x 4 trials = 192 trials).

To identify the point of maximum discrimination between the two classes around the cue onset, we employed a 2-class shrinkage linear discriminant analysis (sLDA) [17], [18], using overlapping 1.2-second window segments of current and past EEG lags within each participant. We experimented with various window lengths and selected the one that yielded the highest accuracy. The input of the classifier included EEG data (band-pass filtered between 0.5 - 5 Hz, as described in section B.) from both gesture (i.e., 'movement') and resting (i.e., 'rest') trials, which were aligned around one of the four previously mentioned visual cues. To evaluate the performance of the classifier within each window, we applied a trial-based 10x1 fold cross-validation approach. Subsequently, we

selected the window with the highest cross-validated accuracy to train the final classifier, utilizing data from all trials. This classifier was then used to predict offline movement instances during the self-paced paradigm.

To prevent multiple detections during a movement period and to reduce the number of FP during the self-paced paradigm, we adjusted the threshold for the movement class probability and we additionally introduced a dwell time and a refractory period [13], [19]. The dwell time verifies whether there are consecutive detections within a specified timeframe, and only when this condition is met a movement is finally predicted. Once a movement has been predicted, the refractory period skips any further check for movement until a specified amount of samples have passed. For the classification of the asynchronous data the class probability, the dwell time and the refractory period were optimized individually for each subject and varied between 0.6 to 0.99, 0.5 to 2s and 2.5 to 4s, respectively. By adjusting these parameters, we ensured that multiple detections did not occur within a detection window (defined as [-0.5 1] seconds around an actual movement), while maintaining an overall FP count to 2 FP/min. Furthermore, considering that the gesture was performed for approximately 3s, an additional MRCP was produced when the participant returned to the resting state. Therefore, the refractory period was crucial to avoid detecting these movements upon returning to the resting state.

## RESULTS

When assessing the effectiveness of visual cues, the discrepancy between the actual onset of movement and the onset signaled by the cue is of essential importance. Therefore, we defined 'temporal variability' as the difference between the movement onset and the cue onset for each movement performed by every subject. A negative time value indicates that the movement occurred before the cue onset, whereas a positive value indicates that the movement occurred after the cue onset. We investigated this variability for the four different cue types. The temporal variability is shown in the violin plots of Fig. 4. To evaluate statistically the differences between cues, we conducted for each pair of cues a Wilcoxon ranksum test and corrected for multiple comparisons using the Benjamini-Hochberg method. It is evident that the rotation and reference cues exhibit a narrower distribution, whereas the results for the fade and star cue are much more widely dispersed. In Fig.5 the grand average MRCPs produced by the different cues can be seen. The rotation and reference cues display significantly more distinct grand average MRCP patterns, whereas the fade and star cues yield notably more blurred results.

The cross-validated accuracy (%), averaged across all participants, for the 2-class classification task (i.e., 'Movement' vs 'rest') during the cue-based part of the experiment is depicted in Fig. 6. Fig. 6 essentially illustrates the temporal evolution of the accuracy around

the cue onset. The rotation cue achieves its highest accuracy of 68.1% at 0.97 second, matching the accuracy of the reference cue, which also peaks at 68.1% and occurs at 0.77 seconds. In comparison, the Star and Fade cues both demonstrate their peak accuracy at 0.56 seconds, achieving 65.2% and 64.85%, respectively.

The results of the classification comparison of the different classifiers trained on the specific cue and applied on the self-paced data to predict the self-paced movement can be seen in Fig. 7. We show the different true positive rates (TPRs) within the specified detection window (i.e., [-0.5 1] seconds) around the true movement onset for the four cue types. The highest median TPR was achieved with the reference cue and the rotation cue with a value of 60%, while the fade and star cue had a significantly lower TPR of 55%. The highest subject wise accuracy was achieved by the rotation cue with a value of 67%, while the lowest accuracy occurred for the star cue with a TPR of 48%. The accuracy of both rotation and reference cue shows a similar variation, ranging approximately from 67% to 55%. Similarly, the fade and star cue demonstrate accuracy levels ranging from 63.5% to 48% and 65% to 48%, respectively. We performed a Wilcoxon ranksum test and corrected for multiple comparisons using the Benjamini-Hochberg method to examine potential variations in results among the different cues. There were no significant differences between the fade and star cue, as well as between the rotation and reference cue. However, there was a significant difference ( $p < 0.01$ ) between the fade cue and both the rotation and reference cue, as well as between the star cue and these two cues.

The rotation and reference cues exhibit similar ranges of accuracy, both in terms of lower and upper limits. Similarly, the fade and star cues also demonstrate comparable levels of accuracy across their respective lower and upper bounds. Notably, the rotation and reference cues generally show similar performance, as do the fade and star cue. Note that the overall FP count was maintained at 2 per minute for all participants. The median temporal disparity was for all cues the same and the difference between the predicted movement onset and the actual movement onset was 0.3s, indicating that, on median, the movement was forecasted 300 ms after the commencement of the actual movement.

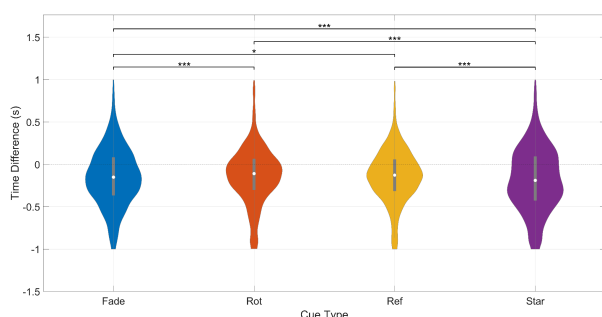


Figure 4: Violin plots depicting the time difference between movement onset and cue onset from all gesture

trials for each of the four cue types. Statistically significant differences between cues are indicated with stars (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). p-values were corrected for multiple comparisons using the Benjamini-Hochberg method.

MRCPs (Training classifier)

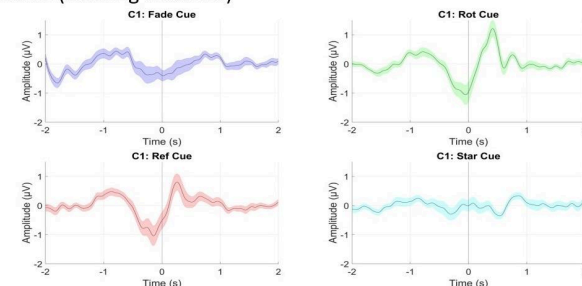


Figure 5: Grand average of MRCPs during different cues. MRCPs are aligned to the cue onset ( $t=0s$ ).

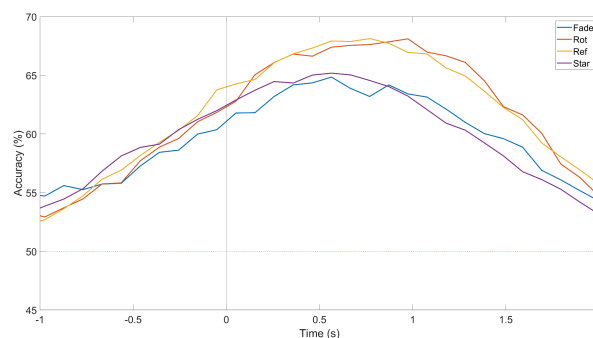


Figure 6: Grand-average cross-validated accuracy (%) for the 2-class classification task (i.e., 'Movement' vs 'rest') and for the different cue types within the duration of a trial ( $t=0s$  corresponds to the cue onset).

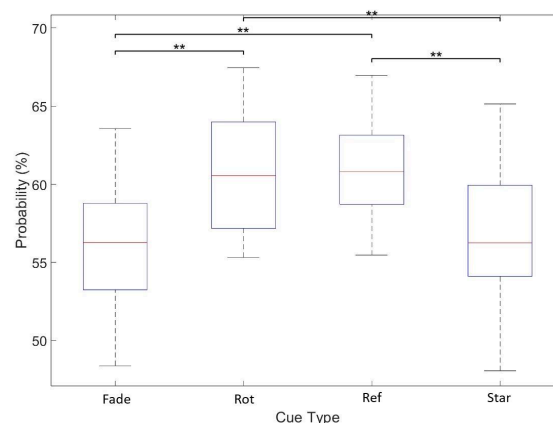


Figure 7: Boxplot showing the subject-specific TPRs in % for the different cue types. Statistically significant differences between cues are indicated with stars (\*\* $p < 0.01$ ). p-values were corrected for multiple comparisons using the Benjamini-Hochberg method.

## DISCUSSION

In this work we focused on movement onset prediction

during self-paced executed movement using classifiers trained on MRCPs triggered by four different visual cues.

In this context we observed that there is a significant difference in the decoding performance among the cues. Specifically the rotation and reference cue yielded the highest TPR in the self-paced data - a result that was expected since the shape of the grand average MRCPs was in general more pronounced for these cues. Additionally, the design of the cue considerably impacted the exact initiation time of participants' movement execution. As a result, we observed a higher temporal variability for the fade and star cue in the movement onsets, which led to blurring in the grand average potentials. These findings indicate the importance of the cue design, meaning that the starting period for the movement needs to be as precise as possible while the graduating period from the start to the 'go' position of the cue minimizes visual evoked potentials influencing the MRCPs.

Overall a TPR for the reference and rotation cue of 61% was obtained, while the fade and star cue achieved a significantly lower TPR of 56%. The total corresponding FP count amounted to 2 FP/min. These findings can be compared to the work done by [7][20], however, it's important to note that their investigations are based on attempted arm movement and motor imagery, respectively.

One reason for the low TPR and quite high FP count could be that the training of the classifier was based on MRCPs triggered on the cue onset, where the presence of temporal variability have adversely affected the MRCP patterns and thus the prediction performance. Additionally, the high imbalance of the dataset between the amount of movement and no movement instances poses a significant challenge.

In terms of movement onset detection we report a median temporal time delay of 0.3s for the movement onset prediction. This is expected since we use a non-causal filter, whereas in an online scenario the need of a causal filter would further increase this delay.

For the decoding performance, the correct settings of the hyperparameters class probability threshold, dwell time and refractory period proved as crucial for achieving the optimal balance between an overall low FP count and the highest TPR. In detail, we tuned the values of the movement class probability threshold between 0.5 to 0.99, the dwell time between 0.5s and 2s and for the refractory period between 2.5s to 4s for each subject. It is worth noting that the refractory period could be set up to 4s because we knew in advance how much time there would be between each movement. For a real world use the refractory period needs to be adjusted to a much lower time to allow a higher communication rate.

Although the study delivered interesting insights into the improvement of cue design in relation to synchronous and asynchronous BCI, there are some aspects to be considered for future work. First, in terms of the study design it would maybe be better to allow

participants to perform the gestures during the self-paced part at their own pace, meaning there is no condition like to hold the gesture for a certain amount of time. This could lead to a more consistent outcome in ignoring the movement offset MRCPs, since now it could be that the participants have a high variance in the time of hold when returning from the end position of the gesture to the rest position which is maybe not always covered by the refractory period. Second, upon analysis, both the rotation and reference cues consistently outperformed the fade and star cue across the evaluated metrics. However, further refinement of their design principles is necessary to enhance the precision of indicating the start of movement execution. Additionally the instructions and test runs for the participants could be improved to explain to them the importance of the exact starting time when they are indicated to do so. This would lead to more pronounced MRCPs on average when triggered on the cue onset. When analyzing the differences between cue onset and actual movement onset, the fade and star cue exhibited high temporal variability inflicted by the nature of their design. This suggests that these cues synchronize movement timing less effectively, which warrants consideration in future studies.

## CONCLUSION

We demonstrated that the variations in MRCP shapes influenced by visual cue types play a crucial role in decoding performance when applying a cue-based decoder on self-paced data. Significantly, the rotation and reference cue yielded the most favorable results in terms of the true positive rate, whereas the fade and star cue exhibited comparatively poorer performance. This discovery aligns with the hypothesis that the rotation and reference cue, with their precise onset timing and more pronounced grand average MRCP patterns, are better suited for training classifiers in online scenarios. Future cue designs may be able to further improve the exact indication timing of the go cue for the movement to enhance the decoding performance even more. For the detection of executed movement in an asynchronous BCI there is a need for further improvements to lower the FP count and increase the TPR when the classifier is trained on MRCPs triggered on a cue onset. The tuning of the three hyperparameters is crucial but future considerations should also involve enhancing the methodological aspects of this work,

## ACKNOWLEDGEMENTS

This project is funded by the European Union's HORIZON-EIC-2021-PATHFINDER CHALLENGES program under grant agreement No 101070939 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00198.

The authors acknowledge Kyriaki Kostoglou and Hannah S. Pulferer of the Graz BCI Lab for their

valuable comments and suggestions on the methodology and interpretation of the data.

## REFERENCES

- [1] J. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. OUP USA, 2012.
- [2] M. Clerc, L. Bougrain, and F. Lotte, *Brain-Computer Interfaces 1: Methods and Perspectives*. John Wiley & Sons, 2016.
- [3] G. Müller-Putz, M. Crell, J. Egger, P. Suwandyjeff, and K. Kostoglou, “Towards Implantable Brain-Computer Interface for Communication in Locked-In Syndrome patients: An introduction to INTRECOM,” *Current Directions in Biomedical Engineering*, vol. 9, no. 2, pp. 1–4.
- [4] L. Deecke, *Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential u. reafferente Potentiale*. [Ausz.] (Mit 7 Textabb.). 1965.
- [5] H. Shibasaki and M. Hallett, “What is the Bereitschaftspotential?,” *Clin. Neurophysiol.*, vol. 117, no. 11, pp. 2341–2356, Jul. 2006.
- [6] G. Pfurtscheller, R. Scherer, G. R. Müller-Putz, and F. H. Lopes da Silva, “Short-lived brain state after cued motor imagery in naive subjects,” *Eur. J. Neurosci.*, vol. 28, no. 7, pp. 1419–1426, Oct. 2008.
- [7] P. Ofner, A. Schwarz, J. Pereira, D. Wyss, R. Wildburger, and G. R. Müller-Putz, “Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury,” *Sci. Rep.*, vol. 9, no. 1, p. 7134, May 2019.
- [8] J. V. Odom *et al.*, “Visual evoked potentials standard (2004),” *Doc. Ophthalmol.*, vol. 108, no. 2, pp. 115–123, Mar. 2004.
- [9] H. Scheel, R. Xu, N. Jiang, N. Mrachacz-Kersting, K. Dremstrup, and D. Farina, “Influence of external cues on synchronized Brain-Computer Interface based on movement related cortical potentials,” in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2015, pp. 45–48.
- [10] S. Pearce, J. Boger, N. Mrachacz-Kersting, D. Farina, and Ning Jiang, “Evaluating the effectiveness of different external cues on non-invasive brain-computer interfaces,” *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2017, pp. 2782–2785, Jul. 2017.
- [11] A. Savić *et al.*, “Movement Related Cortical Potentials and Sensory Motor Rhythms during Self Initiated and Cued Movements,” in *Replace, Repair, Restore, Relieve – Bridging Clinical and Engineering Solutions in Neurorehabilitation*, Springer International Publishing, 2014, pp. 701–707.
- [12] F. Cincotti *et al.*, “Vibrotactile feedback for brain-computer interface operation,” *Comput. Intell. Neurosci.*, vol. 2007, p. 48937, 2007.
- [13] G. R. Müller-Putz, V. Kaiser, T. Solis-Escalante, and G. Pfurtscheller, “Fast set-up asynchronous brain-switch based on detection of foot motor imagery in 1-channel EEG,” *Med. Biol. Eng. Comput.*, vol. 48, no. 3, pp. 229–233, Mar. 2010.
- [14] J. Pereira, A. I. Sburlea, and G. R. Müller-Putz, “EEG patterns of self-paced movement imaginations towards externally-cued and internally-selected targets,” *Sci. Rep.*, vol. 8, no. 1, p. 13394, Sep. 2018.
- [15] P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz, “Upper limb movements can be decoded from the time-domain of low-frequency EEG,” *PLoS One*, vol. 12, no. 8, p. e0182578, Aug. 2017.
- [16] A. Delorme and S. Makeig, “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
- [17] R. Peck and J. Van Ness, “The use of shrinkage estimators in linear discriminant analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 4, no. 5, pp. 530–537, May 1982.
- [18] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, “Single-trial analysis and classification of ERP components--a tutorial,” *Neuroimage*, vol. 56, no. 2, pp. 814–825, May 2011.
- [19] G. Townsend, B. Graimann, and G. Pfurtscheller, “Continuous EEG classification during motor imagery--simulation of an asynchronous BCI,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no. 2, pp. 258–265, Jun. 2004.
- [20] I. K. Niazi, N. Jiang, O. Tiberghien, J. F. Nielsen, K. Dremstrup, and D. Farina, “Detection of movement intention from single-trial movement-related cortical potentials,” *J. Neural Eng.*, vol. 8, no. 6, p. 066009, Dec. 2011.

# Towards a model-based personalization approach for driving a BCI

Adyasha Dash  
Institute of Neural Engineering  
Graz University of Technology  
Graz, Austria  
adash@tugraz.at

Selina Christin Wriessnegger  
Institute of Neural Engineering  
Graz University of Technology  
Graz, Austria  
s.wriessnegger@tugraz.at

**Abstract**—A Brain-Computer Interface (BCI) translates a person's intent, derived from brain signals, into control commands for various applications. This work focuses on Motor Imagery-based BCI (MI-BCI), specifically emphasizing sensorimotor-rhythm (SMR) and MI as the relevant task. While improvements have been made in classification algorithms and signal acquisition, human factors influencing user-BCI compatibility remain underexplored. User performance in MI-BCI systems is impacted by personal, psychological, and neurophysiological factors, leading to a phenomenon termed “BCI illiteracy”. In this work, we aim to address BCI illiteracy through a systematic, standardized study, incorporating various human factors to enhance user performance by developing a neural network model predicting a trainability score and a training regime. To achieve this, the MI-BCI systems use population-specific indicators and task-based modulators, integrating anatomical, psychological, and neurophysiological information (EEG, biosignals). The proposed model-based personalization approach offers reproducible, innovative, and open-source training protocols to boost BCI performance avoiding prolonged and ineffective training sessions. The ultimate goal is to eliminate BCI illiteracy as a barrier to compatibility between users and BCI systems.

**Index Terms**—BCI-Illiteracy, Deep learning, individualization

## I. INTRODUCTION

Past research has identified predictors of performance in Motor Imagery (MI)-based Brain-Computer Interface (BCI) systems, primarily focusing on neurophysiological and psychological factors. Noteworthy neurophysiological predictors were extensively reviewed by [1], with studies like [2] highlighting the predictive value of resting sensorimotor-rhythm (SMR) amplitudes. Psychological factors, such as mood, motivation, focus of control, and fear, have also been linked to MI-BCI performance [3][4]. Additional studies established correlations between attention span, personality, motivation, spatial abilities, and MI-BCI performance [5]. Recent work by [6] associated Event-Related Desynchronization (ERD) with age, education level, management impression, and anxiety, emphasizing the need to consider such factors in designing ERD-based MI-BCIs. A comprehensive review by [7] proposed strategic approaches to address performance variations and enhance BCI reliability. In a distinct effort, [8] investigated the impact of the Most Discriminant Frequency Band (MDFB) selected during MI-BCI calibration. Their findings suggested a

correlation between user-specific frequency band characteristics and classification accuracy, emphasizing the importance of understanding the learning characteristics of both human users and machines. Despite these individual efforts, a systematic approach integrating all identified factors is currently lacking in the pursuit of improving overall user performance in BCI systems.

Extensive research has already concentrated on identifying specific factors that impact the accuracy of performance in Motor Imagery (MI)-based Brain-Computer Interface (BCI) systems. While some investigators have explored potential reasons for suboptimal BCI performance from the user's perspective, others have dedicated their efforts to enhancing machine learning algorithms or diversifying hardware types. Nevertheless, the question regarding the underlying causes of user incompatibility with BCI systems remains unresolved. Therefore, it is imperative to undertake a systematic, standardized, and well-operationalized research project to address inquiries pertaining to the influence of specific human factors on BCI performance. The primary objective of the proposed framework is to embark on this research endeavor, ultimately aiming to establish a person-specific trainability score that can be employed in subsequent studies to enhance BCI performance from its inception. Specifically, our focus is on the development of a neural network model capable of predicting a trainability score for MI-based BCI systems. This prediction will be based on population-specific indicators and task-based modulators. The significant advantage lies in the potential to create an individualized BCI paradigm for each person, leveraging their unique features to ensure optimal outcomes with minimal training time. Essentially, the proposed model represents a groundbreaking step toward innovative activation protocols, introducing a model-based personalization approach for driving a BCI.

## II. GAPS IN MI-BCI RESEARCH

Researchers identified some critical aspects affecting MI-BCIs correct operation in general [9]. These aspects (referred to as components of MI-BCI) include signal measurement (acquisition hardware), classification and recognition algorithms, and user-BCI compatibility. A holistic approach is used to compute the performance in a MI-BCI system, which captures

the resultant of performances of each of its components. Thus, the failure or inefficiency in any individual component could significantly affect the efficiency of the MI-BCI system. Researchers have attempted to identify the problems associated with each of these components and, further, worked towards addressing them to improve the efficiency of a MI-BCI system. Specifically, a great deal of research has been directed towards (1) improving the EEG acquisition system by developing cost-effective, portable, wireless, and easy mounting EEG devices to operate in low power setting, and (2) developing state-of-the-art methods for processing and decoding information from EEG signal. However, the efforts made towards understanding and improving user-BCI compatibility are still very sparse. The pictorial representation of these components are given in Figure.1.

#### A. Poor understanding of BCI-Illiteracy and its influencers

The BCI illiteracy could come under the umbrella of "user-BCI compatibility" and is defined as a condition where users of BCI technology fail to reach proficiency in using a BCI within a standard training period. According to the literature, nearly 15–30% of BCI users could be labeled as BCI illiterate [10][11]. The cause behind the incompatibility due to BCI-illiteracy may not be always because of the deficit innate to the user, rather could also be driven by the incapability of the system to tailor its functionality according to the user. For example, the poor performance of the user can be a result of (1) user being unable to receive input from the system (stimulus, feedback, or information about the state of the BCI) or being potentially scared by the stimulus, (2) user being unable to focus on the required mental task because of a high mental workload or an increase in fatigue, (3) variability in user-centric factors such as mood, stress, engagement, and level of attention etc. Thus, it is imperative to include these fast performance predictors, based on anatomical, psychological, and neurophysiological information of the user, to estimate likelihood of incompatibility. Then a user-specific training protocol could be proposed to alleviate incompatibility situation in case of MI-BCI. As suggested by recent literature [12][13][14][15][16], a BCI paradigm that is compatible with all participants does not exist. Moreover, it must be adapted to the users' needs by following a user-centered design and individual features (e.g., personality, age, mood, motivation, etc.) which should be taken into account.

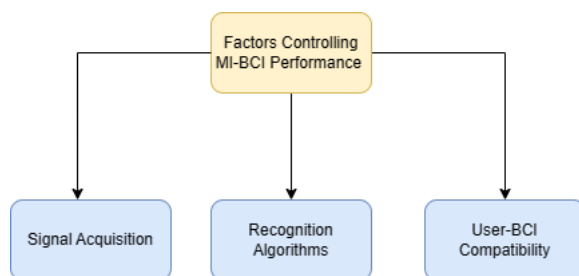


Fig. 1. Components of MI-BCI

#### B. Adherence to a single training protocol for all the participants

Conventionally, the MI-BCI training paradigm adapts a fixed training protocol that is administered uniformly to all subjects. However, considering the significant individual differences in the ability of skill training, necessary for MI-BCI adaptation, it is expected that performance outcomes may vary among individuals when subjected to a fixed training regimen. Thus, an individualized approach in defining training regime could be a viable solution to avoid poor performance in using MI-BCI systems. According to literature, the skill training is often linked to the demographics (e.g., age, gender), cognitive ability and prior experience [7][17]. Thus, it is important to take these factors into consideration for defining an individualized BCI-training approach.

The main goal of this work is to design a framework (1) to predict the likelihood of user incompatibility for a MI-BCI paradigm (referred as trainability score) and (2) to define the intensity of MI-BCI-training in a user-specific manner for maximizing the user performance (and avoiding incompatibility). (3) By collecting data of more than 100 participants, establishing a publicly open database.

### III. PROPOSED METHOD

This work proposes a personalized framework for addressing MI-based BCI training problems by predicting the possibility of incompatibility and then pre-deciding the intensity of training paradigm for the user while considering its demographics and physiological factors into consideration for optimizing his/her MI-BCI performance. The framework, a neural network architecture, takes demographics, neurophysiological, psychological factors as input, and gives trainability score and the personalized training paradigm for user. The pictorial representation of the framework is given in Figure.2.

Here, we will consider a participant pool of 80-100 healthy individuals in the age group of 18-60 years, including all genders. These participants will be exposed to multiple trials over a period of 4 weeks during with their neurophysiological data (EEG, EMG, HR), demographics and psychological information will be collected. Their performance in all the trials will be recorded. All the data collected during the trials will be used to train the neural network model that can take the demographic information, psychological states, physiological measurements and performance in a trial to predict the performance of a subsequent trial. This data collection protocol, given in Fig.3, will be performed over different sessions divided into runs of approximately 7 minutes each [18][19]. To record enough data samples for the DL model one session will include 8 runs, and each run itself is divided into trials, with 30 per class (i.e., per MI-task). One trial typically lasts 8s. Figure 3 illustrates the timings and parts of one trial. The MI task will include 3 classes namely, imagination of foot, right hand and left hand. The study will be conducted with a prior approval from local ethics committee of the University of Technology Graz. The participants will give their written and informed consent prior to their enrollment in the study.

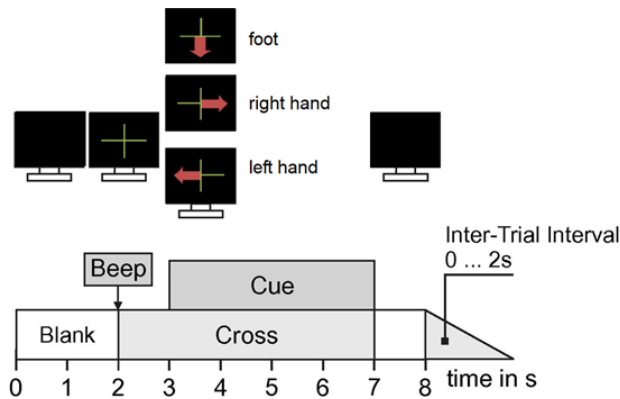


Fig. 2. Data collection paradigm

#### A. Modulators influencing MI-BCI performance

As evident from the literature, several modulators act simultaneously to influence the MI-BCI performance of the user. In this study, we categorize these modulators into three classes, (1) demographic factors, (2) psychological factors and, (3) neurophysiological factors (Figure 3). In demographic factors,

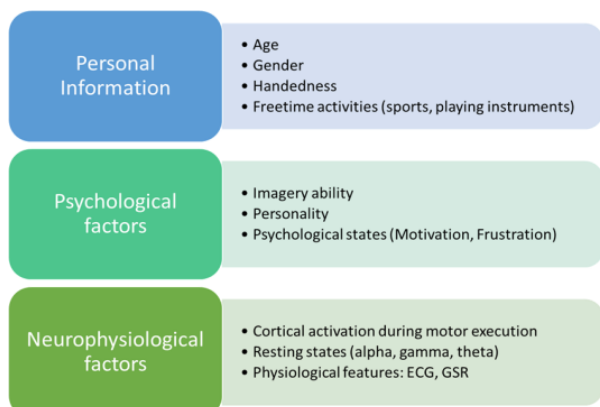


Fig. 3. Modulators of User-BCI compatibility in MI-BCI

age and gender have already been identified to have an impact on driving BCIs or even on the ability to perform a motor imagery task. In a recent study [20], we investigated how handedness impacts brain activity in motor-related areas and found significant differences in brain activity between left- and right-handed participants during MI. In psychological factors, attention, memory load, fatigue, and competing cognitive processes [21][22] [23][24] influence instantaneous brain dynamics. In addition, states like empathy might influence BCI performance as shown by [15]. Motivation is also related to P300-BCI performance [3]. Hammer et al. [25] for example found that abilities in visuo-motor coordination and the ability to concentrate on a task were correlated with BCI performance. Others [26] reported on the correlation between motor imagery ability (measured by questionnaires) and following BCI performance. Psychological information will be collected through different questionnaires. For example, to evaluate the imagery ability

of persons the “Vividness of Movement Imagery Questionnaire (VMIQ-2)” will be used. Personality factors will be retrieved by “B5T Big Five personality test” and with the “STADI”, anxiety and depression can be recorded both as a state and as a trait. Furthermore, the “Intrinsic Motivation Inventory (IMI)” assesses participants’ interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice while performing a given activity, thus yielding six subscale scores. The Perceived Stress Scale (PSS) will be used for measuring the perception of stress. It is a measure of the degree to which situations in one’s life are appraised as stressful. In neurophysiological factors, physiological predictors such as spectral entropy and power spectral density, derived from resting state EEG recordings are correlated with BCI performance [27][28][29][30]. In addition, the baselines of resting state networks (RSNs) are dynamic and modify any cortical signature instantaneously [31]. An efficient BCI system must be robust to such inherent physiological fluctuations over time to enable more generalized systems [32]. Ahn et al. [13] for example reported that high theta and low alpha is the pattern for BCI-illiteracy and that frontal gamma correlated with BCI performance. Another important neurophysiological predictor for a participant’s performance in operating an MI-based BCI was developed by [10]. They found that the alpha rhythm shows a positive correlation with online BCI performance. This so-called Blankertz SMR-predictor is currently one of the most replicated and reliable neurophysiological predictors of MI-BCI performance. Furthermore, Halder et al. [33] observed a correlation between structural integrity and myelination quality of deep white matter structures and BCI performance. We will integrate the execution task to compare related ERDS patterns with those of MI as a further possible predictor from neurophysiology based on EEG.

#### B. Model and training objective

A data-driven exploration of individual information and physiological data will be carried out with the help of sequential processing deep learning methods. We will use neural network model which are designed to process sequential data such as time series, natural language etc. Different variants of the deep neural network will be considered in the study. These architectures can learn from input data and predict values for the future steps. We will train these models to predict performance scores of a participant of MI tasks from the modulators and performance scores of previous tasks. The effects of motor imagery training can be maximized by personalized experimental designs based on the outcome of the NN model. For example, designing individual schedules, choosing adequate task complexity, instructions, and, in clinical populations, adapting the models for individual impairment.

#### IV. CONCLUSION AND FUTURE WORK

In this work, we introduced a novel theoretical framework for addressing MI based BCI user training problems by pre-

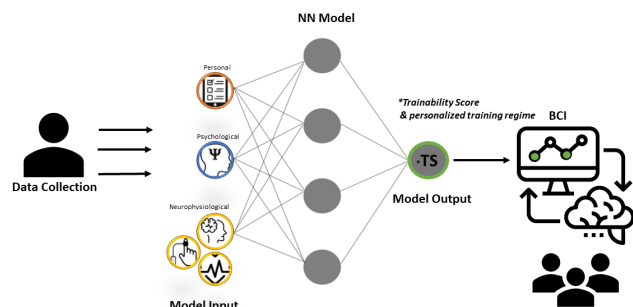


Fig. 4. Flowchart of personalized model for MI-BCI training

dicting the possibility of incompatibility and then pre-deciding the intensity of training paradigm for users. The framework will be a neural network-based architecture and considers demographics, neurophysiological, psychological factors of the user as input to decide a personalized training regime for the user. In the course of execution of this work, we will cover a large sample size with varying age, gender, cognitive ability, handedness and physiological activation. The data collected during the implementation phase will be made available in an open-source platform. Moreover, in future, this endeavor might lay the groundwork for crafting a personalized training paradigm that is both efficient in time utilization and does not compromise user performance.

## REFERENCES

- [1] Ahn, M., Jun, S. C. (2015). Performance variation in motor imagery brain-computer interface: a brief review. *Journal of neuroscience methods*, 243, 103-110.
- [2] Blankertz, B., Sannelli, C., Halder, S., Hammer, E. M., Kübler, A., Müller, K. R., ... Dickhaus, T. (2010). Neurophysiological predictor of SMR-based BCI performance. *Neuroimage*, 51(4), 1303-1309.
- [3] Nijboer, F., Birbaumer, N., & Kübler, A. (2010). The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis—a longitudinal study. *Frontiers in neuroscience*, 4, 1653.
- [4] Hammer, E. M., Kaufmann, T., Kleih, S. C., Blankertz, B., & Kübler, A. (2014). Visuo-motor coordination ability predicts performance with brain-computer interfaces controlled by modulation of sensorimotor rhythms (SMR). *Frontiers in human neuroscience*, 8, 574.
- [5] Jeunet, C., N'Kaoua, B., Subramanian, S., Hachet, M., & Lotte, F. (2015). Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns. *PloS one*, 10(12), e0143962.
- [6] Rimbart, S., & Lotte, F. (2022, March). Relationship between ERD modulations, MI-based BCI performance and users' traits. In *CORTICO 2022: Invasive and non invasive Brain-Computer Interfaces-A handshake over the cliff*.
- [7] Ahn, M., & Jun, S. C. (2015). Performance variation in motor imagery brain-computer interface: a brief review. *Journal of neuroscience methods*, 243, 103-110.
- [8] Benaroch, C., Yamamoto, M. S., Roc, A., Dreyer, P., Jeunet, C., & Lotte, F. (2022). When should MI-BCI feature optimization include prior knowledge, and which one?. *Brain-Computer Interfaces*, 9(2), 115-128.
- [9] Spezialetti, M., Cinque, L., Tavares, J. M. R., & Placidi, G. (2018). Towards EEG-based BCI driven by emotions for addressing BCI-Illiteracy: a meta-analytic review. *Behaviour & Information Technology*, 37(8), 855-871.
- [10] Blankertz, B., Sannelli, C., Halder, S., Hammer, E. M., Kübler, A., Müller, K. R., ... & Dickhaus, T. (2010). Neurophysiological predictor of SMR-based BCI performance. *Neuroimage*, 51(4), 1303-1309.
- [11] Vidaurre, C., & Blankertz, B. (2010). Towards a cure for BCI illiteracy. *Brain topography*, 23, 194-198.
- [12] Edlinger, G., Allison, B. Z., & Guger, C. (2015). How many people can use a BCI system?. *Clinical systems neuroscience*, 33-66.
- [13] Ahn, M., Cho, H., Ahn, S., & Jun, S. C. (2013). High theta and low alpha powers may be indicative of BCI-illiteracy in motor imagery. *PloS one*, 8(11), e80886.
- [14] Kübler, A., Nijboer, F., & Kleih, S. (2020). Hearing the needs of clinical users. *Handbook of clinical neurology*, 168, 353-368.
- [15] Kleih, S. C., & Kübler, A. (2013). Empathy, motivation, and P300 BCI performance. *Frontiers in human neuroscience*, 7, 642.
- [16] Lotte, F., & Jeunet, C. (2018). Defining and quantifying users' mental imagery-based BCI skills: a first step. *Journal of neural engineering*, 15(4), 046030.
- [17] Zich, C., Debener, S., Thoene, A. K., Chen, L. C., & Kranczioch, C. (2017). Simultaneous EEG-fNIRS reveals how age and feedback affect motor imagery signatures. *Neurobiology of aging*, 49, 183-197.
- [18] Neuper, C., & Pfurtscheller, G. (2010). Neurofeedback training for BCI control. *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*, 65-78.
- [19] Pfurtscheller, G., & Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7), 1123-1134.
- [20] Crotti, M., Koschutnig, K., & Wriessnegger, S. C. (2022). Handedness impacts the neural correlates of kinesthetic motor imagery and execution: A fMRI study. *Journal of Neuroscience Research*, 100(3), 798-826.
- [21] Di Rienzo, F., Collet, C., Hoyek, N., & Guillot, A. (2012). Selective effect of physical fatigue on motor imagery accuracy.
- [22] Baykara, E., Ruf, C. A., Fioravanti, C., Käthner, I., Simon, N., Kleih, S. C., ... & Halder, S. (2016). Effects of training and motivation on auditory P300 brain-computer interface performance. *Clinical Neurophysiology*, 127(1), 379-387.
- [23] Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biological psychology*, 102, 118-129.
- [24] Jeunet, C., Jahanpour, E., & Lotte, F. (2016). Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study. *Journal of neural engineering*, 13(3), 036024.
- [25] Hammer, E. M., Halder, S., Blankertz, B., Sannelli, C., Dickhaus, T., Kleih, S., ... & Kübler, A. (2012). Psychological predictors of SMR-BCI performance. *Biological psychology*, 89(1), 80-86.
- [26] Vuckovic, A., & Osuagwu, B. A. (2013). Using a motor imagery questionnaire to estimate the performance of a brain-computer interface based on object oriented motor imagery. *Clinical Neurophysiology*, 124(8), 1586-1595.
- [27] Kang, J. H., Youn, J., & Kim, J. (2021). Effects of frontal theta rhythms in a prior resting state on the subsequent motor imagery brain-computer interface performance. *Frontiers in Neuroscience*, 15, 663101.
- [28] Leeuwis, N., Yoon, S., & Alimardani, M. (2021). Functional connectivity analysis in motor-imagery brain computer interfaces. *Frontiers in Human Neuroscience*, 15, 732946.
- [29] Zhang, K., Robinson, N., Lee, S. W., & Guan, C. (2021). Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network. *Neural Networks*, 136, 1-10.
- [30] Zhang, Y., Xu, P., Guo, D., & Yao, D. (2013). Prediction of SSVEP-based BCI performance by the resting-state EEG network. *Journal of neural engineering*, 10(6), 066017.
- [31] Casimo, K., Weaver, K. E., Wander, J., & Ojemann, J. G. (2017). BCI use and its relation to adaptation in cortical networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10), 1697-1704.
- [32] Saha, S., & Baumert, M. (2020). Intra-and inter-subject variability in EEG-based sensorimotor brain computer interface: a review. *Frontiers in computational neuroscience*, 13, 87.
- [33] Halder, S., Varkuti, B., Bogdan, M., Kübler, A., Rosenstiel, W., Sitaram, R., & Birbaumer, N. (2013). Prediction of brain-computer interface aptitude from individual brain structure. *Frontiers in human neuroscience*, 7, 105.

---

---