Proceedings of the 8th Graz Brain-Computer Interface Conference 2019

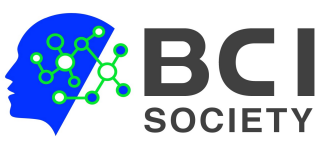
Bridging Science and Application

September 16-20, 2019
Graz University of Technology, Austria

Edited by
Gernot R. Müller-Putz, Jonas C. Ditz, Selina C. Wriessnegger

Officially endorsed by the BCI Society

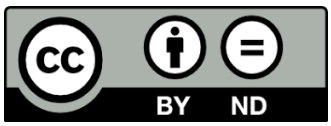
Verlag der Technischen Universität Graz 2019



 DOI: 10.3217/978-3-85125-682-6-00

Verlag der Technischen Universität Graz

http://ub.tugraz.at/Verlag



ISSN 2311-0422

ISBN 978-3-85125-682-6

DOI 10.3217/978-3-85125-682-6

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

Welcome Note

Bridging Science and Application

The reason for the choice of the title for our conference was to strengthen new ways and push for translation.
New insights naturally lead to open questions which we want to investigate and find solutions. However, we have to keep in mind the high goal of this research – to find solutions for people with disabilities, end users and not to forget their relatives. This year, the Conference has, besides workshops on specific topics one very special side event: The CYBATHLON BCI Series 2019. Pilots with physical impairments, who control computer avatars with their thoughts, will demonstrate how far brain-computer interface research has come. The BCI race will take place in front of a live audience and – challenging for all involved, pilot as well as developing team: far away from any lab conditions. As a special keynote, we present Prof. Robert Riener (ETH Zurich, Switzerland). He is full professor of sensory motor systems and the initiator of CYBATHLON. With his talk he will make a link between the BCI field, his research disciplines and CYBATHLON.

Furthermore, participating teams are invited to present their technologies, methods and algorithms in form of posters throughout the whole conference – hopefully provoking stimulating discussions. This 8th Graz Brain-Computer Interface Conference (GBCIC2019) offers the opportunity for extensive discussions and exchange of ideas among BCI experts from more than 20 countries. We received 76 scientific contributions from roughly 250 authors. The scientific contributions have been peer-reviewed by at least two reviewers (acceptance rate 87%) and collected in form of open access conference proceedings.
For the Conference itself, we have been able to setup a colorful and multifaceted program. We are very happy that the GBCIC2019 has been officially endorsed by the BCI Society. Further, we are lucky that outstanding experts in the field, Dr. Damien Coyle (Ulster University, Northern Ireland, UK), Prof. Moritz Grosse-Wentrup (University of Vienna, Austria), Dr. Robert Gaunt (University of Pittsburgh, PA, USA), Dr. Mariska Vansteensel (University Medical Center Utrecht, The Netherlands), and Prof. Pim Haselager (Radboud University Nijmegen, The Netherlands) accepted our invitation to present keynote addresses at the Conference.

We hope that this conference contributes towards a strong scientific cooperation among our field, and we wish all participants an exciting, stimulating and productive Graz BCI Conference 2019!

Gernot R. Müller-Putz
Conference Chair

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

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 Technoloy 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 18 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 six international Brain-Computer Interface Conferences over the last 13 years in Graz and chairing the 8th Conference in Sept. 2019. He was also in the Programm Committees of the 7th International BCI Meeting 2018, 10th NeuroIS Restreat, ICCHP 2018. He is Review Editor of Frontiers in Neuroscience, special section Neuroprosthetics, Associate Editor of IEEE Transactions in Biomedical Engineering and Associate Editor of the Brain-Computer Interface Journal. Since August 2019 he is Speciality Chief Editor of Frontiers in Human Neuroscience: Brain-Computer Interfaces. He has authored more than 156 peer reviewed publications and more than 180 contributions to conferences which were cited more than 14300 times (h-index 60). Recently he was awarded with an ERC Consolidator Grant “Feel your Reach” from the European Research Council. In May 2017 he received the Ludwig-Guttman 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 2010 he received the Science Award from the State of Styria.

Jonas Christian Ditz is university assistant at the Institute of Neural Engineering (BCI-Lab), Graz University of Technology, Austria. He received his M.Sc. in Bioinformatics from the Eberhard Karls University Tübingen in 2018. From 2016 to 2018 he worked as a research assistant at the Max Planck Insitute for Biological Cybernetics in the Cognition & Control in Human-Machine Systems group. Currently he is working towards his PhD degree in computer science.

Selina Christin Wriessnegger is assistant 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. 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 subliminal visual information processing, neural correlates of motor imagery, novel applications of BCIs for healthy users, passive BCIs and embodiment of language acquisition.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

Organizing Committee I

Conference Chair

Univ.-Prof. Dipl.-Ing. Dr.techn.
Gernot R. Müller-Putz
Institute of Neural Engineering
Graz University of Technology
Austria
(BCI Society Member)

Papers and Proceedings

MSc
Jonas C. Ditz
Institute of Neural Engineering
Graz University of Technology
Austria

Ass.Prof. Mag.rer.nat. Dr.phil.
Selina C. Wriessnegger
Institute of Neural Engineering
Graz University of Technology
Austria

Students Awards

ing. Dr. MSc
Andreea I. Sburlea
Institute of Neural Engineering
Graz University of Technology
Austria

Social Media & Graphics

MSc
Joana Pereira
Institute of Neural Engineering
Graz University of Technology
Austria
(BCI Society Member)

Industrial Sponsoring

Bsc
Aleksandra Krajnc
Institute of Neural Engineering
Graz University of Technology
Austria

Poster Sessions

ing. Dr. MSc
Andreea I. Sburlea
Institute of Neural Engineering
Graz University of Technology
Austria

MSc
Catarina Lopes Dias
Institute of Neural Engineering
Graz University of Technology
Austria

Administration

Petra Still
Institute of Neural Engineering
Graz University of Technology
Austria

Bsc
Aleksandra Krajnc
Institute of Neural Engineering
Graz University of Technology
Austria

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

Organizing Committee II

Exhibition & Infrastructure

Dipl.-Ing.
Reinmar Kobler
Institute of Neural Engineering
Graz University of Technology
Austria

BSc
Aleksandra Krajnc
Institute of Neural Engineering
Graz University of Technology
Austria

Program Book

MSc
Catarina Lopes Dias
Institute of Neural Engineering
Graz University of Technology
Austria

MSc
Joana Pereira
Institute of Neural Engineering
Graz University of Technology
Austria
(BCI Society Member)

Talks and Venue

Dipl.-Ing.
Reinmar Kobler
Institute of Neural Engineering
Graz University of Technology
Austria

Sience Slam

Ass.Prof. Mag.rer.nat. Dr.phil.
Selina C. Wriessnegger
Institute of Neural Engineering
Graz University of Technology
Austria

Volunteers

Dipl.-Ing.
Andreas Schwarz
Institute of Neural Engineering
Graz University of Technology
Austria

BSc
Aleksandra Krajnc
Institute of Neural Engineering
Graz University of Technology
Austria

Additional Local Staff

Dipl.-Ing.
Lea Hehenberger
Institute of Neural Engineering
Graz University of Technology
Austria

Bsc
Philipp Raggam
Institute of Neural Engineering
Graz University of Technology
Austria

Dr.
Valeria Mondini
Institute of Neural Engineering
Graz University of Technology
Austria

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

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	University Medical Center Utrecht

B

Bauernfeind Günther	Medizinische Hochschule Hannover
Bianchi Luigi	University of Rome Tor Vergata

C

Chavarriaga Ricardo	Ecole Polytechnique Federale de Lausanne
Clerc Maureen	INRIA
Congedo Marco	CNRS
Cohen Ori	Bar Ilan University
Congedo Marco	University Grenoble Alpes
Coyle Damien	University of Ulster

D

Desain Peter	Radboud University
Ditz Jonas Christian	Graz University of Technology

F

Faller Josef	Columbia University
Farquhar Jason	Radboud University
Friedman Doron	IDC Herzliya

G

Grosse-Wentrup Moritz	Ludwig Maximilians University Munich
Guan Cuntai	Nanyang Technological University
Gutiérrez Dania	Center for Research and Advanced Studies (Cinvestav)

H

Halder Sebastian	Julius Maximilians University of Würzburg
Hehenberger Lea	Graz University of Technology

I

Ibanez Pereda Jaime	University College London

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

International Program Committee and Review Board II

J

Jeunet Camille	University Toulouse Jean Jaurès
Jin Jing	East China University of Science and Technology

K

Kanoh Shin'Ichiro	Shibaura Institute of Technology
Kobler Reinmar	Graz University of Technology
Krusienski Dean	Old Dominion University
Käthner Ivo	Julius Maximilians University of Würzburg
Kübler Andrea	Julius Maximilians University of Würzburg

L

Lopes Dias Catarina	Graz University of Technology
Lotte Fabien	INRIA Bordeaux Sud-Ouest
López-Larraz Eduardo	Eberhard Karls University of Tübingen

M

Mattia Donatella	Fondazione Santa Lucia, IRCCS
Mattout Jérémie	Université de Lyon
Melinscak Filip	University of Zürich
Modini Valeria	University of Bologna
Müller-Putz Gernot R.	Graz University of Technology

O

Ortner Rupert	g.tec medical engineering Spain S.L.

P

Pereira Joana	Graz University of Technology
Poel Mannes	University of Twente
Prasad Girijesh	Ulster University

R

Ricco Angela	Fondazione Santa Lucia
Ron-Angevin Ricardo	University of Malaga
Rutkowski Tomasz M.	University of Tokyo

S

Sburlea Andreea	Graz University of Technology
Schwarz Andreas	Graz University of Technology
Si-Mohammed Hakim	INRIA
Solis-Escalante Teodoro	Delft University of Technology

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

International Program Committee and Review Board III

Sorger Bettina	Maastricht University
Steyrl David	University of Vienna

T

Tangermann Michael	University of Freiburg
Tonin Luca	Ecole Polytechnique Fédérale de Lausanne

V

Volosyak Ivan	Rhine-Waal University of Applied Sciences

W

Wagner Johanna	University of California, San Diego
Wriessnegger Selina Christin	Graz University of Technology

Z

Zhang Yu	East China University of Science and Technology

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

List of Authors I

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

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

List of Authors II

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

List of Authors III

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

List of Authors IV

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

List of Authors V

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

List of Authors VI

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-00

List of Authors VII

Published by Verlag der TU Graz
Graz University of Technology

Table of Contents I

Published by Verlag der TU Graz
Graz University of Technology

Table of Contents II

Published by Verlag der TU Graz
Graz University of Technology

Table of Contents III

Published by Verlag der TU Graz
Graz University of Technology

Table of Contents IV

Published by Verlag der TU Graz
Graz University of Technology

Table of Contents V

Published by Verlag der TU Graz
Graz University of Technology

Table of Contents VI

Published by Verlag der TU Graz
Graz University of Technology

Table of Contents VII

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-01

NIRS-BASED NEUROFEEDBACK TRAINING TO TREAT DYSPHAGIA

S.E. Kober[1,2], S. Weiss[3], P. Grieshofer[3], G. Wood[1,2]
[1] Institute of Psychology, University of Graz, Graz, Austria
[2] BioTechMed-Graz, Graz, Austria
[3] Klinik Judendorf-Straßengel, Gratwein-Straßengel, Austria

E-mail: silvia.kober@uni-graz.at

ABSTRACT: Neurological patients often show difficulties in swallowing, so-called dysphagia. Here we investigated whether motor imagery of swallowing during neurofeedback activates brain areas, which are involved in swallowing, and consequently fosters neuronal plasticity. Based on findings in healthy individuals, we performed a NIRS-based neurofeedback training (4 training sessions over 2 weeks) with one male stroke patient with dysphagia, in which the patient tried to modulate the hemodynamic response over the swallowing motor cortex by motor imagery while receiving real-time feedback of changes in the NIRS signal. Before and after the training, brain activation patterns during active swallowing were assessed offline. A healthy male control subject performed the same training. Both subjects were able to modulate the NIRS signal in a desired direction during neurofeedback training. The stroke patient also showed changes in maladaptive brain activation patterns elicited by active swallowing due to neurofeedback training. These results indicate the potential value of NIRS-based neurofeedback using motor imagery of swallowing to treat dysphagia.

INTRODUCTION

Dysphagia, which is a difficulty in swallowing, is a common consequence of brain lesions since swallowing activates a large network of cortical and subcortical brain areas [1, 2]. More than two thirds of stroke patients show dysphagia symptoms [3]. Swallowing difficulties are also prevalent in neurologic healthy elderly (13.5%-16%) [4, 5]. Dysphagia often causes chest infection and malnutrition and is associated with a slower rate of recovery, poorer rehabilitation potential and mortality [6, 7]. Natural recovery of post-stroke dysphagia can range between 6 months and 4 years after stroke onset [3, 8].
Here we investigated whether motor imagery of swallowing can be used to activate brain areas, which are involved in the swallowing process, to foster neuronal plasticity. A large portion of brain-computer interface (BCI) and neurofeedback (NF) training studies successfully showed that receiving real-time feedback about the activity in motor brain areas while imagining limb movements can be used to foster neuronal plasticity and consequently improve motor functions [9–25]. Prior near-infrared spectroscopy (NIRS) studies showed that healthy individuals as well as stroke patients generally show comparable brain activation patterns during executing and imagining swallowing movements [26–29]. Both tasks lead to the strongest NIRS signal change over the inferior frontal gyrus (IFG), which is part of the swallowing network [1, 28]. Healthy young adults are also able to modulate voluntarily the NIRS signal in a desired direction during NF training [27, 30].
The aim of the present proof of concept study was to investigate whether a stroke patient with dysphagia is also able to modulate the activity in the swallowing motor cortex when imagining swallowing movements during NIRS-based NF training. To reveal possible neuronal plasticity processes due to NF training, we examined brain activation patterns elicited during active swallowing before and after NF training, too.

MATERIALS AND METHODS

Participants: A 70-years old male stroke patient with multiple brain lesions in the right hemisphere (4 months after stroke) participated in this study during his stationary stay in the rehabilitation clinic Judendorf-Straßengel. The patient showed no psychiatric symptoms or cognitive deficits. The patient showed a moderate dysphagia at the beginning of the NF training (Bogenhauser Dysphagia Score BODS of 3) [31, 32]. During his stay in the rehabilitation clinic, he additionally received traditional logopedical treatment. Additionally, we investigated a 78-years old male subject with no neurological deficits and no dysphagia symptoms as control subject. Both participants gave written informed consent. The study was approved by the Ethics Committee of the University of Graz, Austria (reference number GZ. 39/25/63 ex 2013/14) and is in accordance with the ethical standards of the Declaration of Helsinki.

NIRS-based NF training: To assess the NIRS signal change (relative concentration changes in oxygenated oxy- and deoxygenated deoxy- hemoglobin Hb) over the bilateral IFG, a NIRSport 88 system from NIRx Medical Technologies (Glen Head, NY) consisting of 8 photo-detectors and 8 light emitters resulting in a total of 20 channels was used. The sampling rate was set to 7.81 Hz and the distance between the optodes was 3 cm. Both

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-01

subjects performed four NF training sessions within two weeks. The task was to increase deoxy-Hb over the bilateral IFG [27, 30] while receiving visual real-time feedback (color changes over the bilateral IFG area depicted on a three-dimensional head model) of relative concentration changes in deoxy-Hb over the bilateral IFG when imagining swallowing movements (imagining how it feels to swallow saliva). In each NF training session, 20 feedback trials were performed (each trial lasted 17-23 s). Pause intervals were presented between the feedback trials with a duration of 30 s.

Pre-post assessment: Before and after the NF training, brain activation patterns elicited by executing swallowing movements (swallowing saliva) were assessed offline (no real-time feedback was provided). Participants performed 10 execution trials with a duration of 15 seconds. Between the motor execution trials, a variable pause of 28-32 seconds was presented.

NIRS data analysis: Relative concentration changes in deoxy-Hb during NF training were analyzed as well as changes in oxy- and deoxy-Hb during the offline measurements (pre-post assessment while executing swallowing movements). Data preprocessing included an artifact correction (visual inspection by a trained expert in NIRS data analysis), band-pass filtering (0.01 Hz high-pass filter, 0.90 Hz low-pass filter), and baseline correction (5 s baseline interval prior to task-onset, seconds -5 to 0). The NIRS signal change was averaged task-related.

For statistical analysis of the NF training data, the 20 feedback trials per session were divided in five blocks (B1 – B5) á four trials: the NIRS signal of the first four feedback trials (trial 1-4: B1), trial number 5-8 (B2), trials 9-12 (B3), trials 13-16 (B4), and the last four trials (trial 17-20: B5) were averaged, respectively. For statistical comparisons, the NIRS signal was averaged for the time interval 5-10 s after task onset (NF task). To analyze changes in deoxy-Hb over the NF training course, regression analyses were performed (predictor variable = block number B1-B5; dependent variable= average of deoxy-Hb for second 5-10 after start of the NF task). The NIRS signal of the left IFG and right IFG was used for statistical analysis.

RESULTS

Both subjects were able to linearly increase deoxy-Hb over the right IFG, while deoxy-Hb over the left IFG did not change significantly during NF training (Fig. 1).
During the offline pre-assessment, the stroke patient showed a stronger NIRS signal change when executing swallowing movements over the right IFG (affected hemisphere) compared to the left IFG, while the healthy control subject showed a bilateral activation pattern over the IFG during the execution task (Fig. 2). After the NF training, both subjects showed a bilateral NIRS activation pattern during active swallowing (Fig. 2).

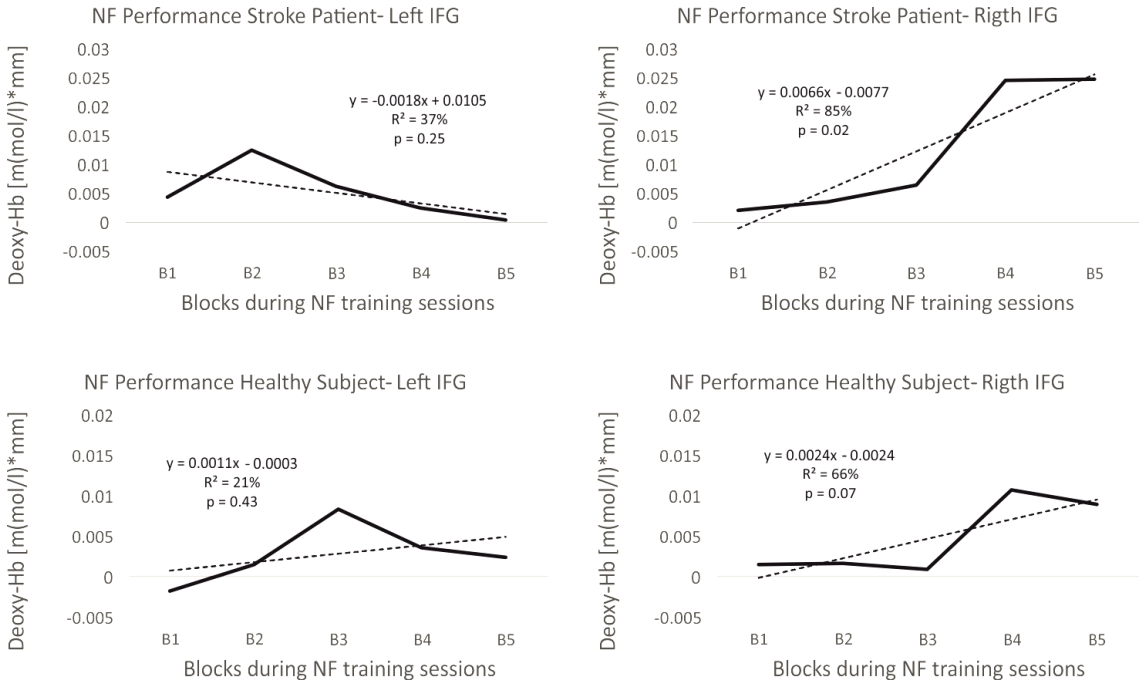


Figure 1: Neurofeedback performance of the stroke patient (upper panel) and the healthy control subject (lower panel). Depicted are changes in deoxy-Hb over the five blocks (B1-B5) within one training session averaged across all 4 NF training sessions and the results of the regression analysis.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-01

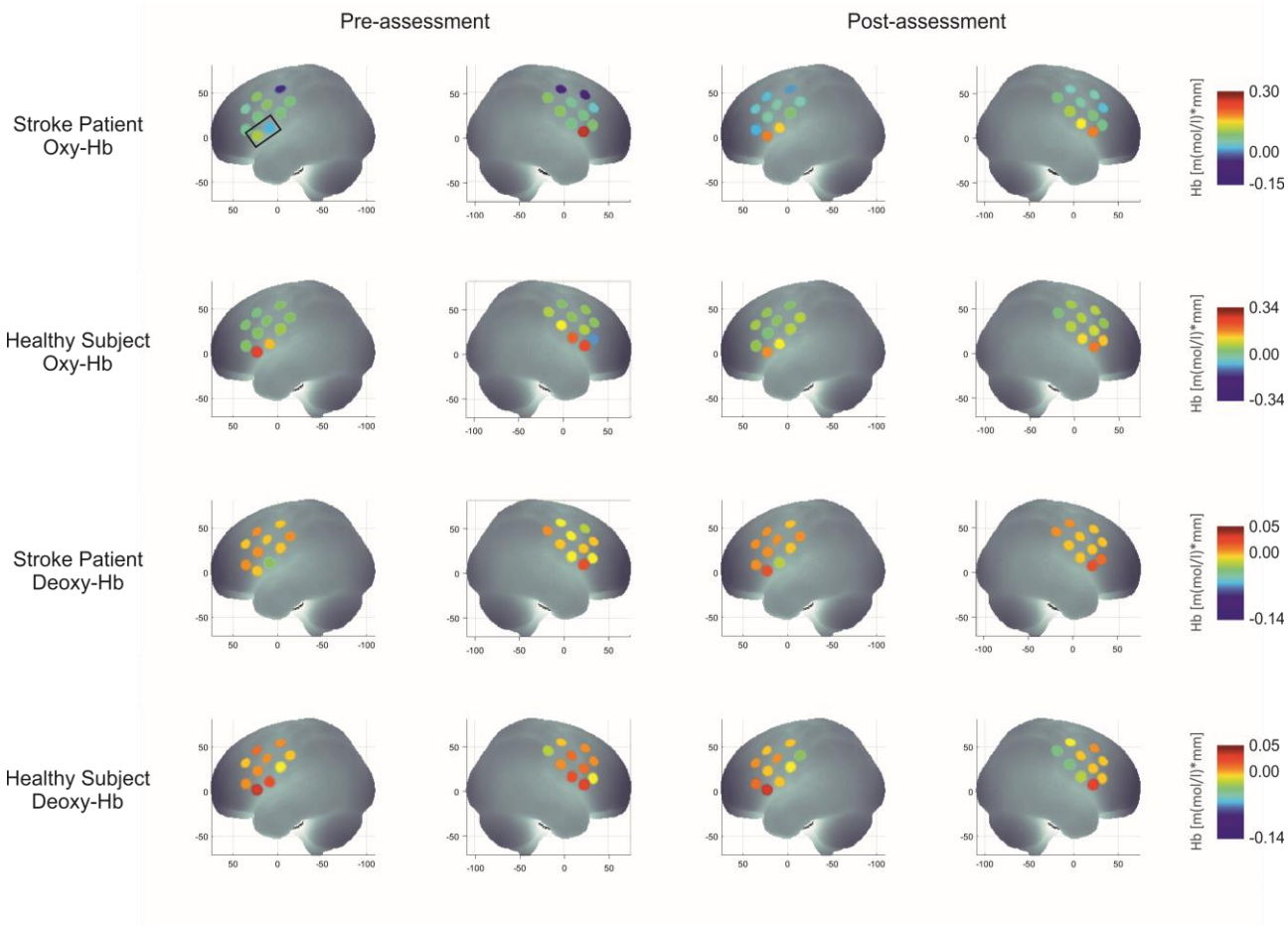


Figure 2: Activation patterns during the motor execution offline task. Oxy- and deoxy-Hb is depicted separately for the stroke patient and the healthy control subject for the 20 NIRS channels. The IFG is marked with a black rectangle in the upper left picture.

DISCUSSION

Here we showed that a stroke patient with dysphagia is able to modulate voluntarily the hemodynamic response in the swallowing motor cortex by imagining swallowing movements and that this NF training leads to neuronal plasticity processes in the swallowing motor cortex after stroke.

A stroke patient with dysphagia and a healthy control subject were able to increase deoxy-Hb linearly during NF training. This indicates that stroke patients with multiple brain lesions can learn to modulate voluntarily activity in the swallowing motor cortex when imagining swallowing movements. Our results are in line with prior findings of NIRS-based NF training studies in healthy young adults [27, 30]. Kober et al. (2018) used the same NIRS-based NF training paradigm as in the present study [30]. In this prior study, healthy young adults (mean age between 23-27 years) trained to modulate the NIRS signal in a desired direction by imagining swallowing movements. One group of healthy young adults also trained to increase deoxy-Hb over the bilateral IFG, comparable to the NF paradigm of the present study. Interestingly, the healthy young adults were also able to increase deoxy-Hb linearly over the right IFG but not over the left IFG, which is in line with the present findings in a stroke patient and a healthy elderly control subject [30]. As discussed in [30], it might be that participants concentrated more on signal changes over one hemisphere when receiving visual feedback of activation changes over the left and right IFG on a three-dimensional head model simultaneously, which might have caused the observed hemisphere differences.

When executing swallowing movements, the stroke patient showed a more unilateral brain activation pattern (stronger activation of the affected right hemisphere compared to the left hemisphere) while the healthy control subject showed a more bilateral distribution of brain activity before the start of the NF training. This is in line with prior findings that dysphagia patients often show a stronger activation of the affected hemisphere during swallowing, while healthy individuals show a more bilateral activation [3, 26–28, 33–36]. After NF training, the stroke patient showed a bilateral activation of the IFG during active swallowing. This activation pattern was comparable to the brain activation pattern of the healthy control subject [26, 28]. Hamdy et al. (1996) also found that the cortical activation of dysphagia

Published by Verlag der TU Graz
Graz University of Technology

patients becomes more bilaterally distributed with recovery of swallowing [36]. Hence, brain activation patterns over swallowing motor areas elicited by executing swallowing movements seem to "normalize" in the stroke patient with dysphagia after NF training.

CONCLUSION

This first proof of concept study shows that a stroke patient with dysphagia can benefit from NIRS-based NF training in which motor imagery of swallowing movements is used as mental strategy to activate the swallowing motor cortex. There is evidence that external stimulation or inhibition of the swallowing motor cortex using repetitive transcranial magnetic stimulation (rTMS) leads to recovered swallowing function in dysphagia patients [3, 35, 37–39]. With NIRS-based NF, dysphagia patients might learn to increase or decrease voluntarily the activation level in specific swallowing related brain areas, without the need of external stimulation such as rTMS [28].
Our results indicate that future NF training studies with larger samples of dysphagia patients might reveal the usefulness of NF training in dysphagia rehabilitation.

REFERENCES

[1] Sörös P, Inamoto Y, Martin RE. Functional brain imaging of swallowing: An activation likelihood estimation meta-analysis. Hum. Brain Mapp. 2009;30:2426–39. doi:10.1002/hbm.20680.

[2] Humbert IA, Robbins J. Normal Swallowing and Functional Magnetic Resonance Imaging: A Systematic Review. Dysphagia. 2007;22:266–75. doi:10.1007/s00455-007-9080-9.

[3] Khedr EM, Abo-Elfetoh N. Noninvasive Brain Stimulation for Treatment of Post-Stroke Dysphagia. Neuroenterology. 2013;2:1–9. doi:10.4303/ne/235663.

[4] Eslick GD, Talley NJ. Dysphagia: epidemiology, risk factors and impact on quality of life--a population-based study. Aliment Pharmacol Ther. 2008;27:971–9. doi:10.1111/j.1365-2036.2008.03664.x.

[5] Ashford J, McCabe D, Wheeler-Hegland K, Frymark T, Mullen R, Musson N, et al. Evidence-based systematic review: Oropharyngeal dysphagia behavioral treatments. Part III--impact of dysphagia treatments on populations with neurological disorders. J Rehabil Res Dev. 2009;46:195–204.

[6] Haider C, Zauner H, Gehringer-Manakamatas N, Kadar K, Wood G., Wallner K, Gassner A. Neurogenic dysphagia: Nutrition therapy improves rehabilitation. Journal für Ernährungsmedizin. 2008;10:6–11.

[7] Bours GJJW, Speyer R, Lemmens J, Limburg M, Wit R de. Bedside screening tests vs. videofluoroscopy or fibreoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: systematic review. Journal of Advanced Nursing. 2009;65:477–93. doi:10.1111/j.1365-2648.2008.04915.x.

[8] Bath PMW, Bath FJ, Smithard DG. Interventions for dysphagia in acute stroke. The Cochrane Library. 2002;1.

[9] Mihara M, Hattori N, Hatakenaka M, Yagura H, Kawano T, Hino T, Miyai I. Near-infrared Spectroscopy-mediated Neurofeedback Enhances Efficacy of Motor Imagery-based Training in Poststroke Victims: A Pilot Study. Stroke. 2013;44:1091–8. doi:10.1161/STROKEAHA.111.674507.

[10] 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;104:774–83. doi:10.1152/jn.00257.2010.

[11] Yágüez L, Nagel D, Hoffmann H, Canavan AGM, Wist E, Hömberg V. A mental route to motor learning: Improving trajectorial kinematics through imagery training. Behavioural Brain Research. 1998;90:95–106.

[12] Mulder T, Zijlstra S, Zijlstra W, Hochstenbach J. The role of motor imagery in learning a totally novel movement. Experimental Brain Research. 2004;154:211–7. doi:10.1007/s00221-003-1647-6.

[13] Liew S-L, Rana M, Cornelsen S, Fortunato de Barros Filho M, Birbaumer N, Sitaram R, et al. Improving Motor Corticothalamic Communication After Stroke Using Real-Time fMRI Connectivity-Based Neurofeedback. Neurorehabil Neural Repair. 2016;30:671–5. doi:10.1177/1545968315619699.

[14] Goebel R, Zilverstand A, Sorger B. Real-time fMRI-based brain–computer interfacing for neurofeedback therapy and compensation of lost motor functions. Imaging in Medicine. 2010;2:407–15. doi:10.2217/iim.10.35.

[15] Watanabe T, Sasaki Y, Shibata K, Kawato M. Advances in fMRI Real-Time Neurofeedback. Trends Cogn Sci (Regul Ed). 2017;21:997–1010. doi:10.1016/j.tics.2017.09.010.

[16] Sitaram R, Caria A, Veit R, Gaber T, Rota G, Kuebler A, Birbaumer N. fMRI Brain-Computer Interface: A Tool for Neuroscientific Research and Treatment. Computational Intelligence and Neuroscience. 2007;2007:1–10. doi:10.1155/2007/25487.

[17] Sitaram R, Ros T, Stoeckel L, Haller S, Scharnowski F, Lewis-Peacock J, et al. Closed-loop brain training: The science of neurofeedback. Nat Rev Neurosci. 2017;18:86–100. doi:10.1038/nrn.2016.164.

[18] Linden DEJ, Turner DL. Real-time functional magnetic resonance imaging neurofeedback in motor neurorehabilitation. Curr Opin Neurol. 2016;29:412–8.

[19] Kober SE, Wood G, Kurzmann J, Friedrich EVC, Stangl M, Wippel T, et al. Near-infrared

Published by Verlag der TU Graz
Graz University of Technology

spectroscopy based neurofeedback training increases specific motor imagery related cortical activation compared to sham feedback. Biological Psychology. 2014;95:21–30. doi:10.1016/j.biopsycho.2013.05.005.

[20] Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans. Neuroscience Letters. 1997;239:65–8. doi:10.1016/S0304-3940(97)00889-6.

[21] Weiskopf N. Real-time fMRI and its application to neurofeedback. NeuroImage. 2012;62:682–92. doi:10.1016/j.neuroimage.2011.10.009.

[22] Wriessnegger SC, Kurzmann J, Neuper C. Spatio-temporal differences in brain oxygenation between movement execution and imagery: A multichannel near-infrared spectroscopy study. International Journal of Psychophysiology. 2008;67:54–63. doi:10.1016/j.ijpsycho.2007.10.004.

[23] Faralli A, Bigoni M, Mauro A, Rossi F, Carulli D. Noninvasive Strategies to Promote Functional Recovery after Stroke. Neural Plasticity. 2013;2013:1–16. doi:10.1155/2013/854597.

[24] Hardwick RM, Caspers S, Eickhoff SB, Swinnen SP. Neural Correlates of Motor Imagery, Action Observation, and Movement Execution: A Comparison Across Quantitative Meta-Analyses. bioRxiv. 2017:1–50. doi:10.1101/198432.

[25] Lorey B, Naumann T, Pilgramm S, Petermann C, Bischoff M, Zentgraf K, et al. How equivalent are the action execution, imagery, and observation of intransitive movements? Revisiting the concept of somatotopy during action simulation. Brain Cogn. 2013;81:139–50. doi:10.1016/j.bandc.2012.09.011.

[26] Kober SE, Bauernfeind G, Woller C, Sampl M, Grieshofer P, Neuper C, Wood G. Hemodynamic Signal Changes Accompanying Execution and Imagery of Swallowing in Patients with Dysphagia: A Multiple Single-Case Near-Infrared Spectroscopy Study. Front. Neurol. 2015;6:1–10. doi:10.3389/fneur.2015.00151.

[27] Kober SE, Gressenberger B, Kurzmann J, Neuper C, Wood G. Voluntary modulation of hemodynamic responses in swallowing related motor areas: A near-infrared spectroscopy based neurofeedback study. PLoS ONE. 2015;10:1–17.

[28] Kober SE, Wood G. Changes in hemodynamic signals accompanying motor imagery and motor execution of swallowing: A near-infrared spectroscopy study. NeuroImage. 2014;93:1–10. doi:10.1016/j.neuroimage.2014.02.019.

[29] Kober SE, Wood G. How to Exercise by Imagining Movements. Front. Young Minds. 2017;5:247. doi:10.3389/frym.2017.00042.

[30] Kober SE, Hinterleitner V, Bauernfeind G, Neuper C, Wood G. Trainability of hemodynamic parameters: A near-infrared spectroscopy based neurofeedback study. Biol Psychol. 2018;136:168–80. doi:10.1016/j.biopsycho.2018.05.009.

[31] Bartolome G, Schröter-Morasch H. Bogenhausener Untersuchungsprotokoll für die Klinische Schluckuntersuchung (KSU). In: Bartolome G, Schröter-Morasch H, editors. Schluckstörungen – Diagnostik und Rehabilitation. 5th ed. München: Elsevier; 2014.

[32] Bartolome G, Schröter-Morasch H, editors. Schluckstörungen – Diagnostik und Rehabilitation. 5th ed. München: Elsevier; 2014.

[33] Kober SE, Wood G. Hemodynamic signal changes during saliva and water swallowing: a near-infrared spectroscopy study. 2018;23 IS -:015009-23-7.

[34] Ertekin C, Aydogdu I. Neurophysiology of swallowing. Clinical Neurophysiology. 2003;114:2226–44. doi:10.1016/S1388-2457(03)00237-2.

[35] Khedr EM, Abo-Elfetoh N, Rothwell JC. Treatment of post-stroke dysphagia with repetitive transcranial magnetic stimulation. Acta Neurologica Scandinavica. 2009;119:155–61. doi:10.1111/j.1600-0404.2008.01093.x.

[36] Hamdy S, Aziz Q, Rothwell JC, Singh KD, Barlow J, Hughes DG, et al. The cortical topography of human swallowing musculature in health and disease. Nature Med. 1996;2:1217–24.

[37] Verin E, Leroi AM. Poststroke Dysphagia Rehabilitation by Repetitive Transcranial Magnetic Stimulation: A Noncontrolled Pilot Study. Dysphagia. 2009;24:204–10. doi:10.1007/s00455-008-9195-7.

[38] Michou E, Raginis-Zborowska A, Watanabe M, Lodhi T, Hamdy S. Repetitive Transcranial Magnetic Stimulation: A Novel Approach for Treating Oropharyngeal Dysphagia. Curr Gastroenterol Rep. 2016;18:10. doi:10.1007/s11894-015-0483-8.

[39] Cabib C, Ortega O, Kumru H, Palomeras E, Vilardell N, Alvarez-Berdugo D, et al. Neurorehabilitation strategies for poststroke oropharyngeal dysphagia: From compensation to the recovery of swallowing function. Ann N Y Acad Sci. 2016;1380:121–38. doi:10.1111/nyas.13135.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-02

THE RIEMANNIAN MINIMUM DISTANCE TO MEANS FIELD CLASSIFIER

M Congedo, P.L.C. Rodrigues, C. Jutten

GIPSA-lab, Université Grenoble Alpes, CNRS, Grenoble-INP

E-mail: Marco.Congedo@Gipsa-Lab.fr

ABSTRACT: A substantial amount of research has demonstrated the accuracy of the Riemannian minimum distance to mean (RMDM) classifier for brain-computer interface (BCI). This classifier is simple, fully deterministic, robust to noise, computationally efficient and prone to transfer learning. The use of the geometric mean in the Riemannian manifold of symmetric positive definite matrices has proved fundamental to obtain these characteristics. Recently the general family of power means living on this manifold, which includes the geometric mean, has been defined. In this article we extend the RMDM algorithm in an unsupervised and adaptive fashion using a sampling of power means, named *means field*. We show that the resulting Riemannian minimum distance to means field (RMDMF) classifier features superior performance. Our conclusion is supported by the analysis of 17 public databases covering two BCI paradigms, for a total of 335 individuals, using the open-source MOABB (Mother of all BCI Benchmark) framework. In order to promote reproducible research, our full code is released.

INTRODUCTION

Riemannian geometry is a branch of differential geometry that studies smooth manifolds, curved spaces with peculiar geometries. In these spaces notions of angles, shortest path between two points, distances, center of mass of several points, etc., allow to study analytic properties of mathematical operators from a geometric perspectives, making them accessible to intuition [1]. In the field of brain-computer interface (BCI) the manifold of symmetric positive-definite (SPD) matrices [2] has proved very useful, since multivariate electroencephalography (EEG) data in finite time windows can effectively be mapped as points onto this manifold through the estimation of some form of their covariance matrix [3-6]. This approach has led to the introduction of classifiers with remarkable characteristics as compared to the state-of-the-art [6]; for a formal introduction to the SPD Riemannian manifold the reader is referred to [2], while for a primer and review of its use in the BCI context to [7,8].

Riemannian classifiers have proved accurate, general and robust to noise, largely superior to state-of-the-art competitors [9], winning five international BCI machine learning competitions in which they have competed [7]. In particular, the Riemannian minimum distance to mean (RMDM) classifier, while not the most accurate among Riemannian classifiers, stands out for its simplicity, computational efficiency and universality (it applies to all BCI paradigms). It is a fully deterministic and parameter-free classifier, thus no parameter needs to be tuned by cross-validation or other methods that may jeopardize its generalization. Further, it can be easily extended to the multi-user scenario [10], has proved accurate also in individuals affected by clinical conditions [11] and has proved apt to integrate transfer learning [12, 13] and adaptation strategies [14], as demonstrated by the calibration-less P300-based BCI video game *Brain Invaders* [14-16].

In its non-adaptive (test-training) form, the RMDM works as it follows: a training provides a set of SPD matrices encoding BCI trials for the available classes. For each class a center of mass of the available trials is estimated. Finally, in test mode, a BCI trial to be classified is encoded in the same way as an SPD matrix and is assigned to the class whose center of mass is the closest according to a suitable distance function acting on the manifold [7]. In adaptive mode instead, the centers of mass are initialized by a database of previous users for a naive user and/or a database of the same user for a non-naive user thanks to transfer learning strategies [12, 13], then the centers of mass are adapted to the user while the BCI is operated [14]. In any case, the good performance of the RMDM classifier derives from the adoption of an appropriate metric for the SPD manifold. The metric in turn determines both the distance function between two points and the definition of a center of mass for a cloud of points, which is also a function of the distance since it is defined as the point on the manifold minimizing the dispersion of the cloud around itself [7, 17]. So far, the hyperbolic (geometric) distance and the geometric mean as a center of mass, which arise adopting the Fisher-Rao (affine-invariant) metric, have been preferred, due to a number of desirable invariance properties they possess. As explained in [7], those are the extension to SPD matrices of the usual hyperbolic distance and geometric mean for scalars. Simply stated, their use instead of the much more common Euclidean distance and arithmetic mean has engendered the success of Riemannian classifiers dealing with covariance matrices [7, 8].

In [18] the authors have defined a one-parameter family of means generalizing to SPD matrices the power means

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-02

for scalars. Given a set of K scalars $\{c_1,\ldots,c_K\}$, the power mean with real parameter $h \neq 0$ is given by

$$g_{(h)} = \left(\frac{1}{K}\sum_k c_k^h\right)^{1/h}. \quad (1)$$

As for the scalar power means, the SPD matrix power means of [18] (see also [19-21]) interpolate between the harmonic mean $(h=-1)$ and the arithmetic mean $(h=1)$, while the geometric mean that we have discussed corresponds to the limit of h evaluated at 0, from either side. This generality of power means is appealing in the BCI context; as suggested in [22], in EEG data the sensor measurement is affected by several noise components and varying the order h one can find an optimal mean depending on the signal-to-noise-ratio. In [22] we have tested the accuracy of the RMDM algorithm using 13 power means with h={±1, ±0.8, ±0.6, ±0.4, ±0.2, ±0.1, 0} (see Fig. 7 therein). The classical RMDM classifier corresponds to the power mean with h=0 (geometric mean). We have found that the value of h offering the maximum accuracy gravitated around zero, but h=0 was optimal only for three out of the 38 tested subjects. Instead, the optimal value of h was highly variable across individuals. Also, there was a significant positive correlation between the maximal accuracy and the value of h allowing such maximum. Thus, the higher the accuracy, which is an indirect measure of signal-to-noise ratio, the higher the optimal value of h.

Finding the optimal value of h for a given subject and session, as we have done in [22], is a supervised procedure. Therefore in seeking the optimal value we roll back to the problem of obtaining a classifier that is prone to overfitting, that lacks transfer learning and that is not capable of adaptation. Let us instead name a *means field* a sampling of power means in the interval $h\in$[-1, 1] such as the one used in [22]. Then, in this article we propose the Riemannian minimum distance to means field (RMDMF) classifier. It uses in an unsupervised and adaptive fashion all the means in the field for classifying. In particular, for a given unlabeled datum, the closest power mean, regardless of its class, is found, then the MDM is applied using the power means with the value of h that corresponds to the closest mean. Such method is unsupervised, in that it can be used blindly to any datum without any learning and is adaptive, in that the preferred value of h is allowed to change during the session.

We employ MOABB (Mother of All BCI Benchmark) [9] for testing the RMDMF against the RMDM classifier on 17 databases covering two BCI paradigms (motor imagery and P300), for a total of 335 individuals. MOABB is an open-source framework for objectively assessing the performance of BCI classifiers on large amount of data. The use of MOABB ensures that exactly the same processing pipeline is applied to all databases of the same type and that both the cross-validation procedure and the Riemannian classifiers operate exactly in the same way for all databases, regardless the BCI type.

MATERIALS AND METHODS

Table 1 reports the main characteristics of the 17 databases we have used for testing. 12 concerns a motor imagery (MI) BCI, five concerns a P300 BCI. For some databases several sessions are available, therefore the actual number of EEG recordings analyzed is superior to the total number of subjects. Also, one may notice that the number of electrodes used in the experiments is highly variable, ranging from three to 128.

Table 1: Main characteristics of the databases used for the analysis. Legend: Ch.=number of channels; Sess=number of sessions; Ss=number of subjects. For extended names of the databases see Table 2.

Name	*Type*	*Ch*	*Trials*	*Sess*	*Ss*
Zhou 2016	MI	14	100	3	4
BNCI 2014-009	P300	18	4200	1	10
BNCI 2015-001	MI	13	200	2 or 3	13
BNCI 2014-002	MI	15	160	1	14
BNCI 2015-003	P300	10	5400	1	10
BNCI 2014-004	MI	3	120-160	5	9
BNCI 2015-004	MI	30	70-80	2	10
BNCI 2014-008	P300	10	4200	1	8
Alexandre MI	MI	16	40	1	9
Weibo 2014	MI	60	160	1	10
Brain Inv 2013a	P300	16	480	1 or 8	24
Cho 2017	MI	64	200	1	49
EPFL P300	P300	32	800	4	8
GW 2009	MI	128	300	1	10
Physionet MI	MI	64	40-60	1	109
Shin 2017a	MI	25	60	3	29
BNCI 2014-001	MI	22	144	2	9

The pipeline for MI databases included: filtering in the 8-32Hz band-pass region, computing the sample covariance matrix for all trials and evaluating the classifiers using (5-fold) cross-validation [3]. The pipeline for P300 included filtering in the 1-24Hz band-pass region and then, during (5-fold) cross-validation, estimating on the training set a spatial filter specifically conceived to enhance the signal-to-noise ratio of event-related potentials (ERPs) [23, 24] retaining the best eight discriminative components, filtering all the trials in the training and test set using this filter, computing the 16x16 extended sample covariance matrix used for ERP data [3, 14] on all trials and finally evaluating the classifiers.

Using these two pipelines we have run statistical tests in MOABB to compare the ROC-AUC classification accuracy of the RMDMF classifier vs. the RMDM classifier for all databases. The power means for the RMDMF classifier where computed for h={±1, ±0.75, ±0.5, ±0.25, ±0.1, ±0.01}. If several sessions for the same subject were available, the ROC-AUC score was averaged across-sessions to provide a unique score for each subject. For each database a paired permutation one-sided t-test has been carried out, enumerating all raw data permutations if the number of subjects was <20, yielding in this case an exact test [25,26], otherwise employing the Wilcoxon signed-rank test, which basically is equivalent to a permutation test performed on the ranked

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-02

data. The *p*-values thus obtained for each database have been combined using the weighted *Liptak combination function* [26] (also known as Stouffer's combination function when expressed in terms of standard normal variables), which is given by

$$p = 1 - \Phi\left(\frac{\sum_{i=1}^{n} w_i \ \Phi^{-1}(1-p_i)}{\sum_{i=1}^{n} w_i^2}\right), \qquad (2)$$

where Φ is the standard normal cumulative distribution function, Φ^{-1} its inverse function, n is the number of databases (17 in our case), p_i the *p*-value observed for the i^{th} database and w_i weights taken as the square root of the number of subjects in each database. This returned a single *p*-value for the global one-sided comparison RMDMF vs. RMDM. Such *p*-value is to be interpreted as *the probability to observe n p-values under the Omnibus hypothesis*, i.e., given that the null hypothesis is true for all of them. Notice that the Liptak combination function (2) assumes that the *p*-values to be combined result from all pair-wise independent hypotheses, which in our case is verified since all the databases are independent. Notice also that this combination function is optimal (most powerful) when all tests have the same effect size, which, although rarely verified in practice, is a desirable property. The effect sizes were also determined, akin to meta-analysis studies, computing the *standardized mean difference* (SMD) for each database and combining them by the weighted arithmetic average using the same weighs as those used for Liptak's *p*-value combination method. All statistical analysis tools here above described are already embedded in MOABB [9].

RESULTS

Table 2 reports the *p*-values and SMDs obtained on the 17 databases along with the combined *p*-value and the average SMD. Figure 1 depicts the SMDs and their 95% confidence interval obtained on the 17 databases. One can see that for five databases the RMDMF classifier significantly outperforms the RMDM classifier. Among the remaining 12 databases, in nine of them RMDMF tends to perform better than RMDM (as seen by the positive SMD or, equivalently, by a *p*-value smaller than 0.5), while the opposite happens in three. There is no evidence that the RMDM significantly outperforms RMDMF for any databases. The weighted average SMD was 0.3 and the weighted combined *p*-value was extremely low (p=0,0000377), allowing a firm rejection of the Omnibus hypothesis. Notice finally that, as expected, the confidence interval for the SMDs tends to be inversely proportional to the sample size.

DISCUSSION

Using MOABB we have presented results on 17 databases for a total of 335 individuals. Those results are therefore solid and powerful. For doing this we have added several databases to MOABB, including one of our own database on the P300 video-game *Brain Invaders* [14-16] and we will continue this effort. MOABB is an ideal framework for testing classifiers objectively and we invite the community to contribute to it in terms of development and data. More in general, we urge the BCI community to promote publications on machine learning for BCI where either real on-line accuracy is reported or the classifiers are tested on large and diverse data. The Python code of this analysis and our data are available along with many other contributions[1].

Table 2: *p*-values and standardized mean differences (SMD) obtained on the 17 databases along with the combined *p*-value and SMD. Small p-values (underlined for p<0.05) indicate that the accuracy is significantly higher for the RMDMF as compared to the RMDM.

Name	***Type***	***Ss***	***p***	***SMD***
Zhou 2016	MI	4	0,188	0,679
BNCI 2014-009	P300	10	0,271	0,201
BNCI 2015-001	MI	13	0,004	0,964
BNCI 2014-002	MI	14	0,041	0,499
BNCI 2015-003	P300	10	0,374	0,136
BNCI 2014-004	MI	9	0,648	-0,120
BNCI 2015-004	MI	10	0,322	0,150
BNCI 2014-008	P300	8	0,078	0,659
Alexandre MI	MI	9	0,285	0,182
Weibo 2014	MI	10	0,023	0,711
Brain Invaders 2013a	P300	24	0,744	-0,132
Cho 2017	MI	49	0,085	0,259
EPFL P300	P300	8	0,836	-0,372
Grosse-Wentrup 2009	MI	10	0,011	0,687
Physionet MI	MI	109	0,109	0,090
Shin 2017a	MI	29	0,052	0,304
BNCI 2014-001	MI	9	0,010	0,998
Combination			***p=0,0000377***	***0,302***

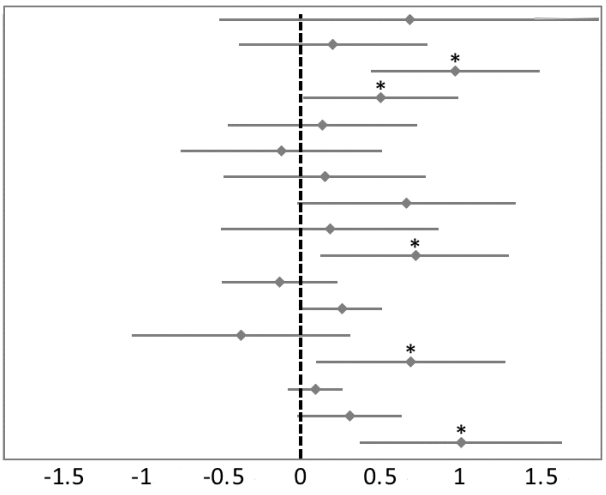


Figure 1: Standardized mean differences (diamond) and their 95% confidence interval (horizontal lines) for the 17 databases (from top to bottom in the same order as in table 1 and 2). A positive SMD value indicate that the accuracy of the RMDMF classifier is higher as compared to the RMDM, the opposite for a negative SMD value. *= significant *p*-value (see table 2).

[1] https://sites.google.com/site/marcocongedo/science

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-02

CONCLUSION

In this article we have proposed an improvement to the Riemannian minimum distance to mean classifier for BCIs, leveraging on recent advances on mathematics (the definition of power means for symmetric positive definite matrices [18-21]) and signal processing (an efficient algorithm for estimating them [22]). Comparing the proposed RMDMF to the RMDM classifier on 17 databases using MOABB has yielded an average SMD equal to 0.3 and a combined *p*-value equal to 0,0000377. As it can been seen in Table 2, the effect is driven by MI databases, whereas for no P300 database the effect is significant. We believe this is due to the fact that in our pipeline for P300 an optimal spatial filter has been estimated during cross-validation, thus the signal-to-noise ratio for these data has been optimized; under these circumstances the benefit of using a means field populated by power means is likely lost. If this is confirmed by further analysis, the RMDMF classifier would prove better apt for working on raw data as compared to the RMDM, but of course it can be used also on spatially filtered data without losing performance, as our analysis suggests. Then the RMDMF, while preserving the other desirable properties of the RMDM (simplicity, computational efficiency, universality, ease of extension to the multi-user scenario, good attitude for transfer learning and for adaptation), truly would support parameter-free classification pipelines.

The computational cost of the RMDMF as compared to the RMDM is increased proportionally to the number of means used to populate the means field. Since the computational complexity of Riemannian classifiers is cubic on the size of the covariance matrices used to encode the EEG data, this does not represent a substantial additional cost. When the dimension of the covariance matrices is high, it can be reduced by well-known methods such as principal component analysis or by methods inspired by Riemannian geometry [27, 28]. As reported in [6], the performance of the MDM drops in high dimension, therefore in these situations a dimensionality reduction step in practice is necessary. The RMDMF may turn more robust than the RMDM also with respect to the dimensionality. Further research is necessary to verify these hypotheses.

For defining the RMDMF we have found effective a ‘closest mean’ approach. However, better strategies may exist to exploit the richness of the means field, such as, for example, pooling and majority voting. Further research is therefore needed to find optimal strategies for exploiting the Riemannian means fields. Not only the strategy can be optimized, but also the definition of the means field itself: while in this work we have populated the means field with power means, other means not belonging to this family may be added to the field, such as the log-Euclidean mean and a sample of the α-divergence means, to which the Bhattacharyya mean belongs [29], making the means field even more rich.

REFERENCES

[1] Levi-Civita T. Lezioni di Calcolo Differenziale Assoluto, Alberto Stock, Roma (1925)
[2] Bhatia R. Positive Definite Matrices, Princeton University Press, New Jersey (2007)
[3] Barachant A, Bonnet S, Congedo M, Jutten C. Multi-class Brain Computer Interface Classification by Riemannian Geometry. IEEE Trans. Biomed. Eng. 2012; 59(4):920-928
[4] Barachant A, Bonnet S, Congedo M, Jutten C. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. Neurocomputing. 2013;112:172-178
[5] Congedo M. EEG Source Analysis, Habilitation à Diriger des Recherches, University Grenoble Alpes. (2013)
[6] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, et al. A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update. J. Neural Eng. 2018; 15(3): 031005
[7] Congedo M, Barachant A, Bhatia R. Riemannian Geometry for EEG-based Brain-Computer Interfaces; a Primer and a Review. Brain-Computer Interfaces 2017; 4(3):155-174
[8] Yger F, Berar M, Lotte F. Riemannian approaches in brain-computer interfaces: a review. IEEE Trans. Neural Syst. Rehabil. Eng. 2017; 25(10):1753-1762
[9] Jayaram V, Barachant A. MOABB: trustworthy algorithm benchmarking for BCIs. J. Neural. Eng. 2018;15(6):066011
[10] Korczowski L, Congedo M, Jutten C. Single-Trial Classification of Multi-User P300-Based Brain-Computer Interface Using Riemannian Geometry. In Proc. IEEE EMBS, Milano, Italy, Aug 2015, 1769-72
[11] Mayaud L, Cabanilles S, Van Langhenhove A, Congedo M, Barachant A, Pouplin S, et al. Brain-computer interface for the communication of acute patients: a feasibility study and a randomized controlled trial comparing performance with healthy participants and a traditional assistive device. Brain-Computer Interfaces 2016; 3(4):197-215
[12] Zanini P, Congedo M, Jutten C, Said S, Berthoumieu Y. Transfer Learning: a Riemannian geometry framework with applications to Brain-Computer Interfaces. IEEE Trans. Biomed. Eng. 2018; 65(5):1107-1116
[13] Rodrigues PLC, Jutten C, Congedo M. Riemannian Procrustes Analysis: Transfer Learning for Brain-Computer Interfaces. IEEE Trans. Biomed. Eng. 2019; in press
[14] Barachant A, Congedo M. A Plug&Play P300 BCI Using Information Geometry, arXiv, 2014; 1409.0107
[15] Congedo M, Goyat M, Tarrin N, Varnet L, Rivet B, Ionescu G, et al. “Brain Invaders”: a prototype of an open-source P300-based video game working with

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-02

the OpenViBE platform. In Proc. Int. BCI Conf., Graz, Austria, 2011, 280-283
[16] Andreev A, Barachant A, Lotte F, Congedo M. Recreational Applications in OpenViBE: Brain Invaders and Use-the-Force. In: Clerc M, Bougrain L, Lotte F (Ed.). Brain-Computer Interfaces 2: Technology and Application. Wiley-iSTE, New York 2016, pp. 241-258
[17] Moakher M. A differential geometric approach to the arithmetic and geometric means of operators in some symmetric spaces. SIAM J. Matrix Anal. Appl. 2005; 26(3):735-747
[18] Lim Y, Pálfia M. Matrix Power means and the Karcher mean. J. Funct. Anal. 2012; 262:1498-1514
[19] Pálfia M. Operator means of probability measures and generalized Karcher equations. Adv. Math. 2016; 289:951-1007
[20] Lawson J, Lim Y. Weighted means and Karcher equations of positive operators, PNAS 2013; 110(39):15626-32
[21] Lawson J, Lim Y. Karcher means and Karcher equations of positive definite operators. Trans. Am. Math. Soc. 2014; 1:1-22
[22] Congedo M, Barachant A, Kharati Koopaei E. Fixed Point Algorithms for Estimating Power Means of Positive Definite Matrices. IEEE Trans. Sig. Process. 2017; 65(9):2211-2220
[23] Rivet B, Souloumiac A, Attina V, Gibert G. xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. IEEE Trans. Biomed. Eng. 2009; 56(8): 2035-2043
[24] Congedo M, Korczowski L, Delorme A, Lopes da Silva F. Spatio-temporal common pattern: A companion method for ERP analysis in the time domain. J. Neurosci. Methods. 2016; 267: 74-88
[25] Edgington ES. Randomization tests, 3rd ed., M. Dekker, New York, USA (1995)
[26] Pesarin F. Multivariate Permutation Tests with applications in Biostatistics, John Wiley and Sons, New York, USA (2001)
[27] Rodrigues PLC, Bouchard F, Congedo M, Jutten C. Dimensionality Reduction for BCI classification using Riemannian geometry. In Proc. Int. BCI Conf., Graz, Austria, Sept. 2017, 80-85
[28] Congedo M, Rodrigues PLC, Bouchard F, Barachant A, Jutten C. A Closed-Form Unsupervised Geometry-Aware Dimensionality Reduction Method in the Riemannian Manifold of SPD Matrices. In Proc. Int. EMBS Conf., Jeju Island, South Korea, July 2017, 3198-3201
[29] Congedo M, Afsari B, Barachant A, Moakher M. Approximate Joint Diagonalization and Geometric Mean of Symmetric Positive Definite Matrices. PLoS One. 2015; 10(4): e0121423

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-03

DECODING SSVEP ON TIME AND FREQUENCY DOMAIN USING CONVOLUTIONAL NEURAL NETWORK

R. Gala[1], J. Xu[#1,2], Z.Wu[1,2], A. Nürnberger[1]

[1]Data and Knowledge Engineering Group, Faculty of Computer Science, Otto von Guericke University, Magdeburg, Germany
[2]Institut für Medizinische Psychologie, Otto von Guericke University, Magdeburg, Germany

E-mail:jiahua.xu@med.ovgu.de

ABSTRACT: A brain-computer interface (BCI) is an highly interdisciplinary research topic which involved psychology, signal processing, machine learning, medical engineering etc. A steady-state visually evoked potentials (SSVEP) paradigm-based Brain Computer Interface systems are the most promising and reliable communication systems for people with disabilities on clinical purposes. This paper proposes the use of Deep Neural Network models for decoding the data acquired through visual-based Electroencephalography (EEG). in this paper, we concentrate on two points with the Convolutional Neural Network The first is decoding of the EEG data using the raw signal (time domain) and the extracted frequency features of the signal after Fourier analysis. The second point explored in this paper is the evaluation of the selection of electrodes on the performance of the Deep Neural Network models for raw signals (time domain data). We have also compared the performance of Canonical Correlation analysis of the data with the Deep Neural Network models.

INTRODUCTION

Brain-Computer Interface (BCI) systems provide direct communication between the human brain and an external device. BCI systems are applicable in different areas such as spellers [1], for operating wheelchairs [2], computer games, virtual reality and controlling home environment [3]. BCI systems make use of different types of signals comprising of slow cortical potentials, P300, sensorimotor rhythms and SSVEPs [1]. SSVEPs are recorded over the scalp from repetitive external stimulations. SSVEPs are of the same frequency as of the stimulus. In other words, looking at an image of frequency 12Hz with 0° phase angle will generate a signal of the same value and angle in the occipital region of the brain. High Signal-To-Noise (SNR) ratio, high Information Transfer Rate (ITR), little user training and less susceptibility to eye movements are the advantages leading to the usage of SSVEPs in BCIs over other types of signals [4].

Essentially, a visual BCI has four sections, signal acquisition, signal processing, signal classification, and output device which realizes the selected commands by the user of BCI indicated by Figure 1. The principal task in this process is signal classification with reliable performance in regards to accuracy and response time. A number of methods for detection of SSVEPs have been used such as Minimum-Energy Combination (MEC), Canonical Correlation Analysis (CCA), Support Vector Machine (SVM) and Power Spectral Density Analysis (PSDA). These methods would require the data to be refined by human intervention. The use of Deep Neural Networks in BCIs has gained popularity due to its robust nature and the capability of extracting high-level features from the data without prior knowledge [7][11]. Use of Convolutional Neural Networks (CNN) for image classification and recognition provides high accuracy, automatic feature extraction with the multiple convolutional layers thus making it the perfect choice for EEG data. The EEG data received from a variety of participants having different features makes CNN an apt solution for classification of the signals. Despite of high performance of CNN, it has a necessity to have a large training dataset so that no or less over-fitting occurs in the model, which is a challenge in the EEG data, given it's relatively smaller size.

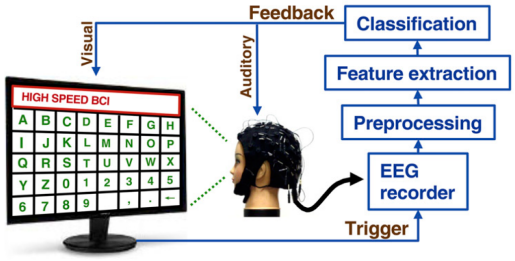


Figure 1: Working of SSVEP-based BCI [8]

In this paper, there are two parts of the implementation, the first is the impact on performance, when giving raw signal as input and extracted frequency features as input. The second part compares the performance of the selection of electrodes and without the selection of electrodes in the input.

MATERIALS AND METHODS

EEG Dataset: This paper uses data published in [5], which is recorded using a 40-target BCI speller. The BCI

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-03

speller consists of a 5x8 matrix of 40 characters, out of which 26 are English alphabets, 10 digits, and 4 symbols. A total number of 35 participants (27 naive, 8 experienced) participated in the experiment. The experiment had 6 blocks, each comprising of 40 trials corresponding to the 40 symbols on the screen. One trial total length is 6 s, with 0.5 s target cue and last 0.5 s was blank. There was a gap of several minutes between two consecutive blocks for the participant to rest. The 40 frequencies recorded are from 8 Hz to 15.8 Hz with an interval of 0.2 Hz. The whole-head EEG data is recorded using 64 electrodes according to the 10-20 system at a sampling rate of 1000Hz as shown in Figure 2. The recorded data epochs were downsampled to 250 Hz as the upper limit of the SSVEP range is 90 Hz. No pre-processing has been performed on the data. At the end of the experiment, the data available for each participant is 240 trials i.e. 6 blocks x 40 trials. Each trial consisted of 64 channels x 1500 time points.

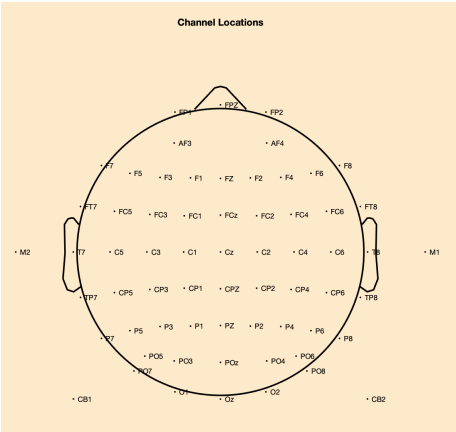


Figure 2: Electrode placement

Canonical Correlation Analysis: Canonical Correlation Analysis is a multivariable statistical method which finds a correlation between two sets of variables. It was introduced in the field of EEG analysis by Lin et. al in [9]. In the case of SSVEP-based EEG signals, there are two variables, X which is the recorded multi-channel EEG signal and Y refers to the reference signals. The frequency recognition is obtained by calculating the canonical correlation between multi-channel SSVEP and the reference signals. The frequency which has the maximum correlation value in the reference signals is the same as the frequency of the multi-channel SSVEP signal. CCA helps in reducing a large amount of information into useful information by maximizing the correlation.

Convolutional Neural Network: Recently CNNs have attained success and popularity in so many different fields. The different layers of CNN help in dimensionality reduction, in turn, reducing the number of training parameters which will increase the training speed and improve performance. A CNN consists of an input layer, convolutional layer, activation, pooling layer, and fully connected layer. Various parameters such as Dropout and Batch Normalization can be used for further optimization

Table 1: CNN for Time Domain

Layer / Parameter	Number / Size
Conv1D	32 filters
MaxPooling1D	2
Dropout	0.5
BatchNormalization	-
Conv1D	64 filters
MaxPooling1D	2
Dropout	0.5
BatchNormalization	-
Flatten	-
Dense	256 filters
Dense	512 filters, 40 output

in the CNN model. The convolution layer slides a filter of a particular size over the given input to produce a feature map. As done in any Neural Network, we use activation function on the output of the convolution layer after which we have a non-linear output. The dimensionality reduction is performed by the pooling layers, which extracts useful parameters and thus reduce the number of parameters to compute and to be learned. This will also prevent overfitting of the model. There can be such multiple layers, each comprising of convolution layer, activation function, and pooling layer. The last few layers are fully connected layers which are similar to the regular neural networks. But before we pass the data from convolutional layers to the fully connected layers, we need to flatten the data as fully connected layers understand only 1-dimensional data. The main part of the training is the configuration of the above-described layers.
In the past few years, the use of CNN in the area of SSVEP has increased. In [6], the accuracy of CNN was 69.03% 256 channel SSVEP recordings of 11 subjects, whereas in [7] the accuracy of CNN classification is as high as 99.27% for static and 94.03% for ambulatory SSVEP data of 7 subjects. In these papers, a power spectral density analysis has been performed to extract frequency features and then feed it to the model for classification. Here in this paper, we will compare the performance of the model between such extracted features and raw signal.

CNN Model for raw signal: The raw signal is in form of time domain as indicated by 64 channel x 1500 time points is given as input to the CNN model. To evaluate the performance of the CNN model on the raw signal as input, we make use of Convolutional Neural Network in 1 dimension. The structure is described in Table 1, 1 Conv1D layer with 32 filters having 'ReLU' as activation function and a Max Pooling layer with size 2. The second layer comprises of 1 Conv1D layer with 64 filters having 'ReLU' as activation function and a Max Pooling layer with size 2. The Dropout value is taken as 0.5 and BatchNormalization has been applied. Fully connected layers of size 256 and 512 have been used. The last fully connected layer uses Softmax as the activation function for giving the output. Adam optimization algorithm is used for updating the network weights in training data. The training and validation accuracy indicated in Figure 3

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-03

Table 2: CNN for Frequency Domain

Layer / Parameter	Number / Size
Conv2D	32 filters
MaxPooling2D	3
Dropout	0.5
BatchNormalization	-
Conv2D	64 filters
MaxPooling2D	3
Dropout	0.5
BatchNormalization	-
Flatten	-
Dense	512 filters, 40 output

shows the learning of the model.

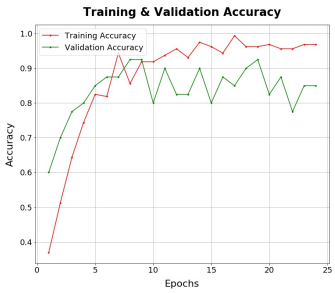


Figure 3: Training & Validation Accuracy - Time Domain

CNN Model for extracted frequency features: Power Spectral Density Analysis (PSDA) is performed on the signal before giving input to the CNN model. Here we use Convolutional Neural Network in a 2 dimensions matrix. The structure is described in Table 2, 1 Conv2D layer with 32 filters having 'ReLU' as activation function and Max Pooling layer with size 3x3. The second layer consists of 64 filters with 'ReLU' as activation function and Max Pooling layer with a filter of size 3. The Dropout value is taken as 0.5 and BatchNormalization has been applied. The last layer which is fully connected layer is of size 512 with Softmax as an activation function. Adam optimization algorithm is used just like the previous for updating the network weights in training data. The training and validation accuracy indicated in Figure 4 shows the learning of the model.

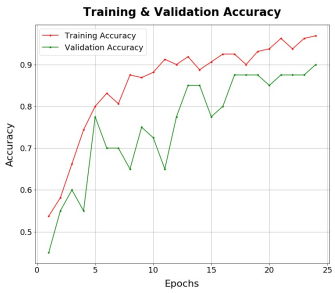


Figure 4: Training & Validation Accuracy - Frequency Domain

Selection of electrodes: The SSVEP signals decoding have a much better quality when the electrodes placed

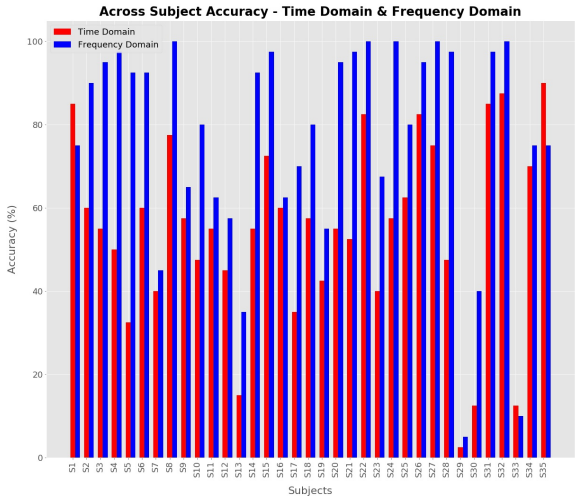


Figure 5: Overall Accuracy of Subjects in both domains

over the occipital and parietal areas are thus removing other background activities [5]. From the 64 electrodes shown in Figure 2, the electrodes selected in this paper are 9 i.e. Pz, PO5, PO3, POz, PO4, PO6, O1, Oz and O2 which give a better performance in classification of the signal [5]. The data from those electrodes is selected for classification. The same model is given two different inputs of raw signal (time domain) and the result is compared to check the performance of the model in both scenarios. The structure of the model is just the same as the one described in Table 1. The channel selection is tested in the case of the time domain.

RESULTS

Comparison of CNN models for raw signal and extracted frequency features: In our models, we have attained an accuracy of 76.5% in the time domain, 80% accuracy in the frequency domain. The CNN model with time domain performs almost as well as the model with frequency extracted features. The accuracy of the model in time domain is quite higher in certain subjects than in frequency domain as shown in Figure 5 indicating that the extraction of the frequency component is alternative. The idea behind using CNN models is an automatic feature extraction and learn the signal specific oscillation in the hidden layers which is achieved in the model used for raw signal. From the given data, we have used has 240 trials per subject, the number of samples is quite less for the model to train in the time domain.

Channel Selection results: The model achieves 76.5% accuracy when data from only 9 channels is taken into consideration and achieves 22.5% accuracy without selection of 9 channels. In [5], it is mentioned that utilization of electrodes positioned over occipital and parietal will give better accuracy in classification of the given signals. Our CNN model verifies that the channel selection proves beneficial in the classification as shown in Fig-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-03

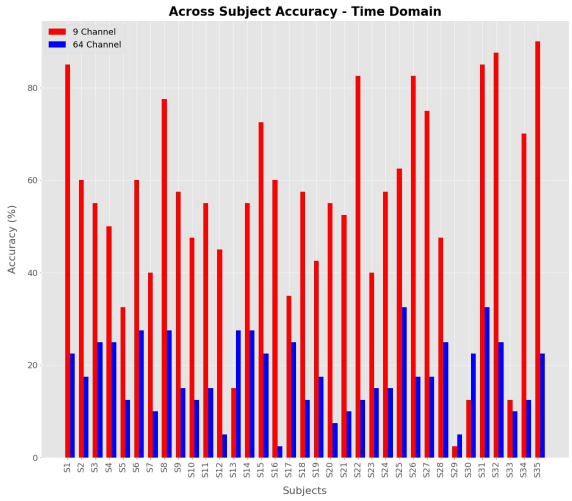


Figure 6: Overall Accuracy of Subjects with 9-channel and 64-channel data

ure 6.

Comparison of CCA with CNN models: The accuracy of CCA for all the subjects can be compared with the accuracy of CNN models in various conditions. The CCA model requires hand-crafted data for processing and getting the maximum correlation values. While the CNN models automatically detect the high-level features with the help of hidden layers. It can also be observed that the CCA and CNN give quite similar results on certain subjects, however as the most successful algorithms in SSVEP decoding for many years,CCA generated the overall higher accuracy than the time and frequency domain. The CCA and CNN results are shown in Figure 7 and Table 3.

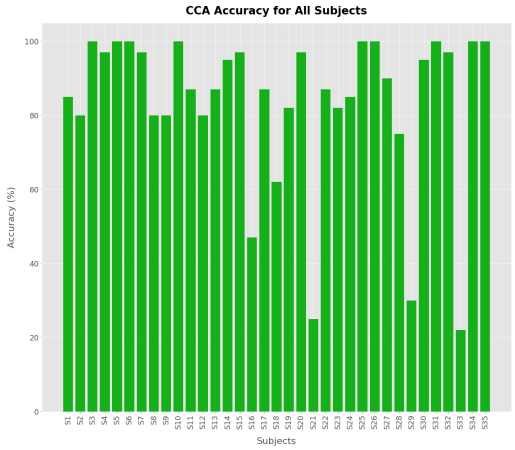


Figure 7: Overall Accuracy of Subjects using CCA

Table 3: CCA,time and frequency domain accuracy

Sub\Acc	CCA	Time Domain	Frequency Domain
Subject 1	85	85	77.5
Subject 3	100	57.5	95
Subject 5	100	52.5	92.5
Subject 6	100	60	92.5
Subject 7	97	40	42.5
Subject 9	80	58.5	62.5
Subject 10	100	58.5	80
Subject 11	87	55	62
Subject 12	80	47.5	58.5
Subject 13	87	17.50	35
Subject 14	95	60	90
Subject 17	87	37.5	70
Subject 19	82	45	50
Subject 20	97	58.5	90
Subject 23	82	40	70
Subject 25	100	65	80
Subject 26	100	82.5	85
Subject 29	30	2.5	5
Subject 30	95	12.5	40
Subject 31	100	85	97.5
Subject 33	22	12.5	10
Subject 34	100	70	77.5
Subject 35	100	90	77.5
Mean	87.21	51.02	65.55
Sub\Acc	**CCA**	**Time Domain**	**Frequency Domain**
Subject 2	80	60	90
Subject 4	97	50	97.5
Subject 8	80	78.5	100
Subject 15	97	72.5	98
Subject 16	47	60	62
Subject 18	62	58.5	80
Subject 21	25	60	95
Subject 22	87	82	100
Subject 24	85	60	100
Subject 27	90	77.5	98.5
Subject 28	75	50	98.5
Subject 32	97	90	100
Mean	76.83	65.5	92.90
Sub\Acc	**CCA**	**Time Domain**	**Frequency Domain**
Overall	83.65	56.51	76.30

DISCUSSION

The evaluation of the CNN model on time and frequency domain with different architectures were performed. from the result, we reconfirmed the method of electrode selection gave a lot higher accuracy than the usual data without selection with deep neural networks. In the time domain data, the CNN model itself learns the low-level and the high-level features and makes the decision of discarding the unnecessary data unlike in the extracted frequency features. the CCA was implemented to compare the traditional algorithm and deep neural networks. We conclude two points from the above experiments. The first being that given the raw signal and adjusted parameters, CNN can classify the input data without prior feature extraction. Initially, the classification accuracy did not increase as the number of parameters like Dropout, BatchNormalization, Regularizers were continuously updated and the model's accuracy was checked.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-03

The second conclusion is that in a certain kind of data, the data taken from selected electrodes would assist the model in better classification than the whole data.

CONCLUSION

Two different deep neural network structure on time and frequency domain have been put forward and tested in this paper. The accuracy of the model in the time domain is quite higher in certain subjects than in the frequency domain which may indicate that the extraction of the frequency component is alternative. Generally on all subjects, frequency domain show the overall higher accuracy than the time domain. The traditional CCA algorithm still achieved the best accuracy in the case of small data on time domain which is not unexpected. The selection of channels play an important role in identifying the brain rhymes. Despite having good accuracy, each of the model can be further tested and configured to achieve even more accuracy given more amount of data. Data augmentation method will be applied for the next round test. 240 trials per subject is not enough for CNN to learn the features and also might generate the overfitting problems in the model. Hence, a reiteration was performed and then selected values for the parameters of regularization to avoid overfitting. Furthermore, CNNs can be combined with Recurrent neural network (RNN)to explore Brain physiological data on time domain data as well as on frequency domain.

ACKNOWLEDGMENT

Jiahua.Xu and Riddhi.Gala contributed this paper equally, Jiahua.Xu proposed the framework of methods,scripted the core code and revised the draft, Riddhi.Gala implemented the code and drafted the paper, thanks Andreas.Nürnberger and Zheng.wu for feedbacks and comments. The authors gratefully acknowledge financial support from China Scholarship Council.

REFERENCES

[1] Chen X, Chen Z, Gao S, Gao X. A high-ITR SSVEP-based BCI speller. Brain-Computer Interfaces. 2014;1:3-4, 181-191.
[2] Leeb R, Friedman D, Müller-Putz GR, Scherer R, Slater M, Pfurtscheller G. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with tetraplegics. Computational Intelligence and Neuroscience. 2007; 79642:1-8.
[3] Gao X, Xu D, Cheng M, Gao S. A BCI-based environmental controller for the motion-disabled. IEEE Trans Neural Syst Rehabil Eng. 2003; 11:137-140.
[4] Wu Z, Lai Y, Xia Y, Wu Y, and Yao D. Stimulator selection in SSVEP-based BCI. Medical Engineering & Physics. 2008;30:8,1079-1088.
[5] Wang Y, Chen X, Gao X, and Gao S. A Benchmark Dataset for SSVEP-Based Brain-Computer Interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. 2016; 25:10, 1746?1752.
[6] Thomas J, Maszczyk T, Sinha N, Kluge T, Dauwels J. Deep Learning-based Classification for Brain-Computer Interfaces, in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, October 5-8, 2017. 234-239.
[7] Kwak N, Müller K, Lee S. A convolutional neural network for stead state visual evoked potential classification under ambulatory environment. PLoS One. 2017; 0172578.
[8] Chen X, Wang Y, Nakanishi M, Gao X, Jung T, Gao S. High-speed spelling with a noninvasive brain-computer interface, in Proceedings of National Academy of Sciences, 2015;112:44, E6058-67.
[9] Lin Z, Zhang C, Wu W and Gao X 2006 Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Transactions Biomedical Engineering, 2006; 53:6, 1172-1176.
[10] Rezeika A, Benda M, Stawicki P, Gembler F, Saboor A and Volosyak I. Brain-Computer Interface Spellers: A Review. Brain Sciences. March 30 2018; 8:57.
[11] Lee H, Choi Y. A Convolution Neural Network Scheme for Classification of Motor Imagery EEG based on Wavelet Time-Frequency Image.IEEE 2018 International Conference on Information Networking. 2018. 906-909.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-04

LOOKING FOR CORTICAL PATTERNS OF SUCCESSFUL MOTOR IMAGERY-BASED BCI LEARNING

M.-C. Corsi[1,2], M. Chavez[3], D. Schwartz[4], N. George[4], L. Hugueville[4], A.E. Kahn[5], S. Dupont[2], D. S. Bassett[5,6,7,8], F. De Vico Fallani[1,2]

[1] Inria Paris, Aramis project-team, F-75013, Paris, France

[2] Institut du Cerveau et de la Moelle Epinière, ICM, Inserm, U 1127, CNRS, UMR 7225, Sorbonne Université, F-75013, Paris, France

[3] CNRS, UMR 7225, F-75013, Paris, France

[4] Institut du Cerveau et de la Moelle Epinière, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, Ecole Normale Supérieure, ENS, Centre MEG-EEG, F-75013, Paris, France

[5] Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA

[6] Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[7] Department of Physics and Astronomy, College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA

[8] Department of Electrical and Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA

E-mail: marie.constance.corsi@gmail.com

ABSTRACT:
Non-invasive Brain-Computer Interfaces (BCIs) based on motor imagery (MI) tasks represent a valuable tool both from a societal and a clinical perspective. Nevertheless, performances vary inconsistently across subjects and the mechanisms underlying a successful skill acquisition is poorly understood. In this longitudinal study performed with the electroencephalography (EEG), we show that BCI training can be characterized by patterns that rely on the neurophysiology. We observed that the desynchronization effect increases significantly over the sessions within the α and β sub-bands for subjects who showed a significant improvement of their BCI scores. Notably, we observed that they also presented a decrease of the functional connectivity in regions beyond those targeted during the BCI experiments, whereas the subjects who did not improve their performances did not show any significant change over sessions. Taken together, these results give additional insights about the skill acquisition process during MI-based BCI trainings.

INTRODUCTION

Non-invasive BCIs are largely used to produce thought-provoked action, by exploiting the ability of subjects to voluntary modulate their brain activity through mental imagery. Despite its societal and clinical applications [1], [2], voluntarily modulating brain activity to control a BCI appears to be a learned skill [3], [4], based on the feedback presented to the user, and, in general, several weeks or even months may be needed to reach relatively high-performance (> 90 %) in BCI control [1],[15]. Furthermore, between 15 and 30 % of the users [6] face difficulties in controlling a BCI even after several training sessions, eliciting a high inter-subject variability. This point shows the difficulty to understand the key-aspects, or subject-related patterns, of an efficient learning process in MI-based BCI more specifically.

In previous studies, two main types of predictors of MI-based BCI success have been elicited [7]. The first category corresponds to the behavioral or psychological predictors. Among them, kinesthetic imagination score, mental rotation ability, self-reliance, visuo-motor coordination and concentration have shown significant correlations with BCI scores [8]–[11]. The second category corresponds to the neurophysiological predictors. Most of them relied on resting state or pre-stimulus recordings and are associated with power spectra. More specifically, these activations involve mainly sensorimotor areas (μ, low α and high θ power [12], [13]) and the fronto-parietal axis (within the γ band [14], [15]).

However, while there is a large number of inter-subject comparisons, less longitudinal studies have been conducted and little is known about how the communication between brain areas may differ during the learning process and between subjects.

In this work, we propose an original approach that aims at eliciting patterns that come from functional connectivity (FC) of successful learning process based on a longitudinal study performed with EEG. We hypothesize that the FC changes over the training,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-04

involving areas beyond those targeted during the BCI experiment, and that the associated properties present a specific pattern of learning efficiency.

MATERIALS AND METHODS

BCI protocol and participants

Twenty healthy subjects (aged 27.45 ± 4.01 years, 12 men), right-handed and BCI-naïve, participated in the study. It consisted of a longitudinal EEG-based BCI training composed by four sessions (i.e. twice in a week for two weeks). None presented medical or psychological disorders. A written informed consent was obtained from subjects after explanation of the study, approved by the ethical committee CPP-IDF-VI of Paris. All participants received financial compensation at the end of their participation. The BCI task consisted of a standard 1D, two-target box task [16] in which, to control the vertical position of a cursor moving from the left to the right side of the screen, the subjects modulated their brain activity. To hit the target-up, the subjects imagined a sustained movement of the right hand (MI condition) and they remained at rest to hit the target-down (Rest condition). Thus, the BCI scores are defined here as the proportion of hit targets. Each session was composed by 6 runs of 32 trials. Each trial lasted 7 s and consisted of a 1 s of inter-stimulus, followed by 2 s of target presentation, 3 s of feedback and 1 s of result presentation (Fig 1). At the beginning of each session, BCI features (EEG channels and frequency) were selected in a calibration phase in which the subjects were instructed to perform the BCI tasks without any visual feedback. We selected the features within the α-β band and localized within the motor area contralateral to the movement. The online classification was performed with a Linear Discriminant Analysis, suited for a two-class paradigm [17].

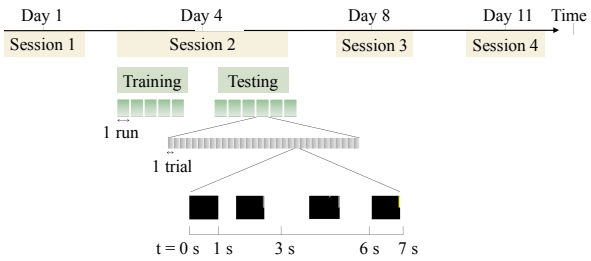


Fig 1 Experimental design. The different timescales (session, run, trial) are presented.

Following the last inclusion, we selected two groups of five subjects each to elicit the most extreme profiles in terms of learning process. The selection was based on the difference in terms of BCI scores between the first and the last session. The five subjects who showed the best improvement in terms of BCI scores (aged 27.40 ± 2.07 years, 4 men) were gathered within the Group 1 (G1). The five subjects who improved the less (aged 27.60 ± 3.36 years, 1 men) were gathered within the Group 2 (G2).

Materials

EEG signals were collected with a 74-channel BrainAmp system (referenced to mastoids signals). Left and right electromyogram (EMG) signals were recorded to ensure that subjects were not moving their forearm while performing the tasks. Recordings were performed with a 1 kHz frequency sampling and an offline filter applied between 0.1 and 300 Hz. BCI sessions were performed from EEG signals transmitted via the Fieldtrip buffer [18] to the BCI2000 toolbox [19].

After the fourth session, individual T1 scans were obtained by using a 3T Siemens Magnetom PRISMA. The experiment consists in a 15 minute-resting-state task to obtain an accurate head model for the source reconstruction.

EEG processing

After having downsampled the signals to 250 Hz, we performed an Independent Components Analysis with the Infomax approach [20] using the Fieldtrip toolbox [18] to remove potential ocular and/or cardiac artifacts.

Once the signals epoched and average referenced, we performed source reconstruction by computing the individual head model with the Boundary Element Method (BEM) [21], [22]. BEM surfaces resulted from three layers associated with the subject's MRI (scalp, inner skull, outer skull) with 1922 vertices each. The weighted Minimum Norm Estimate [23] method was used to estimate the sources via the Brainstorm toolbox [24]. Here, the identity matrix was defined as the noise covariance matrix. The minimum norm estimate corresponds in our case to the current density map. The regions of interest (ROIs) relied on the use of the Destrieux atlas [25].

To compute the power spectrum density of the signals at the source level, we used the Welch method with a window length of 1 s and a window overlap ratio of 50 % applied during the feedback period that ranges (from t = 3 s to t = 6 s) within the individual anatomical space.

Functional connectivity analysis

Functional connectivity has already been shown to be a valuable tool in the BCI domain [26]–[28]. Here, we were particularly interested in eliciting patterns of successful learning that rely on FC. Thus, we used the imaginary coherence, weakly affected by volume conduction and spatial leakage [29], [30], between each pair of ROIs. In this study, we used a concise metric that could elicit highly connected hubs: the node strength, defined as the sum of the connectivity weights of the edges linked to each node i.

Metrics and statistical analysis:

To take into account the subjects' specificity, we defined our frequency bands according to the Individual Alpha Frequency (IAF) [31]. We restricted our study within the α-β frequency bands for two reasons. First, because it corresponds to the frequency bands targeted during the BCI experiments (i.e. modulated during motor imagery task). Secondly, because both α and β bands have been

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-04

shown to be involved in learning processes [32], [33]. Here, the IAF corresponds to the first peak comprised between 6 and 12 Hz. The α_1 band ranges from IAF – 2 Hz to IAF, the α_2 band from IAF to IAF + 2 Hz, β_1 from IAF + 2 Hz to IAF + 11 Hz and β_2 from IAF + 11 Hz to IAF + 20 Hz.

We computed statistical differences among activations (i.e. power spectra) recorded in the MI and the rest conditions at the subject level via a paired t-test. Statistics were corrected for multiple comparisons using the cluster approach [24], [34], with a statistical threshold to 0.05, a minimum number of neighbors of 2 and a number of randomization of 500. Cluster-level statistics are obtained by using the sum of the t-values within every cluster. Besides the correction for multiple comparisons, this method avoids any spatial a priori about areas that show a significant desynchronization effect.

To obtain a single value that takes into account the desynchronization effect, we worked with the relative power, ΔP, defined as follows:

$$\Delta P = 100 \times \frac{P_{MI} - P_{Rest}}{P_{Rest}}$$

where P_{MI} and P_{Rest} correspond, respectively to the averaged power calculated across the cluster from MI and Rest trials. Similarly, we computed the relative node strength ΔN as the relative difference in terms of node strength between the conditions.

RESULTS

Behavioral performances

G1 showed a significant improvement of the BCI scores (one-way ANOVA, $F_{3,12} = 15.1$, $p_{FDR} = 2.36 \times 10^{-4}$) whereas the scores obtained by G2 did not show a session effect ($F_{3,12} = 0.14$, $p_{FDR} = 1$) (Fig. 2A). These results enabled us to check that we did have two groups that strongly differed. Notably, this difference became prominent from session 3 (Fig. 2A).

Brain activation

From the activation perspective, we observed a strong decrease of the relative power from session 3 for G1 (Fig. 2B). Nevertheless, a significant session effect was observed only within the α_2 band (one-way ANOVA, $F_{3,12} = 5.16$, $p_{FDR} = 0.02$). ΔP did not present a session effect for G2 both in α_2 and β_1 frequency bands. Notably, this trend is similar to the one previously observed with behavioral performances (Fig. 2A). As expected, with the practice, the desynchronization effect will increase only within G1. This result is in line with [6] where subjects with poor performances show fewer significant features than the other subjects.

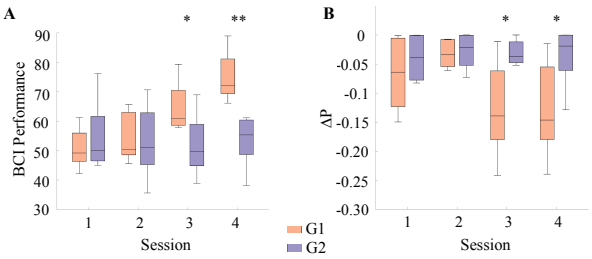


Fig. 2 Trends over sessions obtained for G1 and G2. (A) BCI performances. (B) Evolution of the relative power (ΔP) within the α_2 band. *p < 0.05; **p < 0.05 (Mann-Whitney test).

Brain connectivity

A way to assess potential differences in terms of connectivity is to compare the relative node strength values, obtained during the fourth session when the behavioral results are the most discriminative between the two groups of subjects. First, as shown in Fig. 3 in absolute values, the ΔN values were larger within G1 than within G2, meaning that there was a larger discrimination between conditions within G1 than within G2 from the FC perspective. Secondly, within G1, the strongest negative values of ΔN involved the pre-frontal, parietal and occipital areas (bilateral) whereas the strongest positive ones were located in pre-and postcentral gyri (bilateral) in α_1 and α_2. Thus, at the

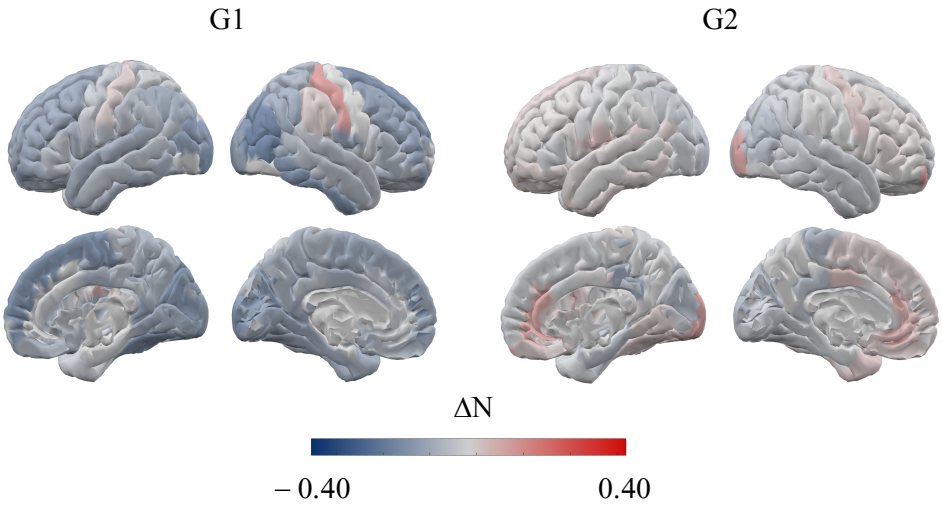


Fig. 3 Averaged relative node strength, obtained during the fourth session, across the subjects from respectively G1 and G2 within the α_2 band.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-04

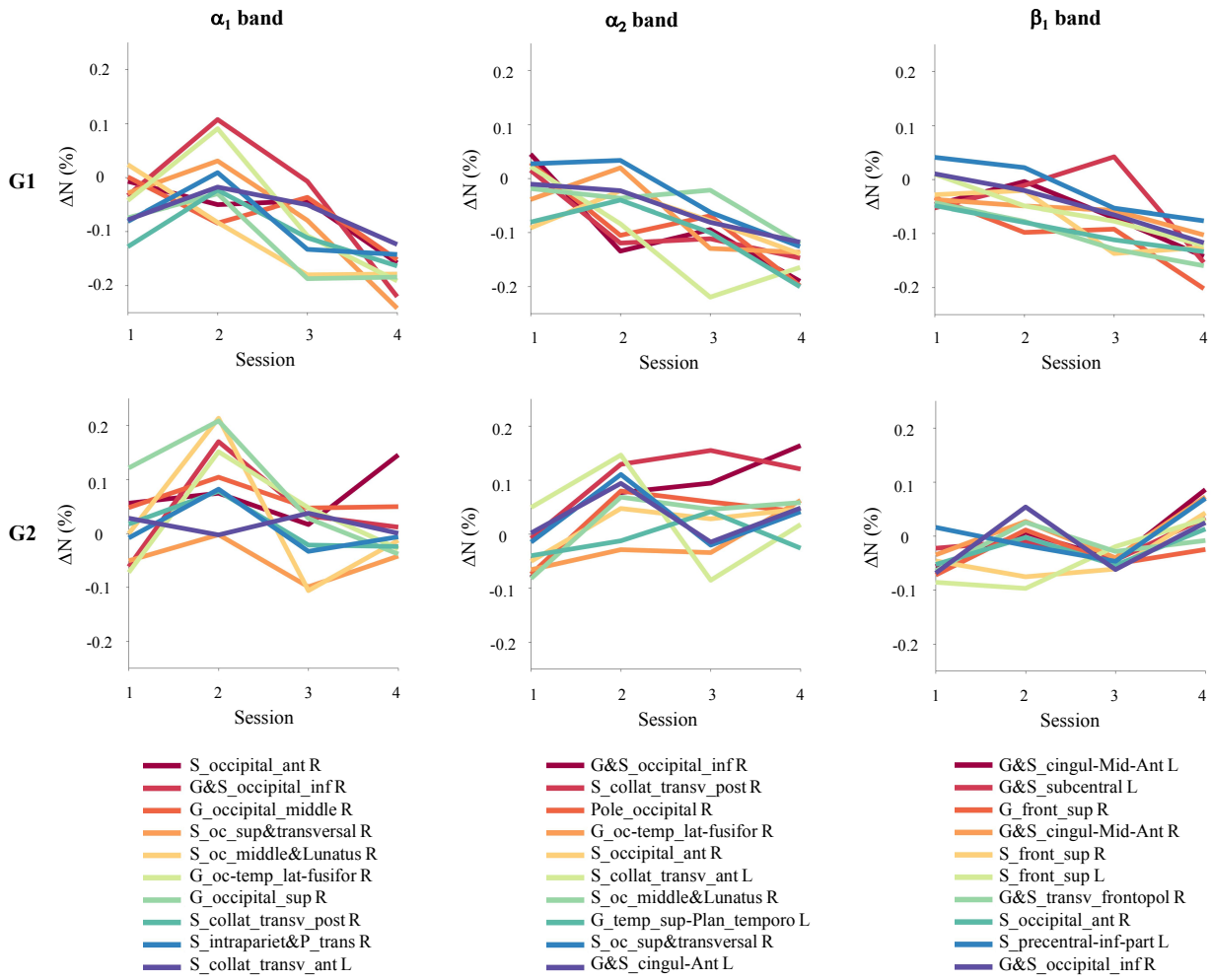


Fig. 4 Evolution of the relative node strength, averaged across subjects from respectively G1 and G2, over the sessions. For each frequency band (i.e. column), we selected the 10 ROIs that discriminate the best the two groups during the last session. The only ROI that shows a significant session effect is the subcentral gyrus (central operculum) and sulci left within G1 and the β_1 band (one-way ANOVA, $F_{3,12} = 4.30$, $p = 0.03$).

end of a successful training, performing a MI task tends to engender a lower connectivity, with respect to the Rest task, within frontal, parietal and occipital areas and engenders higher connectivity in primary somatosensory cortex and primary motor cortex.

To take into account the training effect, we studied the evolution of the node strength over the sessions by focusing our work on the ROIs that showed the highest difference between the groups (Fig. 4). Within the α_1 and the α_2 bands, most of the selected ROIs were located in areas involved in visual perception; in the retinotopic representation of the peripheral visual field (posterior transverse collateral sulcus), the visual attention (intraparietal sulcus and transverse parietal sulci), the peripheral vision (anterior transverse collateral sulcus) and preparation of the movement (inferior occipital gyrus and sulcus). Within G1, the ΔN values tended to decrease with the training whereas they were stable (or increased in the case of the α_2 band) within G2. The trends tended to be similar to those obtained with ΔP.

Notably within the β_1 band, the selected ROIs are known to be involved in the preparation of the movement (inferior occipital gyrus and sulcus), and in working memory (superior frontal gyrus, superior frontal sulcus). Among the selected ROIs some of them belong to the salience network (middle-anterior part of the cingulate gyrus and sulcus), or to the sensorimotor network (subcentral gyrus and sulci, inferior part of the precentral sulcus). Once again, within G1, the FC tended to decrease over the sessions whereas they are stable within G2.

DISCUSSION & CONCLUSION

Understanding the variations of BCI performances between subjects is crucial to improve the reliability of BCI systems. To reduce the inter-subject variations in terms of BCI performances, different approaches have been described, going from the improvement of feature extraction [35] and of the classification algorithm [36] to the adaptation to the user's profile [8]. Another element could be the combination of different modalities to take advantage of their complementarity in terms of

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-04

sensitivity towards noise for instance. The present study relies on EEG signals only. Nevertheless, it has been previously shown that combining for instance EEG and magnetoencephalography could strongly improve the classification accuracy [37].

From the subject-based perspective, one approach consists of identifying factors that can influence BCI scores to adapt the training to the subject's profile [7]. If psychological and power-based predictors have been elicited [7], functional connectivity has been less considered so far whereas it has been proved to provide reliable monitoring of cognitive function such as working memory for instance [38] and to be used as alternative features [35]. In a previous study, we demonstrated that neurophysiological and FC information could be used as predictors of BCI performances [39]. Here, we focused our study on the changes at the neurophysiological level associated with a successful learning process. For that purpose, we used simple metrics to elicit FC behavior from longitudinal study. Our results suggest that an efficient training is accompanied by a decrease of the FC. Notably, as expected, this phenomenon affects areas involved in the sensorimotor network but also areas involved in the movement preparation and the working memory, especially within the β_1 band. In the case of subjects who showed the weakest improvement, the relative node strength tends to remain stable or to increase within these areas. Nevertheless, we acknowledge that increasing the number of sessions and the number of subjects would enable to get a more reliable idea of the neural mechanisms underlying the learning process, in terms of potential re-organization of the brain network but also in terms of learning speed.

Previous studies reported a lack of learning metric in sensorimotor rhythm-based BCI [3] and therefore, a lack of evidence of learning. Here, common trends between BCI scores, relative power and FC can be elicited. Indeed, in the case of G1, after the session 2, BCI scores start to increase significantly, accompanied by an increase of the desynchronization effect and a decrease of the node strength within the α_1, the α_2, and the β_1 bands. In the case of G2, none of the tested metric showed a session effect. Thus, combining power-based and FC markers, which rely on separability of brain features, to identify learning metrics based on neurophysiological information and not only on the BCI accuracy could be a tool to better assess individual learning.

In this study, by using metrics based on FC we could elicit specific patterns that involve regions beyond the sensorimotor areas, mainly characterized by a decrease of the node strength over a successful training. Further experiments and analysis relying on more sophisticated methods, based notably on graph theory, will certainly help to consolidate this result, and to understand the mechanisms underlying the learning process.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002.

[2] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egypt. Inform. J.*, vol. 16, no. 2, pp. 213–230, Jul. 2015.

[3] S. Perdikis, L. Tonin, S. Saeedi, C. Schneider, and J. del R. Millán, "The Cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users," *PLoS Biol.*, vol. 16, no. 5, May 2018.

[4] 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, Jul. 2018.

[5] D. J. McFarland, A. T. Lefkowicz, and J. R. Wolpaw, "Design and operation of an EEG-based brain-computer interface with digital signal processing technology," *Behav. Res. Methods Instrum. Comput.*, vol. 29, no. 3, pp. 337–345, Sep. 1997.

[6] B. Z. Allison and C. Neuper, "Could Anyone Use a BCI?," in *Brain-Computer Interfaces*, D. S. Tan and A. Nijholt, Eds. Springer London, 2010, pp. 35–54.

[7] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain–computer interface: A brief review," *J. Neurosci. Methods*, vol. 243, pp. 103–110, Mar. 2015.

[8] C. Jeunet, B. N'Kaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting Mental Imagery-Based BCI Performance from Personality, Cognitive Profile and Neurophysiological Patterns," *PLoS ONE*, vol. 10, no. 12, p. e0143962, 2015.

[9] E. M. Hammer *et al.*, "Psychological predictors of SMR-BCI performance," *Biol Psychol*, vol. 89, no. 1, pp. 80–86, Jan. 2012.

[10] A. Vuckovic and B. A. Osuagwu, "Using a motor imagery questionnaire to estimate the performance of a Brain-Computer Interface based on object oriented motor imagery," *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 124, no. 8, pp. 1586–1595, Aug. 2013.

[11] E. M. Hammer, T. Kaufmann, S. C. Kleih, B. Blankertz, and A. Kübler, "Visuo-motor coordination ability predicts performance with brain-computer interfaces controlled by modulation of sensorimotor rhythms (SMR)," *Front Hum Neurosci*, vol. 8, Aug. 2014.

[12] B. Blankertz *et al.*, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, 2010.

[13] 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, 2013.

[14] 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.

[15] M. Grosse-Wentrup and B. Schölkopf, "High γ-power predicts performance in sensorimotor-rhythm

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-04

brain-computer interfaces," *J Neural Eng*, vol. 9, no. 4, p. 046001, Aug. 2012.
[16] J. R. Wolpaw, D. J. McFarland, T. M. Vaughan, and G. Schalk, "The Wadsworth Center brain-computer interface (BCI) research and development program," *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 11, no. 2, pp. 204–207, Jun. 2003.
[17] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J Neural Eng*, vol. 4, no. 2, pp. R1–R13, Jun. 2007.
[18] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data," *Comput. Intell. Neurosci.*, vol. 2011, 2011, p. e156869, Dec. 2010.
[19] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
[20] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.
[21] M. Fuchs, M. Wagner, and J. Kastner, "Boundary element method volume conductor models for EEG source reconstruction," *Clin. Neurophysiol.*, vol. 112, no. 8, pp. 1400–1407, 2001.
[22] A. Gramfort, T. Papadopoulo, E. Olivi, and M. Clerc, "OpenMEEG: opensource software for quasistatic bioelectromagnetics," *Biomed. Eng. OnLine*, vol. 9, p. 45, 2010.
[23] 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, May 2006.
[24] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: A User-Firendly Application for MEG/EEG Analysis," *Comput. Intell. Neurosci.*, vol. 2011, Jan. 2011.
[25] 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, Oct. 2010.
[26] A. Mottaz *et al.*, "Modulating functional connectivity after stroke with neurofeedback: Effect on motor deficits in a controlled cross-over study," *NeuroImage Clin.*, vol. 20, pp. 336–346, Jul. 2018.
[27] D. S. Bassett and A. N. Khambhati, "A network engineering perspective on probing and perturbing cognition with neurofeedback," *Ann. N. Y. Acad. Sci.*, vol. 1396, no. 1, pp. 126–143, May 2017.
[28] F. De Vico Fallani and D. S. Bassett, "Network neuroscience for optimizing brain–computer interfaces," *Phys. Life Rev.*, Jan. 2019.
[29] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clin Neurophysiol*, vol. 115, no. 10, pp. 2292–2307, Oct. 2004.
[30] K. Sekihara, J. P. Owen, S. Trisno, and S. S. Nagarajan, "Removal of Spurious Coherence in MEG Source-Space Coherence Analysis," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 11, pp. 3121–3129, Nov. 2011.
[31] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res. Rev.*, vol. 29, no. 2, pp. 169–195, Apr. 1999.
[32] R. Sigala, S. Haufe, D. Roy, H. R. Dinse, and P. Ritter, "The role of alpha-rhythm states in perceptual learning: insights from experiments and computational models," *Front. Comput. Neurosci.*, vol. 8, Apr. 2014.
[33] E. G. Antzoulatos and E. K. Miller, "Increases in functional connectivity between prefrontal cortex and striatum during category learning," *Neuron*, vol. 83, no. 1, pp. 216–225, Jul. 2014.
[34] R. Oostenveld *et al.*, "FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data, FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data," *Comput. Intell. Neurosci. Comput. Intell. Neurosci.*, vol. 2011, 2011, p. e156869, Dec. 2010.
[35] T. Cattai, S. Colonnese, M.-C. Corsi, D. S. Bassett, G. Scarano, and F. De Vico Fallani, "Characterization of mental states through node connectivity between brain signals," presented at the European Signal Processing Conference (EUSIPCO 2018), 2018, pp. 1391–1395.
[36] 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, p. 568, 2013.
[37] M.-C. Corsi *et al.*, "Integrating EEG and MEG signals to improve motor imagery classification in brain-computer interface," *Int. J. Neural Syst.*, Apr. 2018.
[38] J. Toppi *et al.*, "Different Topological Properties of EEG-Derived Networks Describe Working Memory Phases as Revealed by Graph Theoretical Analysis," *Front. Hum. Neurosci.*, vol. 11, 2018.
[39] M.-C. Corsi *et al.*, "Spatiotemporal neural correlates of brain-computer interface learning," bioRxiv, p. 487074, Dec. 2018.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-05

ADVERSARIAL FEATURE LEARNING IN BRAIN INTERFACING: AN EXPERIMENTAL STUDY ON ELIMINATING DROWSINESS EFFECTS

O. Özdenizci*, B. Oken†, T. Memmott†, M. Fried-Oken†, D. Erdoğmuş*

*Northeastern University, Boston, MA, USA
†Oregon Health & Science University, Portland, OR, USA

Corresponding author: oozdenizci@ece.neu.edu

ABSTRACT: Across- and within-recording variabilities in electroencephalographic (EEG) activity is a major limitation in EEG-based brain-computer interfaces (BCIs). Specifically, gradual changes in fatigue and vigilance levels during long EEG recording durations and BCI system usage bring along significant fluctuations in BCI performances even when these systems are calibrated daily. We address this in an experimental offline study from EEG-based BCI speller usage data acquired for one hour duration. As the main part of our methodological approach, we propose the concept of adversarial invariant feature learning for BCIs as a regularization approach on recently expanding EEG deep learning architectures, to learn nuisance-invariant discriminative features. We empirically demonstrate the feasibility of adversarial feature learning on eliminating drowsiness effects from event related EEG activity features, by using temporal recording block ordering as the source of drowsiness variability.

INTRODUCTION

Majority of the limitations for electroencephalography (EEG) based brain-computer interface (BCI) systems are caused by across- and within-recording variabilities of the neural activity. While at one end daily psychological states or longitudinal motivational factors can influence BCI performance [1], on the other hand, gradual changes in fatigue and drowsiness levels during long neural activity recordings are also known to have significant influence on the recorded EEG and hence BCI performances [2,3,4,5]. Such vigilance state fluctuations over long durations of BCI usage can eventually hinder the neural intent inference and decision making pipelines which are usually calibrated daily. Hence, they further require updating or re-calibration of the system during long standing usage. In order to address this infeasibility in brain interface system designs, to the contrary of the significant amount of work exploring neural indicators of drowsiness levels [6,7,8], we highlight an approach towards drowsiness-invariant EEG feature extraction.

In this context, deep neural networks offer significant promise as recently popularized in the form of generic EEG feature extractors [9,10,11,12,13]. Despite being mostly investigated in offline studies [14], deep and complex architectures of such networks are usually assumed to be capable of capturing invariant neural activity to learn generalizable EEG feature representations. However going further, there exists a significant line of work on structured adversarial model training methods for invariant representation learning in many application areas of image processing, both for generative modeling to learn attribute-invariant encoded latent representations for data augmentation [15,16,17], as well as for discriminative models to learn task-distinguishing representations which simultaneously conceal some attribute-specific information [18,19]. In the light of this work, beyond regularizing conventional EEG feature extraction pipelines, we explore whether the recent progress in EEG deep learning architectures could be potentially extended to discover the underlying nuisance (e.g., drowsiness) invariant and BCI task-discriminative neural activity with structured, adversarial feature learning approaches.

In this study we present and assess the feasibility of adversarial feature learning on eliminating drowsiness effects (i.e., drowsiness-invariance) from EEG data to operate a BCI. We use experimental data recorded from 19 healthy participants while using an event-related potential (ERP) based BCI speller, namely the RSVP Keyboard™ [20], in offline copy-spelling mode for five consecutive calibration blocks, for a total of approximately one hour. We use temporal calibration block ordering as a discretized variable of the source of variability in the recorded neural activity as supported by concurrently collected introspective measures of boredom and sleepiness from the participants. Accordingly, we adversarially censor a conventional convolutional neural network (CNN) architecture to obtain drowsiness-invariant ERP detection models. Finally, we present our results regarding the decoding performance gains with features extracted via adversarial model training for drowsiness-invariance.

ADVERSARIAL FEATURE LEARNING

We propose discriminative model training under an adversarial learning framework that aims to learn nuisance-invariant features. Particularly in this application, we aim to learn an EEG feature extractor that exploits information which is ideally independent of the drowsiness levels of the participant during calibration recordings, as well as discriminative of the participant's target versus non-target

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-05

visual stimuli attention (i.e., ERP vs non-ERP). The overall purpose is to use such a task-specific invariant EEG feature extractor where drowsiness levels can fluctuate arbitrarily (e.g., later courses of the experiment session), and/or be different than the drowsiness level of the participant when the calibration EEG data is being recorded for model training. Here, we use within-session EEG recording blocks as a synthetically discretized measure of drowsiness levels. Based on the application context, this nuisance variable can vary beyond drowsiness levels, to calibration session day identification numbers (IDs) or subject IDs as well. Our notation is as follows:

- $\{(\boldsymbol{X}_i, l_i, b_i)\}_{i=1}^{n}$: Model training data set,
- $\boldsymbol{X}_i \in \mathbb{R}^{C\times T}$: Raw EEG data recorded from C channels for T discretized time samples at epoch i,
- $l_i \in \{0,1,\ldots,L-1\}$: Class label corresponding to the EEG epoch for the classification problem,
- $b_i \in \{1,2,\ldots,B\}$: Recording block identification (ID) number as a source of drowsiness variability.

A regular discriminative CNN architecture can be composed of a convolutional feature extractor (i.e., *encoder*) and a subsequent fully-connected output *classifier* layer with softmax units which generates an L dimensional output of normalized log-probabilities across classes. Specifically, a deterministic *encoder* $f = g(\boldsymbol{X};\mu_e)$ with parameters μ_e generates feature representations f which are further used as input by a *classifier* modeling the likelihood $p_{\mu_c}(l|f)$ with parameters μ_c to estimate l. Training of such a CNN is performed by minimizing the cross-entropy loss (i.e., maximizing the log-likelihood).
Adversarial training of this network will be performed by censoring the features f to contain as less information as possible regarding the nuisance variable b. Specifically, an *adversary* network modeling the likelihood $p_{\mu_a}(b|f)$ with parameters μ_a will aim to recover the recording block IDs b from the learned features in parallel. However while training the *adversary* in parallel to maximize the log-likelihood for b, the overall objective of CNN will have an additional adversarial censoring term to force the *encoder* to conceal information regarding b in the learned representations f. During this adversarial training game, the *encoder* will ideally learn features that will not be able to successfully recover b, but will also retain sufficient discriminative information regarding l. The overall objective function will be as follows:

$$\min_{\mu_e,\mu_c}\max_{\mu_a}\mathbb{E}[-\log p_{\mu_c}(l|f) + \lambda \log p_{\mu_a}(b|f)], \quad (1)$$

where the features $f = g(\boldsymbol{X};\mu_e)$. Note that a higher adversarial weight $\lambda > 0$ indicates enforcing stronger invariance, hence trading-off with discriminative power of the features. Here, setting $\lambda = 0$ corresponds to training a regular CNN model. Optimization of the *encoder-classifier* and *adversary* networks based on this objective were done alternatingly with stochastic gradient descent.

EXPERIMENTAL STUDY

Participants and Data Acquisition: Nineteen healthy subjects participated in this single-session study (six male, mean age = 33.42±13.18). Before the experiments, all participants gave their informed consent after the experimental procedure was explained to them. This study was reviewed and approved by the Oregon Health & Science University Institutional Review Board (#15331).
During the experiments, a DSI-24 dry-electrode EEG headset (Wearable Sensing, San Diego, CA) was used for data recordings. EEG data were recorded at 20-channels placed on the scalp according to the International 10-20 system with 300 Hz sampling frequency, Butterworth bandpass filtered at 1.5 to 42 Hz. Two ear-clip electrodes were used for reference placed on both earlobes, and FPz electrode location was used as ground.

Study Design: In this offline study, during the experiments, five consecutive calibration blocks were recorded while the participants were comfortably sitting on a chair placed in front of a computer screen and were using the RSVP Keyboard™, an EEG-based BCI speller that relies on the rapid serial visual presentation (RSVP) paradigm to visually evoke event-related brain responses [20]. Stimuli presentations are performed using twenty-eight characters including the English alphabet letters, and two additional symbols for backspace and space. During calibration block recordings, participants were instructed to visually attend to the center of the screen.
Each calibration block constituted of 100 *trials* and lasted 11 minutes. As a result, pure BCI system usage duration of each participant was approximately one hour. In each trial, the task was to attend to a particular *target* letter randomly determined at the beginning of the trial. Prior to each trial, the target letter cue for that trial was shown on the screen for two seconds, followed by a red cross presented for one second to indicate preparation for rapid serial visual presentation. Subsequently, trials included a series of 15 successive stimuli presentations (including the target letter) at a rate of 5 Hz at the center of the screen. Followed by a blank-screen inter-trial break, the next trial began with presentation of the new target letter. For each participant, per calibration block, experimental data constituted of 1500 stimuli presentations (100 target, 1400 non-target). EEG recording signal quality was re-examined between calibration blocks.
After each calibration block, introspective measures of boredom and sleepiness are obtained based on two self-rated questionnaires: the 6-item Boredom Questionnaire [21] and the Karolinska Sleepiness Scale (KSS) [22]. Fig. 1 presents the questionnaire results, demonstrating an overall increase in drowsiness and sleepiness levels of the participants as the experiments progressed. More experimental analyses on the neurophysiological correlates and implications of these states were previously studied based on the same experimental data set in [5].

EEG Data Analysis: EEG signal epochs were extracted from [0-600] ms post-stimuli time intervals. This

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-05

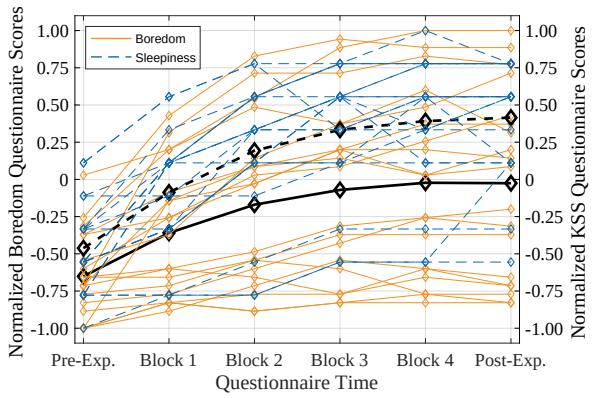


Figure 1: Normalized values of the documented introspective measures of boredom and sleepiness throughout the experiment by the 6-item Boredom Questionnaire and the Karolinska Sleepiness Scale (KSS) questionnaire. Each subjects' scores are represented by a pair of dashed blue and solid orange lines. Black and bold curves indicate the means across subjects. Greater values indicate higher boredom and sleepiness.

resulted in EEG data matrices of dimensions C=20 channels times T=180 discretized time samples. As the only pre-processing step, all trial channels were normalized to have zero mean and scaled to [-1,1] range by dividing with the absolute maximum value [12]. No offline channel selection or artifact correction was performed.

Adversarial Network Architecture: Proposed approach is applicable to any existing neural network based EEG feature extraction paradigm. For our demonstrations, the convolutional *encoder* we implemented was adapted from the EEGNet architecture [13]. We modified the architecture based on the input representations of our data set (e.g., sampling rate, number of channels, etc.). We used 8 temporal convolution units and two depthwise spatial convolution units per each temporally convolved feature map (i.e., the EEGNet-8,2 convention from [13]). Convolutional *encoder* specifications were as follows:

- Input EEG data dimensionality: (20,180)
- Convolution Block 1:
  - 8 x Temporal Conv1D (1,90)
  - Batch Normalization
  - 2 x Spatial DepthwiseConv1D (20,1)
  - Batch Normalization + ReLU Activation
  - Average Pooling: (1,3)
  - Dropout (0.25)
- Convolution Block 2:
  - 16 x Spatio-Temporal Conv1D (1,15)
  - Batch Normalization + ReLU Activation
  - Dropout (0.25)
  - Flatten
- Output feature vector dimensionality: (1,240)

Afterwards, the *encoder* output feature vectors f are used as input both to a *classifier* and an *adversary* network which were simply constructed as single fully-connected layers with L and B softmax output units respectively.

Implementation: We analyzed each participant's data individually and generated model learning and testing splits based on their calibration blocks. Specifically, the last 2 calibration blocks of the recording session was determined as the *testing set*, and models were trained based on the remaining first part of the data set (i.e., first 3 calibration blocks). From that part of the data set, we split 10% of the trials (randomly stratified by $b \in \{1,2,3\}$, $l \in \{0,1\}$ and pooled) as the *validation set*, and the remaining as the *training set*. This resulted in 4050 training, 450 validation and 3000 testing epochs.

At this stage we were not interested in investigating an optimal choice of adversarial regularization weight parameter λ, and we simply chose various arbitrary values ($\lambda \in \{0, 0.01, 0.05, 0.1\}$) for our pilot demonstrations. An intuitive way to determine is to repeat the model learning step while observing validation set *classifier* accuracies since higher adversarial regularization weights will start to impact discriminative performance.

All analyses were performed with the Chainer deep learning framework [23]. Networks were trained with 50 trials per batch for 200 epochs with early stopping based on validation loss, and parameter updates were performed once per batch with Adam [24]. Total number of network parameters to be estimated were 6,160. On average, total elapsed time for GPU model trainings were 57.0±20.7 seconds for $\lambda = 0$ and 57.8±16.5 seconds for $\lambda > 0$.

EXPERIMENTAL RESULTS

We observe relative contributions of adversarial feature learning with respect to its non-adversarial counterpart (i.e., $\lambda = 0$) in learning drowsiness-invariant features from three calibration blocks of data, to utilize the BCI speller in later phases of the experiments when drowsiness levels further increase based on introspective measurements. After model trainings were complete, three-class decoding *adversary* network performances on the *validation set* averaged across subjects were observed as in Tab.1. In the non-adversarial case, an *adversary* is only trained in parallel (i.e., with no adversarial loss feedback) to observe the amount of exploited drowsiness-specific information (i.e., leakage). High deviations for the non-adversarial model points to the fact that deep models can indeed exploit subject- and drowsiness-variant activity in decoding (up to 93% leakage with a participant). Stronger adversarial regularization can reduce block ID decoding accuracies to the 33% chance level.

Table 1: Three-class calibration block identification accuracies of the *adversary* network after model training on the *validation set* with varying λ values, averaged across subjects.

Adversarial Weight	**Validation Adversary Acc.**
$\lambda = 0$	$48.8\% \pm 13\%$
$\lambda = 0.01$	$44.0\% \pm 5\%$
$\lambda = 0.05$	$39.7\% \pm 5\%$
$\lambda = 0.1$	$38.2\% \pm 4\%$

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-05

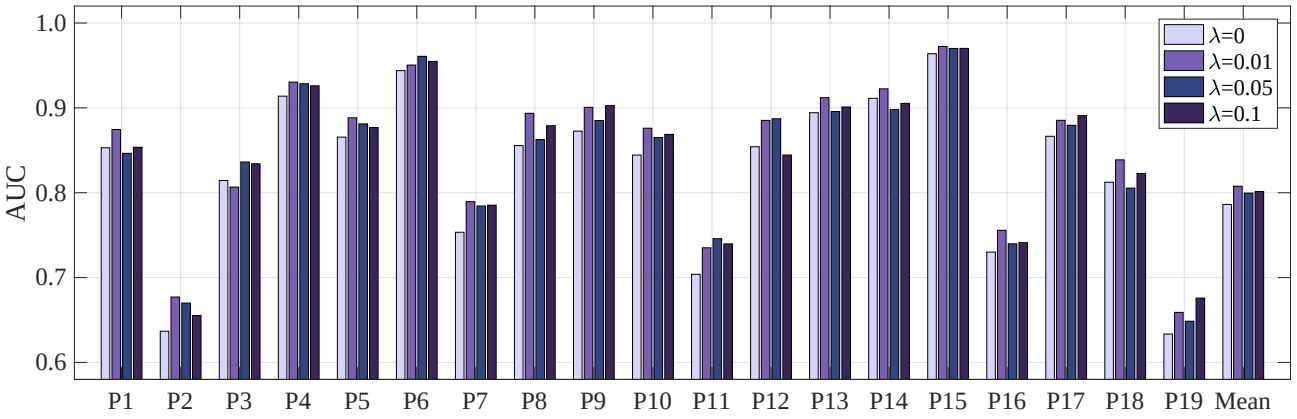


Figure 2: Area-under-the-curve (AUC) values calculated via the ROC curves of target detection on the testing set, for each subject, for varying values of λ. Regular CNN training corresponds to $\lambda = 0$, whereas $\lambda > 0$ indicates adversarial feature learning schemes.

An ideal BCI speller classifier would allow high sensitivity with a low false alarm rate. Hence, using the complete testing set EEG epoch samples, performance was estimated based on the area-under-the-curve (AUC) values calculated through the receiver operating characteristic (ROC) curves. Calculation of the ROC curves were based on sensitivity for target class detection versus false alarm rates on target detection. Fig. 1 demonstrates an overall summary of the AUCs for each participant's data with varying feature learning schemes. The mean bars denote the average AUCs across participants as $78.6\% \pm 20.7\%$ with $\lambda = 0$, and $80.7\% \pm 20.9\%$, $79.9\% \pm 20.7\%$ and $80.1\% \pm 20.7\%$ with $\lambda \in \{0.01, 0.05, 0.1\}$ respectively. Both on average and on an individual level, exploiting drowsiness-invariant features via the adversarial feature learning framework yields better AUCs than the $\lambda = 0$ bars. These differences were also found statistically significant by non-parametric Wilcoxon signed-rank tests on the null hypothesis of zero median across paired differences between non-adversarial and adversarial model AUCs: $\lambda = 0$ versus $\lambda = 0.01$ ($p < 0.0002$), $\lambda = 0.05$ ($p < 0.002$), and $\lambda = 0.1$ ($p < 0.0004$). This finding strongly relates to the nature of the regular deep learning frameworks where the models can arbitrarily exploit nuisance-specific, vigilance based temporal information to build the *classifier*. Structured adversarial learning yields improvements, but in some cases trading-off with discriminative performance when adversarial censoring is strong. Cross-validating λ based on the *validation set classifier* and *adversary* accuracies is a feasible option.

DISCUSSION

We present an adversarial invariant representation learning framework in the context of the recent progress on EEG deep learning studies. We argue that such an adversarial regularization approach could extend discriminative EEG feature extractors to further censor nuisance-specific variability in the data. In our pilot experimental studies, we assess the feasibility of adversarial feature learning particularly on eliminating drowsiness effects from event related EEG features extracted via a CNN-based encoder. Preliminary results demonstrate the feasibility of our approach in learning drowsiness-invariant representations on decoding performance.

In this work, it is important to highlight that we do not present a deep neural network model for feature learning in BCIs, but rather an adversarial regularization scheme that has not been considered in the EEG deep learning studies. We demonstrate our approach based on a recent CNN architecture (i.e., EEGNet [13]). Our approach is applicable to any existing neural network-based EEG feature extraction paradigm for BCIs. On a broader interpretation, one can explore deep invariant latent representation learning by disentangling other specific attributes (i.e., nuisance variables) from the representations via adversarial censoring such as variables based on calibration session days, or subject identifiers for across-subject transfer learning. Recently, we explored a similar approach using adversarial variational autoencoders in our preliminary work on subject-to-subject transfer feature extractor models [25], and for across-recording EEG-based biometric identification models [26]. Future work on our approach requires investigating the neurophysiological interpretations of the learned invariant features within the networks and what it promises for BCIs.

ACKNOWLEDGMENT

Our work is supported by NSF (IIS-1149570, CNS-1544895, IIS-1715858), DHHS (90RE5017-02-01), and NIH (R01DC009834).

REFERENCES

[1] Nijboer F, Birbaumer N, Kübler A. The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis–a longitudinal study. Frontiers in Neuroscience. 2010;4(55).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-05

[2] Ergenoglu T, et al. Alpha rhythm of the EEG modulates visual detection performance in humans. Cognitive Brain Research. 2004;20(3):376-383.

[3] Käthner I, Wriessnegger SC, Müller-Putz GR, Kübler A, Halder S. 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. 2014;102:118-129.

[4] Myrden A, Chau T. Effects of user mental state on EEG-BCI performance. Frontiers in Human Neuroscience. 2015;9:308.

[5] Oken B, Memmott T, Eddy B, Wiedrick J, Fried-Oken M. Vigilance state fluctuations and performance using brain computer interface for communication. Brain-Computer Interfaces. 2019 (to appear).

[6] Shi L-C, Lu B-L. EEG-based vigilance estimation using extreme learning machines. Neurocomputing. 2013;102:135-143.

[7] Martel A, Dähne S, Blankertz B. EEG predictors of covert vigilant attention. Journal of Neural Engineering. 2014;11(3):035009.

[8] Wei C-S, Lin Y-P, Wang Y-T, Lin C-T, Jung T-P. A subject-transfer framework for obviating inter-and intra-subject variability in EEG-based drowsiness detection. NeuroImage. 2018;174:407-419.

[9] Cecotti H, Graser A. Convolutional neural networks for P300 detection with application to brain-computer interfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011;33(3):433-445.

[10] Stober S, Sternin A, Owen AM, Grahn JA. Deep feature learning for EEG recordings, in Proc. International Conference on Learning Representations, 2016.

[11] Bashivan P, Rish I, Yeasin M, Codella N. Learning representations from EEG with deep recurrent-convolutional neural networks, in Proc. International Conference on Learning Representations, 2016.

[12] Schirrmeister RT, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping. 2017;38(11):5391-5420.

[13] Lawhern V, et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. Journal of Neural Engineering. 2018.

[14] 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.

[15] Edwards H, Storkey A. Censoring representations with an adversary, in Proc. International Conference on Learning Representations, 2016.

[16] Louizos C, Swersky K, Li Y, Welling M, Zemel R. The variational fair autoencoder, in Proc. International Conference on Learning Representations, 2016.

[17] Lample G, et al. Fader networks: Manipulating images by sliding attributes, in Proc. Annual Conference on Neural Information Processing Systems (NIPS), 2017, 5967-5976.

[18] Xie Q, Dai Z, Du Y, Hovy E, Neubig G. Controllable invariance through adversarial feature learning, in Proc. Annual Conference on Neural Information Processing Systems (NIPS), 2017, 585-596.

[19] Louppe G, Kagan M, Cranmer K. Learning to pivot with adversarial networks, in Proc. Annual Conference on Neural Information Processing Systems (NIPS), 2017, 981-990.

[20] Orhan U, Hild KE, Erdoğmuş D, Roark B, Oken B, Fried-Oken M. RSVP keyboard: An EEG based typing interface, in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2012, 645-648.

[21] Markey A, Chin A, Vanepps EM, Loewenstein G. Identifying a reliable boredom induction. Perceptual and Motor Skills. 2014;119(1):237-253.

[22] Gillberg M, Kecklund G, Akerstedt T. Relations between performance and subjective ratings of sleepiness during a night awake. Sleep. 1994;17(3):236-241.

[23] Tokui S, Oono K, Hido S, Clayton J. Chainer: a next-generation open source framework for deep learning, in Proc. Workshop on Machine Learning Systems in the Annual Conference on Neural Information Processing Systems (NIPS), 2015, 1-6.

[24] Kingma DP, Ba J. Adam: A method for stochastic optimization, in Proc. International Conference on Learning Representations, 2015.

[25] Özdenizci O, Wang Y, Koike-Akino T, Erdoğmuş D. Transfer learning in brain-computer interfaces with adversarial variational autoencoders, in Proc. IEEE EMBS Conference on Neural Engineering, 2019.

[26] Özdenizci O, Wang Y, Koike-Akino T, Erdoğmuş D. Adversarial deep learning in EEG biometrics. IEEE Signal Processing Letters. 2019;26(5):710-714.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-06

SENSOR TYING, OPTIMAL MONTAGES FOR VEP-BASED BCI

S. Ahmadi[1], M. Borhanazad [1], D. Tump [2], J. Farquhar [1], P. Desain[1]

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[2]MindAffect, Nijmegen, The Netherlands

E-mail: s.ahmadi@donders.ru.nl

ABSTRACT: As Brain computer interfaces (BCI's) attract more attention in the customer market, the need for easy-to-use headsets increases. In this paper we propose an optimized fixed montage EEG headset for VEP BCIs based on covering the most relevant areas on the skull with simulated large sensors. To obtain large sensors, we propose to tie (that is short-circuit) multiple sensors and we simulate the tying by averaging the signals. We show that a circular center-surround configuration with proper sensor tying can provide high performance and good robustness to misplacement of the headset while the number of required channels can be as few as eight. We also provide an alternative cheap design requiring only two channels, which can still achieve acceptable average performance.

INTRODUCTION

Recently, Brain Computer Interfaces (BCIs) have attracted more attention in the general consumer market. Contrary to current BCIs used in a lab setting, with long and precise cap fitting, this new group requires short cap fitting with a robustness against slight displacement [5]. Therefore, there is an increased need for easy-to-use headsets.

One way to make headsets easier to use is with a fixed montage. However, this makes the headset dependent on the type of brain response expected. Visual evoked potential (VEP), especially broad band VEP BCIs (BBVEP) are found to be the fastest BCIs [1], and therefore the most suitable to bring to users. The spatial activity patterns reported for VEP systems clearly show that the brain response is spatially localized to the occipital lobe [2, 7]. The full EEG cap can therefore be replaced with a headset measuring only around the area of interest, reducing the number of channels and simplifying the design. The reduction in number of channels not only reduces amplifier costs, it also reduces computation costs. This is especially important when the BCI systems need to be more compact, such as embedded on a chip, or run on batteries.

One aspect of usability on the consumer market is that the headset needs to be robust to slight misplacement. This would imply that we need to measure the significant signals accurately, even when the EEG sensors might be positioned at slightly different location. One way to achieve this without requiring additional channels is simply to use physically larger EEG sensors spaced further apart. As moving such a large sensor only has a small effect on the measured area of the scalp, the output should be insensitive to misplacement. As large sensors are not commonly available, we suggest to construct such a sensor by electrically tying together multiple sensors into one channel. In this paper, we simulate sensor tying by means of averaging their signals. Under some conditions, like roughly similar impedances, such averaging of individual sensor measurements can be shown, at least to a first approximation, to give a similar result as physical electrical tying [6].

In this paper we develop an optimized fixed montage for VEP BCIs by comparing different montages in a BBVEP visual speller [7]. Montages are evaluated in terms of their accuracy for an average and worse subject. Robustness to misplacement is also evaluated in terms of the worst case drop in accuracy for a given displacement. We conclude by identifying the 'optimal' montage for VEP BCIs which demonstrate reliable performance whilst maximising ease-of-use in a home setting.

MATERIALS AND METHODS

Design ideas: The average spatial filter of the BBVEP speller, found by [7], suggests that the most important areas for decoding during the BBVEP task is a classic Laplician-derivation with circular center-surround shape at the back of the head (Fig.1a). Therefore we propose a circular design, covering the active area with a center-surround configuration, an invention described in [3], with an inner circle and outer ring (Fig.1b). To allow for the fact that not all subjects have an ideal circularly symmetric response, like the one captured by the atypical spatial filter of [7] shown in Fig.1c, we propose different ways of breaking each area into sub-regions (Fig.1d). Thus, to identify the 'optimal' center-surround montage we need to identify the correct size for the center and the ring, the number of sub-regions and the sub-regions shapes.

Experimental setup: Physically making and evaluating all possible montages would require excessive subject recording hours. Thus we evaluated possible mon-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-06

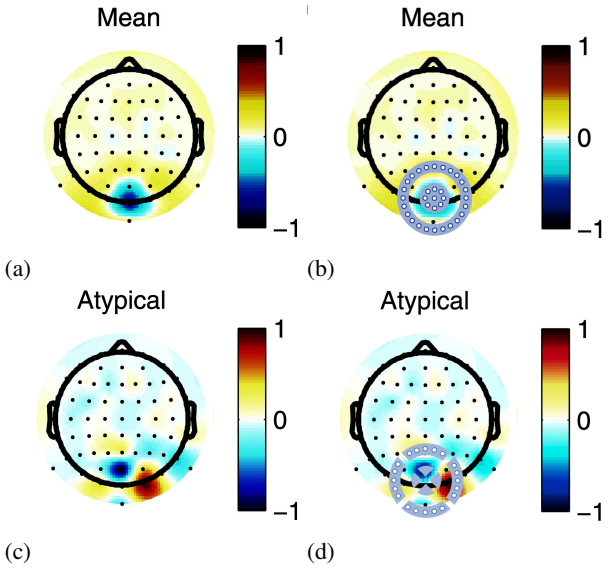


Figure 1: a: The grand average spatial filter of the BBVEP visual speller. (Figure adapted from [7]). b: proposed circle design c: an example of circularly asymmetric spatial filter d: Breaking the circle design to deal with the possible asymmetric spatial response.

tages in off-line simulation by averaging the sensor measurements in each sub-region from an off-line experiment recorded with a high-density EEG system (256 sensors with 20mm sensor spacing).

Participants were seated at a distance of 60 cm from the screen. EEG data was recorded with a 256 Biosemi cap with gel electrodes at a rate of 360 Hz. A Biosemi-active2 amplifier was used and the data was recorded based on a HQ reference system. The BBVEP speller consisted of 3 blocks of 36 trials. In each block the subject was instructed to attend to 36 characters in a random order. The order was different for each training block but the same across subjects. Each trial had a length of 4.2 seconds, with an inter trial time of 0.5 second.

Stimuli: Stimuli consisted of a flickering matrix-layout keyboard presented at 60Hz. All buttons flickered from black to white colors according to a Gold code [4]. The modulated Gold codes ensure small cross-correlations between flickering patterns. The flickering pattern consists of two event types; long and short flashes.

Participants: 5 healthy adults (age: 37.6 ± 16.6, mean ± standard deviation, 1 female) participated in the study after having signed written informed consent. All participants had normal to corrected-to-normal vision and no history of epilepsy. The study was approved by and conducted in accordance with the guidelines of the Ethical Committee of the Faculty of Social Sciences at the Radboud University.

Analysis: To simulate various montages, we consider each individual measuring point on the skull as one sensor, which in our experiment setting is one of the 256 electrodes. To form one channel covering a wider area (consisting of multiple sensors), we tie multiple sensors together by averaging the signals of the sensors.

We apply sensor tying in the circle design in two ways:

1. C-C montage. Two channels comprising of one filled circle in the middle surrounded by another circle (ring) which diameter is between 12.5 to 45 mm larger. Each circle is formed by tying the sensors together into one channel located in the area covered by that circle (Fig.2a).
2. CS-CS montage. One filled circle in the middle divided into 4 segments, surrounded by another circle (ring) which diameter is between 12.5 to 45 mm larger. The second circle is also divided into 4 segments and altogether form 8 channels (Fig. 2b)

The resulting signals after sensor-tying were then fed into our standard BBVEP analysis pipeline, which is given in [7]. See [7] for detailed information, but in outline this consisted of; spectral filtering with a pass band of 2-48Hz followed by model-fitting using CCA to identify subject-specific spatial filters and pulse responses. During testing the pulse responses are re-convolved with the stimulation pattern for each possible output to generate per-output template responses. The predicted target is then identified as the output with maximal correlation between the spatially filtered data and the template response. Accuracy, defined as the percentage of trials where the predicted output was correct, is used in this paper to assess the performance of a montage.

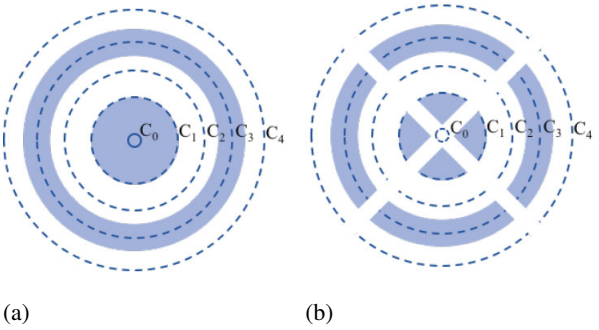
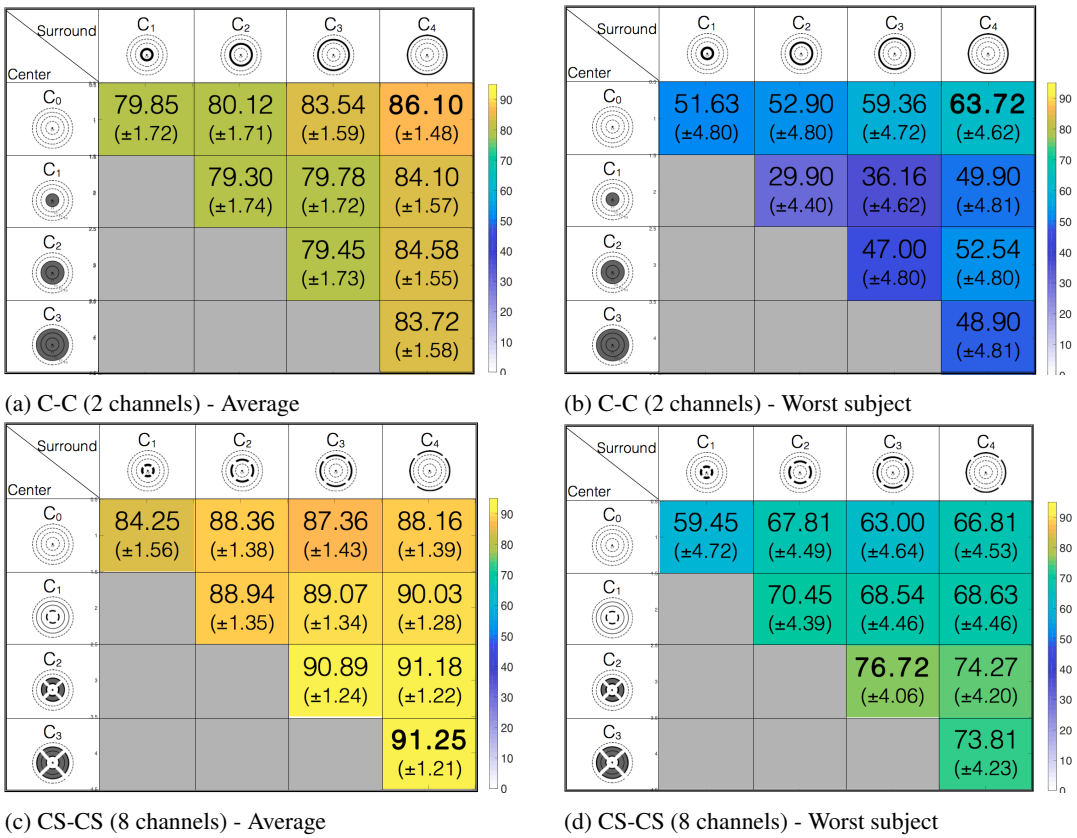


Figure 2: example of a:the C-C montage (C1-C3) and b: CS-CS montage (C1S-C3S). The dashed circles show possible circle sizes.

To study the effect of circle diameter and find the optimum diameter for the inner (center) and outer (surround) circles, we vary the sizes within four values. For the center circle, one of the four following circles is used: C_0 ⌀2mm (a single sensor), C_1 ⌀40mm, C_2 ⌀65mm, C_3 ⌀95mm. Similarly, the surround circle has one of the following four sizes: C_1 ⌀20mm, C_2 ⌀65mm, C_3 ⌀95mm and C_4 ⌀120mm (See Fig.2). The step size in circle diameters is of order of the distance between two adjacent sensors in the Biosemi-256 cap. In the rest of this paper we use these circle names to address various configurations of each montage. For example C_0-C_3 is the 2 channel montage with a single sensor as the center circle and C_3 ⌀95mm as the surround circle. Similarly C_1S-C_3S is the 8 channel montage with C_1 ⌀20mm as center circle and the C_3 ⌀95mm as the surround one.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-06



(a) C-C (2 channels) - Average

(b) C-C (2 channels) - Worst subject

(c) CS-CS (8 channels) - Average

(d) CS-CS (8 channels) - Worst subject

Figure 3: Average and worst performance obtained by simulating the two center configurations with varying center and surround circles. (The numbers in brackets are the 95% confidence interval)

Furthermore, effects of slight displacement of a headset are simulated by choosing different center points. We shifted the center point by two steps horizontally and two steps vertically. Each step size was almost 20mm (one sensor on the cap). The simulation was repeated for each center point by tying the sensors that would be covered by the circles centered at that position. We take the maximum performance drop after 20mm shift as the indicator of the sensitivity to misplacement.

For the analysis presented here 1 second trials are used to avoid saturation effects where all subjects hit perfect performance, and hence increase sensitivity to changes in the montage.

RESULTS

We report the average classification accuracy as well as the accuracy of the worst subject. The worst subject is included to show how each configuration works in a challenging condition.

Figure3 shows the average and worst accuracy, as the percentage of correctly typed characters, for the C-C and CS-CS montages for all the possible center and surround configurations. The numbers in brackets show the 95% confidence interval.

In Fig.4 The maximum relative performance drop as a result of moving the central point by 20mm is visualized for all the configurations of both C-C and CS-CS montages.

DISCUSSION

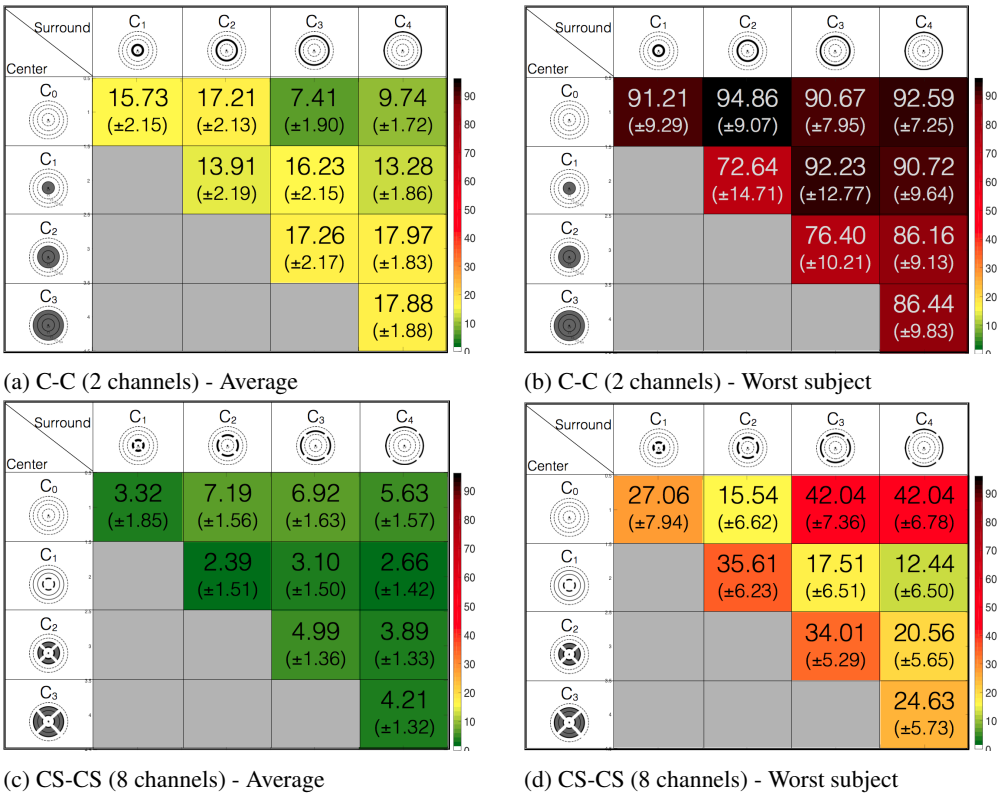
As can be seen in Fig.3a and 3b, increasing the surrounding circle in size in the C-C montage results in a higher performance. Increasing the size of the center circle only decreases the performance. For the CS-CS montage however, as Fig.3c and 3d indicate, an increase of the center circle size increases the performance. Moreover, the performance increase as a result of larger surround circle is less compared to the C-C montage. In this case, both 95mm and 120mm surround circles give similar high performance (to within the 95% confidence intervals).

According to Fig.3a and 3b, the best C-C two channel configuration in terms of average and worst performance is when the configuration has a single point center with a 120mm surrounding circle (C_0-C_4).

The 8 channel configuration (CS-CS montage) as can be seen in Fig.3c and 3d, shows higher performances with maxima reached at 95mm center circle and 120mm surround circle (C_3S-C_4S) for average performance and at 65mm center circle and 95mm surround circle (C_2S-C_3S) for the worst subject. The difference in performance, however, between the 65mm and 95mm center circle and 95mm and 120mm surround circle is not statistically significant. The average accuracy obtained using the full 256 channels was 92.49±1.21%. Clearly, the difference between the performance of our optimal CS-CS montage (C_3S-C_4S) and the full cap is insignificant. The same

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-06



(a) C-C (2 channels) - Average

(b) C-C (2 channels) - Worst subject

(c) CS-CS (8 channels) - Average

(d) CS-CS (8 channels) - Worst subject

Figure 4: % Relative drop in average and worst subject accuracy as a result of 20mm shift of the central point for both C-C and CS-CS montages (The numbers in brackets are the 95% confidence interval).

holds for the worst subjects performance (78.54±3.95% obtained by the full cap compared to 76.72±4.06%) of the C_2S-C_3S montage

It is clear that the 8 channel configuration is superior to the 2 channel one, especially for the bad subject. However, the average performance of the 2 channel configuration is good enough to make such a montage an attractive option for making a cheap EEG headset.

The difference between the 2 channel C-C montage and the 8 channel CS-CS one becomes more visible by looking at the sensitivity of each montage to misplacement. Fig.4 shows the percentage drop in accuracy as the center point of each montage moves 20mm away from the ideal point. The drop reported is the maximum drop in performance of the four possible displacement directions. Comparing Fig.4a and 4c shows that the CS-CS montage is overall less sensitive to misplacement (2.4% up to 7.2% relative drop in performance depending on the circle sizes, compared to 7.4% up to 17.97% of the C-C montage). While the results in Fig.3c indicate that the best center size for achieving maximum average performance of the CS-CS montage is 95mm (C_3), according to Fig.4c, there seem to be a slight preference for a 40mm center circle (C_1) in terms of robustness to misplacement, however, this difference is not statistically significant. According to Fig.3a, the best result of the C-C montage is obtained when there is a single electrode inner-circle (C_0) with a far outer circle. However, this effect is strongly caused by the worst subject. Moreover, in terms of robustness to misplacement, in this case, the 95mm surround circle (C_3) has a slight advantage to the 120mm one.

As illustrated in Fig.4b, the performance of the C-C montage for the worst subject can drop dramatically by shifting the headset 20mm to the side. However, the CS-CS montage seem to handle such a condition very well (Fig.4d). This can be an example of a circularly asymmetric response due to a non-radial dipole (see Fig1c). In this case, having very large sensors increases the chance that a single sensor is placed across the boundary between the positive and negative components and averages the signals that need to be kept separate for detection. This can also explain why the best C-C configuration for the worst subject is the one with a single point sensor in the center. By splitting the channels, the risk of cancelling out the activities reduces and therefore the placement of the sensors become less difficult. Hence, in this type of design, on one hand large sensors are desirable since they save the number of required channels and on the other hand the sensor should not be too large to mix up different activities and cancel them out. The 8 channel CS-CS montage seem to provide a good trade off.

In short, the 8 channel CS-CS montage seems to provide a higher average accuracy than the 2 channel C-C montage. Importantly for a subject with poor performance, the difference between the CS-CS and C-C montage is even larger. An further advantage of the 8 channel CS-CS montage is that it is relatively robust against misplace-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-06

ment. This robustness increases the usability of a headset for a consumer. The only disadvantage is the increase in channel count and associated costs.
In this paper, sensor tying is defined as an offline averaging over the individual signals of the separate electrodes. In future work we will validate the reliability of this off-line simulation by performing further measurements for the identified best-performing montages with physical sensor tying in an on-line test. This research does, however, show the possible benefits of sensor tying. The robustness against slight displacement, together with a reduction in the number of channels, would make a headset more suitable for home users.

CONCLUSION

In this paper we proposed a fixed montage EEG acquisition headset idea for VEP-based BCI with a focus on home-use customer product design. Our idea was based on covering the most important areas of the skull with a few large sensors in an optimal layout. To evaluate the idea, we did a high density EEG recording and averaged signals from the multiple adjacent sensors to simulate larger sensors in various layouts. We concluded that an 8 channel circular center-surround montage with center size of 95 and surround size of 120mm, results in maximum average classification accuracy which is almost identical to what can be achieved using the full high density measurements. Such a design also provides very good robustness to misplacement such that moving it by 20mm causes only 4% relative drop in performance. The design is also very robust for subjects with asymmetric spatial responses. We also showed that it is possible to reduce the number of channels to 2 and still have acceptable average performance. Even though the placement of such a headset has to be more accurate, it still makes an attractive design for low budget devices.

REFERENCES

[1] Bin Guangyu, Gao Xiaorong, Wang Yijun, Hong Bo, Gao Shangkai. VEP-based brain-computer interfaces: time, frequency, and code modulations [Research Frontier]. IEEE Computational Intelligence Magazine. 2009;4(4).
[2] Creel Donnell. Visually evoked potentials. 2012.
[3] Desain Peter. Bioelectrical sensing electrode assembly. Patent P6071656NL (NL). 2018.
[4] Gold Robert. Optimal binary sequences for spread spectrum multiplexing (corresp.) IEEE Transactions on Information Theory. 1967;13(4):619–621.
[5] Kübler Andrea et al. The user-centered design as novel perspective for evaluating the usability of BCI-controlled applications. PLoS One. 2014;9(12):e112392.
[6] Oostendorp Thom F., Van Oosterom Adriaan. The surface Laplacian of the potential: Theory and application. IEEE Transactions on Biomedical Engineering. 1996;43(4):394–405.
[7] Thielen Jordy, Van Den Broek Philip, Farquhar Jason, Desain Peter. Broad-band visually evoked potentials: Re (con)volution in Brain-computer interfacing. PLoS ONE. 2015.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-07

INTERPRETABLE RIEMANNIAN CLASSIFICATION IN BRAIN-COMPUTER INTERFACING

Jiachen Xu[1,2], Moritz Grosse-Wentrup[1,3,4], Vinay Jayaram[1]

[1]Max Planck Institute for Intelligent Systems, Department of Empirical Inference, Tübingen, Germany
[2]Technical University of Munich, Department of Electrical and Computer Engineering, Munich, Germany
[3]Ludwig Maximilian University of Munich, Department of Statistics, Munich, Germany
[4]University of Vienna, Faculty of Computer Science, Vienna, Austria

E-mail: {jiachen.xu, moritzgw, vjayaram}@tuebingen.mpg.de

ABSTRACT: Riemannian methods are currently one of the best ways of building classifiers for EEG data in a brain-computer interface (BCI). However, they are computationally complex and suffer from a lack of interpretability. Since the full covariance matrix is used for each classification, it is not immediately possible to see what underlying signals are generating the classified changes in variance. Particularly in a rehabilitation context, where it is essential to control which brain signals are used for classification, this can be a severely limiting factor. Further, the requirement to perform a matrix logarithm can become prohibitively complex for real-time computation. In this work, we explore a method for extracting spatial filters from a solution in the Riemannian tangent space and compare it against common spatial patterns. We show via comparisons on multiple open-access datasets that it is possible to generate filters that approach the performance of the full Riemannian solution while maintaining interpretability.

INTRODUCTION

One of the major reasons that spatial filtering has enjoyed decades of popularity within the field of brain-computer interfacing is its interpretability. In contrast to most other methods, spatial filters can be directly evaluated to see whether they represent sources of brain activity or correlated noise sources, which is crucial in the fields of rehabilitation and neurofeedback. As they serve the dual purposes of dimensionality reduction and explainability, there has been a large amount of research on how to calculate optimal ones for a given task. From the original paper on common spatial patterns (CSP) [1] to modern divergence-based approaches [2], hundreds of manuscripts have been written on how to find them.
In parallel to this development, in the last few years, there has been an upsurge of interest in methods based on Riemannian methods [3, 4]. By considering the sensor covariance matrix as a whole, instead of trying to isolate particular sources of interest, it is possible to overcome many of the non-stationarity issues that have plagued machine learning in EEG, resulting in tangent-space methods out-performing all others in a recent meta-analysis [5]. In contrast to the often complex optimization methods required for modern approaches to spatial filtering, and the uncertain gains they provide, Riemannian tangent space methods are amenable to standard machine learning algorithms, which significantly increases their ease of use. The downside is that, by using the entire covariance matrix, it is no longer possible to determine what exact sources are generating a difference between experimental conditions. This is of little concern in consumer BCIs, where the goal of the device is usability for healthy subjects, but in the case of neurofeedback or patient groups, this is a severely limiting factor. Proof that a brain signal, and not a correlated physiological signal such as muscular activity, is the driving force for a BCI is necessary for these contexts, and so it is not yet possible to use Riemannian methods there.
Our contribution is the following: Based on recent findings [6], we validate a method for decomposing a linear function in the Riemannian tangent space into a set of spatial filters. We show in tests on multiple datasets that these spatial filters approach the performance of the full Riemannian method and further that they are more robust than CSP to different numbers of filters used. We also show associated spatial patterns to validate that the filters capture relevant brain regions. In summary, our contributions render Riemannian classification methods applicable to BCIs for rehabilitation and neurofeedback.

METHODS

BCI paradigms commonly influence the amplitude of brain rhythms. To create better features, spatial filters have been used to isolate the component of the recorded signal corresponding to a rhythm of interest. These filters are often found by maximizing variance differences, as the variance is a mathematically convenient proxy for amplitude.
One of the most representative and common data-driven spatial filters is CSP [1, 7], which is computed by find-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-07

ing the sources that maximize the variance ratio between two conditions. CSP has made a great contribution to the stable increase of BCI's performance and has over the years spawned many improved versions, such as filter bank CSP [8], regularized CSP [9] and divergence-based CSP [2]. In addition to increased decoding performance, another important reason behind the popularity of CSP is its interpretability. By observing the spatial patterns associated with discovered spatial filters, it is possible to verify whether they are related to neural or other sources. Unlike spatial filtering, which uses variance-based features, Riemannian methods are based on directly comparing sensor covariance matrices. In this paper, the adopted Riemannian metric is called the affine-invariant Riemannian metric (AIRM) [10], where the "affine-invariance" means that the distance between two symmetric positive definite (SPD) matrices is invariant to the distance after applying an affine transformation to both matrices [11], i.e.,

$$\delta_R(\mathbf{TC}_1, \mathbf{TC}_2) = \delta_R(\mathbf{C}_1, \mathbf{C}_2) = \left\| \mathrm{Log}\left(\mathbf{C}_1^{-1}\mathbf{C}_2\right) \right\|_F, \quad (1)$$

where $\|\cdot\|_F$ represents the Frobenius norm, $\mathrm{Log}(\cdot)$ means the logarithm of a matrix, $\mathbf{C}_{(.)}$ are SPD matrices, and $\mathbf{T}$ is any affine transformation matrix. While this method returns a very robust estimate of distance, most machine learning algorithms require a Euclidean space to work. This can be done via projecting manifold points onto the tangent space to a given point.

The tangent space is defined as a real vector space centered around a reference point on the manifold. In practice, this point is normally selected as the mean point across all trials [10]. After fixing the reference point, the covariance matrices on the manifold are mapped onto this space using the logarithmic map (Eq. 2) and then vectorized into the tangent vectors (TVs).

$$\mathrm{Log}_{\mathbf{C_1}}(\mathbf{C}_2) = \mathbf{C}_1^{\frac{1}{2}} \mathrm{Log}\left(\mathbf{C}_1^{-\frac{1}{2}}\mathbf{C}_2\mathbf{C}_1^{-\frac{1}{2}}\right)\mathbf{C}_1^{\frac{1}{2}}, \quad (2)$$

$$\mathrm{Exp}_{\mathbf{C_1}}(\mathbf{C}_2) = \mathbf{C}_1^{\frac{1}{2}} \mathrm{Exp}\left(\mathbf{C}_1^{-\frac{1}{2}}\mathbf{C}_2\mathbf{C}_1^{-\frac{1}{2}}\right)\mathbf{C}_1^{\frac{1}{2}}, \quad (3)$$

where $\mathrm{Exp}(\cdot)$ means the exponential of a matrix. Eq. 2 represents the logarithm mapping of $\mathbf{C}_2$ from the manifold to the TS based on the point $\mathbf{C}_1$ and Eq. 3 shows the inverse projection, i.e., the exponential mapping. For the affine-invariant and log-Euclidean metrics, the logarithmic map requires computing the matrix logarithm of its argument, as a way of mapping the eigenvalues of the input matrices from the strictly positive space of SPD matrices to a real-valued space. As the tangent space represents a linearization of the manifold, distances between points approximate true manifold distances, and geometry on the manifold is preserved, which allows for linear machine learning methods to be used.

A major drawback of this procedure is the difficulty in interpreting the resulting classifier because the tangent space embedding cannot be easily linked to any intuitive physical process. We address this problem by introducing a procedure to extract spatial filters from a linear function in the tangent space, which allows for a unified methodology in the computation of spatial filters that can also take advantage of the large body of work on machine learning in linear spaces.

The procedure can be summarized in three steps: Classification, spatial filter generation, and filter selection. For classification, we follow the approach in [12] (linked to point 2 to point 5 in Algorithm 1, abbreviated as p2 to p5 afterwards) and find the optimal SVM solution in the affine-invariant tangent space (p6). Though any algorithm is possible, we chose a regularized SVM as maximum-margin classifiers can deal well with the high dimensionality of the tangent space. Once a linear classifier is computed, we project the weight vector in the tangent space back to the SPD manifold via the exponential map (Eq. 3) (p8 to p9). From here, we compute spatial filters by taking the solution to the generalized eigenvalue (GE) problem between the tangent space reference point and the exponential map of the weight vector (p10).

Algorithm 1 Extraction of Tangent Space Spatial Filters

INPUT: Band-pass filtered raw signal $\mathbf{X}_i,\ \forall i = 1, ..., T$
OUTPUT: Spatial filter after ordering $\mathbf{F}$

1: **procedure** I. CLASSIFICATION
2: $\mathbf{C}_i = \mathbf{X}_i\mathbf{X}_i^T\ \forall i = 1, ..., T$
3: $\mathbf{C_m} = \underset{\mathbf{C}\in C(n)}{\arg\min} \sum_{i=1}^{T} \delta_R^2(\mathbf{C}, \mathbf{C}_i)$ {AIRM mean}
4: $\mathbf{S}_i = \mathrm{Log}_{\mathbf{C_m}}(\mathbf{C}_i)$ {Mapping from manifold to TS}
5: $\overrightarrow{S}_i = \mathrm{vec}(\mathbf{S}_i)$ {Vectorize the upper triangular matrix}
6: $\overrightarrow{W} = \mathrm{SVM}\left(\overrightarrow{S}_i, y_i\right)$ {Weight vector on the TS}
7: **procedure** II. SPATIAL FILTER GENERATION
8: $\mathbf{S}_w = \mathrm{devec}\left(\overrightarrow{W}\right)$ {Convert the vector to SPD matrix}
9: $\mathbf{C}_w = \mathrm{Exp}_{\mathbf{C_m}}(\mathbf{S}_w)$ {Mapping from TS to manifold}
10: $\mathbf{C}_w\mathbf{V} = \mathbf{C_m}\mathbf{V}\Gamma$ {Mapping from TS to manifold}
11: **procedure** III. FILTER SELECTION
12: $\widetilde{\mathbf{X}}_i = \mathbf{V}^T\mathbf{X}_i$ {Filter raw signal using full components}
13: $\overrightarrow{f_i} = \log(\mathrm{var}(\widetilde{\mathbf{X}}_i))$ {Compute the log-variance}
14: $\overrightarrow{\beta}$ = LinearRegression $\left(\overrightarrow{f_i}, y_i\right)$ {Weight vector of fitted classifier}
15: $\overrightarrow{id}$ = sort $\left(\left|\overrightarrow{\beta}\right|\right)$ {Sort the absolute value of weight coeficient in descending order}
16: $\mathbf{F} = \mathbf{V}\left[:, \overrightarrow{id}\right]$ {Get the re-ordered spatial filter}

The eigenvectors that solve a GE problem simultaneously diagonalize both input matrices, which can be interpreted as finding a space where the weight SPD matrix and the input SPD matrices are all roughly diagonal. This independence means that the logarithm of the matrices in this space is close to the log-variances of each of the filtered dimensions, allowing us to approximate the tangent space linear function as a linear function of the log-variances

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-07

Table 1: Summary of all MOABB datasets with left-hand versus right-hand motor imagery. The dataset marked in red color has only 3 channels and is hence excluded from this analysis.

Dataset Name	#Channels	#Subjects	#Sessions
BNCI 2014-001	22	9	2
BNCI 2014-004	3	9	5
Cho et al. 2017	64	49	1
Munich Motor Imagery	128	10	1
Physionet Motor Imagery	64	109	1
Shin et al. 2017	25	29	3
Weibo et al. 2014	60	10	1
Zhou et al. 2016	14	4	3

in the filtered space. However, the associated eigenvalues have no relationship to the linear function, requiring another criterion in order to determine which of the computed filters are most relevant to classification. In order to do this, we fit a linear function to the log-variances of our projected training data (12 to 14) and attempt to reconstruct the function values of the true tangent space function. Based on the absolute values of the coefficients from this regression problem, we rank the GE solution filters (15 to 16).

To evaluate the performance of our method in a rigorous and reproducible fashion, we adopted the datasets and framework from the open source benchmark *Mother of all BCI Benchmark (MOABB)* in our work [5]. For simplicity, we choose left- versus right-hand motor imagery, which is one of the most common paradigms in EEG-based BCIs. We restrict our analysis to the α- and β-bands (8Hz $\sim$ 32Hz). We use all available channels except those used for electrooculography (EOG). We exclude the dataset *BNCI 2014-004* from our analysis (marked in red in Table 1), because this data set has only three channels.

After bandpass filtering the raw EEG signal, the corresponding covariance matrices are estimated via the empirical covariance estimator. Afterward, the spatial filter is extracted according to the pseudocode shown in Algorithm 1, and the classifier is fit using the log-variances of the filtered signal. As for the classification algorithm, we choose Fisher's Linear Discriminant Analysis (LDA) for simplicity. Decoding accuracy is estimated via five-fold cross-validation within each data set.

Two statistical parameters are utilized to compare our method against CSP and the full Riemannian approach: the p-values and the effect sizes. The p-value for the one-sided test is computed within each dataset, whose null hypothesis is that the median accuracy of using CSP is not larger than that of applying the spatial filter designed by us. Effect sizes show the standardized mean difference (SMD) between the accuracies of the two compared methods. All statistical tests are implemented as described in [5].

RESULTS

Meta-analysis:

As a spatial filtering technique, the natural comparison for TS SF is against CSP. Therefore, we compared them under three typical scenarios: applying very few filters (for speed), using filters corresponding to half the number of channels (for the compromise between speed and accuracy) and using all filters (for accuracy). To ensure the results are rigorous, the comparison is done across seven datasets from MOABB [5] (as shown in Table 1).

When merely applying two spatial filters (Fig. 1a), TS SF is generally better than CSP, but the positive significance is only present in two datasets: *Shin* and *Munich*. After increasing the filter number to the half the channel dimensionality (Fig. 1b), the positive significance appears in three more datasets *Weibo*, *Physionet* and *Cho*, in addition to an increased meta-effect. Moreover, the SMD of data set *Munich* has increased to 1.5 which implies a strong effect. Lastly, when using all filters (Fig. 1c), the effect sizes are lessened in the datasets with significant performance, except for *Weibo*.

Classification accuracy w.r.t. the number of applied filters:

In addition to comparing against CSP, we are also interested in seeing how well TS SF can approximate the accuracy of the full tangent space classifier, as well as how this approximation evolves as a function of the number of filters. Therefore, we choose three representative datasets *Munich*, *Cho*, *Physionet* and plot their accuracy w.r.t. the number of applied filters in Fig. 2, showing the tangent space accuracy as well as an upper bound.

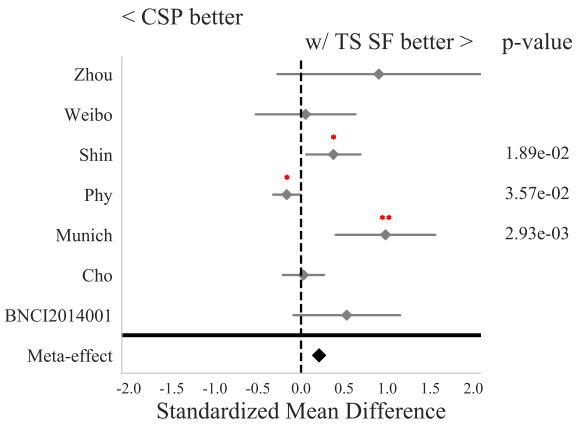
When using data set *Munich* (Fig. 2a), TS SF (green line) first outperforms CSP (red line) when the number of applied filters is less than 16. Afterward, their accuracies are tied with each other until using more than 40 filters. Subsequently, TS SF maintains its stable performance, but the accuracy of CSP starts to decrease.

The information conveyed by Fig. 2b and Fig. 2c is quite similar to each other. TS SF and CSP perform very comparably up until around 13 filters are used. After this point, using more CSP components hurts the performance, while the performance stays much more stable when using more TS SF filters. However, in Fig. 2c, the accuracy drops when using TS SF at high numbers of filters.

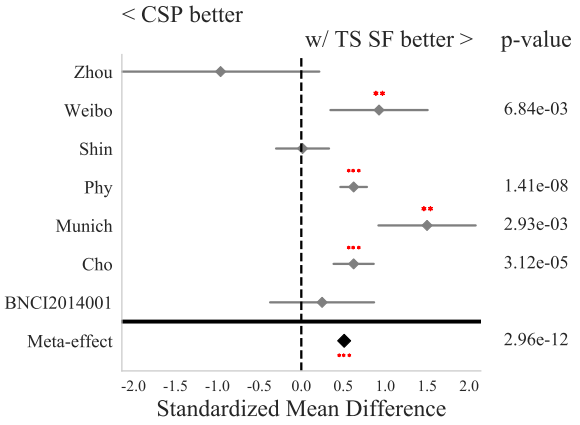
Overall, the performance of the full Riemannian method (Full TS, brown line) is always superior to both spatial filtering methods because the full approach utilizes the full information decoded in the covariance matrices.

Published by Verlag der TU Graz
Graz University of Technology

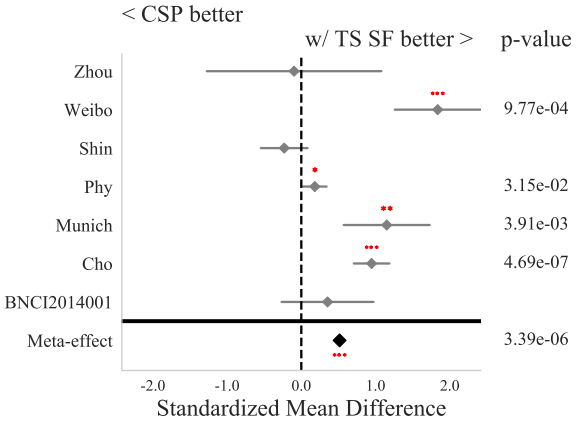
 DOI: 10.3217/978-3-85125-682-6-07



(a) Applying the first two filters



(b) Applying the first half of all filters



(c) Applying all filters

Figure 1: Meta analysis of accuracy after applying CSP or TS SF as spatial filter. Parameters: p-value p and SMD are computed within each dataset. Red *, ** and *** represent $p < 0.05, 0.01, 0.001$ respectively. Grey diamonds signify the SMD, while grey bars show the confidence interval of the mean.

Spatial patterns:
From Fig. 1 and Fig. 2 we know that TS SF on average outperforms CSP in terms of accuracy. For a spatial filtering method, however, its interpretation is as crucial as its classification accuracy. Therefore, we compared the corresponding spatial patterns of both TS SF and CSP in order to see what sources were chosen. We choose the data set *Munich* because it has the best overall data quality [5] and highest channel number (128), and plot the two best filters using both CSP and TS SF. In order to convert spatial filters into spatial patterns, we use the approach proposed by [13].

When comparing Fig. 3a and 3b there are some interesting results. First, the spatial patterns from CSP tend to be more patchy than from TS SF, e.g., S1, S2, S10, in particular for those subjects noted as having low data quality [14]. Second, for the subjects who have excellent signal quality, the spatial patterns from TS SF are highly consistent with from CSP, e.g., S3, S4, and S6. Third, when looking at S2, the pattern from TS SF is more interpretable comparing to the patchy pattern from CSP. When solely focusing on Fig. 3b, it is obvious to see that for S2, S3, S4, S5 and S6, they represent strong signal sources surrounding the sensorimotor cortex which is consistent with this paradigm. This is in contrast to, e.g., S7 and S10, where CSP clearly picks up ocular artifacts.

DISCUSSION

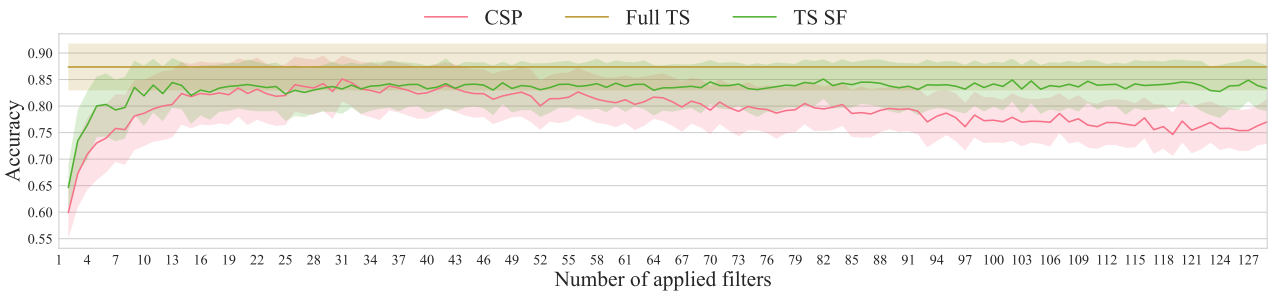
We show that it is possible to construct spatial filters that can approximate the performance of tangent space linear functions, and validate this result on a variety of open-source datasets. In our simulations, it can be seen that the TS SF out-performs CSP, in particular in the cases where the full tangent space function out-performs CSP. Further, we show in the *Munich* dataset that the spatial filters recovered by TS SF are more robust to noisy data than those from CSP. With these advances, it is possible to use Riemannian methods in neurofeedback and clinical work. To evaluate the spatial filter, the first two essential indicators are the classification accuracy and the quality of spatial patterns.

From the perspective of performance, TS SF is still imperfect. First, as shown in Fig. 2, TS SF is constantly worse than the full Riemannian methods. Hence, to move forward, it is important to diminish the approximation error as much as possible, especially when using very few filters. The choice of how to determine filter relevance, though effective, can also be improved. Regularized linear regression approaches, such as LASSO or ridge regression, may better isolate good filters, or using another target variable.

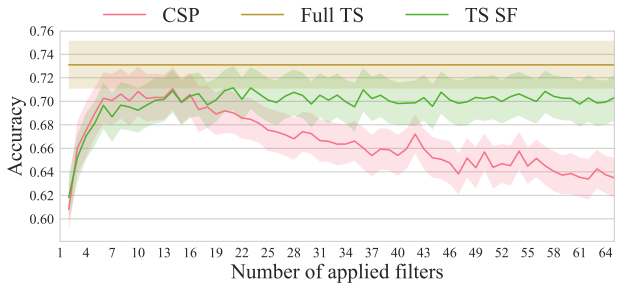
The spatial patterns associated with TS SF are robust to artifacts. As shown in Fig. 3, the reflected neural signal sources after filtering by TS SF is more reasonable and interpretable than by CSP, especially for the low-quality signal. For example, for S2 of the data set *Munich* where the percentage of trials contaminated by artifacts reaches 55% [14], CSP merely collects the artifacts whereas TS SF presents the strong sources around sensorimotor cortex which is consistent with the chosen paradigm. Nonetheless, for the subjects with high data quality, TS SF is still as good as CSP (e.g., for S3 and S4,

Published by Verlag der TU Graz
Graz University of Technology

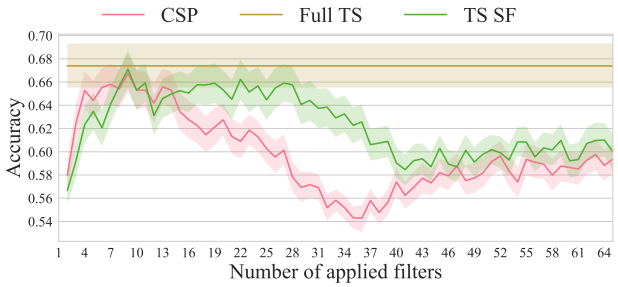
 DOI: 10.3217/978-3-85125-682-6-07



(a) Dataset: Munich MI



(b) Dataset: Cho.



(c) Dataset: Physionet

Figure 2: Accuracies comparison among three different pipelines and for each number of applied filters, the central line and the error band are computed across all subjects and sessions. **1. CSP**: Choose CSP as the spatial filter and classified by LDA. **2. Full TS**: Use the standard Riemannian classification framework with AIRM without any spatial filtering and classified by SVM. **3. TS SF**: Apply the spatial filters which are extracted from the tangent space and classified by LDA. Error band setting: confidence interval = 68%. Central line: the mean accuracy. It is not difficult to notice that TS SF has a better and more robust performance than CSP w.r.t. the increase of the number of applied filters.

where 5.0% and 6.3% of trials are affected by artifacts respectively [14]). Therefore, as far as the interpretability is concerned, TS SF is more robust than CSP.

As shown in Fig. 2, it is obvious that TS SF is always more robust than CSP considering the tendency of the accuracy w.r.t. the filter numbers. However, when focusing at Fig. 2c, we could find an evident fluctuation for TS SF. Accuracies drop at high numbers of filters in contrast to the other datasets, and the overall range of mean accuracies is far lower as well. Much of this can likely be explained by the low data quality of data set *Physionet*, which would lead to worse solutions of the initial tangent space problem that likely translate into poorer spatial filters. However, it is important to understand what conditions lead to better or worse TS SF solutions, especially when applying full filters. Also, it should be explored more that why is TS SF still more robust than CSP to filter number even when TS SF performs badly.

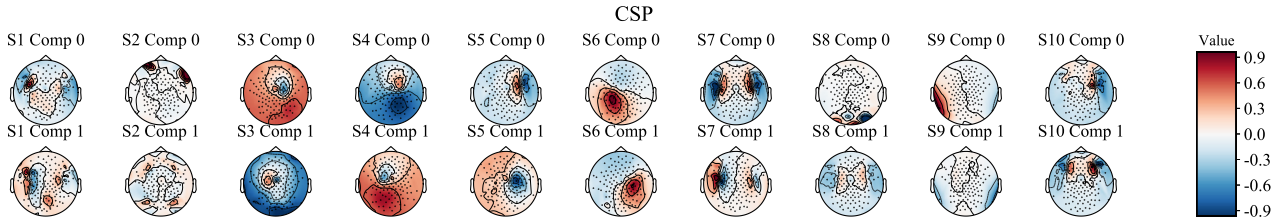
Usually, the more robust performance demand more complex computation, but this argument does not always hold. Although TS SF requires a tangent space function to be computed as an initial step which means the computation of the filters themselves is far more intensive than that of CSP, however, in the online application, the procedure is identical: the data is filtered via a linear projection and then the log variances of the resulting time series are used as features for a final classifier. When comparing to the standard Riemannian classification framework, TS SF leads to an approximation error. However, the computation for online classification would be significantly reduced, especially with large channel numbers. Given that the loss in accuracy on average is relatively small, this may represent a feasible way of using Riemannian methods in high-dimensional, online settings, and further of using them to increase the efficacy of clinical neurofeedback.

CONCLUSION

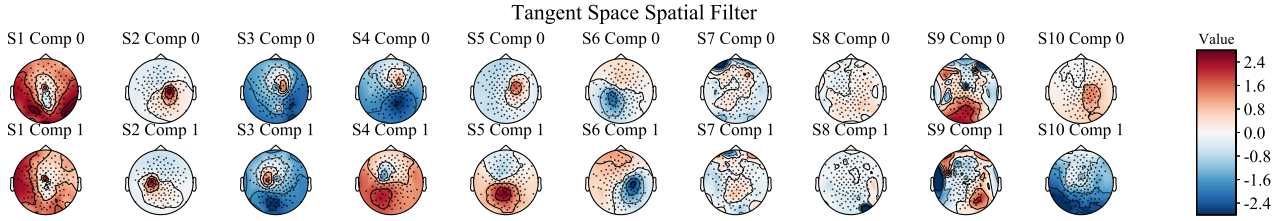
In this paper, we propose a new method to generate spatial filters from the Riemannian classification framework, Tangent Space Spatial Filters (TS SF), and after comparing between CSP and TS SF, we found that TS SF has better and more robust performance than CSP with the same computational cost online. By extracting spatial filters, the Riemannian approach becomes more interpretable without much loss of accuracy.
However, we still do not have a strong mathematical argument for why this simplification works and the interpretation of the weight vector after projection to the SPD manifold. Nonetheless, the undeniable fact is that the performance of TS SF reveals to us the potential of interpretable Riemannian approaches in BCIs.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-07



(a) Applying CSP as spatial filter



(b) Applying tangent space spatial filter (TS SF) as spatial filter

Figure 3: The spatial patterns of data set *Munich* associated with applying the first two spatial filters.

REFERENCES

[1] Zoltan J Koles, Michael S Lazar, and Steven Z Zhou. “Spatial patterns underlying population differences in the background EEG”. In: *Brain Topography* 2.4 (1990), pp. 275–284.

[2] Wojciech Samek, Motoaki Kawanabe, and Klaus-Robert Müller. “Divergence-based framework for common spatial patterns algorithms”. In: *IEEE Reviews in Biomedical Engineering* 7 (2014), pp. 50–72.

[3] Marco Congedo, Alexandre Barachant, and Rajendra Bhatia. “Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review”. In: *Brain-Computer Interfaces* 4.3 (2017), pp. 155–174.

[4] Florian Yger, Maxime Berar, and Fabien Lotte. “Riemannian approaches in brain-computer interfaces: a review”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.10 (2017), pp. 1753–1762.

[5] Vinay Jayaram and Alexandre Barachant. “MOABB: Trustworthy algorithm benchmarking for BCIs”. In: *Journal of Neural Engineering* (2018).

[6] Alexandre Barachant, Jason Carmel, Kathleen Friel, and Disha Gupta. “Extraction of motor patterns from joint EEG/EMG recording: A Riemannian Geometry approach”. In: *6th International Brain-Computer Interface Meeting*. 2016.

[7] Herbert Ramoser, Johannes Muller-Gerking, and Gert Pfurtscheller. “Optimal spatial filtering of single trial EEG during imagined hand movement”. In: *IEEE Transactions on Rehabilitation Engineering* 8.4 (2000), pp. 441–446.

[8] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. “Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b”. In: *Frontiers in Neuroscience* 6 (2012), p. 39.

[9] Fabien Lotte and Cuntai Guan. “Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms”. In: *IEEE Transactions on Biomedical Engineering* 58.2 (2011), pp. 355–362.

[10] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. “A Riemannian framework for tensor computing”. In: *International Journal of Computer Vision* 66.1 (2006), pp. 41–66.

[11] Xavier Pennec and Nicholas Ayache. “Uniform distribution, distance and expectation problems for geometric features processing”. In: *Journal of Mathematical Imaging and Vision* 9.1 (1998), pp. 49–67.

[12] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. “Riemannian geometry applied to BCI classification”. In: *International Conference on Latent Variable Analysis and Signal Separation*. Springer. 2010, pp. 629–636.

[13] Stefan Haufe et al. “On the interpretation of weight vectors of linear models in multivariate neuroimaging”. In: *Neuroimage* 87 (2014), pp. 96–110.

[14] Moritz Grosse-Wentrup, Christian Liefhold, Klaus Gramann, and Martin Buss. “Beamforming in noninvasive brain–computer interfaces”. In: *IEEE Transactions on Biomedical Engineering* 56.4 (2009), pp. 1209–1219.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-08

KEEP AN EYE ON IT: REPRESENTATION OF EYE MOVEMENTS IN THE SENSORIMOTOR CORTEX

M.P. Branco[+], J.J. Luppi[+], S. Leinders, Z.V. Freudenburg, B.H. van der Vijgh, E.G.M. Pels, E.J. Aarnoutse, N.F. Ramsey and M.J. Vansteensel

UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, Utrecht, The Netherlands

+ these authors contributed equally to this work

E-mail: M.PedrosoBranco@umcutrecht.nl

ABSTRACT: The Utrecht NeuroProsthesis system is a fully implanted electrocorticographic (ECoG) Brain-Computer Interface (BCI) that aims to provide independent control of a computer to people with locked-in syndrome. An incidental finding of the study showed that goal-oriented eye movements generated patterns of activity in the primary motor cortex (M1) that were similar to those elicited by attempted hand movement. In order to examine the overlap between eye and hand representation in M1 we compared responses elicited by both movements using ECoG and electrical cortical stimulation in epilepsy patients, and functional magnetic resonance imaging in healthy volunteers. Results from all modalities showed indications of eye movement-related activity in M1, but not consistently across subjects. Even though this overlapping representation is not a universal feature, the occurrence of eye-movement related activity in the M1 hand area bears relevance for refining the user-specific accuracy of BCI applications in people with severe paralysis, especially in cases where eye movements are (also) used to communicate.

INTRODUCTION

The aim of the Utrecht NeuroProsthesis (UNP) project is to develop viable BCI implants that provide individuals with locked-in syndrome (LIS) with a means of independent communication (www.neuroprosthesis.eu). The UNP implant employs electrocorticography (ECoG) arrays placed directly on the surface of the cortex. The first UNP participant was implanted with electrode strips over the hand region of the sensorimotor cortex (SMC) as well as over the dorsolateral prefrontal cortex [1]. By attempting to move her right hand, the participant is able to generate reproducible patterns of brain activity in the SMC. These activity patterns are used to extract signal features for generating 'brain clicks', which the participant uses to independently control a spelling program.

The participant has been successfully using the UNP BCI system for communication at home, supplementing and at times replacing the eye tracker she uses for communication as well. Interestingly, it was found that during the use of the eye tracker the goal-directed eye movements generated unintentional brain clicks in the UNP system. While there are several cortical areas associated with eye movements, such as the frontal eye field (FEF), supplementary eye field (SEF) and the parietal eye field (PEF), none of these are conventionally described to involve the primary motor cortex (M1) [2]. The observation of these 'false positives clicks' raised, therefore, the questions *whether the representation of eye movements in the M1 is a universal feature a*nd *if there is an overlap between the cortical representations of the eyes and hand in this region.* Answering these questions is important for understanding the mapping of cortical activity and refining the UNP system in order to improve its accuracy by reducing interference. Hence, in this study we used three neural signal recording modalities to address these questions, namely ECoG, electrical cortical stimulation (ECS) and functional magnetic resonance imaging (fMRI), all with a separate group of participants.

MATERIALS AND METHODS

All participants gave informed consent to participate in this research, which was approved by the Medical Ethical Committee of the UMC Utrecht in accordance with the Declaration of Helsinki (2013).

ECS and ECoG

The clinical ECS records of 34 epilepsy patients (14 female, mean age 28.4 ± 11.7) were investigated. Reports of eye movement responses elicited by electrical stimulation of the SMC using ECoG electrode arrays were collected. Mapping was performed by applying brief currents to neighboring pairs of electrodes using an IRES 600 CH electrical stimulator, with similar clinical settings to those described in [3]. For each electrode pair, all behavioral responses to the stimulation were noted by a trained clinician. Only data of patients with an electrode grid coverage that included the SMC was used. For the cases where eye movement responses were found in or close to the M1 region of interest, the associated electrodes were plotted on a

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-08

reconstruction of the cortical surface and compared to the location of ECS *hand* responses. The localization of the electrodes was determined using a combination of cortical surface reconstructions based on pre-operative structural MRI scans and post-operative CT (computed tomography) scans, as described in [4-5]. Additionally, in order to provide a comparison and assess the degree of overlap between eye and hand representations on M1, the ECS results were compared with ECoG activity elicited during a hand motor task performed by the same patients. For that, ECoG data acquired during the hand motor task was filtered to remove 50 Hz noise. Channels with extremely poor-quality signal were excluded from the analysis. Signed R^2 values for the active versus rest condition in the high-frequency range 60-130 Hz were plotted on the cortical surface reconstructions. The high-frequency band (HFB) was chosen as a target for the analysis due to it being part of the control signal of the UNP participant [1] as well as for its known relevance for cortical activation [6-7].

High-Density ECoG

Data from one patient (E1, female, 31 years old) temporarily implanted with subdural ECoG grids for removal of the focus of epilepsy was used. The subject was also implanted with a high-density (HD) ECoG grid over the sensorimotor cortex (PMT Corporation, MN, USA; 4 mm inter-electrode spacing, sampled at 2000 Hz). We defined an ROI for the M1 hand region based on anatomical landmarks. E1 performed one eye movement and one finger movement localizer task with 15 s blocks of rest and active conditions (10 trials in total). In the active condition of the eye localizer the subject was asked to follow a red circle moving along the edges of a square with her eyes. The target circle could start in any of the corners of the square and move in either clock- or counter-clockwise direction. The active trials of the finger localizer consisted of continuous finger tapping movements. Data were processed in similar way as described in the previous section.

fMRI

Ten healthy (F1-10), right-handed volunteers participated in the study (6 female, mean age 26.9 ± 10.7). The fMRI scanning was carried out using a 7 Tesla Philips Achieva MRI system (Philips Healthcare, Best, Netherlands). Participants performed an eye-movement task while fMRI data was acquired using an Echo Planar Imaging (EPI) sequence. The fMRI task employed a different paradigm than the ECoG tasks, in that an event-related design was used instead of a block design. The participants were instructed to fixate on a target at the center of the screen. Once every 10.5 s, the target moved upwards, downwards, to the left or to the right (32 trials in total). When this happened, the subject had to execute a saccade to fixate on the target at the new position. Additionally, the subjects performed a hand localizer task, from which movement of all fingers (against rest) was used to look for overlap between hand and eye activation. fMRI data was preprocessed using standard protocol and co-registered to the T1-weighted anatomical scan of the participant. FreeSurfer (https://surfer.nmr.mgh.harvard.edu/) was then used to create a surface reconstruction based on the anatomical images. First level statistics were carried out using SPM12. A custom procedure called CGRID was employed to visualize the results and facilitate between-subject comparisons [8]. CGRID transforms the flattened surface reconstructions of the SMC into a grid of 84 x 28 tiles with standardized x- and y-coordinates. In this grid an ROI corresponding to the M1 hand area which consisted of 210 tiles.

RESULTS

ECS and ECoG

Eye responses to stimulation of M1 were present in some subjects. Of the 34 clinical ECS records available for analysis, 10 contained reports of eye-related motor or sensory responses. Of these, only 3 reports described eye-related responses in the proximity of the region of interest (Figure 1). One individual had eye responses close to M1 but with no overlap with hand movement. Two individuals had eye-responsive electrodes exactly on M1 and partly overlapping with electrodes that elicited hand movement during stimulation. One of these subjects (S3) also showed overlapping with electrodes that showed significant HFB power changes during the finger tapping localizer task.

High-Density ECoG

For subject E1, multiple electrodes on M1 responded to the eye task with high and significant signed R^2 values (Figure 2). Many of these electrodes were located on M1, even partly overlapping with electrodes that showed increases in HFB activity in response to a hand movement task. A total of 7/42 electrodes in the M1 ROI with signed R^2 values > 0.25 ($p < 0.05$) in response to the eye task also had HFB increases in response to the hand task with at least $R^2 > 0.25$ ($p < 0.05$).

fMRI

Movement of the fingers elicited clear and recognizable patterns of activity in the contralateral SMC, with the majority of activity increases in M1 being observed in the hand area, located relatively high on M1 (Figure 3). While the responses to eye movements were variable, in most subjects increases in activity were observed in either the hand region or very near it. Furthermore, in some subjects, such as F1, F3 and F7, more than 20% overlap between the eye and hand activation was present, measured as the number of tiles with both eye and hand activity out of all active tiles. Even in the case of F2, F5 and F6, who had little or no actual overlap, distinct regions of eye and hand activation in the ROI were present in close proximity to each other. On average, out of the total active tiles in the ROIs of all participants, 12% had overlap, while 25% were exclusively eye-responsive (Figure 4).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-08

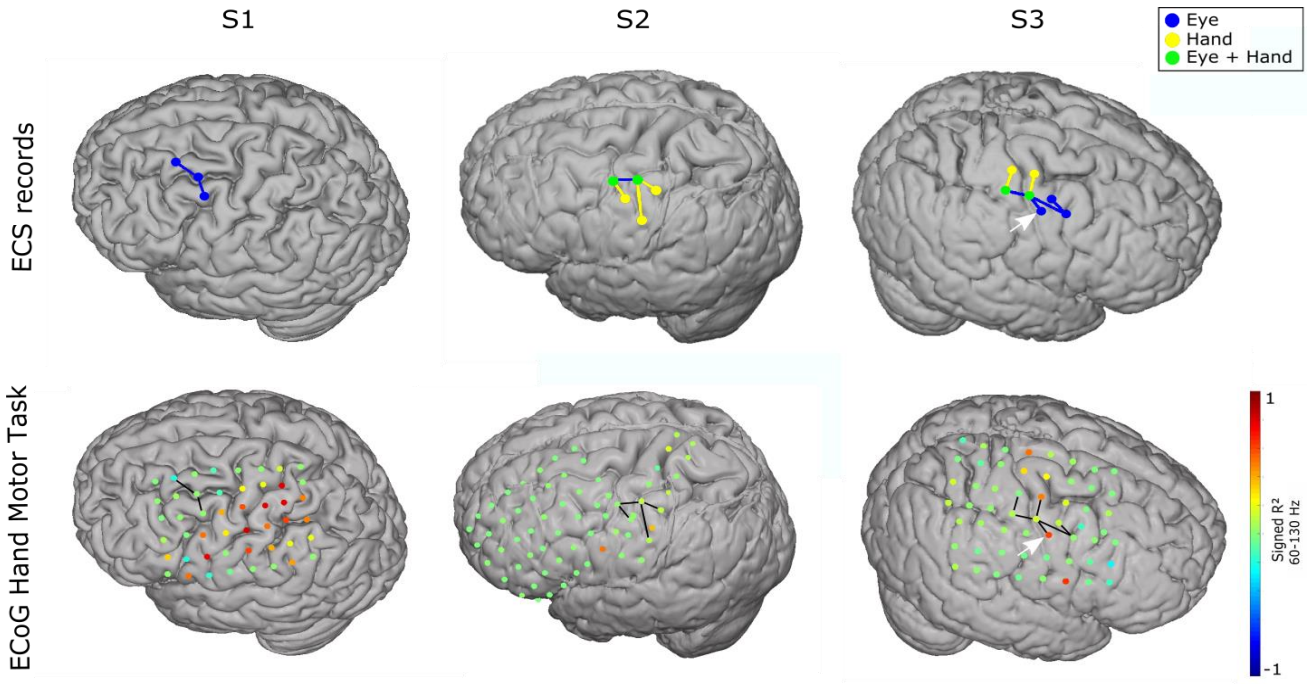


Figure 1: Top row - three of the investigated clinical ECS records (coded S1, S2 and S3) contained reports of eye responses (sensation for S1, movement for S2 and S3) in response to electrical stimulation of the SMC near the M1 hand region. These electrodes are plotted on the cortical surface reconstructions of the specific patients. Pairs of electrodes with eye-related responses are connected by blue lines, whereas yellow lines indicate pairs associated with hand responses. Electrodes that were associated with both eye and hand responses are coloured green. Bottom row - the signed R^2 values for the 60-130 Hz frequency band from an ECoG hand motor task of the same patients. The electrodes of interest (based on the ECS results) are connected by black lines. One electrode of S3 showed both eye ECS responses and ECoG responses to a hand task (white arrow).

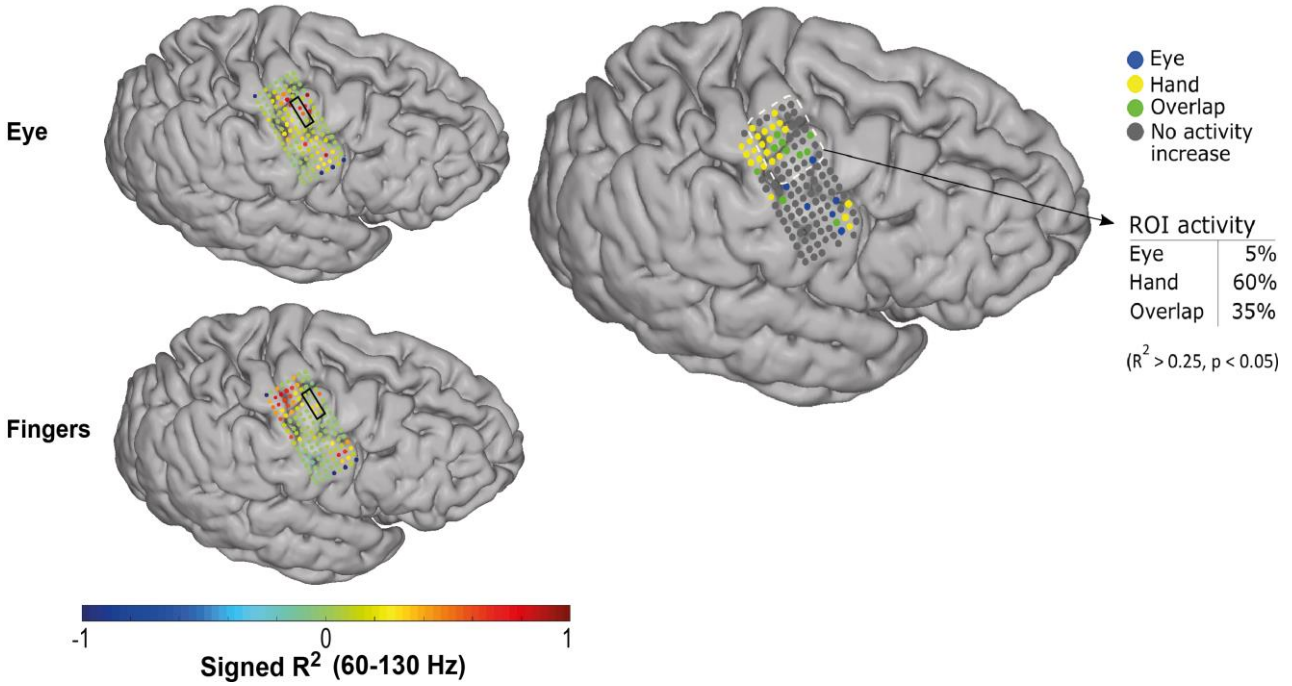


Figure 2: HD-ECoG grid plotted on a surface reconstruction of the cortex of E1 (Left column). Each electrode is color-coded with its signed R^2 values (HFB; 60-130 Hz) during an eye movement task (top-left) or a finger movement task (bottom-left). Electrodes that had a significant $R^2 > 0.25$ for the eye task, finger task or both were plotted onto the surface reconstruction of the subject's cortex (right). The hand region ROI is outlined by white dashed lines and contained 42 electrodes. Of the 20 active electrodes, 35% contained overlap between eye and hand responses. 60% were activated exclusively by the hand task, while one electrode (5%) showed an activity increase only during the eye task.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-08

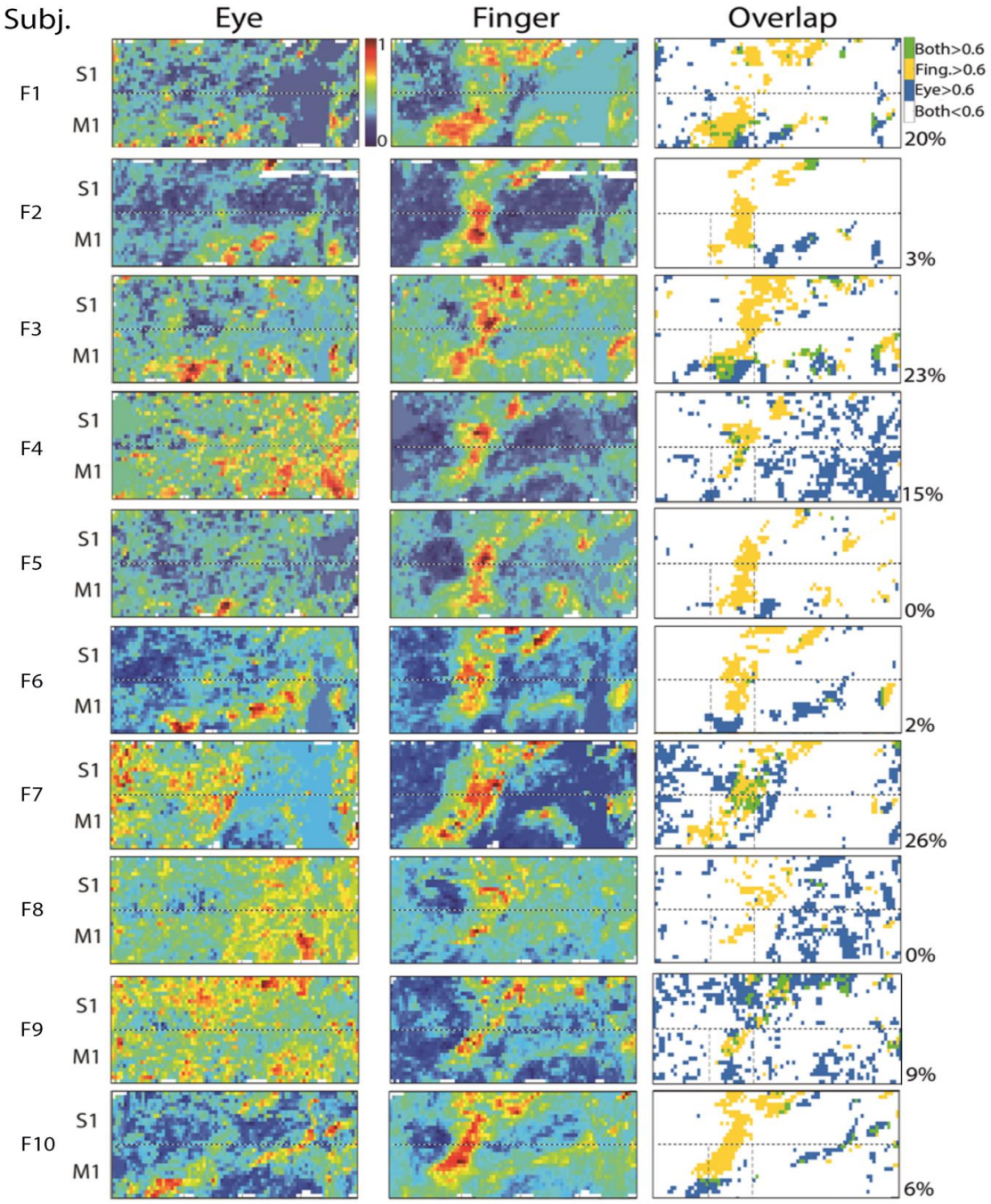


Figure 3: The complete CGRID maps of all 10 fMRI subjects. Dorsal direction is to the left on the panels. The three columns visualize the maps of eye movement activation, contralateral hand movement activation and the overlap between the two, respectively. The t-maps to be mapped on the CGRID were normalized between 0 and 1, and 0.6 was set as a threshold for a tile to be considered active. Blank tiles in the actual CGRID maps are tiles that were excluded in order to conform to the shape of the CGRID. In the overlap maps, the ROI is outlined by dashed lines. The percentage of overlapping tiles out of the total number of activated tiles within the ROI is given as a percentage to the right of the overlap maps.

DISCUSSION

Overlap between hand and eye activity

An incidental finding in the first participant of the UNP project showed that goal-directed eye-movements elicited similar activity to that elicited during attempted hand movement. Here, we used three neural signal recording methods to assess whether this finding was an isolated case or a general feature of the SMC. Overall, we found evidence of eye movement-related activity in the M1 hand area, although not consistently across subjects.

The results of the fMRI study were likely the most

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-08

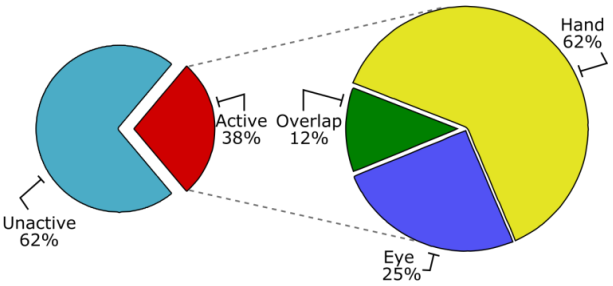


Figure 4: Percentage of tiles inside the ROI that were active during both eye and hand task (out of 210) averaged across subjects. From the active tiles, 25% were activated by eye movement only and 12% were activated during eye and hand movement.

convincing of the three methodologies and were able to demonstrate eye movement-related activity in M1 in multiple subjects. However, for a few of the subjects the peak t-values for the activity increases in response to the eye task were somewhat lower than might be expected, resulting in more scattered activation maps, likely due to noise. This is largely explained by the fact that event-related designs elicit shorter neuronal activation and result in lower t-values than block designs due to the intrinsic slow temporal resolution of the fMRI recording. In the future, a block design may be an interesting consideration that may further increase the robustness of the results obtained. The limited ECS outcome could be explained by the fact that the records that were investigated were diagnostic recordings, not experimental ones. The clinicians making the records may not have focused on the eyes as much as, for example, on the movement of the limbs. The number of subjects implanted with HD-ECoG grids covering the region-of-interest and who performed the eye localizer task was very limited (one subject). In the future, the results of the ECS and ECoG measurements should be replicated in more subjects in order to establish the prevalence of this finding.

The functional boundary of the FEF

Literature typically associates the frontal eye fields (FEF) with the generation of eye movements. The FEF is generally located around the intersection of the superior frontal sulcus and the precentral sulcus, which is in the proximity of M1. As an area, it is defined by its capacity to generate eye movements in response to stimulation or to display activity increases during the execution of eye movements [9]. Due to this functional definition, it does not have strict anatomical boundaries. In humans, it is generally described to be located frontal to the SMC, with some studies also describing it extending to M1 [10-12]. The high degree of variability in the localization of eye movement-related responses around the area of interest reported by the diverse studies (Figure 5) is in line with the results here found with ECS, ECoG and fMRI.

Inter-individual variability

While the small sample size in the current study prevents measuring the prevalence of overlap between eye and hand representation in M1, the results so far strongly indicate that such overlap is present in some individuals. If, indeed, the degree of eye representation on M1 is something that varies greatly from one person to another, it could help explain why it is not discussed in literature at large, but it also begs the question why such inter-individual variation exists. While disease-related neurophysiological phenomena and neural plasticity induced from BCI feedback training could help explain the results of the UNP participant, especially when considering that ALS is known to influence the arrangement of cortical representations [13], the fact that similar results were also found in subjects without the condition make this explanation insufficient. As such, the presence of eye-hand overlap found on the motor cortex in the UNP participant could not be explained solely by disease-related changes in cortical function representation.

CONCLUSION

In this study, we investigated the possible overlap between the eye and hand representations in the primary motor cortex using three methodologies. Results show that eye movements activate patches of M1 very close to or overlapping with the hand region. Due to the variability of the results, it is proposed that such overlap exists in the overall population but its presence and extent vary greatly on an individual basis. Based on existing literature it is likely that the eye representation in question belongs to the FEF extending beyond its conventionally defined borders. Further research into the phenomenon is encouraged in order to gain an understanding of its localization as well as the potential effects of eye movements on BCI signal acquisition from the SMC. This will help inform the implantation procedure for BCI systems in order to avoid false positives caused by interference in the signal, therefore improving overall accuracy and usability of its applications.

ACKNOWLEDGEMENTS

This study was funded by the ERC-Advanced ‘iConnect’ project (grant ERC-Adv 320708) and the Dutch Technology Foundation STW NeuroCIMT project (grant 14906). The authors would like to thank Carlijn Bakker for acquiring part of the fMRI data, Wouter Schellekens for designing the fMRI tasks and helping with preprocessing of the fMRI data, and Mark Bruurmijn and Mathijs Raemaekers for developing the CGRID tool.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-08

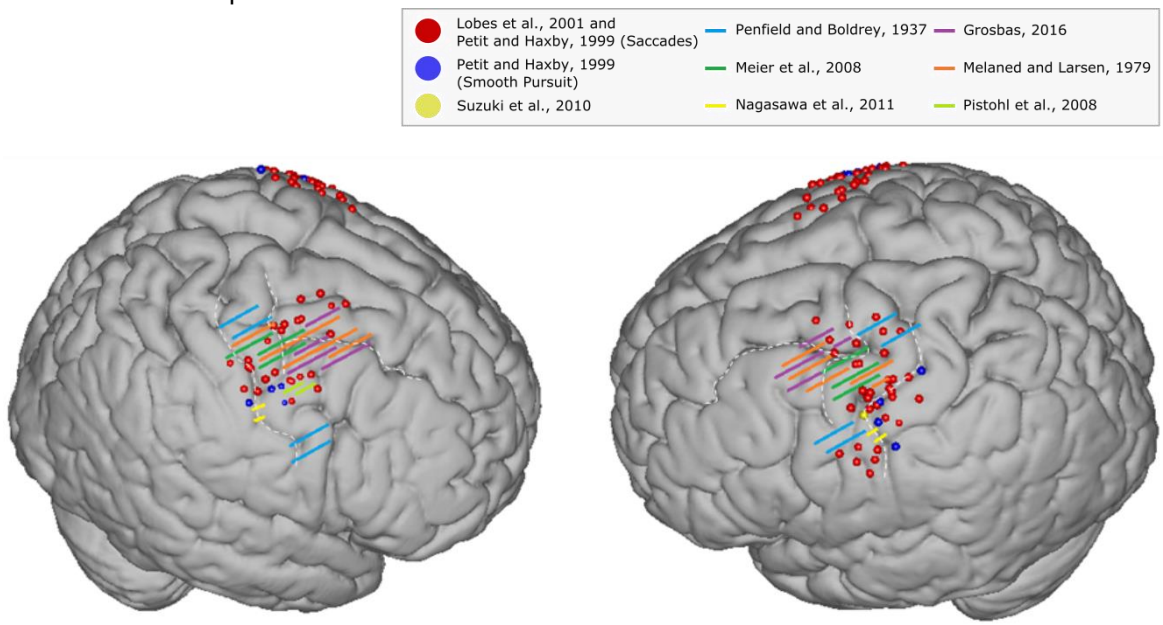


Figure 5: Reports of eye movement-related activity in the vicinity of the SMC found in the literature mapped onto a standardized MNI brain. Spheres correspond to specific coordinates reported in studies, with red coding for saccades, blue coding for smooth pursuit movements and yellow coding for closing and opening of the eye. Striped regions correspond to areas reported in the figures or descriptions of each study. The white dashed lines indicate the central sulcus as well as the precentral sulcus and superior frontal sulcus.

REFERENCES

[1] Vansteensel, M. J., Pels, E. G. M., Bleichner, M. G., Branco, M. P., Denison, T., Freudenburg, Z. V., et al. (2016). Fully Implanted Brain–Computer Interface in a Locked-In Patient with ALS. New England Journal of Medicine, 375(21), 2060–2066.

[2] Pouget, P. (2015). The cortex is in overall control of "voluntary" eye movement. Eye (London, England), 29(2), 241–5.

[3] Vansteensel, M. J., Bleichner, M. G., Dintzner, L. T., Aarnoutse, E. J., Leijten, F. S. S., Hermes, D., & Ramsey, N. F. (2013). Task-free electrocorticography frequency mapping of the motor cortex. Clinical Neurophysiology, 124(6), 1169-1174.

[4] Hermes, D., Miller, K. J., Noordmans, H. J., Vansteensel, M. J., & Ramsey, N. F. (2010). Automated electrocorticographic electrode localization on individually rendered brain surfaces. Journal of neuroscience methods, 185(2), 293-298.

[5] Branco, M. P., Gaglianese, A., Glen, D. R., Hermes, D., Saad, Z. S., Petridou, N., & Ramsey, N. F. (2018). ALICE: A tool for automatic localization of intra-cranial electrodes for clinical and high-density grids. Journal of neuroscience methods, 301, 43-51.

[6] Crone, N. E., Miglioretti, D. L., Gordon, B., & Lesser, R. P. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain: a journal of neurology, 121(12), 2301-2315.

[7] Jacobs, J., & Kahana, M. J. (2010). Direct brain recordings fuel advances in cognitive electrophysiology. Trends in cognitive sciences, 14(4), 162-171.

[8] Bruurmijn, M., Cornelisse, P.A., Schellekens, W., Raemaekers, M.A.H., Vansteensel, M.J., and Ramsey, N.F (2016). Novel 2D standard coordinate space for sensorimotor cortex validated with high-resolution 7T rs-fMRI functional connectivity. Program No. 132.13. 2016 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2016. Online.

[9] Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A., & Valero-Cabré, A. (2014). Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. Frontiers in Integrative Neuroscience, 8, 66.

[10] Petit, L., & Haxby, J. V. (1999). Functional Anatomy of Pursuit Eye Movements in Humans as Revealed by fMRI. *Journal of Neurophysiology*, *82*(1), 463–471.

[11] Grosbras, M.-H., Laird, A. R., & Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. Human Brain Mapping, 25(1), 140–154.

[12] Pierrot-Deseilligny, C., Müri, R. M., Nyffeler, T., & Milea, D. (2005). The role of the human dorsolateral prefrontal cortex in ocular motor behavior. In Annals of the New York Academy of Sciences (Vol. 1039, pp. 239–251).

[13] Lulé, D., Diekmann, V., Kassubek, J., Kurt, A., Birbaumer, N., Ludolph, A. C., & Kraft, E. (2007). Cortical Plasticity in Amyotrophic Lateral Sclerosis: Motor Imagery and Function. Neurorehabilitation and Neural Repair, 21(6), 518–526.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-09

A WEARABLE SSVEP-BASED BRAIN-COMPUTER INTERFACE WITH OFF-THE-SHELF COMPONENTS

L. Angrisani[1], P. Arpaia[1], A. Esposito[2], N. Moccaldi[1], M. Parvis[2]

[1]Department of Electrical Engineering and Information Technology (DIETI), Universita' degli Studi di Napoli Federico II, Naples, Italy
[2]Department of Electronics and Telecommunications (DET), Politecnico di Torino, Turin, Italy

E-mail: pasquale.arpaia@unina.it

ABSTRACT: In this paper, a wearable Brain-Computer Interface is proposed. This BCI exploits a non-invasive single-channel electroencephalogram to measure steady-state visually evoked potentials (SSVEP) from the user's scalp. Dry electrodes are employed. The visual stimuli for SSVEP elicitation are presented on the LCD display of Augmented Reality glasses, and each stimulus can be associated with a command to actuate. The brain signals processing is conducted by means of a simple power spectral density analysis based on an FFT algorithm. Then, the signal features are classified with a Support Vector Machine. The resulting algorithm requires a low computational burden, and it was easily implemented in Android. The final BCI system is built with off-the-shelf components, and it is easily customizable. This work aims to give a contribution to the development of BCI systems for applications in daily life.

INTRODUCTION

A Brain-Computer Interface (BCI) is a powerful tool capable of improving our way of communicating with the external world [1]. Historically, BCI applications have been addressed to help people with sever motor disabilities or other medical diseases [2, 3]. However, in the last decade, other fields have considered the adoption of a BCI, such as gaming, entertainment, education, or robotics [4–6], and today we are assisting to a rapid growth of this technology [7]. Nonetheless, the majority of work on BCI still remains at the level of laboratory research. The reasons for that are technical limitations, such as the problem of motion artifacts [8] or the trade-off between speed and performance of the BCI [9], and practical limitations, such as costs and wearability [10].
Aiming to build a wearable device to use in daily life, a non-invasive technique for brain activity measurement must be considered. Electroencephalography (EEG) is here taken into account. Indeed, EEG can be inexpensive and safe, while providing a spatial resolution in the order of centimeters and temporal resolution in the order of milliseconds [11]. Several paradigms there exist for BCI systems, such as motor-imagery [12, 13], P300 [14], or SSVEP [15–17]. In particular, "steady-state visually evoked potentials" (SSVEP) are highly reliable in terms of accuracy and reproducibility [15, 18, 19] and user training is not mandatory [20, 21]. Hence, a SSVEP-based BCI is a suitable choice for a practical device.
In the present work, a single differential channel with dry electrodes is considered for the acquisition of EEG signals. The resulting BCI is highly wearable and low-cost. Clearly, this choice has some drawbacks, such as a low signal-to-noise ratio (SNR) if compared to more invasive techniques, where electrodes implanted inside the scalp guarantee higher SNR, and few information on brain activity due to the employment of a single channel. These drawbacks are mitigated if SSVEP signals are considered because they guarantee a good SNR with respect to other paradigms. However, an SSVEP-based BCI requires visual stimuli, namely flickering LEDs or icons on a display. A possible solution is the employment of the LCD displays of Augmented Reality (AR) glasses. They can be employed for stimulus presentation without affecting system wearability. The combination of BCI and AR platforms is not new. In [22], research solutions were surveyed, and it was highlighted that most state-of-the-art systems made use of two VEP-based paradigms, SSVEP or P300.
In a preceding phase, our research group demonstrated the feasibility of employing AR glasses in conjunction with a single-channel electroencephalography to build a wearable SSVEP-based BCI system. Moreover, the presented system has been built with components available off-the-shelf. The aim of the current work is to keep giving a contribution to the development of BCI systems that could be employed in everyday life. A particular focus is on the implementation of an algorithm for SSVEP signal classification. The reminder of this paper is organized as follows. The implementation of the system and the classification algorithm are first described. Then, the results of a training and evaluation procedure for assessing the performance of the SSVEP classifier are reported. Finally, the results are discussed and future steps addressed.

MATERIALS AND METHODS

The architecture of the BCI system is here described, along with the off-the-shelf components employed for its implementation. The system of concern is divided into

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-09

three parts, (i) the visual stimuli generator, (ii) the EEG acquisition, and (iii) the signal processing.
As mentioned in the introduction, the stimulation consists of flickering icons on the LCD display of AR glasses. The Android-based "Epson Moverio BT-200" AR glasses were chosen (Fig. 1). These are relatively low-cost (about 600 € per unit) and provide the necessary characteristics for stimuli implementation. Notably, the 60 Hz refresh rate of the display allows for many flickering frequencies to be set. Two flickering icons were chosen as stimuli, at nominal frequencies equal to 10.0 Hz and 12.0 Hz, respectively. This choice is based on studies demonstrating that a good signal-to-noise ratio can be achieved at these frequencies [15], and this was also confirmed with some preliminary experiments. The AR glasses were programmed in Android Studio for the visual stimuli generation.

Figure 1: Epson Moverio BT-200 AR glasses.

The brain activity, containing the elicited SSVEP signals, is acquired with a non-invasive electroencephalography. A single differential channel consisting of two dry electrodes has been employed. These electrodes are placed on the user's scalp according to the 10-20 system [23], at the points "Fpz" and "Oz", as reported in Fig. 2. A third electrode is also needed as a ground. This electrode is usually placed on the forehead or ear location, but it can also be placed on a wrist or a leg [24].

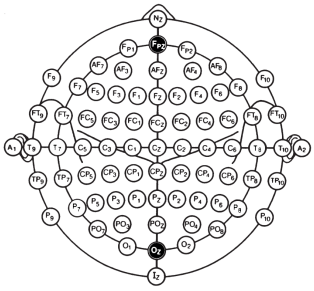


Figure 2: Placement of EEG electrodes in the International standard framework 10-20 [23] (in black): "*Fpz*" and "*Oz*", in the scalp frontal and occipital region, respectively.

An open source EEG device, which is commercially available, was employed for the brain activity measurement, i.e. the Olimex EEG-SMT. This is shown in Fig. 3. Two active electrodes (CH1+ and CH1- respectively), and a passive electrode (DRL) are depicted too. The active electrodes differ from the passive one because of some additional circuitry, based on operational amplifiers, for high-frequency interference rejection. These two electrodes were placed on the scalp, while the passive electrode was placed on the left wrist. The device and the electrodes total cost is less than 200 €. It is worth noting that the unused differential channel (CH2) was connected to the internal reference voltage as suggested by the EEG-SMT manual for achieving less noise. Fig. 3 also shows that some silver pins were soldered on one of the active dry electrodes. This electrode is placed on the occipital area ("Oz"), hence the pins aim to overcome the hair in order to better reach the user's scalp.

Figure 3: Olimex EEG-SMT and dry electrodes.

The acquired EEG signal is converted from analog to digital thanks to the micro-controller being part of the EEG-SMT, and then it is continuously transferred through USB on a computing unit for elaboration. In this regard, the sampling frequency was set to 256 Sa/s, the A/D converter has a 10-bit resolution, and the overall gain was set to 6427 V/V. In the final integrated system, the digital EEG signal is directly transferred to the AR glasses micro-processor, where it is elaborated and the SSVEP signals classified. A command can then be associated to each class, namely the "10 Hz" class and the "12 Hz" class. However, to better study the signal processing, the EEG signal has been first transferred to a PC in order to be elaborated with MATLAB. After this step, whom results are reported in the following, the algorithm was ported in Android. A simple algorithm has been conceived, so that it could results in a low computational burden. This also means that the computational time does not affect the response time of the system, which is limited to some seconds because of the required stimulation/acquisition time. The signal time length is fixed within the algorithm. This acquisition time can be decided for each subject after an algorithm training phase, and it was initially fixed at 10.0 s for all the subjects.
The first processing step is a pass-band digital filtering. The pass-band has been set to (8-28) Hz so that the frequency components at the stimuli frequencies and their respective second harmonic are not corrupted. Meanwhile, the stop-band tries to reduce the effect of artifacts on the signal, especially EOG artifacts at low frequencies [25]. Note that, for example, a regression method

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-09

for artifact removal can not be employed aiming to adopt a single EEG channel. A FIR filter based on the Hamming windowing was designed. The order was set to 100, which resulted as a good compromise between computational burden and desired filter performance. The coefficients of the designed filter have been then employed for filtering in on-line signal analysis. After this step, the signal is zero-padded to the nearest power or 2 and transformed into the frequency domain with an FFT algorithm. In particular, the amplitude spectrum corresponding to the (8-28) Hz interval is derived, and the power spectral density (PSD) in the neighbour of the flickering frequencies can be calculated as signal features to be classified. For each stimulus frequency and their second harmonic, the power density is calculated as the sum of the squared amplitudes associated to corresponding bin and some nearest bins:

$$P(f_i) = \frac{1}{\Delta k} \sum_{n=k_i-\frac{\Delta k}{2}}^{n=k_i+\frac{\Delta k}{2}} A^2(n), \qquad (1)$$

where k_i is the bin associated to the i-th frequency f_i, Δk is the number of bins in the neighbour of k_i, and $A(n)$ the amplitude associated to the n-th bin. In principle, k_i is found dividing f_i by the spectral resolution. However, f_i has an uncertainty due to the refresh rate frequency, which is not exactly 60 Hz. Hence, the algorithm automatically adjusts the k_i value looking for a peak (i.e. a maximum amplitude) in the neighbor of its initial value. This neighbor was chosen as the corresponding of a 0.5 Hz interval. Then, the Δk for the PSD calculation was chosen as the corresponding of a 0.2 Hz interval.
The last processing step is the SSVEP classification. In the present discussion, there are two features to take into account for each stimulus frequency (corresponding to a possible SSVEP frequency). The PSD corresponding to 10 Hz and 20 Hz are considered for the 10 Hz flickering icon, while the PSD corresponding to 12 Hz and 24 Hz are considered for the 12 Hz flickering icon. These pairs of PSD are classified with a simple machine learning supervised algorithm, the Support Vector Machine (SVM) [26]. Notably, a linear kernel was adopted and the SVM model was trained with experimental data from ten subjects consisting in labeled SSVEP signals. Then, further experimental data from the same subjects were employed for the SVM model evaluation, where the model is adopted to classify unlabeled signals. The details of this procedure are presented in the next section.

RESULTS

Ten subjects, 5 males and 5 females between 22 and 29 years old, took part to the experiments. Each subject was asked to seat on a comfortable chair and limit unnecessary movements. Fig. 4 represents one of the subjects wearing the system. At this moment, the system is still a prototype, and the electrodes are placed with the help of tight bands.

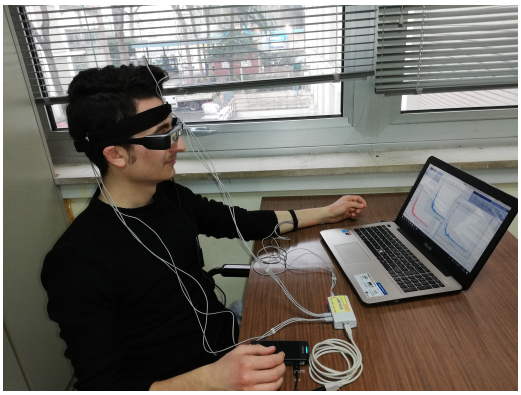


Figure 4: A subject wearing the BCI system during the experimental campaign.

In a first phase, signals were acquired for the training of the SVM model. Hence, the EEG-SMT was connected to a laptop with MATLAB. The laptop was disconnected from the AC power. At the beginning of each test, the raw EEG signal amplitude was checked in the frequency domain in absence of stimulation. A typical signal measured after wearing the EEG-SMT transducer is represented in Fig. 5. In this algorithm training phase, 12 trials were conducted. The brain signal was acquired for 10.0 s, with few seconds between consecutive trials. The subject could chose the icon to stare at, with the constraint that he/she had to choose 6 times 10 Hz and 6 times 12 Hz. The time the user had to stare at the icon corresponded to the acquisition duration, i.e. 10.0 s. Then, the subject declared the choice to the tester in order to assign a label to the trial. The features are extracted for these labeled data as explained in the previous section, and the labeled features are employed for the SVM training thanks to the MATLAB function 'fitcsvm'. The classification accuracy of the trained model was then evaluated with 12 further trials. Labels were necessary for these signals too in order to compare the guessed labels with the actual ones. Moreover, several time windows were considered: with less samples of the signals acquired for 10.0 s, it was possible to analyze the classification accuracy with latencies from 2 s to 10 s. As already mentioned, a linear kernel was taken into account. The employment of a Gaussian ("rbf") kernel was also attempted, but a lower classification accuracy resulted.
The classification accuracy obtained in the evaluation phase is reported in Tab. 1 for the ten subjects. It is to note that, in case of a single subject, one wrongly classified signal corresponds to an accuracy diminishing equal about to 8.3% (1/12). The mean classification accuracy is reported too in Fig. 6, where the accuracy of a random classifier is highlighted: since two classes (two stimuli) are considered, the accuracy of such a classifier is 50%.

DISCUSSION

The trade-off between classification accuracy and latency has been considered. The usual trend is an accuracy diminishing when a shorter time window is taken into ac-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-09

time [s]	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
2	100.0	100.0	58.3	83.3	91.7	83.3	91.7	91.7	50.0	66.7
3	100.0	100.0	91.7	91.7	100.0	83.3	91.7	91.7	50.0	58.3
4	100.0	100.0	91.7	58.3	91.7	91.7	91.7	66.7	50.0	58.3
5	100.0	100.0	100.0	91.7	91.7	91.7	91.7	75.0	91.7	75.0
6	100.0	100.0	100.0	91.7	91.7	91.7	91.7	83.3	83.3	66.7
7	100.0	100.0	100.0	91.7	100.0	91.7	100.0	75.0	91.7	58.3
8	100.0	100.0	100.0	91.7	100.0	100.0	100.0	83.3	83.3	58.3
9	100.0	100.0	100.0	83.3	100.0	100.0	91.7	100.0	83.3	75.0
10	100.0	100.0	100.0	83.3	100.0	100.0	91.7	100.0	83.3	83.3

Table 1: Classification accuracy of SSVEP signals for each subject (S1-S10) at varying time windows (2 s to 10 s).

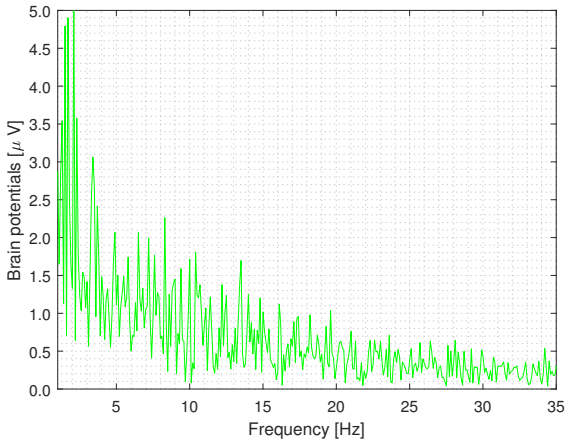


Figure 5: Amplitude spectrum of the raw EEG signal when the user is not stimulated.

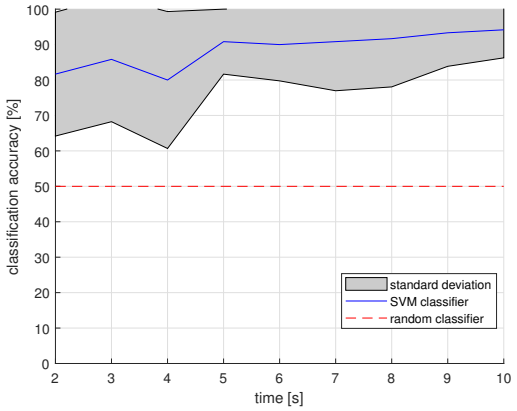


Figure 6: Mean classification accuracy of SSVEP signals calculated across 10 subjects at varying time windows.

count. In our case, the results of the experimental campaign show that, even with a BCI system built with components off-the-shelf and a simple signal processing, the accuracy is as high as 94% with a latency of 10.0 s, dropping to about 80% at 2.0 s. However, some exceptions in the accuracy trend are present depending on the subject. From one side, this could be explained with a varying attention level of the user during the stimulation time or the presence of artifacts localized in time, so that a longer time window is not necessarily better than a shorter one. On the other hand, these variations also depend on the adopted algorithm. Hence, it would be necessary to complicate the algorithm while trying to keep the computational burden low. The algorithm for SSVEP recognition should also be more robust with respect to artifacts. Nonetheless, it is worth noting that artifacts rejection could also be improved by enhancing the mechanical stability of electrodes placing. This implies that there is the need of a first level of engineerization for the adopted prototype.

These results are to be compared with other recent BCI-based systems proposed in literature. In [27], the employment of a single-channel BCI is proposed to build a speller. The user is stimulated for 10.0 s and the reported accuracy is 99.2%. This value was obtained considering 5 subjects. Another example is a speller based again on a single-channel BCI and deep neural networks-based processing [28]. An example of solution integrating AR glasses and BCI is then reported in [29] for the control of a quadcopter. This system employs 14 dry electrodes and 2 reference electrodes, and the achieved accuracy was 85%, by considering 5 subjects executing a flight task. Hence, the performance of the system proposed in this manuscript are compatible with state-of-the art results, though the system was built with off-the-shelf components.

Regarding the final system, it is possible to directly connect the EEG-SMT to the Moverio BT-200. In the Android application developed for the stimuli generation, the possibility to acquire the EEG data from USB and elaborate it was added. This was done to ensure the wearability of the system. However, it has been remarked there is still the need to investigate better signal processing algorithms prior to the porting of this algorithm in Android.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-09

CONCLUSION

A wearable BCI system with off-the-shelf components has been proposed. The BCI relies on SSVEP signals, elicited with flickering icons on AR glasses. It is easy to build such a system with relatively low-cost components as described within the paper, and applications are foreseen in daily life. As an example, the presented device is currently studied for an alternative communication channel with sensors in industrial or domotics applications. An important aspect that was treated is the classification algorithm. Signal features were extracted with a simple power spectral density analysis, and the PSD was calculated, for each icon, at the corresponding flickering frequency and its second harmonic. An experimental campaign was carried out to assess the classification accuracy for varying latencies. Classification was conducted with a Support Vector Machine.
Results demonstrated that the minimum stimulation and acquisition time for the SSVEP signals in our single-channel BCI can be as low as 2.0 s with mean accuracy equal to 80%, going up to 90.0% - 100.0% for some subjects. However, the need of further advances has been highlighted. Future research will deal with the optimization of the integrated system, enhancing the mechanical stability of electrodes placed both on the scalp and on the wrist, and moving on to wireless solutions to avoid the cumbersome wires that were needed in this prototype. An important performance increase is then expected with an improved algorithm, which should however remain enough "light" for an on-line EEG signal processing.

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 et al. Brain-computer interface technology: a review of the first international meeting. IEEE transactions on rehabilitation engineering. 2000;8(2):164–173.
[3] 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.
[4] Kerous B, Skola F, Liarokapis F. EEG-based BCI and video games: a progress report. Virtual Reality. 2018;22(2):119–135.
[5] Perrin X, Chavarriaga R, Colas F, Siegwart R, Millán JdR. Brain-coupled interaction for semi-autonomous navigation of an assistive robot. Robotics and Autonomous Systems. 2010;58(12):1246–1255.
[6] Sanchez-Fraire U, Parra-Vega V, Martinez-Peon D, Sepúlveda-Cervantes G, Sanchez-Orta A, Muñoz-Vázquez A. On the Brain Computer Robot Interface (BCRI) to Control Robots. IFAC-PapersOnLine. 2015;48(19):154–159.
[7] Hu K, Chen C, Meng Q, Williams Z, Xu W. Scientific profile of brain–computer interfaces: Bibliometric analysis in a 10-year period. Neuroscience letters. 2016;635:61–66.
[8] Minguillon J, Lopez-Gordo MA, Pelayo F. Trends in EEG-BCI for daily-life: Requirements for artifact removal. Biomedical Signal Processing and Control. 2017;31:407–418.
[9] Spüler M. A high-speed brain-computer interface (BCI) using dry EEG electrodes. PloS one. 2017;12(2):e0172400.
[10] Penders J. Wearable, Wireless EEG Solutions in Daily Life Applications: What are we missing? IEEE Journal of Biomedical and Health Informatics. 2015;19:6–21.
[11] Tan D, Nijholt A. Brain-computer interfaces and human-computer interaction. In: Brain-Computer Interfaces, 2010, 3–19.
[12] Ahn M, Jun SC. Performance variation in motor imagery brain–computer interface: a brief review. Journal of neuroscience methods. 2015;243:103–110.
[13] Lo C-C, Chien T-Y, Chen Y-C, Tsai S-H, Fang W-C, Lin B-S. A wearable channel selection-based brain-computer interface for motor imagery detection. Sensors. 2016;16(2):213.
[14] Guger C et al. How many people are able to control a P300-based brain–computer interface (BCI)? Neuroscience letters. 2009;462(1):94–98.
[15] Wang Y, Wang R, Gao X, Hong B, Gao S. A practical VEP-based brain-computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2006;14(2):234–240.
[16] Ajami S, Mahnam A, Abootalebi V. Development of a practical high frequency brain–computer interface based on steady-state visual evoked potentials using a single channel of EEG. Biocybernetics and Biomedical Engineering. 2018;38(1):106–114.
[17] Wang Y-T, Wang Y, Jung T-P. A cell-phone-based brain–computer interface for communication in daily life. Journal of neural engineering. 2011;8(2):025018.
[18] Cheng M, Gao X, Gao S, Xu D. Design and implementation of a brain-computer interface with high transfer rates. IEEE transactions on biomedical engineering. 2002;49(10):1181–1186.
[19] Volosyak I, Gembler F, Stawicki P. Age-related differences in SSVEP-based BCI performance. Neurocomputing. 2017;250:57–64.
[20] Cecotti H. A self-paced and calibration-less SSVEP-based brain–computer interface speller. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2010;18(2):127–133.
[21] Angrisani L, Arpaia P, Casinelli D, Moccaldi N. A Single-Channel SSVEP-Based Instrument With Off-the-Shelf Components for Trainingless Brain-Computer Interfaces. IEEE Transactions on Instrumentation and Measurement. 2018.
[22] Si-Mohammed H, Argelaguet F, Casiez G, Roussel N, Lécuyer A. Brain-Computer Interfaces and Aug-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-09

mented Reality: A State of the Art. In: Graz Brain-Computer Interface Conference. 2017.
[23] Klem GH, Lüders HO, Jasper H, Elger C, et al. The ten-twenty electrode system of the International Federation. Electroencephalogr Clin Neurophysiol. 1999;52(3):3–6.
[24] Teplan M et al. Fundamentals of EEG measurement. Measurement science review. 2002;2(2):1–11.
[25] Fatourechi M, Bashashati A, Ward RK, Birch GE. EMG and EOG artifacts in brain computer interface systems: A survey. Clinical neurophysiology. 2007;118(3):480–494.
[26] Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20(3):273–297.
[27] Ajami S, Mahnam A, Abootalebi V. Development of a practical high frequency brain–computer interface based on steady-state visual evoked potentials using a single channel of EEG. Biocybernetics and Biomedical Engineering. 2018;38(1):106–114.
[28] Nguyen T-H, Chung W-Y. A Single-Channel SSVEP-Based BCI Speller Using Deep Learning. IEEE Access. 2019;7:1752–1763.
[29] Wang M, Li R, Zhang R, Li G, Zhang D. A Wearable SSVEP-Based BCI System for Quadcopter Control Using Head-Mounted Device. IEEE Access. 2018;6:26789–26798.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-10

COUNT ON IT: DORSOLATERAL PREFRONTAL CORTEX FOR BCI CONTROL IN LOCKED-IN SYNDROME

S. Leinders[1], E.G.M. Pels[1], B.H. Van der Vijgh[1], M.P Branco[1], Z.V. Freudenburg[1], M.J.E. Van Zandvoort[2], E.J. Aarnoutse[1], M.J. Vansteensel[1], & N.F. Ramsey[1]

[1] UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, Utrecht, the Netherlands

[2] Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, The Netherlands

E-mail: s.leinders@umcutrecht.nl

ABSTRACT: Self-paced Brain-Computer Interfaces (*BCIs*) have traditionally relied on sensorimotor activity for control. Because some individuals with locked-in syndrome (*LIS*) may not be able to control a sensorimotor-based BCI, alternative control strategies should be investigated. As part of the Utrecht NeuroProsthesis project, two individuals with LIS have been fitted with a fully implanted BCI to test feasibility of independent home-use and investigate long term stability of the BCI control features. Neural activity in the UNP system is measured with subdurally implanted electrocorticography electrodes from sensorimotor cortex and left dorsolateral prefrontal cortex (*dlPFC*). Here we present results from the left dlPFC and show for the first time that LIS users are able to reliably activate the left dlPFC at will by mental arithmetic with no training, and that the relevant neural features (high frequency band power) are stable until the last measurement (161 and 61 weeks after implantation, for the two participants). We conclude that dlPFC-based control is a viable control strategy.

INTRODUCTION

Neurodegenerative disease, stroke, or brain injury can lead to locked-in syndrome (*LIS* [1]), characterized by an almost complete paralysis, the inability to speak, and intact cognition. For the subset of LIS people who cannot reliably control traditional augmentative and alternative communication (*AAC*) devices with residual movement [2], Brain-Computer Interfaces (*BCIs*) are the most promising option for self-initiated communication.

Many BCI systems rely on neural activity associated with mental tasks that can be performed at will and in a self-paced manner. Traditionally, self-paced BCI control employs signals measured from sensorimotor areas, allowing people to control a cursor to type [3-8] or control a Windows tablet [9] during research sessions. Moreover, as part of the Utrecht NeuroProsthesis (*UNP*) project, an individual with LIS due to late-stage ALS was able to control commercial AAC software independently and reliably at home [10], without any research staff present, by attempting to move her hand and thereby generate signal changes in the sensorimotor cortex that were converted into brain-clicks.

Although sensorimotor-based BCI has proven its potential to replace lost function, there may be LIS users who cannot reliably control such a BCI. For instance, when paralysis starts at young age, motor imagery may be difficult [11]. Moreover, cortical atrophy due to neurodegenerative disease such as ALS [12] and stroke or injury to specific areas may prevent BCI control based on those areas.

For the subset of LIS people who cannot reliably control a sensorimotor-based BCI, an alternative self-paced control strategy is needed. A promising alternative is to employ signals from the dorsolateral prefrontal cortex (*dlPFC*). First, this region can be activated by a number of self-initiated tasks, such as mental arithmetic [13-15] and random number generation [16-19]. Second, its signals can be measured from the brain surface with electrocorticography (*ECoG),* as demonstrated in an earlier study from our group [15]). In that study, epilepsy patients with temporarily implanted ECoG electrodes over their left dlPFC were able to control a cursor based on high frequency band (*HFB*) power changes generated by mental serial subtraction.

Here, we investigated whether people with LIS are able to reliably activate the left dlPFC. We report data from two participants of the UNP study who were fitted with the fully implantable BCI system.

MATERIALS AND METHODS

Ethics approval: This study was approved by the medical ethics committee of the UMC Utrecht and conducted according to the declaration of Helsinki (2013).

Participants: The first participant (UNP1) was a woman, 58 years old at time of informed consent (September 2015), with late-stage ALS. Her

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-10

communication was limited to eye tracker control and blinking for yes and no, and later to sporadic eye tracker control, lip twitches for yes and no, and the UNP system for typing. The second participant (UNP4) was a woman who was 39 years old at time of informed consent (August 2017). A brainstem stroke in 2004 left her in a locked-in state. She used head movements for control over a switch and joystick, and for yes and no. Both participants gave informed consent with a dedicated procedure described in [10].

Screening and Implantation: A neuropsychologist tested mental arithmetic skills prior to implantation. A pre-surgical fMRI was done to determine the target locations for electrode placement. Based on the fMRI results, ECoG strips (4 electrodes per strip, diameter 4mm; inter electrode distance 10mm; Resume II ®, Medtronic, off-label use) were implanted subdurally through burr holes. Leads were tunneled and connected to an amplifier-transmitter device (Activa ® PC+s, Medtronic, off-label use), which was implanted subcutaneously under the left clavicle. fMRI and neurosurgery details have been described earlier [10].

Task and Data: After implantation, the participants regularly performed a count task, while the neural signal was recorded with the implanted device. The *count task* was a block-design task, comprising alternating active (mental arithmetic) and rest trials of 15 seconds. During active trials, participants had to perform serial subtraction for the duration of the trial. For UNP1, task length was 2 or 5 minutes, and for UNP4, task length was always 3 minutes. Every trial, UNP1 was presented with a starting number and step size, and UNP4 with just a starting number (i.e. she always used the same step size in one run, which was sometimes adjusted between sessions when she indicated it was getting easier). During research sessions at the homes of the participants, an experimenter administered the count task regularly to 1) test whether the participants could regulate HFB (65-95Hz) power in the left dlPFC using mental arithmetic, and 2) test stability of the neural signal features. All count task data presented here was recorded with the same bipolar pair for each of the participants (bipolar pair selection was based on an initial evaluation of all bipolar combinations, using the count task). Count task data were recorded with the implanted device at a sampling frequency of 200Hz, with a high pass filter of 0.5Hz. For analysis, raw voltage was converted offline into power data for frequencies from 1-100Hz in 1Hz bins. The mean HFB power (65-95Hz) was calculated for each 15 second block. The coefficient of determination (r^2) between those mean values and the block design of the task (a binary array) and the respective p-values were calculated for each run.

RESULTS

UNP1 performed 61 runs of the count task, with the last run recorded 161 weeks after implantation. Mean HFB power correlated significantly with the task in all runs, except for two runs in weeks 41 and 42 (Figure 1A). The average r^2 value was 0.68 ± 0.20.

UNP4 performed 31 runs of the count task and her last run was recorded 61 weeks after implantation. Again, HFB power correlated significantly with the task for all runs (Figure 1B), except in 3 runs in weeks 6, 35 and 50. The average r^2 value was 0.61 ± 0.18.

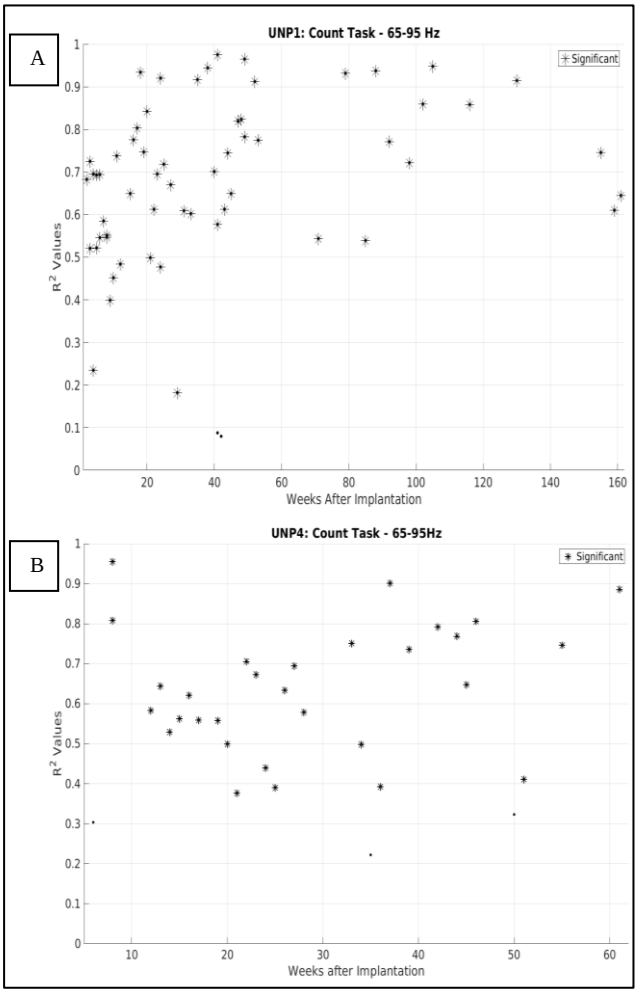


Figure 1: r^2 values calculated between the block-design of the count task and the mean HFB power of active and rest blocks for both participants. X-axis indicates weeks since implantation. Asterisks indicate a significant r^2 value. Due to different task lengths (i.e. sample points), the significance level of r^2 values can differ between sessions or participants (e.g. UNP1 in week 29 was significant, but UNP4 in week 6 was not).

DISCUSSION

Both participants were able to activate their left dlPFC – as measured by HFB power – by mental arithmetic for a period of 161 and 61 weeks (UNP1 and UNP4, respectively). They were able to activate their dlPFC at will from study start, meaning no training was required for volitional HFB power increase with mental

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-10

arithmetic.

Importantly, for both participants, HFB was not always significantly modulated by the task. Moreover, compared to sensorimotor results from UNP1 [10], count task r^2 values were lower (average sensorimotor r^2 value was 0.89 ± 0.16; all r^2 values plotted in [10]). However, count task r^2 values reported here are in the same range as count task r^2 values reported for three able-bodied epilepsy patients in [15] (r^2 = 0.79, 0.53, and 0.65). The lower r^2 in the count task could be explained by the fact that a motor attempt strategy is more consistent than the use of variable starting numbers and step sizes, some of which may be easier than others. Varying degrees of difficulty may be associated with varying degrees of HFB power increase in some trials, and thus, on average, lower r^2 values. It has been shown before that HFB response in left-dlPFC varies between different mental calculation trials [15]. Moreover, the claim that difficulty may affect HFB power modulation is supported by the marked difference in difficulty and r^2 values between UNP4's first (week 6) and following (week 8 and thereafter) sessions. In the first session, UNP4 used a fixed step size (2) which – she reported - was quite easy. In the next session, she switched to a larger step size, which resulted in an increase of the r^2 values in the following sessions.

The dlPFC may provide a useful complement to a sensorimotor-based BCI system. We envision users of such a system could switch to dlPFC control when sensorimotor control does not work optimally, for instance due to interference caused by care or passive movement of limbs. However, it is currently unclear to what extent everyday cognitive tasks would interfere with a dlPFC-based BCI. Similar to sensorimotor areas, dlPFC-based BCI may work well in some situations, but not in others. This topic deserves further investigation. Moreover, HFB modulation during mental arithmetic (as expressed here by r^2 values) fluctuates. To what extent the relative unruly HFB modulation can be used to control a BCI application reliably is unknown.

The next steps in this avenue of BCI research are ongoing in our group, and include using HFB power from dlPFC for cursor control and for generating brain-clicks to control scanning software for communication purposes. It is preferable if the mental tasks used for BCI control can be performed without cues. Therefore, future research should in part focus on training the user to activate the left dlPFC without cues, testing which tasks work best (e.g. random number generation, mental arithmetic, or a combination) and elucidate how this training could be optimized.

CONCLUSION

These results indicate that the dlPFC may be a viable alternative for BCI users, when sensorimotor-based BCI is not an option. Future research should confirm the feasibility of home use of a dlPFC-based BCI by a user with LIS.

FUNDING

This study was funded by grants from the Dutch Technology Foundation STW and European Research Council.

ACKNOWLEDGEMENTS

We thank our participants for their courage, motivation, input, and hospitality.

REFERENCES

1. Smith, E., & Delargy, M (2005). Clinical Review: Locked-in syndrome. BMJ 2005;330:406, https://doi.org/10.1136/bmj.330.7488.406
2. Glennen, L. G., & DeCoste, D. C. (1997). The Handbook of Augmentative and Alternative Communication. Singular Publishing Group, Inc. San Diego, London.
3. Wolpaw, J. R., & McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proceedings of the national academy of sciences, 101(51), 17849-17854.
4. Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., ... & Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature, 442(7099), 164.
5. Kim, S. P., Simeral, J. D., Hochberg, L. R., Donoghue, J. P., & Black, M. J. (2008). Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. Journal of neural engineering, 5(4), 455.
6. Jarosiewicz, B., Sarma, A. A., Bacher, D., Masse, N. Y., Simeral, J. D., Sorice, B., ... & Cash, S. S. (2015). Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. Science translational medicine, 7(313), 313ra179-313ra179.
7. Bacher, D., Jarosiewicz, B., Masse, N. Y., Stavisky, S. D., Simeral, J. D., Newell, K., ... & Hochberg, L. R. (2015). Neural point-and-click communication by a person with incomplete locked-in syndrome. Neurorehabilitation and neural repair, 29(5), 462-471.
8. Nuyujukian, P., Sanabria, J. A., Saab, J., Pandarinath, C., Jarosiewicz, B., Blabe, C. H., ... & Hochberg, L. R. (2018). Cortical control of a tablet computer by people with paralysis. PloS one, 13(11), e0204566.
9. Pandarinath, C., Nuyujukian, P., Blabe, C. H., Sorice, B. L., Saab, J., Willett, F. R., … &

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-10

Henderson, J. M. (2017). High performance communication by people with paralysis using an intracortical brain-computer interface. Elife, 6, e18554 Pandarinath, C., Jarosiewicz, B., Blabe, C. H., ... & Hochberg, L. R. (2018). Cortical control of a tablet computer by people with paralysis. PloS one, 13(11), e0204566..

10. Vansteensel, M. J., Pels, E. G., Bleichner, M. G., Branco, M. P., Denison, T., Freudenburg, Z. V., ... & Van Rijen, P. C. (2016). Fully implanted brain–computer interface in a locked-in patient with ALS. New England Journal of Medicine, 375(21), 2060-2066.
11. Mutsaarts, M., Steenbergen, B., & Bekkering, H. (2007). Impaired motor imagery in right hemiparetic cerebral palsy. Neuropsychologia, 45(4), 853-859.
12. Mohammadi, B., Kollewe, K., Samii, A., Dengler, R., & Münte, T. F. (2011). Functional neuroimaging at different disease stages reveals distinct phases of neuroplastic changes in amyotrophic lateral sclerosis. Human brain mapping, 32(5), 750-758.
13. Ramsey, N. F., van de Heuvel, M. P., Kho, K. H., & Leijten, F. S. (2006). Towards human BCI applications based on cognitive brain systems: an investigation of neural signals recorded from the dorsolateral prefrontal cortex. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14(2), 214-217.
14. Vansteensel, M. J., Bleichner, M. G., Freudenburg, Z. V., Hermes, D., Aarnoutse, E. J., Leijten, F. S., ... & Ramsey, N. F. (2014). Spatiotemporal characteristics of electrocortical brain activity during mental calculation. Human brain mapping, 35(12), 5903-5920.
15. Vansteensel, M. J., Hermes, D., Aarnoutse, E. J., Bleichner, M. G., Schalk, G., van Rijen, P. C., ... & Ramsey, N. F. (2010). Brain–computer interfacing based on cognitive control. Annals of neurology, 67(6), 809-816.
16. Daniels, C., Witt, K., Wolff, S., Jansen, O., & Deuschl, G. (2003). Rate dependency of the human cortical network subserving executive functions during generation of random number series–a functional magnetic resonance imaging study. Neuroscience Letters, 345(1), 25-28.
17. Knoch, D., Brugger, P., & Regard, M. (2004). Suppressing versus releasing a habit: frequency-dependent effects of prefrontal transcranial magnetic stimulation. Cerebral Cortex, 15(7), 885-887.
18. Jahanshahi, M., Saleem, T., Ho, A. K., Dirnberger, G., & Fuller, R. (2006). Random number generation as an index of controlled processing. Neuropsychology, 20(4), 391.
19. Koike, S., Takizawa, R., Nishimura, Y., Marumo, K., Kinou, M., Kawakubo, Y., ... & Kasai, K. (2011). Association between severe dorsolateral prefrontal dysfunction during random number generation and earlier onset in schizophrenia. Clinical neurophysiology, 122(8), 1533-1540.
20. Crone, N. E., Miglioretti, D. L., Gordon, B., & Lesser, R. P. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain: a journal of neurology, 121(12), 2301-2315.
21. Levy, R., & Goldman-Rakic, P. S. (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. In Executive control and the frontal lobe: Current issues (pp. 23-32). Springer, Berlin, Heidelberg.
22. Wilson, F. A., Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. Science, 260(5116), 1955-1958.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-11

ASYNCHRONOUS DETECTION OF ERROR-RELATED POTENTIALS USING A GENERIC CLASSIFIER

C Lopes-Dias[1], A I Sburlea[1], G R Müller-Putz[1]

[1]Institute of Neural Engineering, Graz University of Technology, Graz, Austria

E-mail: gernot.mueller@tugraz.at

ABSTRACT: Error-related potentials (ErrPs) can be used to improve BCIs' performance but its use is often withheld by long calibration periods. We recorded EEG data of 15 participants while controlling a robotic arm towards a target. In 30 % of the trials, the protocol prompted an error during the trial in order to trigger ErrPs in the participants. For each participant, we trained an ErrP classifier using the data of the remaining 14 participants. Each of these classifiers was tested asynchronously on the data of the selected participant. The threshold that maximized the product of the average true positive rate (TPR) and the average true negative rate (TNR) was $\tau = 0.7$. For this threshold, the average TPR was 53.6 % and the average TNR was 82.0 %. These results hint at the feasibility of transferring ErrPs between participants as a reliable strategy to reduce or even remove the calibration period when training ErrP classifiers to be used in an asynchronous manner.

INTRODUCTION

Brain-computer interfaces (BCIs) are a suitable tool to help restoring some autonomy to people with severe motor disabilities [1,2,3]. Most BCIs rely on converting modulated brain activity of a user (often measured using electroencephalography (EEG)) into commands of an external device, such as a robot. Nevertheless, the performance of most BCIs is not optimal and, occasionally, the interface misinterprets the intention of its user and thus a wrong command is executed. The user's awareness of the committed mistake is associated with a neural pattern named error-related (ErrP), which is also measurable by EEG [4] .

Incorporating ErrPs' detection in a BCI can help to improve its performance [5, 6]. A main barrier to its widespread use is the calibration time necessary to train ErrP classifiers: many trials are needed to train a classifier and errors should not occur too often to still be perceived as so. Two main approaches have been proposed to reduce the training duration of ErrP classifiers, based on either transferring information between different tasks or transferring information between different participants. Iturrate and colleagues studied the use of classifiers trained with ErrPs from one observation task and tested in ErrPs from another observation task, using latency correction [7,8]. Kim and colleagues studied the use of an ErrP classifier trained with ErrPs from an observation task and tested with ErrPs from an interaction task (and vice-versa) [9,10]. Nevertheless, Ehrlich and colleagues, did not recommend re-using ErrP classifiers across different experimental tasks [11]. Kim and colleagues also studied the use of an ErrP classifier trained with ErrPs from several subjects and tested in ErrPs from another subject [9]. These studies suggest that transferability of ErrPs is viable in the context of discrete BCIs (in which all events occur in a discrete way).

Recently, an effort has been made to develop BCI paradigms that promote a smoother and more intuitive interaction with their users, by relying on continuous control or actions - continuous BCIs [12,13,14]. In this context, the user's error awareness can occur at any moment and is not, necessarily, time-locked to specific events, requiring an asynchronous detection of ErrPs. The existence of ErrPs in continuous contexts as well as its asynchronous detection has been established [15, 16, 17,18]. Another approach to improve BCIs consists in developing BCIs that closer resemble end-user applications, in which users interact with or observe robots [10,19,20,21,22].

In this work, we developed a paradigm in which the user has continuous control over a robotic arm in a task in which errors are triggered by the paradigm. We studied the electrophysiological signature of the ErrPs in this task and, additionally, investigated the feasibility of using a generic ErrP classifier trained on the ErrPs of 14 participants by testing it asynchronously with data of another participant.

MATERIALS AND METHODS

EEG recording: We recorded EEG and EOG data at a sampling frequency of 500 Hz, using BrainAmp amplifiers (Brain Products, Munich, Germany). We used 61 EEG electrodes and 3 EOG electrodes. The EEG electrodes were placed at positions Fp1, Fp2, AF3, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, 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, PO9, PO7, PO3, POz, PO4, PO8, PO10, O1, Oz, and O2. The ground electrode was placed at position AFz and the reference electrode was placed on the right mastoid. The electrodes were placed above the nasion and below the

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-11

outer canthi of the eyes.

Participants: We recorded 15 right-handed healthy volunteers (5 female). The participants were, on average, 23.4 ± 2.5 years old (mean ± std). Participants were paid 7.50 € per hour and, before the experiment, read and signed an informed consent form that was previously accepted by the local ethical committee.

Experiment layout: Figure 1 depicts the layout of the experiment. Participants sat in front of a table, with their right hand lying flat on the table, covered by a wooden structure. On the ceiling of this structure was a Leap Motion device that tracked their right hand movements (Leap Motion, San Francisco, USA). On the right of the participants was a robotic arm (Jaco Assistive robotic arm - Kinova Robotics, Bonn, Germany). On top of the wooden structure were two violet boxes representing the physical targets, centred in relation to the home position of the robot's hand. Behind the wooden structure was a monitor that displayed information regarding the experiment. During the trials, the participants could control the position of the robot's hand on an approximately horizontal plan, by moving their right hand on the table. We considered robot's hand displacement to be three times bigger than the participants' hand displacement, to reduce the range of the participants' movements.

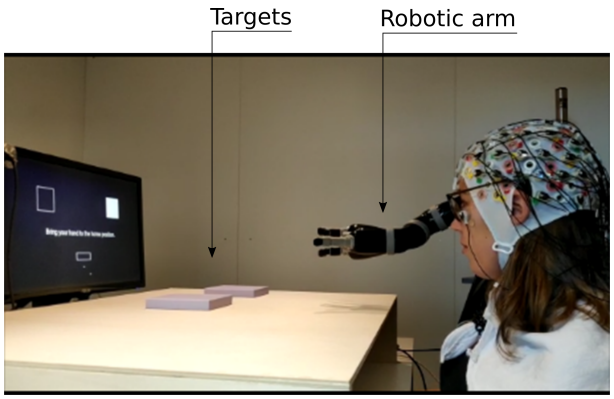


Figure 1: Experimental setup during the pre-trial period. In this image, the robot is at its home position. The squares on the screen indicate that, in the next trial, the participant should move the robot's hand towards the right target (purple box) on the wooden structure. The screen also shows the home position for the participant's hand (rectangle on the bottom part of the screen). The text on the screen (not readable in the picture) states 'Bring your hand to the home position'.

Experiment overview: The experiment consisted of 8 blocks of 30 *trials* each. Each block contained 21 *correct trials* and 9 *error trials* (70% and 30%, respectively). The position of the error trials within the block was randomly generated, using an uniform distribution.

Pre-trial period: During this period, on the upper part of the screen were displayed two squares, representing the two targets on the wooden structure. One of the squares was filled in white, representing the selected target for the next trial, and the other had no fill. On the bottom part of the screen was a rectangle representing

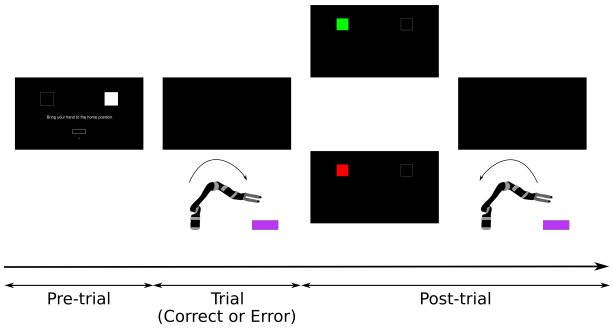


Figure 2: Experimental protocol. During the pre-trial period, the participants can rest for as long as they wish. The pre-trial period ends and a trial starts, when the participant brings his/her right hand to its home position (the bottom rectangle). During the trials, the screen is black. The participants were instructed to bring the robot's hand to the selected target (indicated by the white square). A trial finishes either when the robot reaches the target or after 6 seconds (if the target was not reached). Afterwards (post-trial), the squares reappear on the screen for 1.5 seconds and give feedback regarding hitting the target (a green square indicates that the target was hit and a red square indicates that the target was not hit). Then the screen turns black and the robot automatically returns to its home position and a new pre-trial period starts.

the home position for the participants' hand. The participants' hand was represented by a dot on the screen. A new trial started when the participants' hand entered its home position. The participants could use the pre-trial period to rest for as long as they needed. Participants were instructed to bring their hand to below the home position, to fixate their gaze on the selected physical target and to enter the home position when they felt ready to start a new trial. Participants were also instructed not to move their gaze during the entire trial duration, in order to minimize eye movements.

Trials: The aim of each trial was to move the robot's hand to the selected target. During the trials, the screen was black. A trial finished when the robot's hand was above the target (hit) or after 6 seconds (no hit). Afterwards, as shown in Figure 2, the two squares from the pre-trial period reappeared on the screen for 1.5 seconds and the previously filled square was now filled in either green (hit) or red (no hit). Then, the screen would turn black, the robot's hand would automatically move to its home position and a new pre-trial period would start.

Error trials: In these trials, the paradigm triggered an *error* during the trial. The *error* consisted on halting the participant's control of the robot and adding a 5 cm upwards displacement to the robot's hand.The errors occurred randomly when the robot's hand was within 25 % to 65 % of the minimal forward displacement necessary for the robot to hit the target. Participants perceived the error by noticing the robot stopping and lifting. After the error happened, the participants could not control the robot until the trial ended. Participants were instructed to remain still. The error trials lasted 6 seconds.

Correct trials: In these trials, no error was triggered by the paradigm. The participants reached the selected target

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-11

in 99.7 ± 0.5 % (mean $\pm$ std) of the correct trials. Correct trials lasted, on average, 2.06 ± 0.17 seconds (mean $\pm$ std).

Preprocessing the data: The eye movements and blinks were removed from the EEG data, using the artefact subspace subtraction algorithm [23]. The EEG data was then filtered between 1 and 10 Hz with a Butterworth causal filter of order 4.

Electrophysiological analysis: For the electrophysiological analysis, we cut the EEG data in 1.5 s epochs. For the error trials, we considered the interval [-0.5, 1]s time-locked to the error onset (0 s). Since correct trials have no intrinsic onset, we defined a virtual onset, occurring one second after the start of the trial (at a time-point in which errors could already occur). For the correct trials we considered the interval $[-0.5, 1]$s, time-locked to the virtual onset (0 s).

Asynchronous ErrP classification with a generic classifier: For every participant we trained an ErrP classifier with two classes (correct and error) using the data from the remaining 14 participants. In order to train each of these classifiers, we considered as features for the error class the amplitudes of all EEG channels at all time points within the window $[0.30, 0.75]$s after the error onset. Similarly, we considered as features for the correct class the amplitudes of all EEG channels at all time points within the window $[0.30, 0.75]$s after the virtual onset. Afterwards, in order to reduce the number of features, we applied principal component analysis (PCA) to the feature matrix and kept the components that preserved 99 % of the data variance. These components were used to train a shrinkage LDA classifier [24]. Each of these classifiers was tested asynchronously in the data of the participant not used for training. For that, we slid a 450 ms window through the trials, obtaining an output from the classifier every 18 ms.

For every fixed threshold τ (τ from 0 to 1 in steps of 0.025), we considered an *error detection* when the classifier's probability for the error class (p_e) was greater or equal to the threshold τ for two consecutive windows ($p_e \geq \tau$). As an evaluation metric for the asynchronous classification, we defined as true negative trials (TN trials) the correct trials in which no error detection occurred. We defined as true positive trials (TP trials), the error trials in which no error detection occurred before the error onset and at least one error detection occurred within 1.5 seconds after the error onset. We considered the group performance to be optimal for the threshold that maximized the product of the average TPR and the average TNR.

RESULTS

Figure 3 displays the grand average correct and error signals at channel FCz (green and red solid lines respectively). The 95 % confidence interval for the average curves are represented by the shaded green and red areas. The time-regions in which correct and error signals were significantly different are represented by grey shaded areas (Wilcoxon signed-rank tests, $p < 0.01$, Bonferroni corrected). Figure 3 displays also the topoplots for the correct and error grand average signals at the peaks of the grand average error signal (t =0.354 s and $t = 0.568$ s).

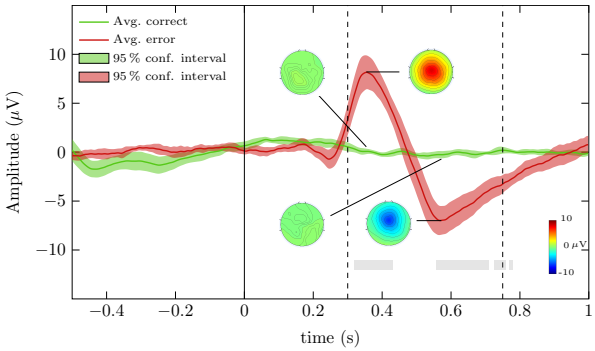


Figure 3: Grand average correct and error signal at channel FCz (green and red solid lines, respectively). The shaded areas represent the 95 % confidence interval for the average signals. The grey regions represent the time-regions in which correct and error signals were significantly different (Wilcoxon signed-rank tests, $p < 0.01$, Bonferroni corrected).The topoplots for the correct and error grand average signals are displayed for t =0.354 s and $t = 0.568$ s. The time point $t = 0$ represents the error onset of error trials and the virtual onset of correct trials

Figure 4 shows the average true negative rate (TNR) and the average true positive rate (TPR) (represented with green and red solid lines, respectively), for all the tested thresholds in the asynchronous ErrP classification with a generic classifier. The chance-level TNR and TPR were calculated by performing the same classification procedure with shuffled training labels. The 95 % confidence intervals for the average curves are represented by shaded areas. The threshold that maximized the group performance was $\tau = 0.700$. For this threshold, the average TNR was 82.0 % and the average TPR was 53.6 %.
Figure 5 depicts the individual TNR and TPR of each participant. It depicts also the threshold that maximizes group performance ($\tau = 0.700$, grey dashed lines) and the thresholds that maximizes the individual performance (blue dashed lines).

DISCUSSION

We developed an experimental task relying on continuous control of a robot towards a target. In 30 % of the trials (error trials), an error was triggered by the paradigm, causing the participants to loose control over the robot during the trial. We then studied the electrophysiological features associated with the error trials. The peaks of the error signal occurred at t =0.354 s and t =0.568 s. Our results are not directly comparable with state-of-the-art literature because we processed the EEG signal with a causal filter, causing the N200 component of the ErrP to shift to around 600 ms. We decided to keep the causal filter because it depicts the ErrP's shape in the scenario of an online ErrP decoder, bringing awareness to the fact that ERP shapes can be influenced by the filter used to

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-11

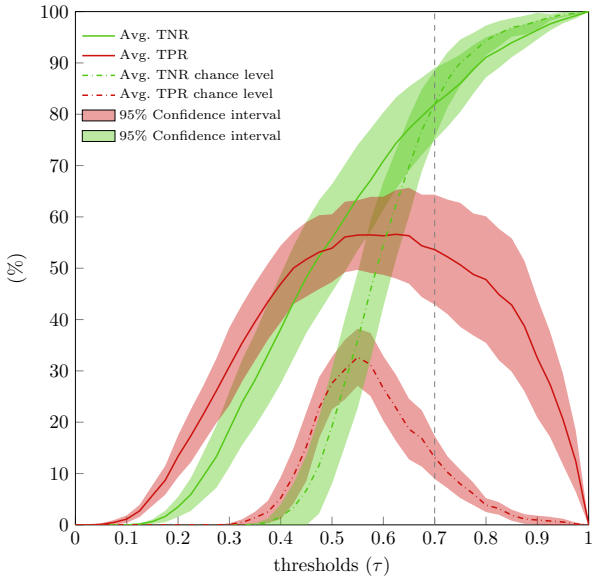


Figure 4: Average TNR and average TPR (green and red solid lines, respectively) for the different thresholds tested in the asynchronous ErrP classification with a generic classifier. The chance-level TNR and TPR are depicted with green and red dashed lines. The shadowed areas represent the 95 % confidence intervals for the average curves. The threshold that maximizes the group performance is represented with a grey vertical dashed line.

process the signal.

Afterwards, we evaluated the feasibility of transferring ErrP information across participants, by training a classifier with the data from 14 participants and testing it with the data of the remaining participant in an asynchronous manner (generic classifier). From Figure 4, we observe that the average TPR is above chance-level for all the thresholds and that the average TNR is increasing with the threshold. This points to the feasibility of using such classifiers as a starting point for an adaptive BCI. In Figure 5, it is possible to compare the individual performance of every participant with the generic classifier. We observe that participants with higher individual threshold present minor or negligible drops in performance with the use of a generic classifier tuned to the group performance (e.g. participants 1, 2 and 3). On the other hand, participants with lower individual threshold can present major performance drops (e.g. participants 5 and 8). This indicates that the performance of such classifier is still determined by individual characteristics of the participants. Nevertheless it seemed a suitable option for the majority of the participants.

CONCLUSION

In this work we showed the feasibility of transferring ErrP information across participants, by training a classifier with the data from 14 participants and testing it with the data of the remaining participant in an asynchronous manner. We then showed that, although the performance of such classifiers is still dependent on individual characteristics of the participants, the majority of them would benefit from such generic approach. Therefore, we believe that transferring ErrP information across participants is a viable alternative to reduce the calibration period in a scenario of asynchronous ErrP classification, as a starting point for an adaptive BCI.

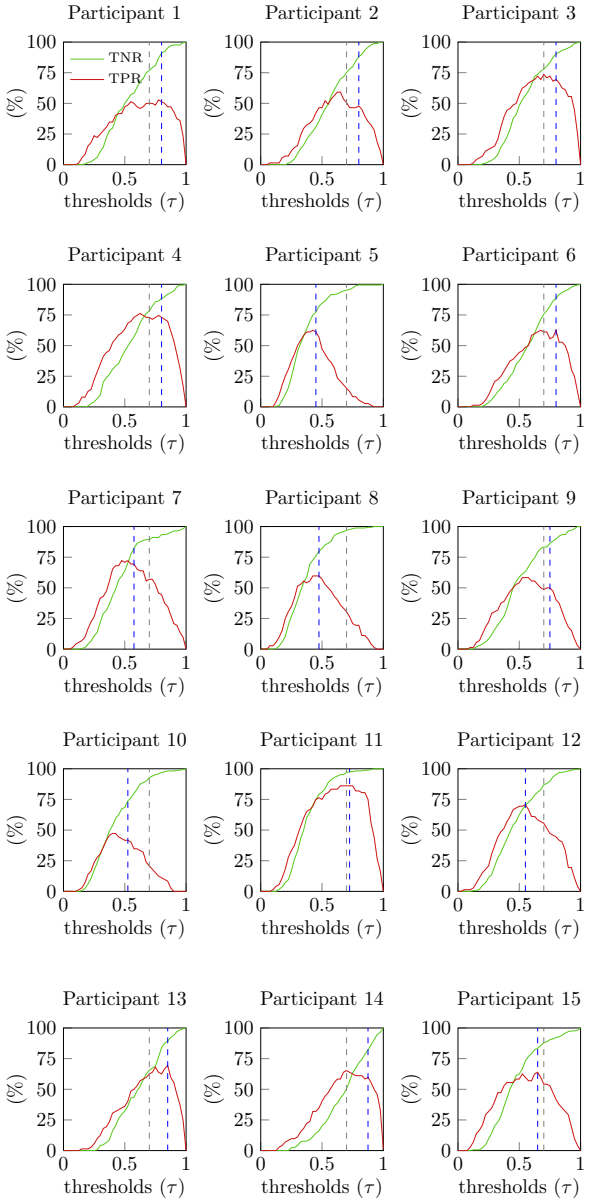


Figure 5: Individual TNR and TPR (green and red solid lines, respectively) for the different thresholds tested in the asynchronous ErrP classification with a generic classifier. The threshold that maximizes the group performance is represented with a grey dashed line ($\tau = 0.7$). The threshold that maximizes the individual performance is represented with a blue dashed line.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the fruitful discussions with the members of 'Feel Your Reach'. This work was supported by Horizon 2020 ERC Consolidator

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-11

Grant 681231 'Feel Your Reach'.

REFERENCES

[1] Millán JdR et al. Combining Brain–Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges. Frontiers in Neuroscience 2010; 4:161.
[2] Müller-Putz G et al. Towards Noninvasive Hybrid Brain–Computer Interfaces: Framework, Practice, Clinical Application, and Beyond, in Proc. of the IEEE, 2015, 103(6), 926-943.
[3] Müller-Putz GR et al. Towards non-invasive brain-computer interface for hand/arm control in users with spinal cord injury, in Proc. 5th International Winter Conference on Brain-Computer Interface (BCI), Sabuk, 2017, 63-65.
[4] 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.
[5] Schmidt NM, Blankertz B, Treder MS. Online detection of error-related potentials boosts the performance of mental typewriters. BMC Neuroscience. 2012; 13(1):13-19.
[6] 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):1388-1337.
[7] Iturrate, I, Chavarriaga R, Montesano L, Minguez J, Millán JdR. Latency correction of error-related potentials reduces BCI calibration time, in Proc. of the 6th International Brain-Computer Interface Conference, Graz, Austria, 2014.
[8] Iturrate I, Chavarriaga R, Montesano L, Minguez J, Millán JdR. Latency correction of event-related potentials between different experimental protocols. Journal of Neural Engineering 2014; 11(3):036005.
[9] Kim SK, Kirchner EA. Handling few training data: classifier transfer between different types of error-related potentials. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2016, 24(3):320-332.
[10] 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, 7:17562.
[11] Ehrlich S, Cheng G. A Feasibility Study for Validating Robot Actions Using EEG-Based Error-Related Potentials International Journal of Social Robotics. 2018.
[12] Doud AJ, Lucas JP, Pisansky MT, He B. Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface. PLOS ONE. 2011; 6(10):e26322.
[13] Coyle D, Garcia J, Satti A, Mcginnity TM. EEG-based continuous control of a game using a 3 channel motor imagery BCI: BCI game. in 2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), Paris, France, 2011, 1-7.
[14] Galán F et al. A brain-actuated wheelchair: Asynchronous and non-invasive Brain–computer interfaces for continuous control of robots. Clinical Neurophysiology. 2008; 119(9):2159 - 2169.
[15] 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.
[16] Omedes J, Iturrate I, Chavarriaga R, Montesano L. Asynchronous Decoding of Error Potentials during the Monitoring of a Reaching Task 2015 IEEE International Conference on Systems, Man, and Cybernetics, Kowloon, 2015, 3116-3121.
[17] 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; 9:155.
[18] 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.
[19] Kreilinger A, Neuper C, Müller-Putz G. Error potential detection during continuous movement of an artificial arm controlled by brain-computer interface.Med Biol Eng Comput. 2012; 60:223.
[20] Iturrate I, Chavarriaga R, Montesano L, Minguez J, Millán JdR. Latency correction of error potentials between different experiments reduces calibration time for single-trial classification, in 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, 2012, 3288-3291.
[21] Salazar-Gomez AF, DelPreto J, Gil S, Guenther FH, Rus D. Correcting robot mistakes in real time using EEG signals, in 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017, 6570-6577.
[22] Ehrlich SK, and Cheng G. Human-agent co-adaptation using error-related potentials. Journal of Neural Engineering. 2018; 15(6):066014.
[23] Kobler RJ, Sburlea AI, Müller-Putz GR. A comparison of ocular artifact removal methods for block design based electroencephalography experiments, in Proc. of the 7th International Brain-Computer Interface Conference, Graz, 2017, 236-241.
[24] Blankertz B, Lemm S, Trede M, Haufe S, and Müller KR. Single-trial analysis and classification of ERP components — A tutorial. NeuroImage; 2011 56(2):814 - 825.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-12

EVALUATING AUTOMATIC ARTIFACT CORRECTION FOR ONLINE HYPOTHESIS TESTING

A. M. Freitas[1,2], G. Sanchez[1], F. Lecaignard[1], E. Maby[1], A. B. Soares[2], J. Mattout[1]

[1] Lyon Neuroscience Research Center, Lyon, France
[2] Federal University of Uberlândia, Uberlândia, Brazil

E-mail: amanda.medeirosf@yahoo.com.br

ABSTRACT: In BCI, artifact removal remains an acute challenge. Filtering must be efficient in removing artifacts while preserving relevant features, e.g. event-related potentials (ERP) like the mismatch negativity (MMN). MMN is a prediction error signal whose modulations reflect human perceptual inference and learning. Characterizing these subtle processes requires fitting non-linear models onto single-trial data. And disentangling between alternative models is challenging because of a low signal-to-noise ratio. We evaluated four methods for online artifact removal. We mimicked online data processing using real electroencephalography (EEG) data from an auditory oddball paradigm. We compared the four approaches with standard offline analysis, in their ability to reveal (i) the MMN, (ii) the MMN modulations by the manipulation of the predictability of a sound sequence and (iii) the most likely learning mechanism at play. Artifact Subspace Reconstruction (ASR) and Empirical Mode Decomposition (EMD) were the most successful. Interestingly, they even proved more sensitive than the offline analysis, likely because they avoid rejecting trials.

INTRODUCTION

Brain-computer interfaces (BCI) measure and process brain activities for control, monitoring or rehabilitation purposes [1]. Whatever the application, an acute challenge is to extract reliable features and to translate them into meaningful information for the machine, in real time. Moreover, the challenge intensifies when brain signals are measured by non-invasive methods, typically electroencephalography (EEG).

EEG, compared with other techniques, has a poor spatial resolution and is often contaminated with electrical activities generated either by endogenous physiological sources (such as eye movements) or externally over the scalp (line noise) [1]. These artifacts, in addition to decreasing the signal quality, directly influence the classification performance of EEG-based BCIs [2] and may add to other possible reasons why some users are unable to control such BCI systems [3].

In the last few years, some approaches have been investigated to detect and remove artifacts in real-time, such as Artifact Subspace Reconstruction (ASR) [4], Fully Online and automated artifact Removal for brain-Computer interfacing (FORCe) [5], online Empirical Mode Decomposition (EMD) [6–8], and online Independent Component Analysis (ICA) [9–11]. However, none of these methods can be acknowledged as the "gold standard" yet, and to our knowledge, no research has been conducted to compare these methods with standard offline preprocessing, nor to enable online hypotheses testing for optimized cognitive neurosciences experiments [12]. Ideally, these filtering methods should provide a clean signal that contains the same relevant information as if the data would have been processed offline.

The mismatch negativity (MMN) is an automatic evoked EEG component that is typically observed during the listening of an oddball tone sequence, when subtracting the average response to deviant (rare) sounds from the average response to standard (frequent) sounds. This negative deflection occurs usually between 150 and 250 at frontal and central scalp electrodes, even in the absence of attention oriented towards the sounds [13,14].

The MMN is viewed as a prediction error signal, that is a measure of the discrepancy between the expected sensation (a standard) and the observed sensory input (a deviant). The computation of prediction errors and their resolution obey hierarchical predictive coding, where cortical processes (higher levels) send predictions to lower hierarchical levels. Whenever the current prediction fails, prediction errors are forwarded up to higher hierarchical levels, following an ascending pathway [15].

Besides, a few offline auditory-based cognitive neuroscience studies demonstrated that MMN modulations reflect an implicit learning process, such that a predictable sequence of auditory stimuli yields a

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-12

decreased MMN. In other words, the more predictable the occurrence of a deviant stimulus, the more reduced the prediction error, hence the lower the MMN amplitude [14,16]. However, revealing these learning processes requires fitting non-linear mathematical models with unknown parameters, at the single trial level and for each subject independently. It has been shown that online adaptive designs could help to optimize the selection and fitting of such learning models, at the individual level [12]. In this aim in particular, obtaining clean single trial data in real-time is highly crucial.

In this paper, we evaluate and compare the performance of ASR, online EMD, online ICA, and FORCe, for online artifact correction using real data. In our experiments, we mimic real-time data processing. Precisely, we compare the performance of those methods with offline data processing, evaluating their ability (i) to reveal a significant MMN to auditory oddball stimuli; (ii) to reveal the more subtle modulation of the MMN by the predictability of the sound sequence and (iii) to identify the right learning model based on single-trial data fitting.

MATERIALS AND METHODS

Experiment: We used the EEG dataset from Lecaignard et al. [14]. These data were obtained in 20 healthy adults (10 female, mean age: 25 $\pm$ 5 years) who passively listened to auditory oddball sequences, where the occurrence of deviant sounds was either predictable or unpredictable. We implemented the same data processing performed in this study and considered it as a standard offline approach to compare with the online artifact correction methods.

Data processing: Three of the four online artifact correction methods need to be calibrated. ASR requires a clean EEG calibration signal, while ICA requires a calibration signal that contains the artifact to be removed. EMD also requires a clean calibration signal with approximately the same window size as the subsequent epochs to be preprocessed. FORCe does not need to be calibrated. The first 30s of each block of stimulations (each subject performed 4 blocks, see [14] for more details) were used to calibrate the different methods when needed. Those initial segments were not used in the subsequent MMN analyses. These signal windows were band-pass filtered at 2-20 Hz using the inverse Fast Fourier Transform (FFT) filtering available in the EEGLAB software environment [17]. In addition, the ASR calibration signal was processed to eliminate eye blinks and samples greater than 50 μV, in other words, samples that exceeded this threshold were not used for ASR calibration. For EMD, only 1s of the 30s of the filtered signal were used.

To compute ICA we used the infomax algorithm from EEGLAB. ICA was applied to all electrodes of calibration data, and the independent components were rejected by visual inspection.

After calibration, the signal was processed over epochs starting 200ms before stimulus onset and ending 500ms after. Each epoch was also band-pass filtered at 2-20 Hz prior to processing (inverse FFT). As FORCe requires at least 1 second until it is capable of removing artifacts, specifically for this method, we used epochs starting 1.2s before stimulus onset.

The performance of ASR depends on the choice of hyper-parameters [18]. Therefore, after empirical testing, with the exception of the window size parameters (window length = 0.7s, step size = 0.5s and look-ahead = 0.2s), we used the default values.

After applying the different online artifact correction methods, we removed the last 200ms from the analysis, ensuring that the signals were not contaminated by edge artifacts introduced by the filter. Furthermore, we only used as feature of interest a specific time windows from 160 to 190 ms after stimulus onset (centered on the MMN response) and channels (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) as in Lecaignard et al. [14], which were found to be the most responsive channels. We calculated the averaged signal in that spatio-temporal region of interest, for deviant and its preceding standard stimuli in order to compute the MMN amplitude.

Models: In this part of the study, we only compared the online artifact correction methods that proved able to reveal the MMN as well as its modulation by predictability. Therefore, we applied a two-tailed *t*-test to test whether the unpredictable and predictable MMN amplitudes come from populations with unequal means, with a 5% significance level. This proved to be the case for ASR and EMD (see below). The following cognitive models were implemented for comparison:

- a null model (M0) assuming no difference between responses to standard and deviant sounds;
- static (non-learning) models including binary change detection (CD), deviant detection (DD) and linear change detection (LD) [19];
- Bayesian learning models (BL) that depend on a forgetting parameter τ, meaning that the larger τ, the wider the memory [19]. We considered four different models of this type, each corresponding to a different value of τ: 2, 6, 10, and 100.

We used the VBA (Variational Bayes Analysis) toolbox [20] for our computational model analyses. To reduce the number of inversions, the EEG signal was down-sampled from 600 Hz to 100 Hz. The VBA toolbox allows to reject specific trials in the evolution model. Therefore, for the offline method, we consider the same rejected

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-12

trials as identified in [14], and for ASR and EMD we declare the calibration period (the first 30s) of each block as rejected trials (as explained before). To find out which models outperform the static null model (M0) at a given latency, we computed the relative Free-energy (as a proxy for relative model evidence) [19].

RESULTS

The left upper panel in Fig. 1a displays the grand average responses computed offline from the identified responsive channels, for the standard, the deviant and the MMN. The right upper panel in Fig. 1a shows the MMN values computed in the window of interest (160-190 ms) for all subjects and all artifact correction methods. Except for ICA, all methods revealed a significant auditory MMN.

Fig. 1b summarizes the methods ability to reveal MMN modulation by predictability. The left lower panel shows the modulation of the MMN obtained offline. The right lower panel displays the MMN amplitude for the predictable and unpredictable conditions, as obtained with each online artifact correction method. The two-tailed *t*-test ($p < 0.05$) showed that, like the offline analysis, ASR and EMD proved able to reveal the significant modulation of the MMN by predictability.

Finally, we further investigated whether online artifact correction with ASR or EMD would yield the same conclusion compared to previous offline analysis, regarding Bayesian model selection of perceptual learning processes as reflected by single trial modulations of auditory evoked responses [21]. Fig. 2 shows the relative Free-energy (model evidence) of every model with respect to the null model (M0). Note that models DD and BL10 prevail compared to other models, at the latency of the MMN (between 160 and 200ms).

DISCUSSION

We compared artifact correction methods with offline data processing and investigated their abilities to reveal a significant MMN, to uncover subtle MMN modulation by a contextual manipulation of the sound sequence (deviance predictability), and to identify the best perceptual model based on single-trial data analysis.

Artifact correction: We explored four methods for online artifact correction on real data, mimicking real-time data processing.

In the course of ASR implementation, we noticed that noisy channels identified during calibration must be excluded. ASR applies Principal Component Analysis (PCA) transformation to filter the signal, therefore the noise present in the calibration data may be propagated to the other channels. Besides, as noise levels (internal and external) change throughout the experiment, re-calibration of the method is advisable. In our study, we performed the re-calibration for each new block and new EEG file, that is, every 674 stimuli (approximately 10 minutes of data).

The FORCe method was designed with the proposal of being fully automated, so no calibration period or parameter setting is required. However, the function requires that at least one second of the signal must be filtered. At first, we used the same time window (-200ms to 500ms) as for the other methods, but results proved worse than when using a 1-second-long window. It is worth mentioning that the method preserved the MMN waveform, but could not reveal a statistically significant modulation by predictability.

To implement the EMD method, since the channel cluster of interest was already well established, we first computed the average time series over these channels and then applied the correction. We chose this approach to avoid processing the intrinsic mode functions (IMFs) for all 62 channels, which would require a prohibitive time for real-time implementation. Although this could decrease the signal quality, since averaging smoothes the signal, the results showed that it did not affect the MMN response. Hence the numbers of channels must be considered when choosing this method.

ICA showed the worst performance to reveal the MMN. Probably because of the simple and rigid manner we applied ICA. Indeed, we simply apply the initially estimated mixing matrix to all subsequent trials. One could imagine updating this matrix or identifying whether a given trial needs to be filtered or not, and which components should be removed. Furthermore, the size of the time window used to decompose the signal can be enhanced to better estimate the independent components.

All methods were applied to epoched data. This strategy is well suited for ERP analysis and alleviates the computational burden.

The artifact correction methods investigated in this study were able to reveal the MMN response. However, ICA excessively smoothed the signal waveform, altering the main component of the MNN peak between 150 and 200ms.

These findings look promising as they open the way to reliable online MMN analyses, that could have far-reaching applications with patients. The MMN has indeed long been investigated to assess impaired cognitive functions in various clinical populations [22].

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-12

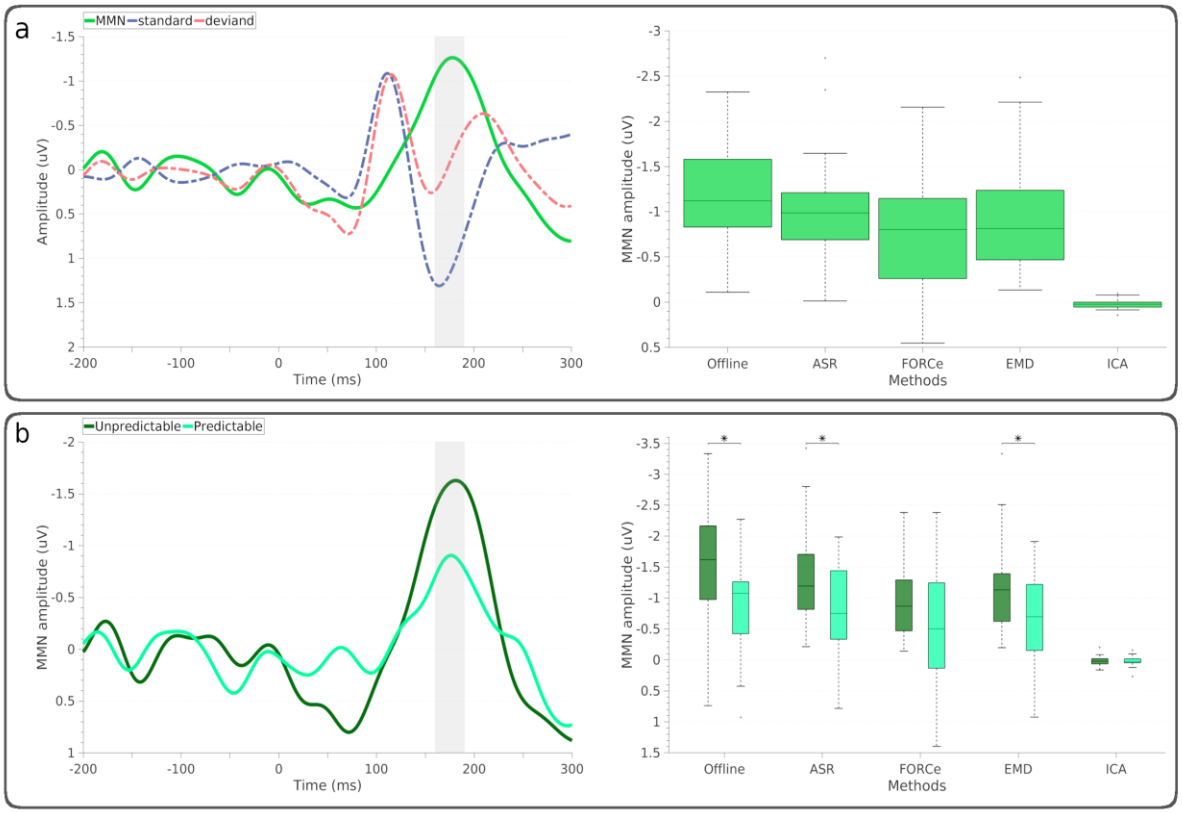


Figure 1 - Mismatch Negativity (MMN) (electrodes of interest: F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2). **(a)** Grand-average ERPs elicited by standard, deviant and their difference (MMN) obtained offline; boxplot of MMN amplitudes (averaged between 160 and 190 ms) obtained with the different online artifact corrections; **(b)** Effect of predictability on the MMN (offline method); MMN amplitudes for each condition (predictable and unpredictable) and each artifact correction approach; the asterisks indicate statistical significance of the modulation by predictability ($p < 0.05$, two-tailed *t*-test).

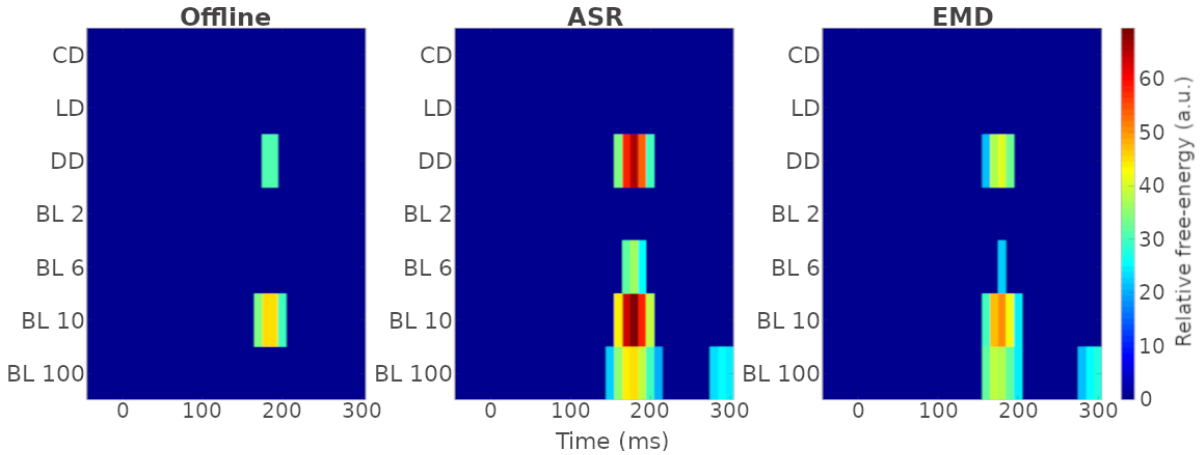


Figure 2 - Relative Free-energy maps. Relative Free-energy or (log) Bayes-factor values obtained offline (left panel), with ASR (middle panel) and EMD (right panel), respectively. In each panel, the relative Free-energy is given for each alternative model relative to the null model (y-axis) and for each peri-stimulus time sample in ms (x-axis). Values greater than 20 indicate significant evidence in favor of the alternative model.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-12

Predictability effect: After this first analysis, we investigated the impact of artifact correction on MMN modulation. As discussed in other studies [14, 23, 24], predictability of the occurrence of a deviant sound yields a decrease in prediction error, hence a decrease in the MMN amplitude. In other words, the greater the sound predictability, the smaller the MMN.

Results depicted in Fig. 1b shows that FORCe and ICA hardly revealed the MMN modulation. Although the MMN in the predictable condition was smaller when using the FORCe approach, it was not significantly different from the one in the unpredictable condition. Whereas for the ICA approach, the estimated amplitudes were the same in the two conditions.

Models: The last part of our study was implemented to analyze if any learning or non-learning models, based on sensor-level signals, could explain the trial-by-trial variations in the ERP signals. At the latency of the MMN (150 to 200ms), the offline method presented favorable results for models DD and BL10, but the DD model took longer to stand out and its window of significance proved shorter. The Free-energy map for ASR and EMD reveal similar latencies of significance. However, the ASR approach yielded slightly longer time windows and presented larger values compared to the other methods. This suggests that ASR was not only able to appropriately correct for artifacts but in doing so, it also avoided rejecting trials and led a more sensitive analysis, better suited to reveal the minutiae of auditory processing dynamics.

CONCLUSION

Our results suggest that ASR and EMD can perform accurate online artifact detection and correction, as they were able to reproduce the results obtained at the group level with classical offline analysis. Interestingly, in terms of model comparison, we also note that the mimicked online approaches based on ASR and EMD yielded more sensitive results than the reference offline approach. However, ASR is computationally more efficient than EMD. Overall, ASR offers great potential for real-time applications, in particular those that would exploit adaptive designs in order to optimize hypothesis testing or clinical diagnosis at the individual level [12].

REFERENCES

[1] L. F. Nicolas-Alonso and J. Gomez-Gil. Brain Computer Interfaces, a Review. *Sensors*, vol. 12, no. 2, pp. 1211–1279, Feb. 2012.

[2] 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, Apr. 2018.

[3] D. Tan and A. Nijholt. Brain-Computer Interfaces and Human-Computer Interaction. Brain-Computer Interfaces. Springer London, 2010, pp. 3–19.

[4] T. R. Mullen *et al.*. Real-time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG. IEEE Trans. Biomed. Eng., vol. 62, no. 11, pp. 2553–2567, Nov. 2015.

[5] I. Daly, R. Scherer, M. Billinger, and G. Müller-Putz. FORCe: Fully Online and Automated Artifact Removal for Brain-Computer Interfacing. IEEE Trans. Neural Syst. Rehabil. Eng., vol. 23, no. 5, pp. 725–736, Sep. 2015.

[6] R. Fontugne, P. Borgnat, and P. Flandrin. Online Empirical Mode Decomposition. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 4306–4310.

[7] A. O. Andrade, S. Nasuto, P. Kyberd, C. M. Sweeney-Reed, and F. R. Van Kanijn. EMG signal filtering based on Empirical Mode Decomposition. Biomed. Signal Process. Control, vol. 1, no. 1, pp. 44–55, Jan. 2006.

[8] Huang Norden E. *et al.*. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci., vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[9] R. N. Vigário. Extraction of ocular artefacts from EEG using independent component analysis. Electroencephalogr. Clin. Neurophysiol., vol. 103, no. 3, pp. 395–404, Sep. 1997.

[10] A. K. Barros, A. Mansour, and N. Ohnishi. Removing artifacts from electrocardiographic signals using independent components analysis. Neurocomputing, vol. 22, no. 1, pp. 173–186, Nov. 1998.

[11] M. Milanesi, N. Martini, N. Vanello, V. Positano, M. F. Santarelli, and L. Landini. Independent component analysis applied to the removal of motion artifacts from electrocardiographic signals. Med. Biol. Eng. Comput., vol. 46, no. 3, pp. 251–261, Mar. 2008.

[12] G. Sanchez, F. Lecaignard, A. Otman, E. Maby, and J. Mattout. Active SAmpling Protocol (ASAP) to Optimize Individual Neurocognitive Hypothesis Testing: A BCI-Inspired Dynamic Experimental Design. Front. Hum. Neurosci., vol. 10, 2016.

[13] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clin. Neurophysiol., vol. 118, no. 12, pp. 2544–2590, Dec. 2007.

[14] F. Lecaignard, O. Bertrand, G. Gimenez, J. Mattout, and A. Caclin. Implicit learning of predictable sound sequences modulates human brain responses at different levels of the auditory

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-12

hierarchy. Front. Hum. Neurosci., vol. 9, 2015.

[15] G. G. Parras, J. Nieto-Diego, G. V. Carbajal, C. Valdés-Baizabal, C. Escera, and M. S. Malmierca. Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. Nat. Commun., vol. 8, no. 1, p. 2148, Dec. 2017.

[16] V. B. Perez *et al.*. Mismatch Negativity is a Sensitive and Predictive Biomarker of Perceptual Learning During Auditory Cognitive Training in Schizophrenia. Neuropsychopharmacology, vol. 42, no. 11, pp. 2206–2213, Oct. 2017.

[17] 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.

[18] C. Chang, S. Hsu, L. Pion-Tonachini, and T. Jung. Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 1242–1245.

[19] D. Ostwald, B. Spitzer, M. Guggenmos, T. T. Schmidt, S. J. Kiebel, and F. Blankenburg. Evidence for neural encoding of Bayesian surprise in human somatosensation. NeuroImage, vol. 62, no. 1, pp. 177–188, Aug. 2012.

[20] J. Daunizeau, V. Adam, and L. Rigoux. VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. PLOS Comput. Biol., vol. 10, no. 1, p. e1003441, Jan. 2014.

[21] F. Lecaignard, O. Bertrand, A. Caclin, and J. Mattout. Evidence for implicit and adaptive deployment of precision weighting during passive listening: a simultaneous EEG/MEG study. Neuroscience, preprint, Dec. 2018.

[22] R. Näätänen, E. S. Sussman, D. Salisbury, and V. L. Shafer. Mismatch Negativity (MMN) as an Index of Cognitive Dysfunction. Brain Topogr., vol. 27, no. 4, pp. 451–466, Jul. 2014.

[23] C. Wacongne, J.-P. Changeux, and S. Dehaene. A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. J. Neurosci., vol. 32, no. 11, pp. 3665–3678, Mar. 2012.

[24] I. Winkler and I. Czigler. Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. Int. J. Psychophysiol., vol. 83, no. 2, pp. 132–143, Feb. 2012.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-13

EXPECTATION MISMATCH DURING A TRACKING TASK: AN EEG ANALYSIS

M. Bevilacqua [1], C. Lopes-Dias[2], A.I. Sburlea[2], G. Müller-Putz[2]

[1]Defitech Chair in Brain-Machine Interface (CNBI), Center for Neuroprosthetics, School of Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[2]Institute of Neural Engineering (INE), Graz University of Technology, Graz, Austria

E-mail: gernot.mueller@tugraz.at

ABSTRACT: Understanding the dynamics of brain areas' activation during BCI tasks is essential to improve BCIs performance and features selection when training a classifier. The role of the cerebellum in the sudden adaptation of the motor plan in response to unexpected perturbations in tasks with continuous motor control and continuous visual feedback has been well reported in the literature. We recorded EEG data of five subjects to study the cerebellar activation in tasks requiring a sudden motor plan update. Using sLORETA inverse source localization method, we observed activation in the cerebellum at around 200 and 400 ms after the deviation onset. Furthermore, EMG recording and analysis at the level of the neck could disprove that the activity observed was due to movement artifacts.

INTRODUCTION

Being able to precisely localize the components of functional networks involved in a given mental task and knowing the real-time dynamics of brain areas activation can provide a better understanding of the neural processes underlying brain functioning. The information about the temporal interaction between components of different neural networks and their specific role in the processing of mental tasks is essential to obtain a complete comprehension of the brain activity and to be able to decode it precisely. This information could for example be exploited by BCI applications in order to improve the data acquisition, the processing and, above all, the feature extraction process used to train a classifier, increasing the performance of the controlled device [1–3]. EEG provides a temporal resolution in the order of the milliseconds [4]. EEG source localization technique thus represents a potential tool to study almost in real-time the dynamics of neural networks involved in mental tasks. EEG source localization is used mostly to analyze cortical sources formed by groups of pyramidal neuron, since they are thought to be the main contributors to the generation of the electrical potential acquired. However, studies show that it can also be a valid method to investigate sources in the deeper cerebral structures, such as the cerebellum [5–7]. Different studies highlighted the involvement of the cerebellum in continuous performance monitoring [8], in supervised learning during motor tasks [9] and in the adaptation of motor plan in response to visuomotor perturbations [10]. Tseng et al. [10], in a study on goal-directed arm movements, concluded that the adaptation to visuomotor perturbations depends on the cerebellum and that it is driven by the mismatch between predicted and actual sensory outcome of motor commands. EEG evidences of cerebellar role in performance monitoring have been reported by Peterburs et al. [8] and by Wolpert et al. [11], who also hypothesized that in the context of movement coordination, the cerebellum applies an internal forward-model to predict the sensory consequences of actions. Furthermore, the study on the computation method of the cerebellum performed by Doya [9] evidenced that the cerebellum is the specialized organism for supervised learning based on continuous motor tasks with continuous feedback. The results of the reported literature elicited interesting speculations about the cognitive involvement of the cerebellum in tasks with continuous motor control and continuous feedback requiring the adaptation of the motor plan or the need for more control on the motor task. The possible activation of the cerebellum during this type of tasks could indeed be due to an attempt performed by the subject to suddenly adapt the motor plan to an sudden perturbation of the expected task. At the moment of the perturbation onset, the defined target-directed motor plan should be quickly interrupted and updated in order to comply with the new goal; in this framework the cerebellum could be involved in the attempt of adapting the motor plan to the mismatch acknowledged through visual feedback. The hypothesis of the active role of the cerebellum in this framework, if confirmed, could represent a novel finding about the involvement of the cerebellum in expectation mismatch tasks. The aim of this study is thus to use EEG source localization to provide insights about the cerebellar involvement in the processing of continuous motor controlled tasks in which a mismatch between visual feedback and expectation is generated, and a sudden change in the motor plan is required. In addition, this experiment aims to study the involvement of muscle artifacts in the cerebellum activation.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-13

MATERIALS AND METHODS

Hardware and data acquisition: EEG data were recorded at a sampling rate of 1000 Hz using BrainAmp amplifiers and an ActiCap system (Brain Products, Munich, Germany) with 59 active electrodes (FP1, FP2, AF3, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8,PO9, O1, Oz, O2, PO10), reference on the right mastoid, the ground electrode in AFz, 3 EOG electrodes (above the nasion and below the outer canthi of the eyes) and 2 EMG electrodes (on the skin, on the left (EMG1) and on the right (EMG2) of the spinal cord; right above the trapezius). EMG recording have been included to the experimental paradigm in order to assess the amplitude of muscular artifacts generated by the movement of the arm and the shoulder during the trials at the level of the neck. In addition to EEG, EOG and EMG channels, one supplementary channel reporting the coordinates of a handle of the joystick used in the experiment was acquired.

Participants and experimental environment: Five volunteers (with ages between 22 and 26 years, 3 male, all right-handed and already experienced in BCI and EEG recording) participated in the experiment, which took place in a darkened shielded room. Participants sat on a comfortable armchair, 1.5 meter away in front of a computer screen displaying the protocol. The right armrest of the chair was replaced by a table with the joystick. EEG data of the 5 subjects were acquired while controlling a cursor on the computer screen with the joystick. The subjects were always able to move the joystick in every direction without dragging the elbow on the table or raising the shoulder.

Experiment overview: The experiment consisted of 10 blocks of 32 trials each. Between each trial there was a 2.5 s break. Between each block the subject could rest as long as wanted, since the beginning of the following block was triggered by pressing a joystick button. 25% of the trials of every block were *deviation* trials, while the resting 75% *no deviation* trials. For all the subjects the exact 3D position of the electrodes on the head was recorded before the beginning of the protocol, using CMS 20 EP system (Zebris Medical GmbH, Isny, Germany). Before the beginning of the protocol, 2 minutes of EEG recording of the subject in a complete resting condition was acquired for every subject.

Trial and task description: As depicted in Figure 1 (top), at the beginning of a trial, 2 equally spaced blue squares were displayed on the upper part of the screen, at the same distance from its center. In addition, a black fixation cross was displayed. On the lower part of the screen there were two circles, one white and one red, vertically aligned. The white circle was automatically moved by the protocol, while the red circle represented the cursor controlled by the subject. The squares, the cross and the two circles were displayed on a grey area, inside which the two circles could move. In each trial, one of the two squares was randomly selected by the protocol as the target, without informing the subject about the decision. The two squares had the same probability to be selected as targets. The subject controlled the cursor through the joystick, and the displacement of the joystick handle was directly proportional to the direction of the movement of the cursor. The task consisted following the white circle with the cursor controlled by the joystick. The white circle was directed towards the target. When the white circle started to move, the subject obtained the control of the cursor and the trial begun. A trial ended when the cursor reached the target defined by the paradigm, when it hit the boundaries of the grey region, or when the time limit of 15 seconds per trial was reached. The subjects were instructed to keep their gaze fixed at the fixation cross and to minimize eye blinking and eye movements.

No deviation trials: In these trials the white circle followed a default trajectory, reaching the center of the target and stopping its movement there (see Figure 1, bottom left). In the presented study, *no deviation* trials have been considered valid only if the participant successfully reached the target. *No deviation* trials that ended because of the cursor hit the boundary of the grey region or the time deadline was reached were not considered in the analysis.

Deviation trials: In these trials, the target changed in the middle of the trial. Therefore, the trajectory of the white circle previously described was subjected to a modification. At the deviation onset, the white circle deviated towards the new target. The new trajectory described an arc of circumference connecting the white circle, at the deviation onset, and the new target. The deviation onset occurred when the white circle was located within the green semicircles depicted in Fig. 1 (top), which were invisible to participants. In *deviation* trials the white circle initially aimed one target, and then deviated to finally reach the other target (see Figure 1, bottom right). Also for *deviation* trials, were considered in the study only the trials that ended with the cursor reaching the final target defined by the white circle.

EEG preprocessing: EEG data was bandpass filtered between 1 and 10 Hz using an IIR Butterworth filter of fourth-order with zero phase. Data was resampled to 250 Hz. EEG, and protocol events were imported on Brainstorm [12] (http://neuroimage.usc.edu/brainstorm). Since individual MRI scans were not available, to every subject was attributed the ICBM152 default anatomy [13]. It has been necessary to define a “virtual onset” for *no deviation* trials ("no deviation onset"), computed as the average of the difference between time of the deviation onset and the start of the *deviation* trial and then added to the starting time of every *no deviation* trial. EEG and EOG data was epoched considering the interval [-700, 1200] ms from deviation and no deviation onset events. Two different types of epochs were thus extracted: *deviation* and *no deviation*. Epochs containing eye blink events were dis-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-13

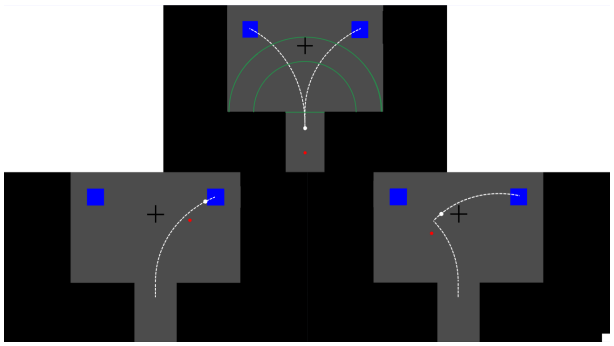


Figure 1: **Top**: starting condition of every trial. Blue squares: fixed targets, white circle: moving target, red circle: controlled cursor, black cross: fixation cross, white lines: default possible trajectories of the moving target (in *no deviation* trials), grey region: area in which the cursor is possible to move, green semicircular lines: area in which a deviation in the trajectory of moving target is possible to occur, green rectilinear line: threshold between upper and lower grey areas. Green and white lines were not visible for the subject during the protocol. **Bottom left**: Moment before a successful end of a *no deviation* trial. **Bottom right**: Moment right after the deviation onset in a *deviation* trial.

carded. On average, the 1% of the trials (*deviation* and *no deviation*) was discarded from every subject. Visual examination of epochs for bad channels was computed, and no channels were removed from any subject.

Joystick coordinates analysis: Together with EEG epochs, also joystick coordinates epochs were extracted. For Subject 1, the acquisition of the joystick coordinates was not available. First, individual averages and then grand averages over 4 subjects of X and Y joystick coordinates have been computed for *deviation* and *no deviation* condition. **Movement Reaction Onset (MRO)**, the time in which the subject reacted to the deviation with a movement of the joystick, has been computed for every subject as the moment in which the derivative of the X coordinate in *deviation* trials changed sign. Then, the average MRO has been computed by averaging the MRO of the four subjects.

EMG analysis: EMG channels were highpass filtered at 1 Hz using an IIR Butterworth filter of fourth-order with zero phase. Also notch filter at 50 Hz was applied to eliminate power line noise. All the frequency information related to muscular artifacts had been kept. EMG channels were resampled to 250 Hz. The moving Root Mean Square (RMS) envelope of EMG channels was then extracted using a time window of 500 ms. Artifact cleaning had been performed using the Signal Space Projection (SSP) method provided by Brainstorm [14]. EMG epochs for *deviation* and *no deviation* conditions were extracted in the same way as EEG and joystick coordinates epochs were.

Scalp level analysis: For every subject, *deviation* and *no deviation* trials were averaged. In order to avoid biases in the SNR (Signal to Noise Ratio) in favor of the condition presenting the highest number of trials (*no deviation*), the same number of epochs were averaged for every condition. The used trials in the *no deviation* condition were chosen from a random uniform distribution from the total number of *no deviation* trials. On average, for every subject, 75 epochs were averaged for both conditions.

Source level analysis: EEG source localization brain imaging was performed with Brainstorm [12]. Since the focus of the source analysis performed was specifically to investigate the source activity at the level of the cerebellum, a *mixed head model* was used to describe the propagation of electrical fields from the cortical surface to the scalp. Cerebral cortex was modeled using a distributed source model, while the cerebellum using an unconstrained distributed model. The ill-posed inverse problem was modeled by means of a distributed source model. Source maps were computed using wMNE (weighted Minimum Norm Estimates) [15] regularization, and normalized according to sLORETA method [16]. The amplitudes of averages for every condition in each subject were normalized by the Global Field Power (GFP) of the baseline period [-600, -100] ms before the event (*deviation/no deviation*) onset. In this way, we normalized the power of the signal across subjects, eliminating the influence of possible inherent higher GFP that could bias the magnitude of source activation. Source analysis on mixed head model was computed on the subject averages of both conditions. Then, grand averages over the 5 subjects were computed at source level for the two conditions. It is important to highlight that the GFP normalization was computed for every subject only on the condition averages, and not on the noise covariance matrix used by sLORETA inverse method. This did not influenced the final source analysis results, but simply scaled in a consistent way the magnitude of the current density of every active source.

RESULTS

EMG analysis: In Fig. 2, the grand averages of the potentials recorded by the two EMG electrodes placed on the neck, for *Deviation* and *No Deviation* conditions are shown. There is no increase in the EMG activity after the cursor deviation onset (black line), neither after the MRO (red dotted line). The only consistent difference is between EMG1 and EMG2, probably because the EMG2 electrode was placed on the right side of neck and the joystick was always moved with the right hand.

Scalp level analysis: Fig. 3 presents the grand averages of the two conditions (*Deviation*/*No Deviation*) at the channels FCz, Cz and Pz. Seven relevant time points have been chosen as a reference in relation with the *Deviation* grand average at FCz: *t=0 ms* (cursor deviation onset), *t=56 ms* (before first small positive deflection), *t=100 ms* (first small positive deflection), *t=168 ms* (negative peak), *t=272 ms* (average MRO), *t=372 ms* (positive peak) and *t=524 ms* (negative deflection after positive peak). The *Deviation* potential shows a negative peak around 168 ms, before the average MRO (red dot-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-13

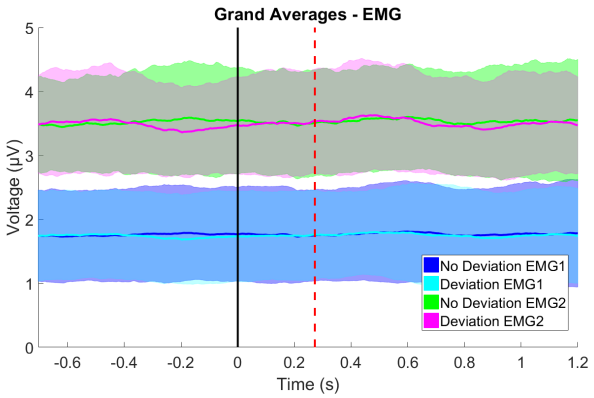


Figure 2: Grand Averages for the 2 conditions (*Deviation*/*No Deviation*) of the 2 EMG channels. The vertical black line represents the time onset of the white circle deviation. The dotted red line represents the average MRO (t=272 ms). The shaded areas indicate the 95% confidence interval for the averaged signals.

ted line), and a positive peak around 372 ms, after the average MRO. *No Deviation* grand average is mostly flat.

Source level analysis: sLORETA source localization inverse method has been computed on the grand averages of the *Deviation* and *No Deviation* trials (Fig. 4). The analysis has been performed on the 7 relevant time points described before. The analysis showed cerebellar activation before (lower, t=168 ms) and after (higher, t=372 ms) the average MRO (t=272 ms). The analysis showed also activation mainly in the pre-SMA (pre-Supplementary Motor Area) during the transition between the two peaks.

DISCUSSION

With this experiment we investigated whether the involvement of cerebellum is present in cognitive processing of expectation mismatch during continuous motor control. The results regarding EMG activation (Fig. 2) did not show noticeable changes after the movement reaction onset of the deviation trials. Therefore, this could disprove the hypothesis that the activation found at the cerebellar level was due to muscular artifacts when handling the joystick. At the scalp level, the shape of the *Deviation* condition grand average resembles the shapes of the most classic error-related potentials (ErrPs) reported in the literature [17–19], with a broad frontocentral negativity at around 200 ms (ERN component) and a broad frontocentral positivity at around 400 ms (Pe component). This can confirm the hypothesis that the potential generated by the expectation mismatch caused by a sudden change in the target-oriented motor plan presents some common features with the ErrPs. The negative peak happened before the reaction movement onset, while the positive one after. Since the Pe component is usually associated with conscious awareness of the expectation mismatch [20, 21], this observation could show that subjects reacted to the deviation before they were consciously aware of it. As expected, grand average of *No deviation* trials appeared almost flat during the entire time interval analyzed. Scalp potential distribution reflected the results provided by literature regarding error-related potential scalp analysis [18, 19]. Source analysis performed on mixed head model with sLORETA inverse method (Fig. 4) confirmed the hypothesis of cerebellum involvement in the processing of expectation mismatches. The cerebellum presented activity in correspondence with the negative and the positive peaks in the deviation condition, while in the interval between the two peaks the pre-SMA and ACC (Anterior Cingulate Cortex) areas appeared more active. This could provide valid insights in relation to the cognitive role of the cerebellum in reinforcement learning, motor performance monitoring and error in continuous motor task. The fact that the cerebellum showed remarkable activation even in correspondence of the negative peak, before the MRO, could confirm the hypothesis that the activity observed was not due to movement artifacts.

CONCLUSION

This study confirmed the possibility of the involvement of cerebellum processing in expectation mismatch tasks with continuous motor control and continuous visual feedback. Furthermore, the hypothesis of the activation being caused by movement artifacts could have been disproved. An open question concerns the activation of the cerebellum in a scenario in which the user is controlling a cursor only through brain signals and without using any muscle activation. It would be interesting to assess if in such a situation the cerebellum would also be involved. Further connectivity studies could be performed on the topic in order to assess the physiological reason underlying the described source activity.

ACKNOWLEDGEMENTS

This work was supported by Horizon 2020 ERC Consolidator Grant 681231 'Feel Your Reach'.

REFERENCES

[1] Goel MK, Chavarriaga R, Millán JdelR. Inverse solutions for brain-computer interfaces: Effects of regularisation on localisation and classification. In: Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on. IEEE. 2017, 258–263.
[2] Peralta MRGde, Andino SG, Perez L, Ferrez PW, Millán JdelR. Non-invasive estimation of local field potentials for neuroprosthesis control. Cognitive Processing. 2005;6(1):59–64.
[3] Noirhomme Q, Kitney RI, Macq B. Single-trial EEG source reconstruction for brain–computer interface. IEEE Transactions on Biomedical Engineering. 2008;55(5):1592–1601.
[4] Baillet S, Mosher JC, Leahy RM. Electromagnetic brain mapping. IEEE Signal processing magazine. 2001;18(6):14–30.
[5] Cebolla AM, Petieau M, Dan B, Balazs L, McIntyre

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-13

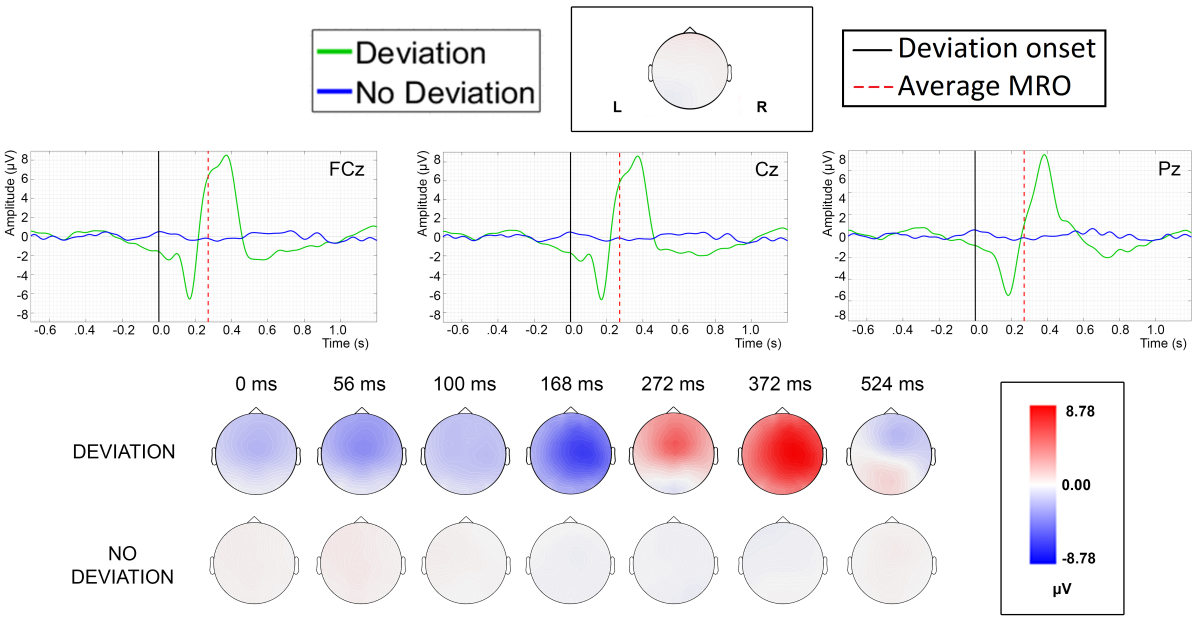


Figure 3: On **top**, the grand averages for the *Deviation* and *No Deviation* conditions at channels FCz, Cz and Pz. The dotted red line is at t=272 ms, and it represents the average MRO. The black continuous line represents the deviation onset. **Below**, the time evolution of scalp distribution of the grand averages for the 2 conditions. The time points t=0 ms, t=56 ms, t=100 ms, t=168 ms, t=272 ms, t=372 ms and t=524 ms are displayed.

J, Cheron G. Cerebellar contribution to visuo-attentional alpha rhythm: insights from weightlessness. Scientific reports. 2016;6:37824.
[6] Attal Y et al. Modeling and detecting deep brain activity with MEG EEG. In: Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE. IEEE. 2007, 4937–4940.
[7] Seeber M et al. Subcortical electrophysiological activity is detectable with high-density EEG source imaging. Nature Communications. 2019;10:753.
[8] Peterburs Jutta et al. A cerebellar role in performance monitoring–Evidence from EEG and voxel-based morphometry in patients with cerebellar degenerative disease. Neuropsychologia. 2015;68:139–147.
[9] Doya K. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural networks. 1999;12(7-8):961–974.
[10] Tseng Y, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ. Sensory prediction errors drive cerebellum-dependent adaptation of reaching. Journal of neurophysiology. 2007;98(1):54–62.
[11] Wolpert DM, Miall RC, Kawato M. Internal models in the cerebellum. Trends in cognitive sciences. 1998;2(9):338–347.
[12] Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. Brainstorm: a userfriendly application for MEG/EEG analysis. Computational intelligence and neuroscience. 2011;2011:8.
[13] Talairach - MRC CBU Imaging Wiki. [Online]. Available: http://imaging.mrccbu.cam.ac.uk/imaging/MniTalairach. [Accessed: 12-Apr-2018].
[14] Uusitalo MA, Ilmoniemi RJ. Signal-space projection method for separating MEG or EEG into components. Medical and Biological Engineering and Computing. 1997;35(2):135–140.
[15] Hämäläinen MS, Ilmoniemi RJ. Interpreting magnetic fields of the brain: minimum norm estimates. Medical biological engineering computing. 1994;32(1):35–42.
[16] Pascual-Marqui RD. Review of methods for solving the EEG inverse problem. International journal of bioelectromagnetism. 1999;1(1):75–86.
[17] 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.
[18] Ferrez PW, Millán JdelR. Error-related EEG potentials generated during simulated brain-computer interaction. IEEE transactions on biomedical engineering. 2008;55(3):923–929.
[19] 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.
[20] Overbeek TJM, Nieuwenhuis S, Ridderinkhof KR. Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. Journal of Psychophysiology. 2005;19(4):319–329.
[21] Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GPH, 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.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-13

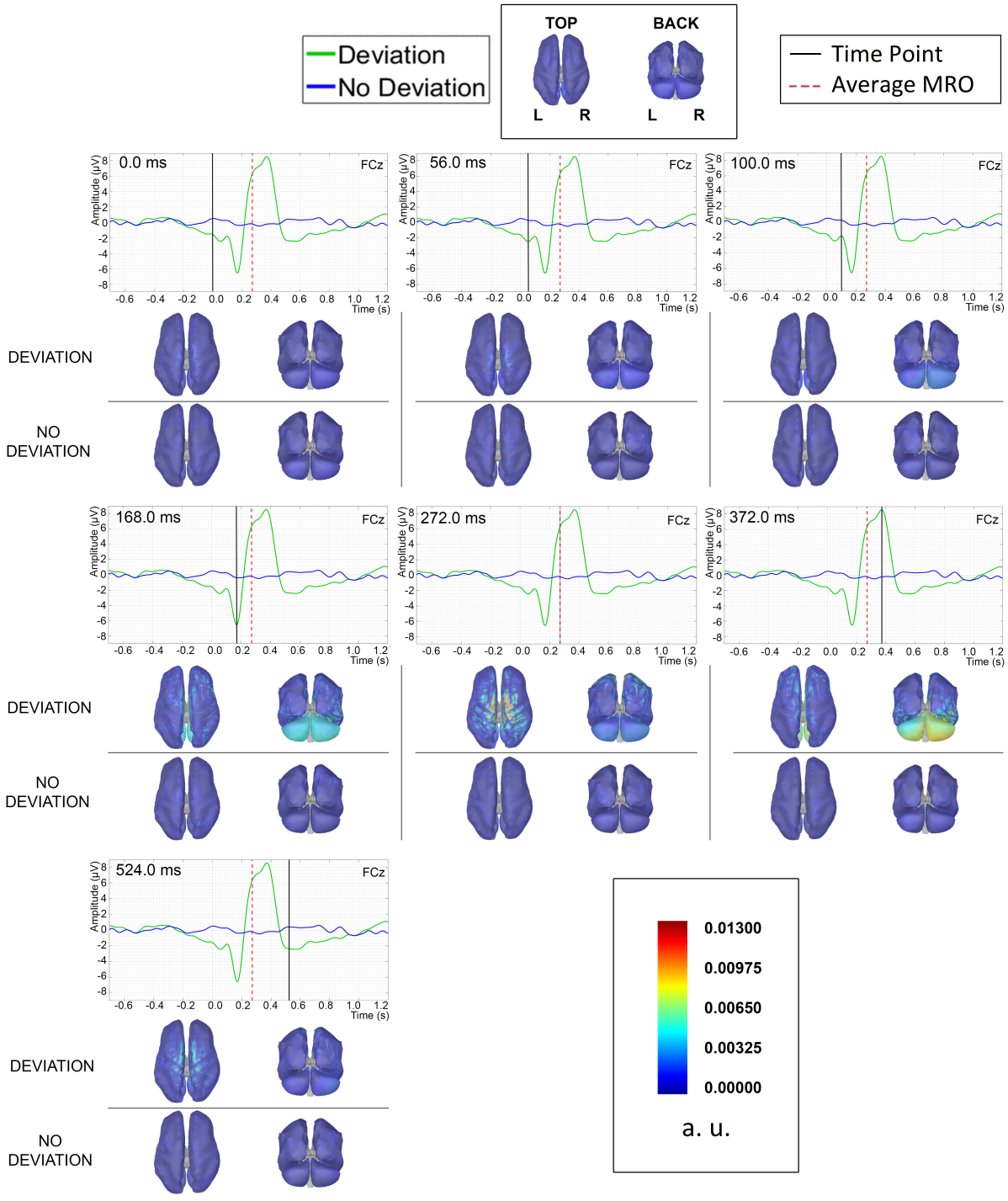


Figure 4: sLORETA source localization computed on the mixed brain surface distributed source model, averaged over the 5 subjects. This mixed brain model includes cerebral cortex and cerebellum surface. Dipoles orientation has been considered constrained to the surface of the cerebral cortex and unconstrained to the cerebellum surface. Seven relevant time points have been taken into account (t=0 ms, t=56 ms, t=100 ms, t=168 ms, t=272 ms, t=372 ms, t=524 ms). For every time point analyzed, the grand averages at channel FCz as been reported as a reference for the cortex views below. The black line indicates the analysed time-points. The dotted red line (t=272 ms) represents the average MRO. Two cerebral and cerebellar cortex surface views for every condition and time point are shown, in this order: top, back. The source activation level on the surfaces is referred to the same scale for all the conditions. The current density of source activation is shown without a unity of measure because of the subject-wise normalization performed before computing the source activity. Nevertheless, this value is proportional to the effective current density.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-14

CASE STUDY: TRADITIONAL MOTOR CORTEX CONTROL FEATURES IN ALS AND BRAINSTEM STROKE

Z.V. Freudenburg[1], M.J. Vansteensel[1], B.H. Van der Vijgh[1], M.P. Branco[1], S. Leinders[1], E.G.M. Pels[1], M.A. Van Den Boom[1], E.J. Aarnoutse[1], N.F. Ramsey[1]

[1] UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, The Netherlands
E-mail: Z.V.Freudenburg@umcutrecht.nl

ABSTRACT: The goal of the Utrecht NeuroProsthesis (UNP) study conducted at the UMC Utrecht is to evaluate the usability of a BCI system for communication in people suffering from Locked-In Syndrome (LIS): people who cannot speak due to a loss of almost all motor control. In the past we demonstrated that a combination of High Frequency Band (HFB) and Low Frequency Band Sensorimotor cortex (SMC) features can be used to provide an individual with LIS due to ALS with a stable and robust communication channel. The recent inclusion of a second LIS participant in the UNP study, a woman who suffered a brainstem stroke, enables us to compare SMC LFB and HFB features in people with LIS caused by two fundamentally different etiologies. We show that while the HFB is a stable feature in both cases, the functional stability of the LFB feature is much less in the brain stem stroke subject.

INTRODUCTION

It is the goal of the BCI research group at the UMC Utrecht to develop and validate devices that will provide people suffering from Locked-In Syndrome (LIS) with a reliable means of communication that works 24/7. People with LIS are quadriplegic, anarthric and are generally only able to move their eyes, but they are conscious of thought [1]. Having a reliable channel for communication has the potential to improve the quality of life and the integration into society of these individuals [2,3].

In a recent report, we demonstrated that a combination of High Frequency Band (HFB) (31-100Hz) and Low Frequency Band (6-30Hz) signal changes, as measured from the motor cortex using subdural electrocorticography (ECoG), can be used to create a robust and stable BCI communication channel in an individual with LIS due to late-stage ALS [4]. The amplitude changes produced in these frequency bands by the participant attempting to move her hand were in close agreement with those traditionally associated with sensorimotor cortex (SMC) function and the generation of movement [5,6]. However, the LIS condition may be caused by multiple clinical conditions, including ALS and brain stem stroke. Whereas multiple studies using EEG [7-8], ECoG [4, 9-10], and intracortical electrodes [11-12] have now proven the feasibility of SMC BCI control in individuals with severe paralysis, little is known about the relation between etiology and SMC neuroelectrical signal characteristics and BCI control ability. The inclusion and implantation of a second participant in the UNP study of an individual who suffers from LIS due a to a brainstem stroke provides a unique opportunity to compare the LFB and HFB features in individuals with different causes of LIS [2,13]. The current study analyses cortical electrical potentials collected regularly over an extended (> 60 week) period from ECoG electrodes implanted over the SMC while participants performed a standardized attempted movement task. Both the functional stability of the LFB and HFB features and the spectral characteristics of the LFB feature are compared and put into the context of the UNP users' specific clinical conditions. We believe that this work can inform the development of future SMC-BCI systems that target people with LIS.

MATERALS AND METHODS

UNP Participants: The medical research ethics committee of the UMC Utrecht approved the UNP study, which was carried out in accordance with the Declaration of Helsinki (2013). Participants gave informed consent through a dedicated procedure (described in detail in [4]: supplementary materials). All work presented here is part of this continuing study. The first participant (UNP1), data of whom were reported in [4], is a woman who was diagnosed with ALS in 2008. She was 58 years old at the time of informed consent in 2015. She was put on invasive ventilation in 2010 and was living with minimal motor control for around 5 years at the time of implantation of the UNP in October of 2015. She had a score of 2/48 on the Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS) [14]. For communication she used, and continues to use, an eye tracker to type and eye blinks, and (lately) small movements of the mouth corner, to answer closed questions. In addition, she now uses the UNP system for communication on a routine basis.

The second UNP participant (UNP4) is a woman who suffered a brain stem stroke in 2004 and who was 39 years old at the time of informed consent in August of 2017. Her motor capabilities are limited to neck

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-14

movements and facial expressions and she had been living with LIS for 13 years at the time of inclusion in the UNP study. She uses a head switch to control scanning software for typing, and horizontal and vertical eye and head movements for answering closed questions. She has a score of 17/48 on the ALSFRS.

UNP ECoG signal: The primary cognitive strategy used to control the UNP device for both subjects is attempted movements of the contralateral (right) hand. Corresponding ECoG signal changes are recorded using subdural electrodes implanted over the 'hand knob' of the sensorimotor cortex (see Figure 1 of [4] for a depiction of the electrode locations in UNP1).
The implant target location was determined prior to surgery using fMRI scans. Subdural electrode strips (Resume II®, Medtronic, 4 electrodes each, 4mm diameter, 1cm distance, off label use) were implanted through burr holes (1cm diameter), over the target areas. An amplifier/transmitter device (Activa® PC+S, Medtronic, off label use) was placed subcutaneously under the clavicle. In addition to the online control mode of this device, in which analog filtered spectral amplitude signal is relayed to a receiving tablet at 5Hz (see [4] Methods for details), the device offers transmission of non-filtered 'raw' time domain signals at 200Hz. This work uses off-line analysis of the time domain signal recorded during repeated attempted hand movements to create time-locked responses in the frequency domain.

Attempted hand movement screening task: In order to track the functional response of the signal features subjects periodically performed a screening task that involved making repetitive attempted hand movements, or relax, for alternating periods of 15s each. In total 50 runs consisting of either 10 (17 runs) or 4 (33 runs) alternated rest and attempted movement trials were performed by UNP1. UNP4 completed 46 runs of the task, each with 6 trials of repeated attempted hand movements and relaxation.

Spectral analysis: The amplitude for each frequency bin from 6 to 100Hz (in steps of 1Hz) was computed offline for every time sample of each time domain data file using the real component of the convolution with a complex gabor wavelet (span 4 cycles at fwhm) [15]. The LFB and HFB responses over time were then computed as the sum of the log of the amplitudes for the frequency ranges 6-30Hz and 31-100Hz, respectively. These ranges were chosen based on the match to LFB and HFB ranges reported in literature [5,6].

Evaluation of- functional response features: Prior analysis of data acquired from UNP1 during the screening task [16] indicated that attempted hand moments produce 3 distinct functional response periods in the signal that are well matched to those reported in ECoG literature [5,6,17]. The 3 periods are: 1) the active period of the task during which the participant attempts to move the hand, 2) a 3 second rebound period directly following the cue to stop making attempted movements. and 3) the rest period, directly following the rebound period, during with participants relax and wait for next cue. Based on this knowledge, the following 3 expected functional responses were quantified: 1) The increase in mean HFB amplitude for active trials vs. rest periods, 2) the decrease in mean LFB amplitude during active trials vs. rest periods, and 3) the increase in LFB mean amplitude during the rebound periods following active trials, vs. rest periods. In the current study, these three functional responses were quantified for each task run using the coefficient of determination (R-squared) statistic over the mean amplitudes of each trail. A significance cut-off of $p < 0.05$ was used to determine significance.
The percentage of task runs with significant responses was also computed and used to compare the two UNP subjects. However, the fact that only two subjects are reported prevents statistical testing of the differences between to subjects.

Oscillatory signal component analysis: To gain a better understanding of the underlying spectral changes leading to the LFB and HFB functional responses both the spectral content of the active trials and rest periods of the screening task and that of separate baseline data sets, during which the UNP subjects simply relaxed and not concentrated on a task, was analyzed.
First, the spectral amplitude over time (see section *Spectral analysis*) was computed for 52 and 32 baseline data sets of 2-5 minutes each- for UNP1 and UNP4 respectively. Then the mean amplitude profile (from 4 to 40Hz) was computed along with the profile of the standard deviation. Here we used a slightly broader LFB range in order to evaluate if the trends in LFB spectral response may have been affected by neighboring frequencies beyond the traditional LFB boundaries.
Next irregular-resampling auto-spectral analysis (IRASA) [18] was used to separate the oscillatory spectral peaks from the scale-free (or fractal) component of the spectra during rest and active periods of the attempted movement task. This procedure allows for a direct comparison of the LFB oscillatory profiles that corrects for differences in spectra profiles due to differences in electrode impedances over runs or signal amplifiers between subjects. Again, a range of 4-40Hz was used. A 3s moving window with step size of 1s was used to divide the data into many time samples of rest and active periods. Windows that crossed the boundary between active and rest periods or that included data from rebound periods were excluded.

RESULTS

As reported earlier [4] the functional stability of the HFB and LFB responses of UNP1 was remarkably high. In fact, 100% of the runs produced HFB responses that were significant at the 0.05 p-value level and 98% of the runs had a significant LFB response. In addition to these, the LFB rebound response (green points) also presents a stable signal feature over the same period (with 90% of the runs having a p-value < 0.05).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-14

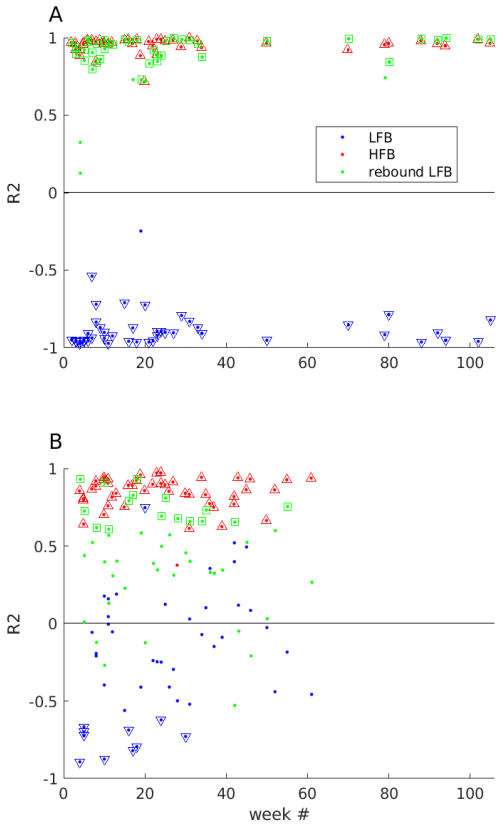


Figure 1: Functional consistency of the HFB and LFB responses of UNP1 and UNP4 over 106 and 61 weeks after implantation. A) The R-squared values of the HFB response (red dots), LFB response (blue dots) and LFB rebound response (green dots) over all task runs for UNP1. Runs with significant responses (p-values <0.05) are indicated with the red upward pointing triangle, blue downward pointing triangle, and green squares for the three response respectively. B) The R-squared values of the HFB response, LFB response, and LFB rebound response over all task runs for UNP4.

In contrast to UNP1, and despite the fact that UNP4's HFB functional response is only slightly less consistent (97.8% of runs showing a significant HFB response) than that of UNP1, the LFB responses generated by UNP4 during the screening task were considerably less functionally stable (21.7% of the runs have a significant LFB response; see Figure 1B). In addition, also the LFB rebound functional feature is much less consistent over runs (34.8% runs show a significant LFB rebound). Importantly, there was no clear trend in improvement in these features over time (Figure 1B).
The lack of significant LFB responses in UNP4 could be due to one or both of two phenomena: 1) a weak or changing oscillatory component during rest leads to low average LFB amplitudes during the rest condition, or 2) poor functional regulation of the LFB oscillatory component which then leads to mean amplitudes of the LFB remaining relatively high during active periods of the task. To gain further insight into the LFB signal we analyzed the spectral content of the LFB feature during the baseline and screening tasks.
When looking at UNP1's LFB signal during the baseline task (Figure 2A left) the mean spectra show a broad peak between 12 and 26 Hz, with the lowest variance within and between runs around 25Hz. This same peak can be seen in the mixed spectra during rest periods of the screening task (blue line Figure 2B left). Indeed, isolated oscillatory components are present in the 10Hz to 27Hz range of the rest periods (Figure 2C left), with the most consistency over runs between 22Hz and 26Hz. This oscillatory component is not present during the active periods of the task. In addition, there is a smaller distinct oscillatory peak around 8Hz that remains present during active periods. Overall these results fit with a model of LFB activity during rest that is diminished with motor cortex activation.
UNP4 demonstrates a spectral bump from 6Hz to around 22Hz in baseline runs (Figure 2A right). However, the peak of oscillatory activity during rest periods of the task is focused between 7Hz and 10Hz and remains present, although to a lesser degree, during active periods (Figure 2B/C right). Thus, while there is evidence of a LFB oscillatory peak in the motor cortex of UNP4, this peak is constrained to the Mu band (6-12Hz), with the least variance at 7-8Hz, and hardly changes during active periods. Both the fact that there is a less distinct peak in the LFB during rest and the fact that this peak remains present to a certain degree during active periods could contribute to the decreased functional difference in the LFB during attempted hand movements for UNP4.

DISCUSSION

In is study we evaluated three distinct functional features of the SMC neuroelectrical ECoG signal in the context of two BCI users in the LIS state with different underlying etiology. Participant UNP1 suffered from ALS which is a neural degenerative disease that affects both upper and lower motor neurons [19]. UNP4 suffered from a brain stem stroke, which caused extensive damage to the brainstem. While both UNP1 and UNP4 have a functionally consistent HFB feature (present in >97% of attempted movement task runs) the work presented here shows a considerable discrepancy in both the functional consistency and oscillatory content of the LFB features between the participants. Here we discuss this result in the context of 1) the traditional SMC LFB functional feature reported in literature and 2) the reported effects of ALS and stroke on these frequency features.

Traditional ECoG SMC functional signal features: The HFB feature has been shown to be ubiquitous to human neocortex [20] and is associated with focal increases in asynchronous neural activity in response to executed, imagined or attempted movement [13-23].

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-14

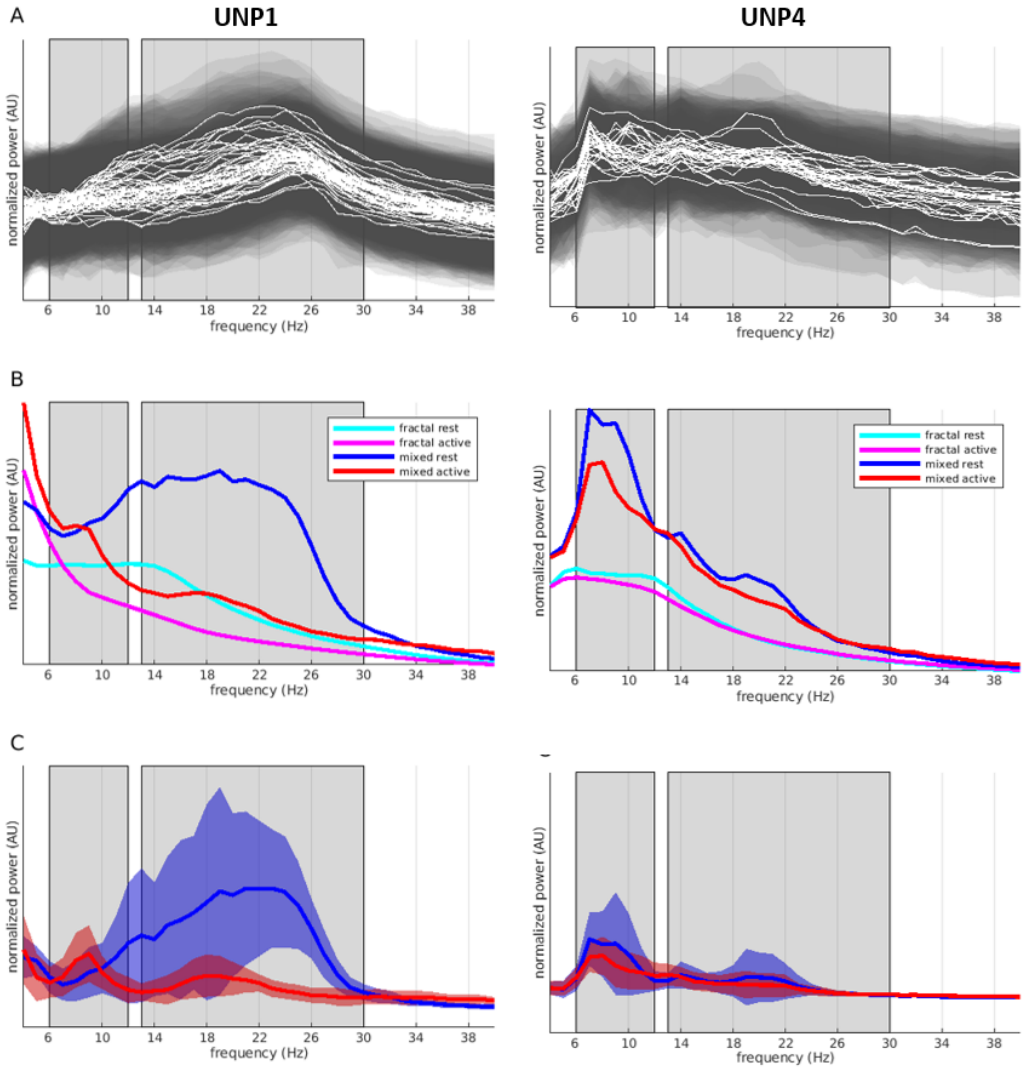


Figure 2: Oscillatory content of the LFB for UNP1 (left column) and UNP4 (right column). A) LFB spectra during baseline task. The mean (white lines) spectrum from 4-40Hz is plot for each run. The semi-transparent grey shaded regions indicate the plus and minus 1 standard deviation boundaries for each run. B) The mean (over runs) fractal spectral profiles for the rest (cyan line) and active (magenta line) periods and the mixed (fractal+oscillatory) spectral profiles for the rest (blue line) and active (red line) periods. C) The mean oscillatory profiles over runs of the rest (blue line) and active (red line) periods. The blue and red shaded regions indicate the rest and active plus and minus 1 standard deviation boundaries respectively.

In the case of the UNP1 it was previously demonstrated that reliable click-based spelling control was dependent on a combination of the HFB feature and an LFB feature [4], the spectral range of which (~6-35Hz) is consistent with that of the classic LFB spectral feature of motor cortex reported in ECoG literature [6,23]. This LFB feature overlaps with the Mu-band (6-12 Hz) motor rhythm often used in EEG-based motor BCI [24], and with other oscillatory features commonly reported in EEG and ECoG literature such as the Theta (4-7Hz) [25], and Beta (12-30Hz) [26] bands. The common term for the observed decrease in amplitude of these bands associated with cortical activation is Event Related Desynchronization (ERD). While the neuronal mechanisms underlying ERD are still being studied and may differ for different LFB sub-bands, the functional significance generally attributed to ERD is the release of inhibitory input on the measured neural population to allow cortical processing, which is reflected in the increase in HFB amplitudes.

The results of UNP1 fit with a model of Beta band desynchronization during motor cortex activation by attempted hand movement. However, there seems to be a lack of ERD in the mu-band. This suggests that Beta and Mu may have different generative sources that are differentially affected by ALS.

UNP4 shows little evidence of baseline Beta band activity suggesting disruption of the generative source for Beta band synchronization. The lack of functional consistency in the LFB rebound is in agreement with a disruption in the Beta band signal generation. Indeed, the Beta band has been most closely linked to Event Related Synchronization (ERS) after movement offset [27-28]. In

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-14

addition, attempted movement of the hand hardly affects LFB power, suggesting weak functional regulation. Although the LFB feature of UNP4 does show evidence of Mu activity, both the functional consistency analysis and oscillatory analysis results indicate that there is a disruption in the ERD of the Mu band as well.

Effects of ALS and Stroke on the LFB functional feature: Since ALS and brainstem stroke are fundamentally different neural motor afflictions it can be expected that the effect of these conditions on the functional features of SMC ECoG signal will differ. Here we discuss our findings in the context of the relatively small body of work addressing LFB (functional) features in stroke and ALS.

One study with 13 individuals who suffered a brainstem stroke reported higher levels of resting state alpha band (comparable to Mu) throughout the cortex, including the central areas [29]. This is interesting in light of our finding that there is a Mu oscillatory peak in UNP1 and UNP4 that persists during active periods. Both decreased Beta band motor ERD [30-31] and decreased Mu band ERD [32] have been reported in chronic stroke subjects. However, these studies focused on cortical and subcortical stroke and did not include any brain stem stoke subjects. While the lack of a Beta peak within the LFB feature, and persistence in Mu band oscillatory activity during active periods are both in agreement with this literature, the considerable differences in affected areas of cortical and brainstem stroke make it likely that other factors underlie the findings presented here than those in cortical stroke patents.

Several studies have addressed the resting state levels of LFB amplitudes in LIS ALS subjects. An EEG study with 8 LIS ALS subjects [33] and an ECoG study with one LIS ALS subject [34] both reported relatively more Theta power and less HFB power in baseline activity. While we did not specifically test for this effect in the baseline signal, a decrease in baseline HFB power could contribute to the robustness of the functional increase of HFB amplitude over rest period levels in our ALS subject. In studies of ALS subjects a decreased ERD [35-36] has been reported and correlated to disease progression [35]. However, another study reports an increase in ERD [37] while two more report no change [38-39]. Thus, the reported effects of ALS on ERD are inconsistent. In contrast, when reported, ERS has been reported to decrease in effect size [38]. However, this study included no ALS subjects who had reached the LIS stage. Our finding that ERS is very robust in the ALS LIS subject may be associated with the far progressed stage of the disease, but this topic deserves further investigation.

CONCLUSION

Our results indicate that while the HFB remains a stable SMC functional feature in two individuals with LIS with fundamentally different etiologies the functional stability of the LFB feature is much less in the brain stem stroke subject. Although these findings need to be confirmed in a larger population, they highlight the need to consider individual etiologies when designing SMC based BCIs targeted at individuals with LIS.

ACKNOWLEDGEMENTS

This work has been funded by the ERC-Advanced ‘iConnect’ project (grant ERC-Adv 320708) and the Dutch Technology Foundation STW (grant UGT7685). We would like to thank the UNP participants for the motivation and feedback provided in the last years.

REFERENCES

[1] Smith and Delargy. Locked-in syndrome. BMJ (Clinical research ed.). 2005; 330: 406–9.

[2] Laureys et al. The locked-in syn-drome : what is it like to be conscious but paralyzed and voiceless. Progress in brain research. 2005; 150: 495–611.

[3] Rousseau et al. Quality of life in patients with locked-in syndrome: Evolution over a 6-year period. Orphanet Journal of Rare Diseases. 2015; 10:88.

[4] Vansteensel et al. Fully Implanted Brain-Computer Interface In a Locked-In Patient with ALS. N Engl J Med 2016;375(21): 2060-66.

[5] Miller et al. Spectral changes in cortical surface potentials during motor movement. J. Neurosci. 2007; 27(9): 2424-32.

[6] Miller et al. Cortical activity during motor execution, motor, imagery, and imagery-based online feedback. PNAS 2010; 107(9): 4430-35.

[7] Kübler et al. Patients with ALS can use sensorimotor rhythms to operate a braincomputer interface. Neurology. 2005; 64: 1775–1777.

[8] Kauhanen et al. EEG-based Brain-Computer Interface for Tetraplegics. CompIntel. & NeuroSci. 2007.

[9] Spuler et al. Decoding of motor intentions from epidural ECoG recordings in severely paralyzed chronic stroke patients. J. Neural Eng. 2014; 11: 066008.

[10] Gomez-Rodriguez et al. Epidural ECoG online decoding of arm movement intention in hemiparesis. Proc. Workshop on Brain Decoding: Pattern Recognition Challenges in Neuroimaging at IEEE ICPR. 2010: 36–39.

[11] Ajiboye et al. Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. Lancet 2017; 389: 1821-30.

[12] Nuyujukian et al. Cortical control of a tablet computer by people with paralysis. PLOS ONE.

Published by Verlag der TU Graz
Graz University of Technology

2018; 13(11): e0204566.

[13] Hayashi and Kato. Total manifestations of amyotrophic lateral sclerosis. J. Neurological Sciences 1989; 93(1): 19–35.

[14] The Amyotrophic Lateral Sclerosis Functional Rating Scale. Assessment of activities of daily living in patients with amyotrophic lateral sclerosis. The ALS CNTF treatment study (ACTS) phase I-II Study Group. Arch Neurol. 1996;53(2):141-7.

[15] Bruns. Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? J. Neurosci Methods 2004; 137.

[16] Freudenburg et al. ECoG control signal optimization for use of a communication BCI implant in a person with Locked-In Syndrome. SfN Ann. Meeting 2016.

[17] Hermes et al. Neurophysiologic correlates of fMRI in human motor cortex. Hum. Brain Map. 2012; 33: 1689–99.

[18] Wen and Liu. Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal. Brain Topogr. 2016;29:13-26.

[19] Kiernan et al. Amyotrophic lateral sclerosis. Lancet. 2011; 377: 942–955.

[20] Miller et al. Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. NeuroImage 2014;85: 711-20.

[21] Miller et al. Power-Law Scaling in the Brain Surface Electric Potential. PLoS Comp. Bio. 2009; 5(12).

[22] Hermes et al. Dissociation between Neuronal Activity in Sensorimotor Cortex and Hand Movement Revealed as a Function of Movement Rate. J. Neurosci. 2012; 32: 9736–44.

[23] Miller et al. Spectral changes in cortical surface potentials during motor movement. J. Neurosci. 2007; 27(9): 2424-32.

[24] Wolpaw and Wolpaw. Brain-Computer Interfaces: Principles and Practice. Oxford University Press, Oxford, England. 2012.

[25] Canolty et al. High gamma power is phase-locked to theta oscillations in human neocortex, Science 2006; 313(5793): 1626-28.

[26] Miller et al. Human motor cortical activity is selectively phase-entrained on underlying rhythms. PLoS Comp. Bio. 2012; 8(9).

[27] Pfurtscheller et al. Post-movement beta synchronization. A correlate of an idling motor area?, EEG & Clin. Neurophys. 1995; 98: 281-293.

[28] Solis-Escalante et al. Cue-induced beta rebound during withholding of overt and covert foot movement. Clin. Neurophys. 2012;123: 1182-1190.

[29] Babiloni et al. Resting state eyes-closed cortical rhythms in patients with locked-in-syndrome: An eeg study. Clin. Neurophys. 2010; 121: 1816-1824.

[30] Rossiter et al. Do movement-related beta oscillations change after stroke? J. Neurophys. 2014; 112: 2053–2058.

[31] Shiner et al. Cortical beta oscillations and motor thresholds differ across the spectrum of post-stroke motor impairment, a preliminary MEG and TMS study. Brain Research. 2015; 1629: 26–37.

[32] Fu et al. Assessment of eeg event-related desynchronization in stroke survivors performing shoulder-elbow movements. Proc. 2006 IEEE ICRA. 2006: 3158–3164.

[33] Jayaram et al. Brain-computer interfacing in amyotrophic lateral sclerosis: Implications of a resting-state EEG analysis. IEEE Proc. EMBC. 2015: 6979–6982.

[34] Bensch et al. Assessing attention and cognitive function in completely locked-in state with event related brain potentials and epidural electrocorticography, J. Neural Eng. 2014; 11: 026006.

[35] Kasahara et al. The correlation between motor impairments and event-related desynchronization during motor imagery in ALS patients. BMC Neurosci. 2012; 13: 66.

[36] Bizovič car et al. Decreased movement-related beta desynchronization and impaired postmovement beta rebound in amyotrophic lateral sclerosis. Clin. Neurophys. 2014; 125: 1689–1699.

[37] Proudfoot et al. Altered cortical beta-band oscillations reflect motor system degeneration in amyotrophic lateral sclerosis. Human Brain Map. 2017; 38: 237–254.

[38] Riva et al. Cortical activation to voluntary movement in amyotrophic lateral sclerosis is related to corticospinal damage: Electrophysiological evidence. Clin. Neurophy. 2012; 123: 1586–1592.

[39] Bai et al. Movement related cortical potentials in primary lateral sclerosis. Ann. Neurology. 2006: 682–690.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-15

RESONANCE - A BCI FRAMEWORK FOR WORKING WITH MULTIPLE DATA SOURCES

Y.O. Nuzhdin

E-mail: nuzhdin.urii@gmail.com

ABSTRACT: Resonance is a framework for creating reproducible BCI experiments with multiple sources of data processed together. It is a cross-platform tool created with C++. It offers ways to create visual environments using QML and perform data processing with R and Python. Resonance has proved its work with multiple brain signal capturing devices including EEG, MEG, Eye tracking and number of other devices. Resonance framework does not exclude developer from the process of experiment creation, but it allows to focus on details of an experiment by using user-friendly languages and concepts. The framework is actively developed and was already used for various experiments in several laboratories. All data-processing code is open sources[6][5].

INTRODUCTION

Many novel non-invasive BCIs are trying to achieve maximum signal efficiency and accuracy with the help of multiple tools. That can be data acquiring devices like electroencephalography (EEG), magnetic resonance imaging (MRI), magnetic encephalography (MEG), eye trackers or stimulators like visual and auditory stimulators, transcranial magnetic stimulation (TMS), electro myostimulation (EMS). Using multiple data source helps understanding brain signal deeply and making better interfaces, however it requires a simultaneous work and a very precise synchronization between different devices.

There is a number of software tools for performing BCI experiments[2], and that list is growing[12]. These tools are created with different programming languages, has different capabilities and features. Previously, we successfully used some of them in experiments[11][1]. However, for new experiments we were looking for a platform that would be capable of processing data from multiple devices simultaneously and online.

Resonance was designed as a framework that can tight up multiple devices and provide an experimenter with an ability to design, acquire, view and analyze data in a user-friendly and unified manner.

The development started in 2014 by an independent team and to date several experiments have been performed with the Resonance framework [3][4][10][13].

FEATURES

Resonance was never meant as a tool that excludes a developer from designing an experiment. Instead it provides a set of tools and general approach that lowers requirements for knowledge and skills. It allows one to concentrate on details of experiment instead of low-level machine interaction details.

Within Resonance system we distinguish following aspects of an experiment: data collection, aggregation, processing and storage, experimenter feedback loop and subject feedback loop. We use the term 'loop' because not only experimentation system affects subject, but a subject may also affect experimentation system. Experimenter can monitor diagnostic data and change settings during an experiment. The subject can affect experiment in two different ways: by being a data source (if we measure parameters, which subject can control) and by changing a procedure and parameters of an experiment (for example by pressing a button when one is ready).

Our goal is to provide an experimenter with implementations for some of these aspects without artificial limitations that simplify creation of the framework but limit abilities to make complicated experiments.

Resonance provides a set of tools to fulfill requirements of named aspects:

- Data collection - Resonance implements drivers for devices. Hence an experimenter has access to all configuration options and working modes provided by a device. These options and modes could be changed during experiment's execution. It is also possible to monitor device failures.
- Data aggregation and data processing - Resonance provides libraries on R[6] and Python[5] that allow to combine and process data from multiple sources.
- Data storage - Resonance provides a facility to store data. It uses its own open data format because it stores every single bit of data produced by the system, which allows for example to detect problems with hardware after an unsuccessful experiment.
- Experimenter and subject feedback loops - Resonance provides libraries for QML to control all aspects of an experiment, including handy widgets like data visualization. In fact, every single UI piece of Resonance is implemented with QML modules and is available open source, so if you find any of existing tools useful you can copy it to your experiment. QML is a declarative language for designing

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-15

user interface-centric applications. Interface's behavior, designed with QML, can be scripted through JavaScript. The tool is multi-platform: available on Linux, Windows, MacOS and Android. It uses video acceleration to produce fast graphical output. From our experience it is not only possible but also very easy to implement tempting features like stimuli presentation for a single frame at desired time without any knowledge of how video system works.

It is important to mention that Resonance can be integrated with a third-party software without any limitations.

SYSTEM USAGE EXAMPLE

We will illustrate abilities of the system at the non-typical BCI experiment performed in Lomonosov Moscow State University[13]. In this experiment we recorded EEG and EMG during TMS stimulation. At each run subject performed two types of mental tasks - one of three motor imagery tasks (experimental condition) and a visual attention task (reference condition). In EMG analysis, the amplitude of TMS-evoked motor potentials (MEP) was measured peak to peak (between negative and positive phases of potential). Potentials with any signs of raised muscular activity during preceding 1000 ms interval were rejected manually. MEP amplitudes in experimental condition (motor imagery) were normalized to mean amplitude of the reference condition ("visual attention" or "blank screen") of the same run. The mean of the subject's MEP amplitudes during fingers flexing imagery task minus 100% was used as a corticospinal excitability increase measure for further analysis.

During TMS measurement online EMG-feedback was displayed at the right side of the screen as a vertical bar with a real-time RMS (root-mean-square) value (300 ms window, 100 ms step)(figure 2). Participants were asked to find hand position with minimal ongoing EMG amplitude and to keep the corresponding bar level constant during the whole run.

To control quality of TMS stimulation during the experiment EMG responses to stimulation with amplitude and latency measurements were shown on the experimenter's screen. Both EEG and EMG data were recorded for offline analysis.

At the center of subject's screen was an image indicating current task. At the right side of subject's screen was a bar showing subject's muscle activity. Value for this bar was calculated from EMG measures online to help a subject to reach a muscle relaxation. On a separate screen an experimenter saw an interface with all parameters of the experiment. At the right-side column TMS-invoked motor potentials (MEP) were displayed. Each row contained EMG data, amplitude and latency values for single stimulation (figure 1).

SYSTEM ARCHITECTURE

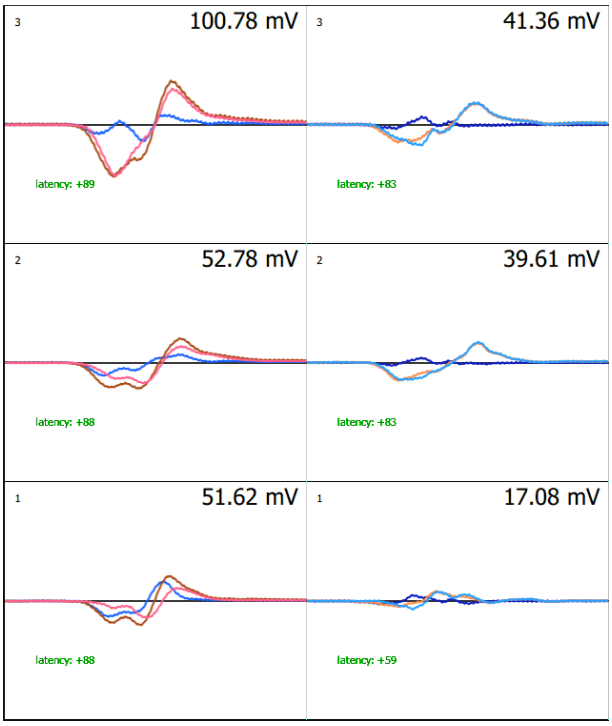


Figure 1: Part of experimenter's screen with MEP, latency and amplitudes measurements. Colors are changed for printing.

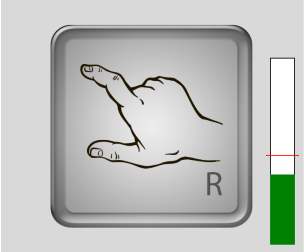


Figure 2: Subject screen. Pictogram of a hand represents mental imagery task cue and bar on the right side shows real-time RMS of subject's EMG from forearm muscles.

The framework provides a general data-flow-centric architecture, which should fit any possible experiment. In this approach we distinguish Resonance services: nodes of data processing and control. Each service can publish data streams, its own parameters and state (e.g. operation mode of a service). All communications between services go through TCP/IP network and services can be started at the same or different machines. Each service can discover other services in a local network, receive data streams, update parameters of other services and send commands to change state (for example start calibration or recording). This approach makes every service independent from others, which brings modularity to system and helps with debugging and analyzing problems.
In the example above were 5 services -

- an EEG driver
- an EMG driver (EEG and EMG were recorded by two different devices)
- a service for subject's screen

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-15

- a service for experimenter's screen
- a service for EMG data processing using R[9] script engine

Data from EEG and EMG devices was collected by experimenter's service and recorded to a file by standard Resonance facility. Data from EMG device was also acquired by R script engine. This service was executing R script to calculate RMS of EMG. This service also gathered myographic data after TMS stimulation to display results on experimenter's screen. Subject service was consuming RMS data from R script engine and stimuli commands from experimenter's service. Experimenter service was responsible for experimentation cycle. It contained a finite state machine to control a stimuli sequence and UI module to display myographic data. Myographic data was updated after each stimulation by detecting corresponding event in a runtime.

ONLINE AND OFFLINE DATA PROCESSING

Making data processing algorithms is not easy, especially when you need to model data processing before or after the experiment. Typical approach to this task is to provide system with a set of data transformation blocks, which you can combine to perform required data processing steps. The positive side is that it is relatively easy to combine and configure such blocks. First problem with this approach is that experimenter is limited by the number of kinds of these blocks. Also creating these blocks requires a good knowledge of the internals of a BCI system and a language with which that system is written (usually C++). Second problem is that it is hard to use these blocks for offline data processing, as they are highly integrated with BCI system, while offline data processing is usually performed by Python, R or Matlab.

Having separate code for online and offline processing makes experiment development very hard. Achieving exactly same computational result in computer systems requires knowledge of how computation is made step-by-step. This means that a developer of an experiment must translate online processing to an offline tool or hope that they will produce results which differ insignificantly. Achieving equal results for offline and online is not a requirement for every application, but that limits in ability to detect irregularities or test additional hypotheses post-hoc. For example if you discover an unusual behavior in an online experiment it will be difficult to figure out if it was due to computational errors or anything else.

In order to overcome this problem Resonance provides a library that can be used for both online and offline data processing. The data flow is still constructed from blocks, but these blocks (as well as all data processing) are written in a language which is more familiar for experimenter (R and python for now) and these blocks are not coupled with the details of online data processing. That means that you can get exactly same results using data from online experiment, and even more, you can model online performance of the system offline, using a language you are familiar with. In fact data processing library is completely abstract from any Resonance-specific data processing, and with minor adaptations could be used with any system which supports R or python language in runtime.

Using Resonance library allows not only to execute online and offline processing, but also helps to visualize it. For example, figure 3 is the automatically generated visualization of data processing plan for EMG processing from the example experiment.

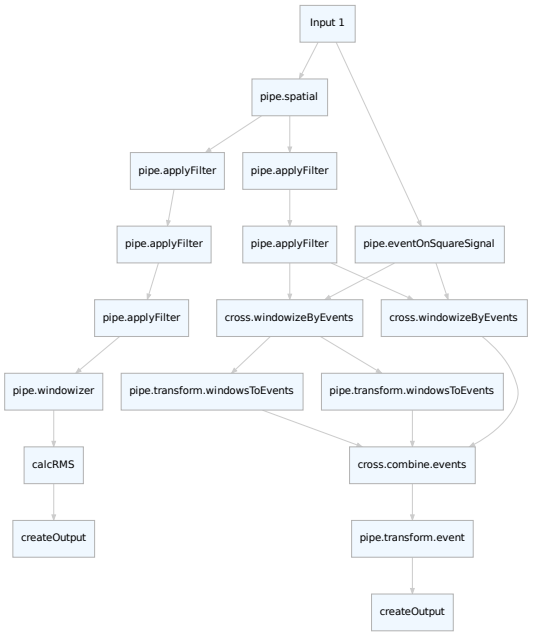


Figure 3: Data processing plan for EMG processing. Arguments for the processing steps are not visualized on this chart.

DATA AGGREGATION AND SYNCHRONIZATION

Data aggregation:

For processing digital signals you can work in two different domains - in sample space and in time space.

- Working in sample space means that you quantify time in your system by a number of samples of the signal you are processing.
- Working in time space means that you measure your time by some timer which is not bound with the signal and that timer has finer resolution than sampling rate of your signal.

Working in the sample space seems very natural when you deal with single signal source. But unfortunately, most of real-world applications never work with the single source of the signal. For example - when you present any stimulus, you have two sources of data - one for EEG

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-15

and another for moments in time when stimulus was presented. In this situation you usually introduce an additional channel with event data or some kind of flags. Such solution works for some applications, but it cannot be generalized. When you try to work with multiple signals even if they have same sampling rate you will face the problem of phase alignment for these signals. Introducing every new signal source will add at least half of the period of this signal to time desynchronization (or even more if the frequencies of signals are not dividable without the remainder). If signals have sampling rates which are not multiples to each other there is a problem of sub sampling - when one sample of signal A corresponds to a variable number of samples of signal B. If you try to process such signals in parallel you will introduce synchronization deviations, and if you bring the signals to a common frequency you will introduce phase, amplitude and spectrum deviations in signal.

Working in time space makes sense if you use timer with resolution smaller than the minimal period of processed signals. In this case applied transformations still can change timestamps of a samples, but as these timestamps are finer than a period of a signal they would not affect your signal. However, working with timestamps of the recording device may introduce unnecessary complexity to the calculation process. Recording devices do not produce data blocks at exact time intervals. For example, EEG recording device may show deviations between timestamps of the blocks which will usually be bigger than a period of the signal.

Resonance framework provides the following solution to synchronization problem:

1. an initial timestamp for the signal is averaged from first few blocks of data
2. timestamps for samples are recalculated from sampling rate
3. during the data processing Resonance also performs a control, so that there are no delays or deviations in data retrieval

Synchronization:

Synchronization between signals is very easy when you work in a time space - you just align closest timings together and your signals are synchronized, and again, by this operation you introduce phase deviation which is smaller than sampling rate, so it is not reflected in your signal as well. Of course, you need to take into account that different computers use different timers and you need to align these timers if you record data with separate machines. However, if you record data with one machine you will have access to a very precise timer (tens of nanoseconds) of the machine, which is enough for any imaginable signal. You can try to achieve time synchronization among different machines via network, but it is more precise and reliable to physically connect recording devices and perform a synchronization based on the same signal recorded by several devices.

In the example experiment there were 3 time-aligned devices - EEG , EMG and TMS. In order to synchronize them EEG and EMG were connected to a TMS trigger channel. TMS sent pulse through this channel each time a stimulation was performed. By that means we achieved not only data-synchronization between EEG and EMG, but also very reliable source of time information when TMS stimulation was performed. That time information was used to extract EMG data related to the stimulation, that was presented on experimenter's screen.

Performance considerations:

When talking about working in signal or time space we usually mean working with separate samples. But for computers calculating separate samples is too inefficient and introduces a big overhead, so we are forced to collect and process data in blocks. From the other side - gathering block of samples introduces latency, and you cannot immediately process collected data. The decision whether blocks are equal in size or not should be made by an experimenter, not dictated by a framework. In order to decide one need to take into account performance requirements of one's application, the way hardware works (for example is it capable to return data blocks of variable size) and what data processing algorithms are used in experiment. Resonance supports blocks of variable size, and each block has a timestamp of creation and a timestamp of receiving. That allows monitoring delays in a network or in blocks' processing. Resonance's R and Python open source libraries provide abstraction over data stream, which is independent of block size and amount. That allows experimenter to try experiment with different block size settings in order to achieve required performance and latency.

PLANS FOR THE FUTURE

For the upcoming year there are few big improvements planned:

- Demo experiments. Currently Resonance distributive contains only basic signal processing tools. Our plan is to share some experiments and tools we made with Resonance. That will provide a basis for other experimenters to build and reproduce their experiments. In order to achieve that existing implementations should be reviewed and documented. These demo experiments will be freely available at the Resonance website[7], as well as instructions about how to install and run them.
- Extend support for devices. Currently there are several recording devices natively supported by Resonance (actiChamp, NVX and openBCI eeg recorders, eyelink and eyetribe eyetrackers, neuromep-4 myostimulator) Many other devices could be connected using fieldtrip[8] protocol, but fieldtrip does not allow to control device parameters and modes. In upcoming year, we will extend native support for at least EEG, MEG and eyetracking devices,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-15

and probably implement other popular protocols like Lab Streaming Layer.

- Code improvements. Resonance claims to be the system that exactly reproduces experiments. This functionality is the core of the system and was extensively tested in simulations and real applications. Last year there was a big improvement in number of automated tests for this functionality. These tests explicitly proof that everything you can do with Resonance have exactly the same behavior online and offline, every time you run application on any machine. For the upcoming year we have a goal to completely cover R and python Resonance libraries with tests.

List of improvements, updates and plans, and other information is available on the website of the framework[7]. All source code for online and offline data processing is freely available on GitHub repositories for R[6] and Python[5].

ACKNOWLEDGEMENTS

Authors would like to acknowledge Anatoly Vasilyev and Sergey Shishkin for providing useful advices and feedback and for their help in testing.

References

[1] A.A. Fedorova et al. A fast "single-stimulus" brain switch. In: Proceedings of the 6th International Brain-Computer Interface Conference 2014. Sept. 2014.
[2] Brunner Clemens et al. BCI Software Platforms. English. In: Toward Practical BCIs: Bridging the Gap from Research to Real-World Applications. Springer, 2013, 303–331.
[3] Nuzhdin Yuri O. et al. Passive Detection of feedback Expectation: towards Fluent Hybrid eye-brain-Computer Interfaces. In: From Vision to Reality - Proceedings of the 7th Graz Brain-Computer Interface Conference, GBCIC 2017, Graz, Steiermark, Austria, September 18-22, 2017. 2017.
[4] Nuzhdin Yuri O. et al. The Expectation Based Eye-Brain-Computer Interface: An Attempt of Online Test. In: Proceedings of the 2017 ACM Workshop on An Application-oriented Approach to BCI out of the Laboratory. BCIforReal '17. Limassol, Cyprus: ACM, 2017, 39–42.
[5] Nuzhdin Yury. *Resonance Python package*. `https://github.com/tz-lom/Resonance-engine-python`.
[6] Nuzhdin Yury. *Resonance R package*. `https://github.com/tz-lom/Resonance-engine-R`.
[7] Nuzhdin Yury. *Website of the Resonance project*. `http://resonance.bcilab.net`.
[8] Oostenveld Robert, Fries Pascal, Maris Eric, Schoffelen Jan-Mathijs. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience. 2011;2011.
[9] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2018.
[10] Shishkin Sergei L. et al. EEG Negativity in Fixations Used for Gaze-Based Control: Toward Converting Intentions into Actions with an Eye-Brain-Computer Interface. Frontiers in Neuroscience. 2016;10:528.
[11] Shishkin S.L., Nikolaev A.A., Nuzhdin Y.O., Zhigalov A.Y., Ganin I.P., Kaplan A.Y. Calibration of the P300 BCI with the single-stimulus protocol. In: Proceedings of the 5th Graz Brain-Computer Interface Conference, GBCIC 2011, Graz, Steiermark, Austria, September 22-24, 2011. 2011, 256–259.
[12] Smetanin Nikolai, Volkova Ksenia, Zabodaev Stanislav, Lebedev Mikhail A., Ossadtchi Alexei. NFBLab—A Versatile Software for Neurofeedback and Brain-Computer Interface Research. Frontiers in Neuroinformatics. 2018;12:100.
[13] Vasilyev Anatoly, Liburkina Sofya, Yakovlev Lev, Perepelkina Olga, Kaplan Alexander. Assessing motor imagery in brain-computer interface training: Psychological and neurophysiological correlates. Neuropsychologia. 2017;97:56–65.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-16

PRESENTATION SPEEDS FOR A N400-BASED BCI

K.V. Dijkstra, J.D.R Farquhar, P.W.M Desain

Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

E-mail: k.dijkstra@donders.ru.nl

ABSTRACT:

The N400 is an ERP sensitive to the semantic content of a stimulus, in relation to the active mental context of a subject. By repeatedly presenting stimuli (i.e., probe words), we can thus infer information about this context through decoding of this relatedness response. This can form the basis of a, so called, semantic Brain Computer Interface, allowing us to directly identify the concept on a users mind without spelling out the letters. The usability of such a BCI depends on how much information can be extracted from a single presentation, but also on how fast stimuli can be presented. Here we report results from a pilot study of an experimental design that aims to determine the effect of the time between such probes on the amplitude of this N400. The preliminary results show that an N400 can still be detected across subjects, even with stimuli presented at nearly twice the rate (1.7x) used in one of our previous studies.

INTRODUCTION

The N400 is an Event Related Potential that is more negative for stimuli that are not related to the person's current mental context [1]. It is thought to reflect the (attempt to) incorporate new semantic information (e.g., from a stimulus) into this existing context. From a Brain Computer Interface (BCI) viewpoint we may be able to exploit this activity to infer information about this hidden mental context. Specifically, we can use the presentation of semantic stimuli (e.g., words) to learn what this active semantic context is, by decoding the N400 from each stimulus presentation, and accumulating information across presentations. In other words, the user thinks of a specific concept (we refer to this as the target word), and the BCI presents multiple consecutive words (referred to as probes) that may or may not be related to this target, and combines the information inferred from the brain responses with knowledge of the relationships between the probe words and any potential target word, to ultimately infer the target word.

This requires a ground truth database that tells us which concepts are or are not related. Such a database can be built by consulting people on whether concepts are related, or by asking people what words come to mind for a given concept (e.g., [2] dutch). However, it is not feasible to ask about all possible combinations of concepts in this way. An alternative approach is to use methods from computational linguistics such as Latent Semantic Analysis (LSA) or the more recently popular word2vec [3]. These methods learn word representations, in the form of n-dimensional vectors, from large text corpora, by looking at which words occur together or which words occur in similar contexts, depending on the method. From these vector representations, the relatedness of two words can then be calculated by computing the *cosine similarity*, i.e., the cosine of the angle between the two vectors. The advantage of these vector-based methods is that the relatedness between any two concepts can be measured, while the methods in which humans are queried result in sparse databases. Importantly, the relatedness scores from such vector-space models have been shown to correlate with the strength of the N400 response, at least as well as the human-elicited association databases [4].

The amount of information that can be extracted about the initial target word depends on both the accuracy with which the relatedness can be decoded, and the amount of stimuli that can be presented in a given time frame. Previous studies suggested the N400 response has high variability and can be decoded with only limited accuracy (50-75% on a binary problem, [5, 6]), making the speed at which stimuli can be presented particularly relevant. In the second of these studies, in which we tested the robustness of this N400 over multiple consecutive probes after a specific target word, we used a Stimulus Onset Asynchrony (SOA) of 1350 ms, i.e., presenting a stimulus every 1.35 seconds.

Here we present an experimental design aimed to determine how much information is lost per stimulus for decreasing SOAs. We test an SOA of 1250 (SLOW), an SOA of 750 ms (MEDIUM) and a 250 ms SOA (FAST). Elicitation of the N400 does not *require* active semantic analysis on the part of the participant, with N400s also having been evoked in passive tasks. However, the N400 is generally stronger in tasks that involve active analysis [7], while reducing the time until the next stimulus will make this harder or impossible for participants. On the other hand a N400 BCI paradigm that does not require an active task, would be preferable from a user point of view. The question therefore is how, for a given time-window, the ability to present more stimuli trades off with a potential reduction in response amplitude and decoding accuracy of a single stimulus.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-16

We report preliminary results of seven participants of a pilot of this experimental design. While the aim for a full scale study would be to determine what, if any, decrease in information transfer rates occurs for faster SOAs, we restrict ourselves here to validating the experimental design. Specifically, we use Grand Average ERPs across subjects to ascertain that the baseline condition (SLOW), elicits a distinguishable N400 response for related and unrelated probes, as some experimental parameters have changed with regard to the previous study (in particular the participants' task and the ground truth model). Further, we aim to determine for the faster SOAs whether either of those elicit a detectable N400. Lastly, we evaluate the performance of participants on the behavioural task to determine whether it is suitable as a behavioural check of the participant's attention to the stimuli.

MATERIALS AND METHODS

In the experiment, participants were presented with a target word to remember, and then shown multiple consecutive probe words that had a varying degree of relatedness to this target word. Words were drawn from the 5000 most frequent words in English (Corpus of Contemporary American English [8]). The speed at which these probe words were presented depended on the condition:

- SLOW: 250 ms stimulus, 1000 ms fixation cross
- MEDIUM: 250 ms stimulus, 500 ms fixation cross
- FAST: 250 ms stimulus, no fixation cross

A behavioural task was added to the experiment to ensure subjects kept the target word in mind during the presentation of the sequence. After a certain number of probe words, the subject was prompted to decide whether the probe word was related or unrelated to the target (see Fig. 1). A total of a 150 of these trials were presented across 6 blocks of ~10 minutes each (50 in each condition).

Participants: Seven participants completed the pilot study, ranging in age between 20 and 30. Despite the availability of a large population of native Dutch speaker at our lab, English was used for the stimuli as there are more good quality vector-space models available. To minimise problems with word comprehension in non-native speakers, participants were recruited who categorized themselves as *"speaking and understanding English" "Well"* or *'Very Well'*. In addition, we used *LexTale* (Lextale.com [9]), as an objective measure of English vocabulary knowledge. Two subjects scored low on this task (50-60%); the remaining five all scored above 80%.

Stimuli and Task: To determine relatedness we obtained pre-trained word2vec vectors, trained on a Google News corpus (GoogleNews-vectors-negative300.bin, from https://code.google.com/archive/p/word2vec/). Target words were selected at random from the 5000 frequent words. For each target word, between 1 and 30 probe words were selected of varying relatedness with the target word. We refer to the combination of a target word together with all of its probes as a trial.

Trials contained a varying number of probes (between 1 and 30), so that participants could not anticipate when a behavioural prompt would appear. Each condition (FAST, MEDIUM and SLOW) was created to have an approximately equal number of probes per trial, but consequently did not have an equal duration; a trial with 30 probes would last ~10s, ~25 or ~40s, for FAST, MEDIUM and SLOW, respectively. Every 5 trials, the condition changed, with the subject being notified of the next presentation speed. The order in which conditions appeared was randomized across subjects.

Participants were given feedback on their relatedness decision in the behavioural task. To account for disagreement between the word2vec model relatedness judgements and their own, they received a short explanation of word-embedding methods, and were tasked to predict the model's judgements, rather than reporting their own. To this end, the continuous cosine similarity scores from the word2vec model were discretized into three bins: -1 to 0.15, 0.15 to 0.3, 0.3 to 1, labeled 'unrelated', 'maybe related' and 'related' respectively. If the subjects choice matched the model's label, they received one point. To keep the task straightforward, participants did not have a 'maybe' option, instead we customized the points for this category: they received a point if they predicted a 'maybe' as related (this suggests they are able to predict the model well), and half a point if they predicted it as unrelated (to penalize them less for not recognizing a relationship). Participants were given a 1250 ms window to respond, and a point was deducted if no choice was made.

Updated design: After three subjects we inspected ERPs and noted that in the FAST condition, the expected location of the N400, in time, coincides with the stimulus response to the subsequent stimulus. To reduce this effect we added a jitter to the stimulus duration of between 1 and 100 ms. We jittered the stimulus duration, rather than the fixation cross duration, as, in the base case, the fastest condition did not have a fixation cross. As, on average, the SOA was now increased by 50ms we reduced the fixation cross duration by this amount to compensate (applicable only to the MEDIUM and FAST conditions).

Analysis:

EEG was recorded with 32 sintered Ag/AgCl active electrodes (BioSemi ActiveTwo, Biosemi, Amsterdam, The Netherlands), at a sampling rate of 256 Hz. Two additional electrodes were placed on the mastoids, and four more electrodes were used to measure horizontal and vertical EOG.

For the analysis the recorded data was loaded and high-pass filtered at 0.1 Hz (4th order Butterworth filter). Data was then sliced into epochs with respect to each probe onset. These epochs were re-referenced to the two mastoid electrodes. To ensure any EEG signals were not contaminated with eye muscle activity, we regressed the signals from the EOG out of the EEG channels [10]. The data was subsequently low-pass filtered at 20Hz. To remove

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-16

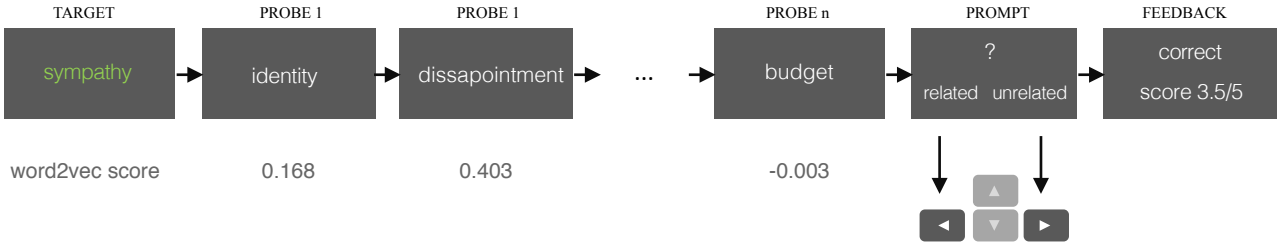


Figure 1: Experimental trials. A target word is presented, followed by several probe words. At the end of a trial, the participant is prompted to determine the relatedness of the most recent presented probe, in relation to the target word. Word2vec cosine similarity scores of the target and the respective probes are shown for this example trial.

any poorly-connected 'bad' channels, EEG channels that had a variance 3.5 standard deviations from the channel mean variance, were rejected, and replaced by an interpolated channel, using a spherical spline interpolation [11]. Epochs with abnormal activity were identified with the same 3.5 standard deviations of variance measure and excluded from further analysis.

RESULTS

Behavioural: Participants were given a behavioural task to allow us to check that they were keeping the target word in mind, and attended the subsequent probe words. We report their scores as a fraction of the maximum obtainable score. Performance on this task can give us an estimate of their attention to the sequences, but is also dependent on their baseline ability to predict the word2vec model relatedness judgements. For this reason we included a 'wordpair' baseline task in which participants received a target with only a single subsequent probe. The behavioural scores for the three conditions and this baseline task can be found in Fig. 2.

The figure shows an increase in (median) score for faster presentation speeds, with participants performing similarly or higher on the fast sequences compared to the baseline task, but achieving lower scores on the MEDIUM and SLOW condition. Keep in mind that across the three conditions sequences were equally long with respect to the number of probes, and not the duration of the trial in seconds.

EEG:

To evaluate the suitability of the experimental design we look at grand average ERPs for each of the three conditions (FAST, MEDIUM and SLOW), contrasting responses to related and unrelated probes. For this purpose we select only epochs with probes that were highly unrelated or highly related according to the model: in the range $[-1, 0.1>, <0.3, 1]$ respectively. The ranges were chosen so that an approximately equal number of probes were assigned to the related as to the unrelated averages (200-250 probes per class). A large portion of stimuli fall outside this range, but have been included in the experiment to more closely resemble the probe distribution in a semantic BCI application, where probes could not be guaranteed to fall on only the 'extreme' sides of the

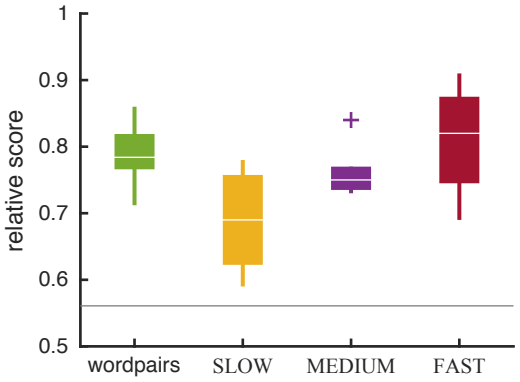


Figure 2: Participant accuracy on the behavioural prediction task. Score is relative to the maximum obtainable score. Scores are reported for each of the three speed conditions, plus a 'wordpair' condition that served as a baseline task. The grey line represents the mean accuracy achieved for pressing buttons randomly.

relatedness spectrum, since the true target concept is unknown.

To test if there was a significant difference between related and unrelated probes for each condition, we used a non-parametric cluster-based permutation test (in Fieldtrip, [12]). The test was performed on all channels, in the 300-800 ms time range, using a bonferroni corrected one-sided alpha-level over conditions ($\alpha = 0.05/3 = 0.0167$). A significant effect was found for both the SLOW and MEDIUM condition ($p = 0.0010$ and $p = 0.0150$, respectively. The Grand Average ERPs across subjects can be found in Fig. 3. Fig. 3A and Fig. 3B show the average ERPs for the SLOW and MEDIUM condition. The grey areas mark the timepoints in the respective significant clusters identified by the permutation test. The average of unrelated probes is more negative around the 400ms range, as expected for the N400. The ERPs for these two conditions look largely similar in shape, though at the end of the 1 second window, for the MEDIUM condition, the (visually) evoked potential to the next stimulus appears to be visible (750ms SOA).

For the FAST condition (Fig. 3C, no significant cluster was identified ($\alpha = 0.0167$, $p = 0.0220$). In this condition four stimulus responses can be observed in the 1 second window, with the second one corresponding to the response to the stimulus triggering the ERP. The stim-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-16

ulus response to these epochs looks similar to those in the other conditions, but the stimulus response to the next stimulus co-occurs with the region in which the N400 was detected in the slower conditions. We introduced the jitter in stimulus duration, of 100ms, noted above, after S03 to reduce this overlap, but have averaged across all participants here (S01-S07).

The topographies reflecting the difference between related and unrelated ERPs, for the three conditions, are shown in Fig. 3D. The topographies represent the difference waves (unrelated − related), for (non-overlapping) periods of 200 ms, starting from 0 s. A negativity across centro-parietal sites, as expected for a N400, is visible in the 400-600ms slice, for all three conditions, though less pronounced for the FAST condition.

To visually investigate the degree of individual variability in ERP responses, per-subject and condition ERPs plots are shown in Fig. 4. Multiple channels in the centro-parietal region are shown. Note that standard deviations are large and not depicted here. Clearly the subject-to-subject variability in the responses is large – something masked by the grand-average responses. S01 shows a consistent negativity around 400ms for all three conditions. Other subjects also show a negativity in the expected timerange, though not necessarily across all conditions. S02, on the other end, shows very little of response that could be interpreted as an N400 at all. For the FAST condition, we would have expected to see a difference before and after introducing the stimulus duration jitter, but any effect is difficult to determine from this data. S06 and S07, for instance, have small stimulus responses even for the one stimulus the average was time-locked to.

DISCUSSION

The goal of this preliminary study was to determine (1), whether the behavioural task was suitable for measuring the degree to which the subject was able to maintain attention during the repeated presentations of probes. (2), whether at the least the baseline condition (SLOW), elicited an N400, as expected. And (3), to give us a glimpse of the results we may expect in the full study with regard to the other two SOAs.

Behaviourally, we see that participants achieved scores well above chance level (though we did not test significance here), and that their performance decreases, rather than increases for slower presentation speeds. This is not what we expected, and suggests that the task in its current state does not benefit from the extra evaluation time of the probe prior to the behavioural prompt appearing. This could mean that conscious evaluation in that window is not important for the task, *or* that participants are not using this time for evaluating the relatedness to the original target (some comments made by participants suggest it is the latter). Additionally, all participants reported finding it harder to sustain attention during the SLOW condition, and some noted that it made them sleepy. This suggests that the decrease in performance is an effect of reduced

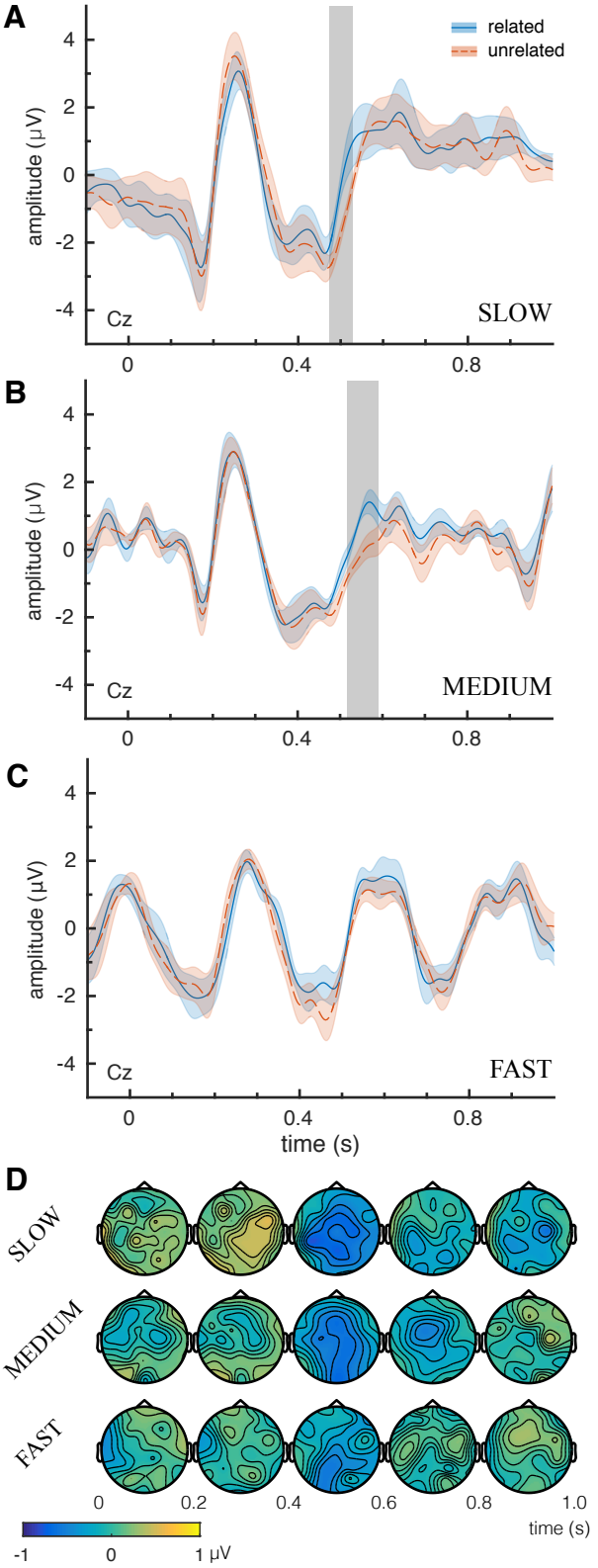


Figure 3: Grand Average ERPs across 7 subjects, for the central midline electrode (Cz), for **(A)**, SLOW, **(B)** MEDIUM and **(C)** FAST condition. Gray boxes denote significant clusters identified by a cluster permutation test. **(D)** topographies of the difference waves for each condition (unrelated − related), from 0-200,200-400,400-600 and 800-1000ms respectively.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-16

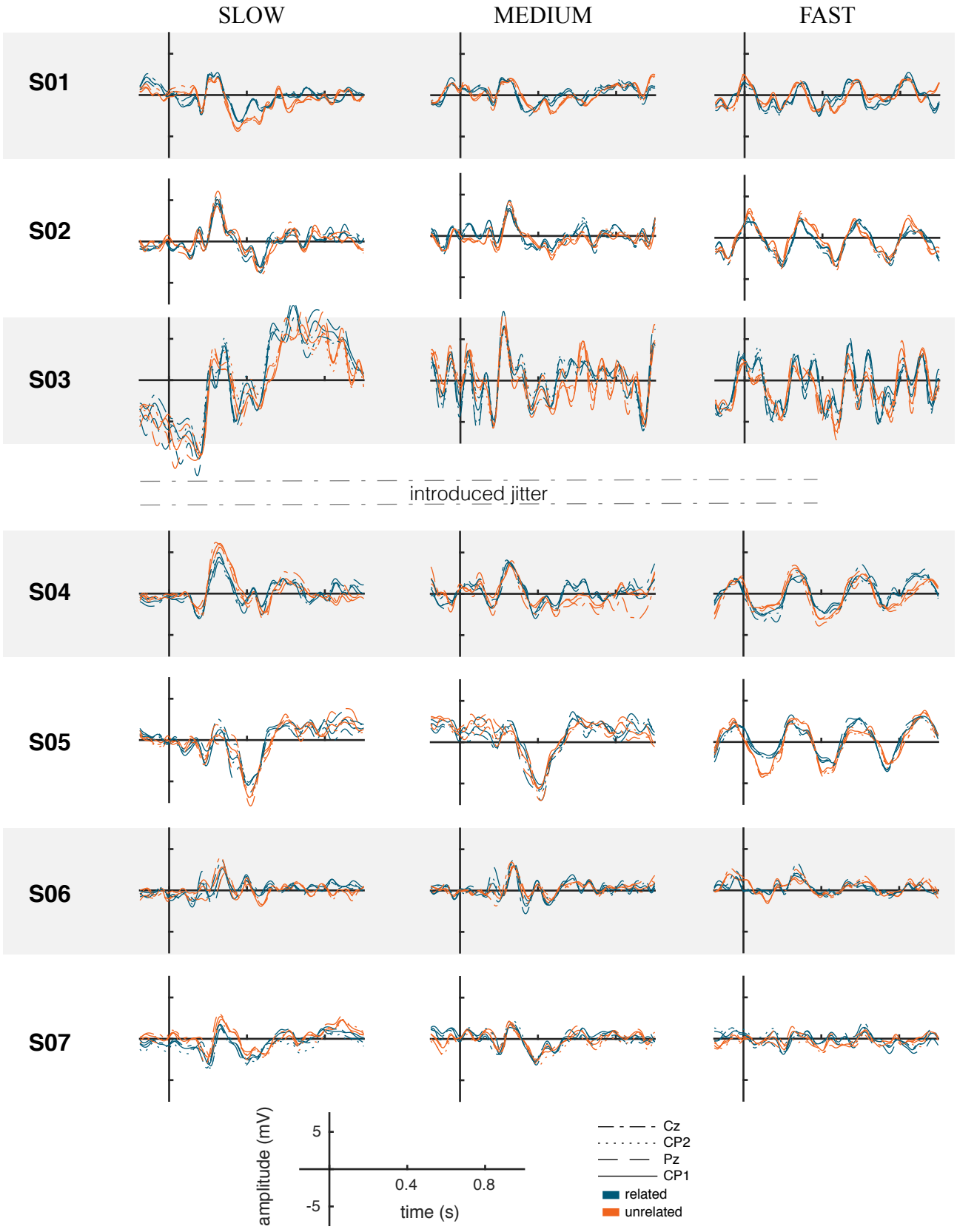


Figure 4: ERPs per individual subject (S01-S07), for the SLOW, MEDIUM and FAST condition, plotted for four centro-parietal electrodes (Cz, CP1, CP2, Pz). The introduction of stimulus duration jittering has been marked in the plot.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-16

attention, however, this may not be intrinsically due to the SOA of this condition, but simply because the total accumulated time between the original target and the behavioural prompt was larger.

Our grand average ERP results show that a difference between related and unrelated probes can be identified from the SLOW condition, in the time range associated with the N400. This confirms that experimental design is able to elicit N400 responses, despite the use of non-native speakers as participants, and the use of the word2vec similarity scores as a relatedness measure. However, the amplitude difference looks small, and the individual ERPs also show only small amplitude differences in the N400 range, if at all. Anecdotally, this ERP difference looks smaller than in our previous study [6](preprint), but this may also be due to the fact that in the other study the task explicitly encouraged subjects to evaluate each probe before a behavioural prompt appeared. Here, we removed this aspect, as we assumed that the shorter SOAs would not give a participant an opportunity to do this, but this potential decrease in N400 amplitude may thus reflect the difference between an active and passive task (as established in other research [7]).

With regard to the other two conditions, we found a significant difference for the MEDIUM condition. This is an encouraging result, though due to the small N we do not yet have the statistical power to compare the size of this response to that of the SLOW condition. We find no related/unrelated difference for the FAST condition, in this pilot data, but statistical power is also a limiting factor here.

CONCLUSION

Overall we can conclude that the experimental design is suitable to answer our question. Furthermore, the ability to present stimuli at the speed of the MEDIUM condition, rather than at the SLOW speed, would increase the rate of stimulation of the potential BCI by 1.7x. The fact that an N400 could still be detected reliably (across subjects), at this speed, is thus an encouraging preliminary result.

A full scale study, together with more sophisticated analysis of single subject data, e.g, regression or classification analysis and determining the information transfer over time, will be required to determine, in detail, the trade-off between accuracy and speed of probe presentation (in terms of SOA). This in turn will determine the speed with which a BCI paradigm that exploits the N400 response can infer the concept on a user's mind, and hence the suitability of this approach for potential applications.

REFERENCES

[1] M. Kutas and K. D. Federmeier, "Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP)", *Annual Review of Psychology*, vol. 62, no. 1, pp. 621–647, 2011.

[2] S. D. Deyne and G. Storms, "Word associations: Norms for 1,424 Dutch words in a continuous task", en, *Behavior Research Methods*, vol. 40, no. 1, pp. 198–205, Feb. 2008. (visited on 10/09/2017).

[3] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality", in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2013, pp. 3111–3119.

[4] C. Van Petten, "Examining the N400 semantic context effect item-by-item: Relationship to corpus-based measures of word co-occurrence", *International Journal of Psychophysiology*, vol. 94, no. 3, pp. 407–419, Dec. 2014.

[5] J. Geuze, M. A. J. van Gerven, J. Farquhar, and P. Desain, "Detecting Semantic Priming at the Single-Trial Level", *PLoS ONE*, vol. 8, no. 4, e60377, Apr. 2013. (visited on 10/06/2014).

[6] K. Dijkstra, J. Farquhar, and P. Desain, "Semantic Probing: Feasibility of using sequential probes to decode what is on a user's mind", [arXiv preprint doi:10.1101/496844, December 2018], Dec. 2018.

[7] D. Cruse, S. Beukema, S. Chennu, J. G. Malins, A. M. Owen, and K. McRae, "The reliability of the N400 in single subjects: Implications for patients with disorders of consciousness", *NeuroImage: Clinical*, vol. 4, pp. 788–799, Jan. 2014.

[8] M. Davies, "The Corpus of Contemporary American English (COCA): 560 million words", 1990-present, 2008. [Online]. Available: `Availableonlineathttps://corpus.byu.edu/coca/`.

[9] K. Lemhöfer and M. Broersma, "Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English", en, *Behavior Research Methods*, vol. 44, no. 2, pp. 325–343, Jun. 2012.

[10] G. Gratton, "Dealing with artifacts: The EOG contamination of the event-related brain potential", en, *Behavior Research Methods, Instruments, & Computers*, vol. 30, no. 1, pp. 44–53, Mar. 1998.

[11] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier, "Spherical splines for scalp potential and current density mapping", *Electroencephalography and Clinical Neurophysiology*, vol. 72, no. 2, pp. 184–187, Feb. 1989.

[12] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG- and MEG-data", *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, Aug. 2007.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-17

TIMEFLUX: AN OPEN-SOURCE FRAMEWORK FOR THE ACQUISITION AND NEAR REAL-TIME PROCESSING OF SIGNAL STREAMS

P. Clisson[1], R. Bertrand-Lalo[2], M. Congedo[3], G. Victor-Thomas[2], J. Chatel-Goldman[2]

[1]Independent researcher, Paris, France
[2]Open Mind Innovation, Paris, France
[3]GIPSA-lab, University Grenoble Alpes, CNRS, Grenoble-INP, Grenoble, France

E-mail: pierre@clisson.net

ABSTRACT: *Timeflux* is an open-source framework for the acquisition and near real-time processing of signal streams. It can be used to build any kind of brain-computer interface and neurofeedback application, allowing high performance for both quick prototyping and final product development. *Timeflux* has been designed to be lightweight, easy to use, scalable, and extensible. It equally suits researchers, companies and neurotechnology enthusiasts. *Timeflux* is developed in Python and released under the permissive and flexible MIT license. It is available at https://timeflux.io.

INTRODUCTION

Research on brain-computer interfaces (BCIs) has been carried out for about half a century, but has been experiencing momentum only recently [1]. It is now attracting researchers—not only in neurosciences, but also in diverse disciplines such as engineering and informatics—as well as private companies and individuals interested in exploring the frontiers of information technology such as hackers [2]. For a long time, advances in BCIs have been hindered by a lack of standardization of data acquisition protocols, data format and data processing. The poor availability of large public databases is another factor that has slowed progress [3]. In Europe, the necessity of introducing standards and sharing data has been pointed out by several consortia of European projects. For instance, the integrative project BNCI Horizon 2020 has created and promoted the first comprehensive repository of public BCI databases [4]. Coupling with this, the Mother of All BCI Benchmarks (MOABB) [5] allows to test the accuracy of BCI decoders on all available databases in an objective and easily reproducible way (*e.g.*, see [6] in these proceedings).

Another factor slowing the progress in BCI research is the lack of a *standard framework for online data acquisition and processing*, that is, the software at the very heart of actual BCI systems. For such a framework several characteristics are desirable: it should be easy-to-use, lightweight, efficient, scalable and extensible. Moreover, it should encourage exploration, allowing quick prototyping and testing as well as facilitating the integration with relevant existing code libraries and external tools, without sacrificing adequacy with the development of reliable products for the public. In this way, it may suit all actors of BCI research and development, both within academic institutions and private companies, allowing global collaboration.

Several frameworks for acquisition and real-time processing of signal streams have been developed (Tab. 1), however none of them satisfy all these criteria. Existing frameworks are either domain-specific—that is, they apply only to specific neuroimaging modalities— or are difficult to extend, or else are made available under a restrictive license that does not favor at the same time free and commercial development. For these reasons we hereby present a new framework, named *Timeflux*.

Timeflux has been developed in Python, which has become the programming language of choice in the data science and machine learning communities, thus already offering a wide panel of code libraries. Python is an extensible and cross-platform language, characteristics that have proved essential for its massive adoption. *Timeflux* is open-source and is released under the MIT license, which is very flexible. It is not specific to neuroimaging or bioelectrical signals, thus it can be used in the more general context of IoT, as well as in geoscience, control engineering, algorithmic trading, and more.

DESIGN PRINCIPLES

Based on the considerations above, the following set of rules has guided the development of *Timeflux*:

Simplicity of use: *The framework should be easy to learn, to use, and to extend.* This implies a clear documentation, a simple descriptive syntax for pipelines, and a minimal effort requirement for writing new plugins.

Lightness: *In order to ensure stability and frictionless maintenance, the core should maintain a small footprint.* Only the essential features belong to the core, everything else is moved to plugins. This leads to a codebase that is easy to apprehend.

Agnosticism: *The user should not be locked into a specific paradigm or set of tools.* From acquisition through processing to recording, every step is configurable and replaceable.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-17

Extensibility: *The architecture should be based on plugins.* This allows custom node development, integration with specialized libraries and hardware, and encourages external contributions without depending on base code merges. In *Timeflux*, plugins are ordinary Python modules, with only one required method to implement.

Reusability: *Resources should not be wasted by rewriting common algorithms and structures. Timeflux* relies on industry standards such as SciPy [7] and Numpy [8] for scientific computing, Pandas [9] for tabular data, Xarray [10] for multidimensional data, Scikit-learn [11] for machine learning, and NetworkX [12] for graph processing. These tools are well maintained, and many scientists are already familiar with them.

Scalability: *Parallelism and concurrency should be automatically enforced when applicable. Timeflux* distributes computing across CPU cores, threads, and even hosts.

Efficiency: *High-level interfacing should not sacrifice performance.* Sensible choices and careful memory management allow near real-time processing.

THEORY OF OPERATION

Timeflux is a framework to create pipelines, called *applications*. It does not matter if they are mere acquisition units, signal processing prototypes or complex BCI and biofeedback solutions. Applications are defined by a set of processing steps, called *nodes*, which are linked together using a simple YAML syntax.

In order to be valid, an application must satisfy the requirements of a directed acyclic graph (DAG) [13, 14], that is, a set of nodes connected by edges, where information flows in a given direction, with no internal loop (Fig. 1). Multiple DAGs are authorized within the same application, optionally communicating with each other using one of the available network protocols. DAGs run simultaneously at their own adjustable rate. Within each DAG, nodes are executed sequentially according to the topological sorting of the graph [15].

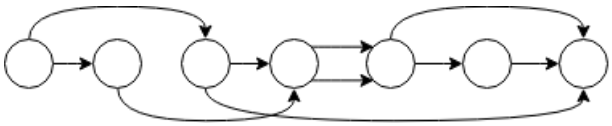


Figure 1: An example directed acyclic graph (DAG), arranged in topological order. Circles represent *nodes*, arrows are *edges*. The connection points between nodes and edges are called *ports*. Information flows from left to right, at a frequency defined by the graph rate.

Nodes may expect one or more inputs and may provide any number of outputs. These I/O, called *ports*, have two main properties. The *data* property is either a datetime-indexed Pandas DataFrame, or an Xarray structure with at least two dimensions: *time* and *space*. The *meta* property is a dictionary containing arbitrary keys, which can be used for example to declare a stream rate or to describe the context associated with an event.

AVAILABLE NODES

Timeflux comes with a growing collection of nodes, either available in core or as plugins.

LSL: The Lab Streaming Layer [16] is a transport and synchronization library compatible with a large range of EEG equipment. *Timeflux* is able to handle both input and output streams.

Publish/Subscribe: The publish/subscribe pattern allows asynchronous messaging between DAGs and/or between external components. Subscribers express interest in topics, and receive data matching these topics. There can be more than one publisher per topic. This protocol is implemented using the ZeroMQ library [17].

OSC: The Open Sound Control protocol [18] is commonly used in media applications. It is useful to create rich interactive environments.

Epoching and windowing: Several nodes to extract epochs or to accumulate data from streams are included.

HDF5: The Hierarchical Data Format [19] is a stable and powerful data storage and query solution. Nodes for recording and playback are incorporated.

Queries and expressions: A small set of nodes are available to extract data matching specific criteria and to execute arbitrary arithmetic operations.

Web User Interface: Visualizing data and sending events is possible directly within a browser. Fig. 2 is a screenshot of a typical session. The current implementation uses the HTML5 canvas object, which runs in the main thread. A new version taking advantage of the WebGL technology is underway. Preliminary tests show that over a million points can be plotted per second.

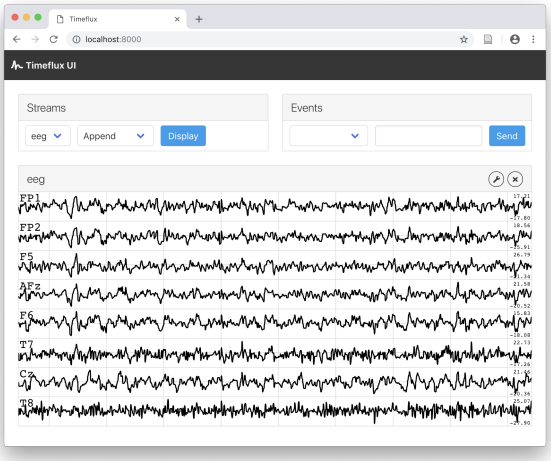


Figure 2: Data stream visualization in a web browser (electroencephalographic data in this example).

Digital Signal Processing: Common signal processing algorithms are already available (for instance, filtering, FFT, covariance matrix estimation, etc.).

Dejittering: A signal periodicity can deviate from its nominal rate, or data may be acquired in chunks. Several mitigating methods are provided.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-17

Machine Learning: Models can be fitted on epoched data, and then used to classify or transform new data. This is made possible by the scikit-learn library.

Branches: The same processing pipelines are often reused from project to project. *Timeflux* will provide a method to define sub-graphs as simple nodes. At the time of writing, this is still a work in progress.

Device drivers: Plugins for direct integration with popular open-source hardware, such as OpenBCI [20] and BITalino [21], as well as proprietary hardware, have been developed and more will be released in the upcoming months.

PERFORMANCES

Good practices: Python is a high-level programming language. While this offers several advantages, it also implies some trade-offs in terms of efficiency. This issue is mitigated in several ways. Special attention is paid to avoid memory copy unless absolutely required. When dealing with numpy-based structures, vectorized functions (written in C) are used whenever possible. At the very least, the Pandas `apply()` function benefits from low-level optimization. These simple measures help maintaining good performances for most applications. In extreme cases, it is possible to speed execution time by rewriting a node's critical functions in Cython [22] or by using Numba [23].

Architecture: *Timeflux* takes advantage of modern CPU architecture and distributes the execution of DAGs across cores. Nodes that require an infinite loop (*i.e.*, device drivers) run parts of their code into a separate thread, so the whole graph is not penalized. Demanding applications can further be optimized by running synchronized *Timeflux* instances on multiple computers.

Real-time: Hard real-time is when missed deadlines are unacceptable and result in system failure [24]. In most cases, hard real-time is not required, and soft real-time (also known as near real-time) is sufficient. *Timeflux* allows a few tens of milliseconds latency (the latter being modulated by the graph rate), as long as the incoming data is timestamped at the point of origin. Time offset correction is achieved either by using a dedicated communication protocol such as LSL, or by repeatedly synchronizing *Timeflux* instances using an algorithm similar to the Network Time Protocol [25], thus ensuring sub-millisecond precision.

EXAMPLE

As an example we illustrate a simple alpha-neurofeedback application, which is schematically represented in Fig. 3. The application consists of three graphs. In the main one, we assume that the EEG data is acquired through a LSL inlet. Data is accumulated into a rolling window, on which the classical Welch's method is applied with default parameters [26]. The frequency bands are then extracted from the periodogram. Finally,

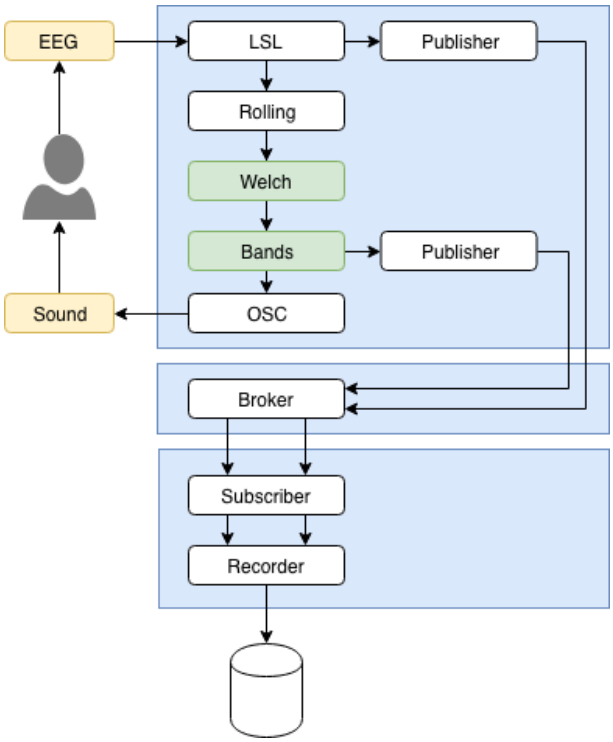


Figure 3: Schematic representation of a basic neurofeedback application. The three blue boxes constitute the *Timeflux* application. The white boxes are core nodes. The green boxes are plugin nodes. Yellow boxes indicate external components.

the relative alpha power is sent to an OSC outlet. An external application receives this data and plays a sound when the feedback signal crosses a defined threshold. The other two graphs are not strictly required, but illustrate some important principles. The graph containing the `Broker` node acts as a proxy. It receives data from publishers (in our example, the raw EEG stream and the computed frequency bands) and redistributes it to subscribers. In the last graph, these two data streams are aggregated and saved to a HDF5 file.
The whole application is expressed in YAML as follows:

```yaml
graphs:

  # The publish/subscribe broker graph
  - id: PubSubBroker
    nodes:
    # Allow communication between graphs
    - id: Broker
      module: timeflux.nodes.zmq
      class: Broker

  # The main processing graph
  - id: Processing
    nodes:
    # Receive EEG signal from the network
    - id: LSL
      module: timeflux.nodes.lsl
      class: Receive
      params:
        name: signal
    # Continuously buffer the signal
    - id: Rolling
      module: timeflux.nodes.window
      class: Window
      params:
        length: 1.5
        step: 0.5
    # Compute the power spectral density
    - id: Welch
```


Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-17

```
    module: timeflux_dsp.nodes.spectral
    class: Welch
  # Average the power over band frequencies
  - id: Bands
    module: timeflux_dsp.nodes.spectral
    class: Bands
  # Send to an external application
  - id: OSC
    module: timeflux.nodes.osc
    class: Client
    params:
      address: /alpha
  # Publish the raw EEG signal
  - id: PublisherRaw
    module: timeflux.nodes.zmq
    class: Pub
    params:
      topic: raw
  # Publish the frequency bands
  - id: PublisherBands
    module: timeflux.nodes.zmq
    class: Pub
    params:
      topic: bands
  # Connect nodes
  edges:
    - source: LSL
      target: Rolling
    - source: Rolling
      target: Welch
    - source: Welch
      target: Bands
    - source: Bands:alpha
      target: OSC
    - source: LSL
      target: PublisherRaw
    - source: Bands
      target: PublisherBands
  # Run this graph 25 times per second
  rate: 25

# The recorder graph
- id: SaveToFile
  nodes:
  # Receive data streams from the broker
  - id: Subscriber
    module: timeflux.nodes.zmq
    class: Sub
    params:
      topics:
      - raw
      - bands
  # Record to file
  - id: Recorder
    module: timeflux.nodes.hdf5
    class: Save
  # Connect nodes
  edges:
    - source: Subscriber:raw
      target: Recorder:eeg_raw
    - source: Subscriber:bands
      target: Recorder:eeg_bands
  # Update file every second
  rate: 1
```

COMPARISON WITH SIMILAR SOFTWARE

Several frameworks are currently available for the online analysis of bioelectrical signals (Tab. 1). Choosing one is a difficult task since it often requires a balance among a variety of criteria, such as desired characteristics, performances, extensibility, license, and ease of use. The characteristics are generally oriented towards specific applications (*e.g.*, BCI using electroencephalography) thus only a few frameworks are versatile enough to be used with heterogeneous time series and in more general contexts. The performance is hard to evaluate without a comprehensive benchmark using comparable objective metrics and a thorough knowledge of the software inner functioning. Regarding extensibility, all open-source solutions are extensible by nature, but the effort required to do so varies according to the architecture and coding experience. Concerning the license, while commercial licensing may incur financial costs and vendor lock-ins, GPL-derived licenses can in some cases also be treacherous since they require derivative works to be distributed under the same copyleft license. The MIT and BSD licenses—and to some extent, the Apache licence—are much more permissive instead. *In fine*, much of these criteria are intrinsically subjective or difficult to evaluate objectively.

Timeflux is a modern framework for real-time processing of signal streams. According to the aforementioned principles we have followed for development, it brings a well-balanced set of features, delivered under a very flexible license.

DISCUSSION

Although *Timeflux* is mature enough to be used in production, there is still room for improvement. Our roadmap foresees significant updates in the following areas:

Performances: Extensive benchmarks will be systematized to identify bottlenecks.

Scalability: For large projects involving many instances spreading over multiple hosts, mechanisms allowing automatic discovery, load-balancing, fail-over, and seamless synchronization (possibly using the Precision Time Protocol [50]), will be implemented.

General usage: We aim for a comprehensive documentation and better development tools. We also intend to provide turnkey BCI paradigms to accelerate prototyping.

User Interface: The web interface is currently being rewritten to significantly increase the number of points that can be plotted per second, and to introduce specialized widgets. In the future, it will be possible to design pipelines directly from a browser. A JavaScript API for stimulation presentation will also be available.

Integration: Nodes are being developed to integrate with more data acquisition devices, stimulation presentation software, and alternative storage solutions.

CONCLUSION

We have presented *Timeflux*, an open, highly flexible and actively developed solution that is meant to accelerate the creation of applications and standardize them. It is naturally suitable for BCI and biofeedback applications.

ACKNOWLEDGEMENT

This project has been supported by Open Mind Innovation.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-17

Table 1: Comparable frameworks

Name	Release year	Language	License	Main domain
BCI++ [27]	2008	C/C++	GPL	Closed loop neuroscience
BCI2000 [28]	2001	C++	GPL	Closed loop neuroscience
BCILAB [29, 30]	2010	Matlab	GPLv2	Closed loop neuroscience
BioEra [31]	2004	Java	Commercial	Closed loop neuroscience
BioSig [32, 33]	2003	C++	GPLv3	Closed loop neuroscience
BrainBay [34]	2014	Python	GPL	Closed loop neuroscience
CloudBrain [35]	2007	Python	AGPLv3	Acquisition
Falcon [36]	2017	C++	GPLv3	Closed loop neuroscience
Fieldtrip [37]	2003	Matlab	GPLv2	Closed loop neuroscience
Gumpy [38, 39]	2018	Python	MIT	Closed loop neuroscience
Midas [40]	2014	Python	MIT	Barebone
Neuromore [41]	2015	C++	Commercial	Closed loop neuroscience
Neuropype [42]	2014	Python	Commercial	Closed loop neuroscience
Nipype [43]	2010	Python	Apache	Neuroimaging
Open Ephys GUI [44, 45]	2011	C++	GPLv3	Extracellular electrophysiology
OpenVibe [46, 47]	2009	C++	AGPLv3	Closed loop neuroscience
PyAcq [48]	2015	Python	BSD 3-clause	Acquisition
Timeflux [49]	2019	Python	MIT	Generic time series

REFERENCES

[1] Gartner. *Gartner identifies five emerging technology trends that will blur the lines between human and machine*. 2018. URL: https://gtnr.it/2N2X0QE (visited on 02/01/2019).

[2] NeuroTechX. *The international community for neurotech enthusiasts*. URL: https://neurotechx.com (visited on 02/01/2019).

[3] I. Obeid and J. Picone. "Bringing Big Data to neural Interfaces". In: *Proc. Fifth Int. BCI Meeting*. Pacific Grove (CA), USA, 2013, p. 180.

[4] B. H. 2020. *Databases*. URL: http://bnci-horizon-2020.eu/database/data-sets (visited on 02/08/2019).

[5] V. Jayaram and A. Barachant. "MOABB: trustworthy algorithm benchmarking for BCIs". In: *Journal of Neural Engineering* 15.6 (2018), p. 066011.

[6] M. Congedo, P. Rodrigues, and C. Jutten. "The Riemannian Minimum Distance to Means field Classifier". In: *Proc. 8th Int. Graz BCI Conference*. Ed. by TUG. Graz, Austria, 2019.

[7] E. Jones, T. Oliphant, P. Peterson, et al. *SciPy: Open source scientific tools for Python*. 2001–. URL: https://www.scipy.org (visited on 02/01/2019).

[8] T. Oliphant. *Guide to NumPy*. 2nd. USA: CreateSpace Independent Publishing Platform, 2015.

[9] W. McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by S. van der Walt and J. Millman. 2010, pp. 51–56.

[10] S. Hoyer and J. J. Hamman. "xarray: N-D labeled Arrays and Datasets in Python". In: *Journal of Open Research Software* 5 (2017).

[11] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12 (Nov. 2011), pp. 2825–2830.

[12] A. A. Hagberg, D. A. Schult, and P. J. Swart. "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by G. Varoquaux, T. Vaught, and J. Millman. Pasadena, CA USA, 2008, pp. 11 –15.

[13] N. Christofides. *Graph Theory: An Algorithmic Approach (Computer Science and Applied Mathematics)*. Orlando, FL, USA: Academic Press, Inc., 1975.

[14] K. Thulasiraman and M. N. S. Swamy. *Graphs: theory and algorithms*. New York: Wiley, 1992. Chap. 5.7 Acyclic Directed Graphs.

[15] U. Manber. *Introduction to Algorithms: A Creative Approach*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1989. Chap. 7.4 Topological Sorting.

[16] C. Kothe. *Lab Streaming Layer*. 2013. URL: https://github.com/sccn/labstreaminglayer (visited on 02/01/2019).

[17] *Distributed Messaging*. URL: http://zeromq.org/ (visited on 02/01/2019).

[18] M. Wright. *Open Sound Control: an Enabling Encoding for Media Applications*. URL: http://opensoundcontrol.org (visited on 02/01/2019).

[19] T. H. Group. *Hierarchical Data Format*. URL: https://www.hdfgroup.org/ (visited on 02/01/2019).

[20] *Open Source Biosensing Tools (EEG, EMG, EKG, and more)*. URL: https://openbci.com/ (visited on 02/01/2019).

Published by Verlag der TU Graz
Graz University of Technology

[21] *BITalino*. URL: https://bitalino.com/ (visited on 02/01/2019).

[22] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. Seljebotn, and K. Smith. "Cython: The Best of Both Worlds". In: *Computing in Science Engineering* 13.2 (2011), pp. 31 –39.

[23] S. K. Lam, A. Pitrou, and S. Seibert. "Numba: A LLVM-based Python JIT Compiler". In: *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. LLVM '15. ACM, 2015, 7:1–7:6.

[24] M. Ben-Ari. *Principles of concurrent and distributed programming*. Prentice Hall international series in computer science. Prentice Hall, 1990. Chap. 16, p. 164.

[25] J. Martin, J. Burbank, W. Kasch, and P. D. L. Mills. *Network Time Protocol Version 4: Protocol and Algorithms Specification*. RFC 5905. June 2010.

[26] P. Welch. "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (1967), 70–73.

[27] L. Maggi, S. Parini, P. Perego, and G. Andreoni. "BCI++: an object-oriented BCI Prototyping Framework". In: *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course* (Jan. 2008).

[28] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw. "BCI2000: a general-purpose brain-computer interface (BCI) system". In: *IEEE Transactions on Biomedical Engineering* 51.6 (2004), 1034–1043.

[29] C. A. Kothe and S. Makeig. "BCILAB: a platform for brain–computer interface development". In: *Journal of Neural Engineering* 10.5 (2013), p. 056014.

[30] T. S. C. for Computational Neuroscience (SCCN. *Open Source Matlab Toolbox for Brain-Computer Interface research.* URL: https://sccn.ucsd.edu/wiki/BCILAB (visited on 02/01/2019).

[31] BioEra. *Welcome to the Home of Bioera!* URL: http://www.bioera.net (visited on 02/01/2019).

[32] C. Vidaurre, T. H. Sander, and A. Schlögl. "BioSig: The Free and Open Source Software Library for Biomedical Signal Processing". In: *Computational Intelligence and Neuroscience* 2011 (2011), 1–12.

[33] BioSig. *The BioSig Project*. URL: http://biosig.sourceforge.net (visited on 02/01/2019).

[34] BrainBay. *BrainBay - an OpenSource Biosignal project*. URL: http://www.shifz.org/brainbay (visited on 02/01/2019).

[35] CloudBrain. *Open-source platform for wearable data analytics.* URL: http://getcloudbrain.com (visited on 02/01/2019).

[36] D. Ciliberti and F. Kloosterman. "Falcon: a highly flexible open-source software for closed-loop neuroscience". In: *Journal of Neural Engineering* 14.4 (2017), p. 045004.

[37] FieldTrip. *Welcome to the FieldTrip website*. URL: http://www.fieldtriptoolbox.org (visited on 02/01/2019).

[38] Z. Tayeb et al. "Gumpy: a Python toolbox suitable for hybrid brain–computer interfaces". In: *Journal of Neural Engineering* 15.6 (2018), p. 065003.

[39] Z. Tayeb et al. *Gumpy: A Toolbox Suitable for Hybrid Brain-Computer Interfaces*. URL: http://gumpy.org (visited on 02/01/2019).

[40] A. Henelius and J. Torniainen. "MIDAS: Open-source framework for distributed online analysis of data streams". In: *SoftwareX* 7 (2018), 156–161.

[41] Neuromore. *Neuromore Studio*. URL: https://www.neuromore.com (visited on 02/01/2019).

[42] Intheon. *NeuroPype: Advanced biosignal processing made simple*. URL: https://www.neuropype.io (visited on 02/01/2019).

[43] Nypipe. *Neuroimaging in Python: Pipelines and Interfaces*. URL: https://nipype.readthedocs.io (visited on 02/01/2019).

[44] J. H. Siegle, A. C. López, Y. A. Patel, K. Abramov, S. Ohayon, and J. Voigts. "Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology". In: *Journal of Neural Engineering* 14.4 (2017), p. 045003.

[45] OpenEphys. *open-source electrophysiology*. URL: http://www.open-ephys.org (visited on 02/01/2019).

[46] Y. Renard et al. "OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments". In: *Presence* 19.1 (2010), 35–53.

[47] INRIA. *OpenVIBE Software for Brain Computer Interfaces and Real Time Neurosciences*. URL: http://openvibe.inria.fr (visited on 02/01/2019).

[48] Pyacq. *Pyacq*. URL: https://github.com/pyacq (visited on 02/01/2019).

[49] P. Clisson et al. *Timeflux*. 2018–. URL: https://timeflux.io.

[50] IEEE Standards Association. *IEEE 1588-2008 - IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*. URL: https://standards.ieee.org/standard/1588-2008.html (visited on 02/01/2019).

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-18

INTRACORTICAL ACTIVITY DECODING OF MOTOR IMAGERY BASED ON DEEP CONVOLUTIONAL NEURAL NETWORK: A PILOT STUDY

Duo Chen[1], Rosa So[2], Yi Ding[1], Neethu Robinson[1], Cuntai Guan[1]

[1]School of Computer Science and Engineering, Nanyang Technological University, Singapore
[2]Institute for Infocomm Research, A*STAR, Singapore

E-mail: chenduo@ntu.edu.sg (Duo Chen)

ABSTRACT: An intracortical activity decoding algorithm based on convolutional neural network (CNN) is proposed for classifying the primate intracortical activity of motor imagery under a series of movement tasks. An intracortical brain-computer interface allows the subjects to continuously drive a mobile robot using its brain activity from the motor cortex. Under the leave-one-day-out cross validation, the CNN-based algorithm achieves reliable classification performance in all the four movement tasks, including, Forward, Left, Right, and Stop. Experimental results indicate that the Cohen's kappa of the two subjects are both above 0.4. Additionally, the comparison shows that the CNN-based method is significantly better than the linear discriminant analysis (LDA) which is a state-of-the-art decoding approach.

INTRODUCTION

Brain–computer interface (BCI), which can provide a new communication channel to humans, has received increasing attention in recent years [5]. BCI allows one to control external devices through direct brain activity recognition by a computer [1]. The use of invasive electrocorticographic (ECoG) signals for BCI applications draws numerous interests recently [13]. ECoG directly records signals originating from brain tissues beneath the electrode surface, thus it has much higher spatial resolution compared with EEG [18].

Hence, the motor imagery based BCI system is based on the fact that there will be a change of activation in certain areas of the brain when a subject imagines movement any part of their body [9]. For example, when a person imagines moving his/her right arm, there will be a desynchronization of neural activities in the primary motor cortex of the left brain [12].

In recent years, researchers have increasingly focused on BCI systems for different practical areas, e.g., brain-wave controlled robot [6]. Multi-array electrodes implanted in motor image area can record stable neural activity for achieving various kinds of objectives in BCI system [8]. With invasive or noninvasive recording method, the BCI uses the neural activity of the brain to control effectors such as robotic arm or wheel chair. The core components of a BCI system are brain signal acquisition, preprocessing, feature extraction, classification, translation and feedback control of external devices [7].

Convolutional neural network (CNN) is a multilayer perceptron with a special topology and contains several hidden layers [12]. CNN has been employed to solve certain computer vision problem, e.g., medical image processing [15]. There is also increasing interest in using deep CNN for end-to-end neural signal analysis. Brain activity decoding can be improved by using batch normalization, dropout and exponential linear units implies which are generally used in deep learning [14].

In this work, we proposed a CNN-based decoding algorithm classify the intracortical activities in the motor cortex of primate under a series of movement tasks, including, "Forward", "Left", "Right", and "Stop". With the robust algorithm, the monkey can control the robot in 360 degree using its brain signal. The decoding problem is transferred into a 4-labels classification problem for annotating long-term activity automatically. Since the cortical activity may vary among subjects and the day-to-day variation may affect the algorithm performance [2], we constructed a robust subject-specific movement decoding algorithm based on CNN using the leave-one-day-out strategy.

METHOD

Data acquisition: The dataset used in this work was collected at A-Star I2R, Singapore, consists of intracortical activities from two healthy monkeys (*Macaca Fascicularis*) [6]. Figure 1 demonstrates the experimental procedure. A total of 96-channel multielectrode array was implanted into the hand/arm region of the left primary motor cortex. Then, the monkeys were first trained to control the platform using a joystick. Specific cues with food stimuli were then given, ordering the monkeys to move toward a certain direction, which is "Forward", "Left", "Right", or "Stop". If the monkey was successful in the trial, the food would be given as a reward. When the monkey can handle the robot skillfully to reach the given targets, the joystick would be disconnected from the platform. Since the joystick was disconnected, the monkey can only control the robotic platform using its intracortical activities. The recorded intracortical activities were transferred to the computational center for fur-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-18

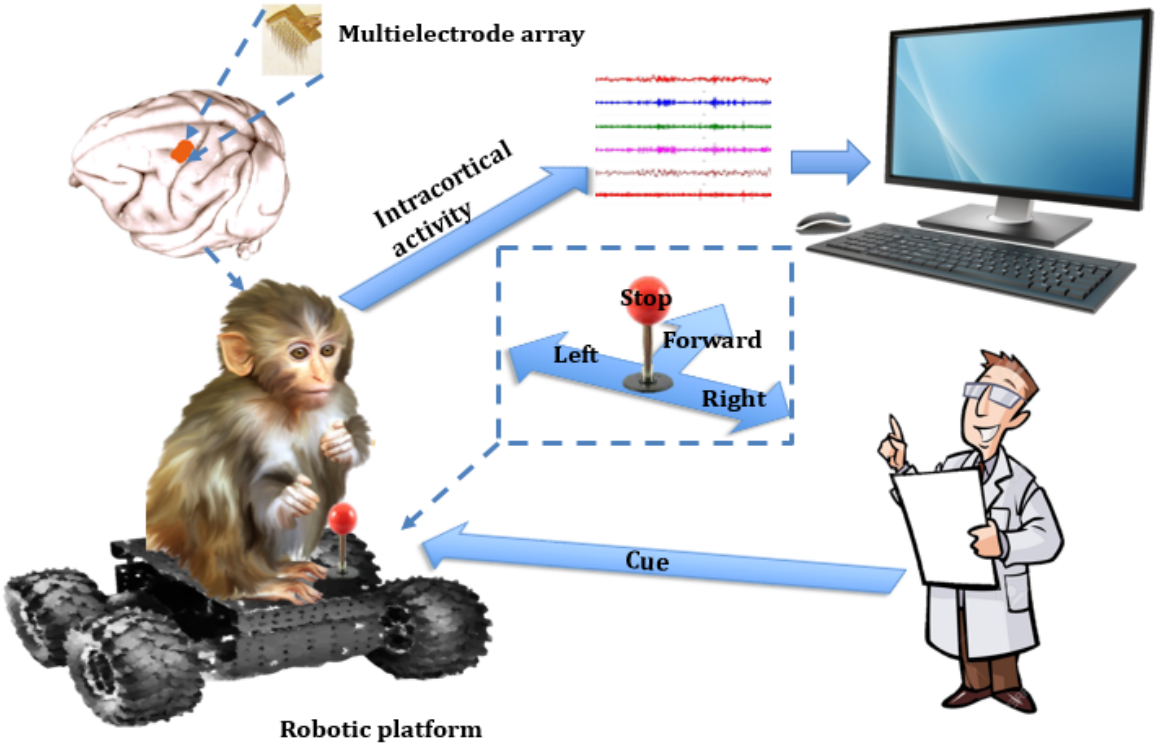


Figure 1: **Experiment Setup.** The implanted array in the hand/arm region records the intracortical activity during the movement tasks. If the monkey moves in the direction correctly, a food reward will be given. The disconnected joystick is treated as the ground truth of the moving willing from the monkey.

ther analysis. Only the data of joystick disconnected were used in the classification problem.

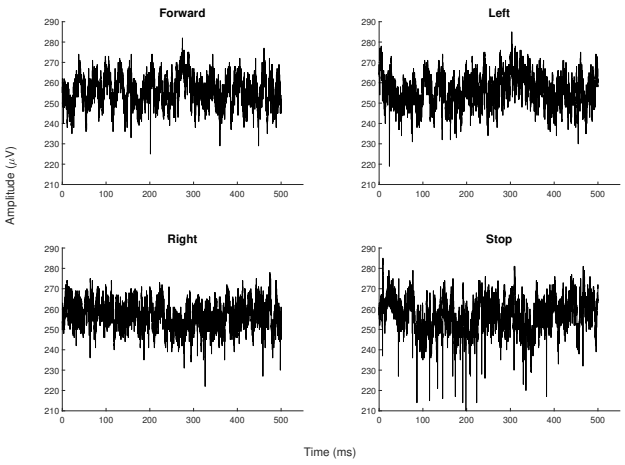


Figure 2: **Examples of the intracortical activities belong to the four movement tasks.**

The intracortical activities were sampled at 12987Hz. Monkey A provided 6 days' data with 1-3 sessions in each day over 10 weeks, depending on how well the monkey cooperated with the trainers. Similarly, Monkey B provided 9 days' data with 1-3 sessions in each day over 10 weeks. Each session lasted about 350 seconds containing 14-16 moving trials. Figure 2 demonstrates an example of the intracortical activity segment for each of the four movement tasks. Each subfigure is a 500ms single-channel intracortical activity segment belongs to a certain movement task. For space sake, only one channel is selected for the illustration.

Preprocessing: The raw data were first filtered through two FIR filters. From a biophysics perspective, much of the cortical activity is known as the origins of the local field potential (LFP). In this work, the intracortical activity in a low frequency band is treated as the LFP for the feature extraction.

High-frequency signals have been proved as putative biomarkers of certain neurological disorder. In BCI system, firing rate of the high-frequency brain wave is usually used as an effective marker for decoding the neural activity. The intracortical activity in a high frequency band is used for extracting the firing rate as a component in the feature matrix.

The features from LFP and high-frequency band are extracted separately and combine together to construct the feature matrix. For the CNN-based classification, the feature matrix is used as the 2D input.

To extract the features from LFP, a 2048-order FIR filter was used with the pass band at $[0.1, 128]Hz$. To extract the features from high frequency band, a 256-order FIR filter was used with the pass band at $[300, 3000]Hz$.

After filtering the signal into low and high frequency bands, the intracortical activities were divided into multi-channel segments using a sliding window. The segmenta-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-18

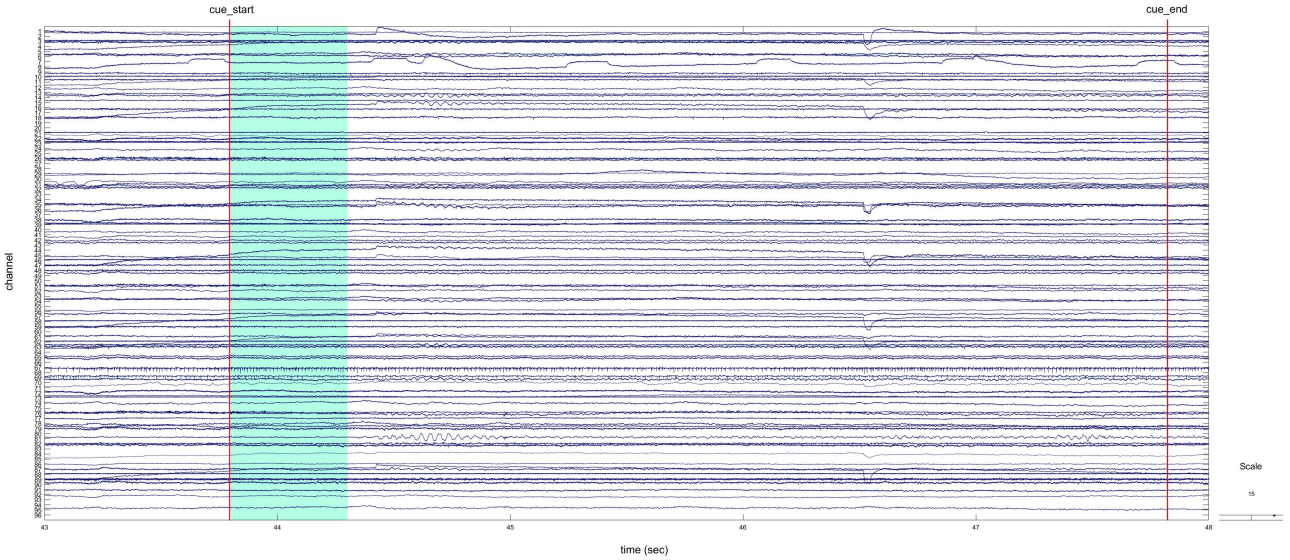


Figure 3: **Segments extracted from long-term multi-channel intracortical activities.** The illustration belongs to a "Right" trial. The time 43.79s is the start cue (cue_start), while the time 49.47s is the end cue of the same trial. The green block is a 500ms slide window with 400ms overlap. In each step, the slide window will move 100ms while the 96-channel signal inside will be treated as one intracortical activity segment.

tion process of one "Right" trial is illustrated in Figure 3. The time 43.79s is the start cue of a right movement task (cue_start), and the time 49.47s is the end cue of the same trial. The green block is a 500ms slide window with 400ms overlap. In each step, the slide window moves 100ms while the 96-channel signal inside is chopped as one cortical activity segment being labelled according to the movement task. With the same sliding window, all the trials of the four movement tasks are transferred into multi-channel intracortical activity segments.

Problem formulation: The purpose of this study is identifying intracortical activity in motor area triggered by various movement tasks, including, "Forward", "Left", "Right", and "Stop".

After the preprocessing, the research problem is transferred into a 4-label classification problem. Algorithm 1 illustrates the process of the intracortical activity decoding.

In the feature extraction, features based on wavelet transform and event-related desynchronization (ERD)/synchronization (ERS) are extracted from LFP while the firing rate is extracted from the high-frequency band. Then, the wavelet feature, ERD/ERS, and firing rate construct the feature matrix.

In the leave-one-day-out cross validation, the feature matrix is treated as a 2D input of the CNN.

Details of feature extraction and CNN-based classification will be given in the following subsections.

Wavelet feature: As a powerful time-frequency analysis toolbox, DWT decomposes a signal into a series of coefficients and features can be extracted from the coefficients to represent the properties of the signal. "sym2" is employed as the mother wavelet here. The decomposition level is set to 5 producing 5 detail bands and 1 approximation band. For each band, 9 features as listed below are extracted to construct the wavelet component in the feature matrix.

1. Max: the maximum coefficient
2. Min: the minimum coefficient
3. Mean: the average of coefficients
4. STD: the standard deviation of coefficients
5. Skewness: the skewness of coefficients
6. Kurtosis: the Kurtosis of coefficients
7. Energy: the squared sum of coefficients
8. nSTD: normalized STD, $\frac{STD}{Max-Min}$
9. nEnergy: normalized energy, $\frac{Energy}{band\ length}$

Firing rate: The signals in $[300, 3000]Hz$ are used for spike detection. Spikes are detected using an automated threshold-crossing criterion selected for each channel. The threshold (Thr) for spike detection follows the formula [11]:

$$Thr = 4\delta_n; \delta_n = median\{\frac{|x|}{0.6745}\} \quad (1)$$

, where x is the filtered signal, and δ is an estimate of the standard deviation of the background noise.

The spike detection are operated in the negative quadrants. Amplitude lower than $-Thr$ will be marked as a spike. For each channel, the number of spikes is recorded as the firing rate in the corresponding channel position.

ERD/ERS: An internally or externally paced event results not only in the generation of an event-related potential (ERP) but also in a change in the ongoing EEG/MEG

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-18

Algorithm 1 Pseudocode for CNN-based intracortical activity decoding

Input: Activities of n Days
Output: Cohen's kappa. precision, recall, f1-score for each movement task

```
Initialization: Kappa = zeros((n,1)). precisionAll, recallAll, f1-scoreAll = zeros((n,4,3)), zeros((n,4,3)), zeros((n,4,3))
Read intracortical activity datasets
Filter data into low and high frequency bands
slide data in low and high frequency bands into 500 ms segments with 400ms overlap
Extract (wavelet, ERD/ERS, firing rate) features ▷ feature extraction
for day in (1 : n) do ▷ leave-one-day-out cross validation
    Set up X_train, Y_train, X_test, Y_test for the ith day
    batchSize = 128; inputRows, inputCols = 96, 56
    lossFun = "sparse_categorical_crossentropy"
    targetNames = ['Forward', 'Right', 'Left', 'Stop']
    model = Sequential()
    model.add(Conv2D(100, (3,3), strides = (1,1), input_shape = (inputRows, inputCols,1)))
    model.add(Activation(actiFun))
    for epoch in (1 : 10) do
        model.add(Conv2D(100, (3,3), strides = (1,1), padding = "same"))
    model.add(Activation("ReLU"))
    model.add(MaxPooling2D(pool_size = (2, 2)))
    model.add(Dropout(0.5))
    model.add(Flatten())
    model.add(Dense(56))
    model.add(Activation("ReLU"))
    model.add(Dropout(0.25))
    model.add(Dense(4))
    model.add(Activation("softmax"))
    model.compile(loss = lossFun, optimizer = "sgd", metrics = "accuracy")
    model.fit(X_train, Y_train, batch_size = batchSize, epochs = 50)
    Y_predict = model.predict_classes(X_test)
    Kappa[day] = cohen_kappa_score(Y_test, Y_predict)
    precisionAll[day], recallAll[day], f1-scoreAll[day] = classification_report(Y_test, Y_predict, target_names = targetNames, output_dict = True)
Cohen's kappa, precision, recall, f1-score = mean(Kappa), mean(precisionAll), mean(recallAll), mean(f1-scoreAll)
return Cohen's kappa, precision, recall, f1-score
```

in form of an event-related desynchronization (ERD) or event-related synchronization (ERS). In this work, ERD/ERS are extracted from the intracortical activity as one component in the feature matrix for the classification. The ERD/ERS follows the definition in [10] as

$$ERD/ERS = \frac{A-R}{R} \times 100\% \tag{2}$$

, where A is the power of signal within the frequency band of interest, R is the power of signal as the preceding baseline in the reference period. Here the reference signal is selected from 2.2 to 0.2 second before each movement cue.

CNN structure: Deep learning is a promising avenue for big data analysis. Some state-of-the-art deep learning algorithms, such as deep neural networks and deep CNN have been successfully applied to image classification [4]. Recently, referenced from previous work in computer vision, CNN has been increasingly employed for EEG signal classification [14]. Here we use the deep CNN as the classifier to decode the intracortical activities under the 4-label movement tasks.
Keras 2.2.4 (`https://keras.io/`) is used to construct the CNN network. Details of the network structure are shown in Algorithm 1.

Leave-one-day-out cross validation: Previous work proved that motor intracortical spiking activity under different movement tasks are easily-classified in one day. However, the problem turns into tricky when facing a inter-day timeline [17]. The mental state, physiological state, individual difference, etc, may affect the motor cortical activity of the subject [2]. In the real practice, a re-calibration for classification on each day will be very time-consuming. Even if the re-calibration can guarantee the prediction performance, ignoring the pre-recorded signals will be quite a waste, abandoning valuable patterns may exist across days.
Considering the individual and the day-to-day differences, the subject-specific leave-one-day-out cross validation is employed for both monkeys. In each step of the cross validation, the feature matrix of one day is treated as the test set, while all the left days construct the training set.

Classification evaluation: Precision, recall, f1-score, and Cohen's kappa are employed for the classification evaluation. Accuracy is not employed here since in daily real action, it is impossible to make a perfect data balance of the 4 movement tasks. The majority of one or more classes may produce a fake high or low accuracy. In contrast, Cohen's kappa are much robust than simple percent agreement.
The definition of Cohen's kappa is:

$$k = \frac{p_o - p_e}{1 - p_e} \tag{3}$$

, where p_o is the empirical probability of agreement on the label assigned to any sample (the observed agreement ratio), and p_e is the expected agreement when both annotators assign labels randomly. p_e is estimated using a

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-18

Table 1: Decoding Performance of CNN on the Two Monkeys.

	Monkey A ($k = 0.4077 \pm 0.1863$)			Monkey B ($k = 0.4754 \pm 0.1297$)		
	precision	recall	f1-score	precision	recall	f1-score
Forward	0.5471 ± 0.1508	0.5017 ± 0.2814	0.4916 ± 0.1972	0.5732 ± 0.1776	0.4640 ± 0.3280	0.4227 ± 0.2016
Left	0.5401 ± 0.2095	0.4065 ± 0.2863	0.4235 ± 0.2162	0.7499 ± 0.2032	0.6719 ± 0.1852	0.6685 ± 0.1376
Right	0.5962 ± 0.2308	0.6832 ± 0.1445	0.5805 ± 0.0966	0.7236 ± 0.1544	$0.5442 \pm 0,2799$	0.5416 ± 0.2028
Stop	0.8271 ± 0.1564	0.5996 ± 0.3142	0.6025 ± 0.2524	0.7179 ± 0.1287	0.7287 ± 0.2235	0.6925 ± 0.1335
Average	0.6276 ± 0.1869	0.5477 ± 0.2566	0.5245 ± 0.1906	0.6912 ± 0.1660	0.6022 ± 0.2542	0.5813 ± 0.1689

Table 2: Decoding Performance of LDA on the Two Monkeys.

	Monkey A ($k = 0.3288 \pm 0.1374$)			Monkey B ($k = 0.4371 \pm 0.1281$)		
	precision	recall	f1-score	precision	recall	f1-score
Forward	0.3287 ± 0.1436	0.2596 ± 0.1803	0.2719 ± 0.1837	0.5632 ± 0.1372	0.3406 ± 0.2621	0.3646 ± 0.2177
Left	0.5830 ± 0.1743	0.5651 ± 0.3044	0.4780 ± 0.2045	0.7693 ± 0.1813	0.5895 ± 0.2846	0.5987 ± 0.2035
Right	0.6184 ± 0.1185	0.4666 ± 0.1869	0.5058 ± 0.1315	0.6614 ± 0.2130	0.5500 ± 0.2715	0.5032 ± 0.1952
Stop	0.6354 ± 0.1577	0.6732 ± 0.2363	0.5957 ± 0.1144	0.6264 ± 0.1178	0.8001 ± 0.2071	0.6751 ± 0.1044
Average	0.5414 ± 0.1485	0.4912 ± 0.2270	0.4629 ± 0.1585	0.6551 ± 0.1623	0.5700 ± 0.2538	0.5354 ± 0.1802

per-annotator empirical prior over the class labels [3]. If the raters are in complete agreement then $k = 1$. If there is no agreement among the raters other than what would be expected by chance, $k = 0$.

In this work, Cohen's kappa is calculated across days in the leave-one-day-out cross validation. For each subject, the performance of CNN in each class is evaluated under precision, recall, and f1-score. Taking class "Forward" for instance, suppose all "Forward" segments are treated as "positive", then, segments belong to "Left", "Right", and "Stop" are clustered into "negative". Therefore, the classifier had 4 possible outcomes: 1. True positive (TP); 2. False positive (FP); 3. True negative (TN); 4. False negative (FN). The definitions of precision, recall, and f1-score are as follows:

1. $f1 - score = \frac{2TP}{2TP+FP+FN}$;
2. $precision = \frac{TP}{TP+FP}$;
3. $recall = \frac{TP}{TP+FN}$;

RESULTS AND DISCUSSION

Here we exhibit the algorithm outcomes and briefly give some heuristics of using the CNN-based intracortical activity decoding algorithm.

Table 1 lists the algorithm performance on the two monkeys. Cohen's kappa is calculated across days in the leave-one-day-out cross validation. The $mean \pm std$ of Cohen's kappa is in the first line. Results indicate that the decoding algorithm produces Cohen's kappa above 0.4 on both monkeys leading to a moderate agreement between the true and the predicted labels of the test data set. To zoom in each of the four movement tasks, for each subject, the performance of CNN is evaluated under precision, recall, and f1- score. The $mean \pm std$ of the cross validation is listed corresponding to "Forward", "Left", "Right", and "Stop". The last line in Table 1 is the average values of precision, recall, and f1-score of the two monkeys.

All the recalls are much higher above the chance level (25.00%). In most cases, the precision, recall, and f1-score are close to or even above 50.00%. The recall in class "Left" of Monkey A gets the lowest value 40.65%, meaning the decoding algorithm more occasionally labelled the positive segments into negative than it performs in the other cases. The high FN also creates the relatively lower f1-score, compared with that in the other three classes. Similarly, the f1-score of class "Forward" on Monkey B is much lower than that in the other classes. This indicates that the individual difference may affect the performance of the CNN-based decoding algorithm. Furthermore, for each subject, the computational results in the four classes reflect the class difference when stripping one from the others. Taking Monkey B for example, the decoding algorithm obviously performs much better in class "Stop" than in "Forward", "Left", or "Right". The precision and recall are both above 70.00% while the f1-score is the highest (69.25%) across all the cases. This suggests that it is relatively easier for the algorithm to make a distinction for class "Stop" than the other classes on Monkey B.

For space sake, we do not illustrate the algorithm performance of each day in the leave-one-day-out cross validation. The standard deviation can partially indicate the day-to-day difference. Both monkeys produce the standard deviation above 10.00% in all the four movement tasks, except the f1-score of class "Right" on Monkey A. This indicates the existence of high day-to-day difference of the intracortical activity on both monkeys.

For the comparison with the state-of-art method for motor imagery decoding, Table 2 illustrates the computational results under linear discriminant analysis (LDA) [16]. In the same feature space, the CNN-based decoding algorithm generates better overall output than LDA. For Monkey A, the average precision, recall, and f1-score got an improvement at .0862, .0565, and .0616, respectively. For Monkey B, the average precision, recall, and f1-score got an improvement at .0361, .0322, and .0459, respectively.

Overall, on both Monkey A and B, the CNN-based de-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-18

coding algorithm produced stable classification results. The average recall is much higher than the chance level while the precision and f1-score are all above 50.00%. Since the method generates robust performance in overcoming the day-to-day and the individual differences, it might be useful in certain BCI system like, brain-wave controlled robot/wheelchair. It is worth to mention that the current precision and recall here are not sufficient for control of the actuator in real practise which leaves much space for decoding performance improvement. In our future work, we will modify the CNN structure and try other classifiers to improve the decoding accuracy.

CONCLUSION

Using multielectrode array implanted in the motor cortex, the BCI system allows the subject to continuously drive a mobile robot in four directions with intracortical activity from the motor cortex. In this work, we present a CNN-based decoding algorithm for classifying the intracortical activities from the motor cortex under a series of movement tasks, including, Forward, Left, Right, and Stop. Considering the individual and the day-to-day differences, we make use of the leave-one-day-out cross validation to attain more robust and thus generalizable results. Experimental results indicate the reliable performance of the CNN-based multi-label classifier. The Cohen's kappa of the two subjects are both above 0.4. The average precision, recall, and f1-score of the 4 classes are all above 50.00%, where the highest is 69.12% and the lowest is 52.45%. Compared with LDA, the new decoding algorithm yields better classification results. In our future work, we will modify the CNN structure and try other classifiers to further improve the decoding outcome.

References

[1] Birbaumer Niels. Breaking the Silence: Brain-Computer Interfaces (BCI) for Communication and Motor Control. Psychophysiology. 2006;43(6):517–532.

[2] Christensen James C., Estepp Justin R., Wilson Glenn F., Russell Christopher A. The Effects of Day-to-Day Variability of Physiological Data on Operator Functional State Classification. NeuroImage. 2012;59(1):57 – 63.

[3] Cohen Jacob. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960;20(1):37–46.

[4] He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian. Deep Residual Learning for Image Recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

[5] Lebedev Mikhail A., Nicolelis Miguel A.L. Brain–Machine Interfaces: Past, Present and Future. Trends in Neurosciences. 2006;29(9):536 –546.

[6] Libedinsky Camilo et al. Independent Mobility Achieved through a Wireless Brain-Machine Interface. PLOS ONE. 2016;11(11):1–13.

[7] McFarland Dennis J., Sarnacki William A., Vaughan Theresa M., Wolpaw Jonathan R. Brain-Computer Interface (BCI) Operation: Signal and Noise during Early Training Sessions. Clinical Neurophysiology. 2005;116(1):56 –62.

[8] Miller Kai J., Schalk Gerwin, Fetz Eberhard E., Nijs Marcel den, Ojemann Jeffrey G., Rao Rajesh P.N. Cortical Activity During Motor Execution, Motor Imagery, and Imagery-Based Online Feedback. Proceedings of the National Academy of Sciences. 2010.

[9] Park C., Looney D., Rehman N. ur, Ahrabian A., Mandic D. P. Classification of Motor Imagery BCI Using Multivariate Empirical Mode Decomposition. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2013;21(1):10–22.

[10] Pfurtscheller G., Neuper C. Motor Imagery and Direct Brain-Computer Communication. Proceedings of the IEEE. 2001;89(7):1123–1134.

[11] Quiroga R. Quian, Nadasdy Z., Ben-Shaul Y. Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Neural Computation. 2004;16(8):1661–1687.

[12] Sakhavi S., Guan Cuntai, Yan S. Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks. IEEE Transactions on Neural Networks and Learning Systems. 2018;1–11.

[13] Schalk G., Leuthardt E. C. Brain-Computer Interfaces Using Electrocorticographic Signals. IEEE Reviews in Biomedical Engineering. 2011;4:140–154.

[14] Schirrmeister Robin Tibor et al. Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. Human Brain Mapping. 2017;38(11):5391–5420.

[15] Shin H. et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Transactions on Medical Imaging. 2016;35(5):1285–1298.

[16] Vidaurre C., Kawanabe M., Bünau P. von, Blankertz B., Müller K. R. Toward Unsupervised Adaptation of LDA for Brain–Computer Interfaces. IEEE Transactions on Biomedical Engineering. 2011;58(3):587–597.

[17] Xu Z., So Rosa, Toe K. K., Ang K. K., Guan Cuntai. On the Asynchronously Continuous Control of Mobile Robot Movement by Motor Cortical Spiking Activity. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014, 3049–3052.

[18] Zhang Dan, Song Huaying, Xu Rui, Zhou Wenjing, Ling Zhipei, Hong Bo. Toward a Minimally Invasive Brain–Computer Interface Using a Single Subdural Channel: A Visual Speller Study. NeuroImage. 2013;71:30 –41.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-19

SIMULTANEOUS DECODING OF VELOCITY AND SPEED DURING EXECUTED AND OBSERVED TRACKING MOVEMENTS: AN MEG STUDY

R.J. Kobler[1,2], M. Hirata[2,3], H. Hashimoto[2,3], R. Dowaki[2,4], A.I. Sburlea[1], G.R. Müller-Putz[1]

[1]Institute of Neural Engineering, Graz University of Technology, Graz, Austria
[2]Endowed Research Department of Clinical Neuroengineering, Osaka University, Suita, Japan
[3]Department of Neurosurgery, Osaka University Graduate School of Medicine, Suita, Japan
[4]Hiroshima University School of Medicine, Hiroshima, Japan

E-mail: mhirata@nsurg.med.osaka-u.ac.jp, {reinmar.kobler, gernot.mueller}@tugraz.at

ABSTRACT: Brain signals carry rich information about voluntary upper-limb movements. Accessing this information to control an end-effector (upper-limb, robotic arm, cursor) has been a central topic in brain-computer interface (BCI) research. To date, non-invasive BCIs based on kinematics decoding have focused on extracting partial information (i.e, single or highly correlated kinematic parameters). In this work, we show that low-frequency magnetoencephalographic (MEG) signals simultaneously carry information about multiple kinematic parameters. Using linear models, we decoded cursor velocity and speed during executed and observed tracking movements with moderate (0.2 to 0.4) correlation coefficients (CCs). Comparing the CCs between executed and observed tracking movements, revealed that the MEG signals carried more information (0.1 higher CCs) about velocity and speed during the executed tracking movements. The higher correlations were mainly explained by increased predictive activity in primary sensorimotor areas. We could, therefore, show that non-invasive BCIs have the potential to extract multiple kinematic signals from brain activity in sensorimotor areas.

INTRODUCTION

Decoding voluntary movement from electrophysiological brain signals has been a central topic in brain-computer interface (BCI) research. In recent years, invasive approaches have demonstrated that individuals who lost control of their upper-limb could successfully control a robotic arm [1] or even regain control of their upper-limb [2]. These invasive BCI systems typically decode various movement parameters from spiking rates of neurons in primary sensorimotor areas [3, 4]. It has long been assumed that non-invasive functional neuroimaging techniques such as electroencephalography (EEG) and magnetoencephalography (MEG) lack the spatial resolution to decode the kinematics of voluntary movements. However, it has been shown otherwise by various studies in the past decade [5–7]. However, the predicted end-effector position of non-invasive decoders typically has a lower signal to noise ratio (SNR) compared to the invasive decoders. We think that the combined decoding of multiple kinematic signals could improve the SNR and, thereby, elevate the performance of non-invasive decoders.

A myriad of non-invasive studies in this field of BCI research has investigated either directional (e.g, velocity, position) or non-directional (e.g, speed) kinematic signals in isolation [6, 8–10]. Bradberry et al. were the first to show that low-frequency EEG signals carried information about positions and velocities during reaching movements [11]. Complementary, Waldert et al. showed that reach direction can be classified from low-frequency EEG and MEG [12]. Jerbi et al. showed that low-frequency MEG signals are coupled to hand speed during a continuous visuomotor (VM) task [8]. Recent invasive studies using Electrocorticographic (ECoG) signals demonstrated that velocity and speed can be decoded simultaneously from low-frequency brain signals in primary sensorimotor areas [13, 14]. Taken together, we surmise that velocity and speed can be jointly decoded from non-invasive M/EEG signals.

We believe that a pursuit tracking task (PTT) is ideally suited to investigate this question. A PTT is characterized by two stimuli - a traget stimulus moving along random trajectories and an end-effector (e.g., cursor). The end-effector is used to track the target stimulus. The PTT has two favorable properties. First, the kinematics vary continuously in a frequency range that can be controlled by the experimenter. Second, the target trajectories can be designed so that specific kinematic signals are jointly uncorrelated. Using a PTT, we showed that the low-frequency EEG originating in premotor and primary sensorimotor areas was preferentially tuned to cursor velocity rather than cursor position [15].

In this study, we investigated the joint decoding of cursor velocity and speed from low-frequency MEG activity during a two-dimensional PTT. Our paradigm separated two conditions. In the first condition (execution), participants tracked the target with their gaze and a cursor. In the second condition (observation), participants tracked the target only with their gaze. This allowed us to address two questions. First, can velocity and speed be decoded simultaneously from non-invasively acquired brain activity during voluntary upper-limb movements? Second, does the decoding performance change if the upper-limb is not involved in the tracking task?

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-19

MATERIALS AND METHODS

23 healthy people participated in this study. 5 were female and 18 male. They were 28.5 ± 2.4 (standard-error of the mean; sem) years old, had normal or corrected to normal vision, and self-reported to be right-handed. The experimental procedure conformed to the declaration of Helsinki and was approved by the ethics committee of the Osaka University hospital. We could not complete the experiment for three participants and identified that one participant was positioned incorrectly inside the MEG scanner during the offline analysis. These four participants were excluded from the final offline analysis.

Figure 1a depicts the experimental setup. The participants were lying in a supine position, with the head resting on a cushion inside an MEG scanner. A projection screen fixed in front of their face presented visual stimuli. We used two visual stimuli, a target and a cursor. The cursor could be controlled by the participants through moving their right hand's index finger on a 2D surface. The tip of the index finger was tracked with a custom optical motion capture system.

After each participant found a comfortable position for his/her right arm and hand, the position of the marker was defined as the resting position. The resting position was mapped to the center of the screen. Finger movements, within a 1.5 centimeter radius around the resting position were mapped to cursor movements within a circle confined by the bounds of a virtual grid. Rightward/forward finger movements were mapped to rightward/upward cursor movements.

As in our previous study [15], the experimental procedure consisted of 4 blocks, lasting for about 3 hours in total. In the first block (10 minutes), the participants could familiarize themselves with the paradigm. In the second block, we recorded eye artifacts and resting activity as described in [16] for about 12 minutes (2 runs; each 6 minutes). During the third block, the main experimental task was performed. The fourth block at the end of the experiment was identical to the second one.

Figure 1b outlines the paradigm of the main experimental task. We investigated a PTT in two conditions. A yellow target stimulus indicated the beginning of a new trial. After 2 s of preparation, the target changed its color to green (execution condition) or blue (observation condition).

In the execution condition, the participants were asked to track the target with their gaze and the cursor. In the first experimental block, each participant trained to minimize the distance to the target and to make smooth cursor movements.

In the observation condition, the participants would only track the target with their gaze and keep their finger in the resting position. In order to achieve similar visual input and tracking dynamics in both conditions, we replayed matching cursor trajectories.

The target moved along pseudo-random trajectories, which were generated from pink noise in the [0.3, 0.6] Hz band. The horizontal and vertical components of the target trajectories were independent and identically distributed. This required the participant to control the cursor in 2 dimensions at the same time. The detailed target trajectory generation and cursor trajectory replay procedures are presented in [15].

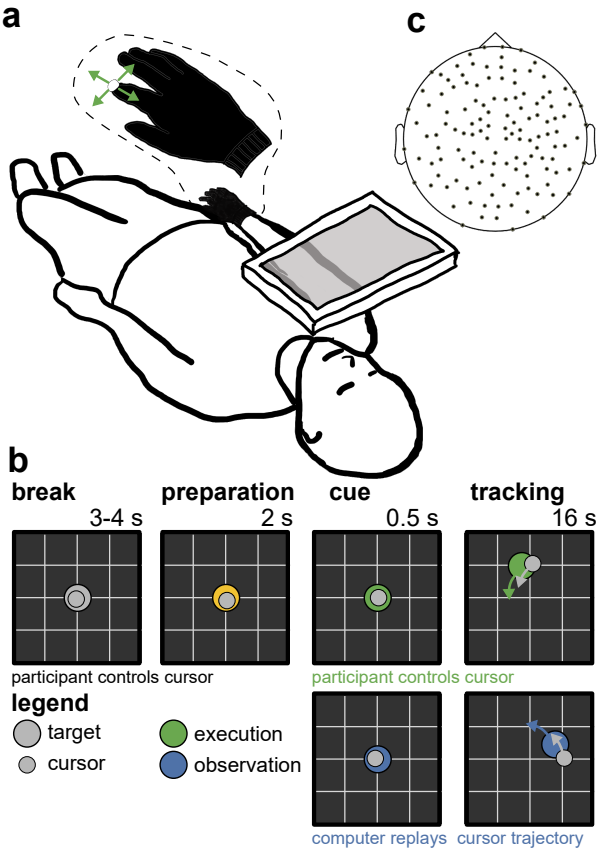


Figure 1: Overview of the experiment. **a**, Experimental setup. The participants were lying in a supine position inside an MEG scanner. They moved their right index finger on a 2D surface to control a cursor on the screen. **b**, Experimental paradigm. Each trial started with a 2 s preparation period. The participants were asked to keep the cursor in the center (i.e, the right index finger in the resting position). After a condition cue, the target stimulus moved along a pseudo-random trajectory for 16 s. In the execution condition (green), the participants tracked the target with both their gaze and the cursor. In the observation condition, they tracked the target only with their gaze. **c**, Topographic distribution of the 129 MEG sensors used in this study.

The paradigm consisted of 160 trials (80 per condition; pseudo-randomly distributed). They were presented in 10 runs (6 minutes each). In between runs, participants could rest for about 2 to 3 minutes. During the experiment, we tested if the finger was in the resting position during the preparation (both conditions) and tracking periods (observation condition). If the position exceeded a threshold, a trial was aborted. On average, 5.2 trials were aborted, resulting in 154.8 complete trials.

Neuromagnetic activity was recorded with a 160 channel whole-head MEG system (MEGvision NEO, Yokogawa Electrip Corp., Kanazawa, Japan) housed in a magnetically shielded room. For this study, we used the signals of 129 sensors (Figure 1c). Electrooculographic (EOG) signals were recorded with 4 electrodes placed at the outer canthi (horizontal EOG) and above/below the left eye

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-19

(vertical EOG). The EOG signals were recorded with a 128-channel EEG system (Neurofax EEG 1200, Nihon Koden Corp., Tokyo, Japan). MEG and EOG signals were recorded synchronously at rate a of 1 kHz.

We asked participants to keep their head and shoulder position fixed during the experiment (blocks 2 to 4). We additionally monitored the head position inside the MEG system with five marker coils, attached to the face. Their position was measured at the beginning of each run.

The custom motion capture and visual stimuli signals were recorded at 60 Hz using the labstreaminglayer (LSL) protocol[1]. We implemented the paradigm in Python 2.7 based on the simulation and neuroscience application (SNAP) platform[2]. Using pilot experiments, we determined the delay between the finger and the cursor movement on the screen. The delays introduced by the hard- and software added up to 190 ms.

We analyzed the recorded data offline with a custom-made pipeline that we implemented with Matlab (Matlab 2015b, Mathworks Inc., USA) and the open source toolboxes EEGLAB [17] (version 14.1.1) and Brainstorm [18] (version 05-Jun-2018). We first synchronized the stimuli and MEG signals by aligning impulses captured by a photodiode. After synchronization, all signals were resampled to 200 Hz. We estimated the cursor velocities by applying a Savitzky-Golay finite impulse response (FIR) differentiation filter (polynomial order 3, 21 filter taps, zero-phase) to the cursor positions.

To compensate small head movements across runs, we spherically interpolated the MEG sensors of all runs (10 tracking, 4 eye) to their average position in relation to the participant's head. If the maximal distance of any channel to the average position was larger than 25 mm, the run was discarded. We discarded 2 runs in total. The grand-average maximal channel distance across accepted runs was 5±0.1 mm (sem). After merging the signals of the tracking runs, we applied high-pass (0.25 Hz cut-off frequency, Butterworth filter, eight order, zero-phase) and band-stop (59 and 61 Hz cut-off frequencies, Butterworth filter, fourth order, zero-phase) filters. To compensate technical and spatially stationary artifacts introduced by equipment, we applied independent component analysis (ICA). In detail, we applied the extended infomax algorithm to decompose the MEG signals (high-pass filter with 0.4 Hz cut-off frequency) into independent components (ICs) that explained 99.9 % of the variance. We visually inspected and marked 8.6±0.2(sem) of 63.5±0.1(sem) ICs for rejection. They were then removed from the 0.25 Hz high-pass filtered signals. We attenuated eye movement and blink artifacts by applying the artifact subspace subtraction algorithm [15, 16]. To extract the low-frequency MEG signals, we applied a low-pass filter to the broadband MEG signals (2 Hz cut-off frequency, Butterworth filter, sixth order, zero-phase) and resampled all signals at 10 Hz.

We then epoched the continuous data into 14 s trials, starting 1.5 s after the condition cue. Trials were marked for rejection, if (1) the broadband MEG signal of any sensor exceeded a threshold (±5 fT), (2) had a abnormal probability, variance or kurtosis (≥ (6/4/6) standard-deviations (stds) beyond the mean), (3) the correlation between the EOG and the target position signals were improbable (≥ 4 stds beyond the mean), or (4) a tracking error happened (i.e., jerky or no cursor movement). All criteria combined resulted in rejecting 26.6±0.4(sem) of 154.8 trials.

Cursor velocity and speed were estimated with a sliding-window, linear regression approach [11, 15, 19]. At single lags, a partial least-squares (PLS) estimator was used to decode a single kinematic signal (horizontal velocity, vertical velocity or speed) from the pre-processed MEG signals. Similar to [15], the PLS estimator considered 10 latent components. The model was evaluated using a 10 times 5 fold cross-validation (CV) scheme with the evaluation metric being Pearson correlation coefficients (CCs) between the recorded kinematic signals and their neural estimates. We estimated chance level performance by shuffling the kinematic signals across trials of the same condition. We then applied 5 fold CV to the shuffled data and repeated the shuffling and CV evaluation 1000 times.

The weights of the linear regression model can be readily transformed to patterns [20]. We computed scaled patterns according to [15].

To ease neurophysiological interpretation, we projected the scaled patterns to the cortical surface of the ICBM152 template boundary element (BEM) head model [21]. We co-registered the template with the head of each participant (and the MEG sensors) by manually fitting the template head model to digitized head points (50 to 60 points per participant) in Brainstorm toolbox. OpenMEEG [22] was applied to compute the forward model for 5011 voxels on the cortical surface. sLORETA [23] was used to estimate the inverse solution for unconstrained sources at the 5011 voxels. The noise covariance matrix was estimated using 5 minutes of resting data (similar preprocessing as the tracking data), recorded during experimental blocks 2 and 4, and applying shrinkage regularization (10% of its largest eigenvalue).

RESULTS

Grand-average results presented here are summarized by the mean and its standard-error across the 19 participants.

We assessed the participants tracking behavior by computing CCs between the target and cursor position signals in the execution condition. The CCs peaked at 0.22±0.01 s for the horizontal component and 0.23±0.01 for the vertical component. That is, the target signal lead the cursor by approximately 225 ms on average. The CCs at the peaks were 0.90±0.01 (horz) and 0.92±0.01 (vert).

We also assessed the visual tracking behavior in both conditions by computing CCs between the horizontal/vertical target position and horizontal/vertical EOG signals. In the execution condition, the CCs were 0.94±0.01 (horz) and 0.79±0.06 (vert). In the observation condition, they

[1] https://github.com/sccn/labstreaminglayer
[2] https://github.com/sccn/SNAP

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-19

were 0.92±0.01 (horz) 0.70±0.06 (vert).

The auto-/cross-correlation curves in Figures 2a-c demonstrate that during the PTT the three signals of interest (horizontal/vertical cursor velocity and cursor speed) were negligibly correlated. The grand-average cross-correlations were below or equal to 0.1 in both conditions. Similar auto-/cross-correlation curves in both conditions show that the cursor trajectories in observation condition (dashed lines) were similar to the executed ones (solid lines).

Figures 2d-f show the grand-average CV test-set CCs between the decoded and recorded kinematic signals for single-lag, sliding-window, linear regression models. We decoded the cursor velocities and speed from the MEG signals of all 129 sensors at lags ranging from [-0.5, 0.5] s in steps of 0.1 s. The MEG sensor signals lead the cursor signals for negative lags.

We used the CCs of the shuffled data to test if the observed results were due to chance. We controlled the false-discovery rate (FDR) at a significance level $\alpha = 0.05$ for $n_{comparisons} = n_{metrics} \cdot n_{lags} = 3 \cdot 11 = 33$ comparisons per subject [24]. The tables at the bottom of Figures 2d-f list the results.

The horizontal cursor velocity decoding results are summarized in Figure 2d. The decoding model performed above chance level for all lags and participants in the execution condition and almost all participants in the observation condition. The execution condition CCs were larger than the observation condition ones for all lags. The paired difference between conditions (exe-obs) peaked at lag -0.3 s.

The vertical cursor velocity decoding results are summarized in Figure 2e. The CCs were above chance level for all lags and participants. Compared to the horizontal velocity results, we observed higher CCs in the execution condition and similar CCs in the observation condition. As a consequence, the paired difference was higher (peaked at lag -0.1 s).

The cursor speed decoding results are displayed in Figure 2f. All CCs were above chance level in the execution condition. In the observation condition, the results varied across lags and ranged from 14 to 18 participants having CCs above chance level. Compared to the velocities, the CCs were lower by approx. 0.1 in both conditions. However, the effect size of condition was comparable to the velocities (paired difference peak of 0.15 at lag -0.3 s).

To identify the spatiotemporal encoding of information about cursor velocities and speed, we transformed the model weights to patterns at the cortical surface. Figures 2g-i show the grand-average patterns at selected lags for execution condition (top), observation condition (middle) and the paired-difference (bottom). The paired-differences for all three kinematic signals show that the predictive pattern activity in contra-lateral primary-sensorimotor (SM1) areas was larger in execution condition at negative lags. Thus, contra-lateral SM1 activity carried information about the upcoming cursor velocities and speed in the execution condition. The difference was maximal at lags -0.3 to -0.2 s and varied across the signals of interest (speed > vertical velocity > horizontal velocity).

We also observed that superior parietal and parieto-occipital areas carried predictive information about the three kinematic signals in both conditions (Figures 2g,h top and middle row). The paired-differences indicate that the activity in these areas was similarly predictive for horizontal cursor velocity in both conditions (Figure 2g), more predictive for vertical cursor velocity in execution condition (Figure 2h), and less predictive for cursor speed in execution condition (Figure 2i).

DISCUSSION

We have demonstrated simultaneous decoding of cursor velocity and speed information by means of low-frequency MEG signals during a PTT. The PTT allowed us to study continuous, uncorrelated cursor velocity and speed signals (Figures 2a-c). During executed index finger tracking movements, contra-lateral SM1 activity was simultaneously predictive for cursor velocity and speed, while superior parietal and parieto-occipital activity was also predictive in observed tracking movements.

Linear, single-lag decoding model performance in terms of CCs was above chance level for all participants during execution condition, and almost all participants during observation condition (Figures 2d-f). The range of CCs is in agreement with the results of previous linear M/EEG kinematics decoding studies [11, 15, 25].

Comparing the two conditions, we found that the MEG signals contained more information about the cursor velocities and speed in the execution condition (Figures 2d-f). The average effect of condition was stronger for the vertical cursor velocity than for the horizontal one. This is in accordance with the findings of our previous EEG study [15].

The differences in the decoder patterns (Figures 2g-i; bottom row) indicate that the activity in contra-lateral SM1 carried more information about the uncorrelated velocity and speed signals in the execution condition. The differences peaked at lags -0.3 s (horizontal cursor velocity; Figure 2,g), -0.2 s (vertical cursor velocity; Figure 2,h) and -0.3 s (cursor speed; Figure 2,i) respectively. Since the difference in decoder CCs in Figures 2d-f were modulated by these peaks, the activity in contra-lateral SM1 must have considerably contributed to the larger decoder CCs in the execution condition.

Considered that the cursor movement was the 190 ms delayed index finger movement, the observed peaks are plausible in terms of neurophysiology [26]. Simultaneous decoding of velocity and speed from contra-lateral SM1 has also been demonstrated by recent invasive studies based on spiking activity [27] and low-frequency ECoG signals [14]. Inoue et al. observed that the firing rates of a large fraction of neurons in SM1 were simultaneously tuned to speed and velocity [27]. Hammer et al. reported a stronger encoding of end-effector speed compared to velocity in low-frequency ECoG signals [14]. This in

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-19

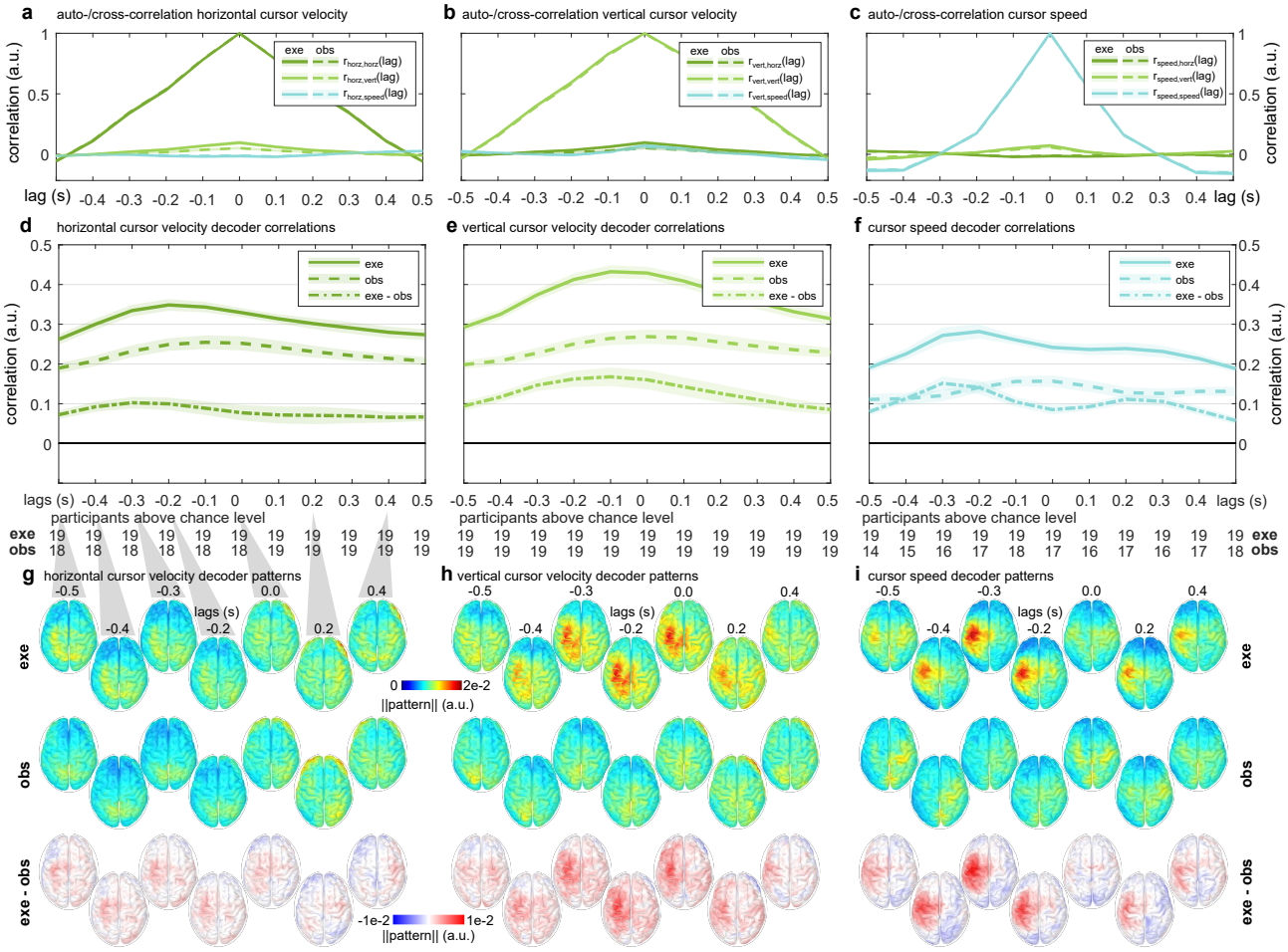


Figure 2: Grand-average results of the experiment. Shaded areas indicate the sem across the 19 participants. **a**, Auto-/cross-correlation curves for the three metrics of interest relative to the horizontal cursor velocity. **b,c**, As in **a** for the vertical cursor velocity (**b**) and cursor speed (**c**). **d**, CV Test-set correlations between the recorded horizontal cursor velocity and its decoded estimate for execution condition (solid line), observation condition (dashed line) and their paired difference (dash-dotted line). MEG sensor signals lead the horizontal cursor velocity signal for negative lags. The table below lists the number of participants whose single-lag decoder CCs were above chance level. **e,f**, As in **d** for the vertical cursor velocity (**e**) and cursor speed (**f**). **g**, Horizontal cursor velocity decoder patterns for selected lags. We projected the patterns to voxels of a template BEM head model and averaged their norms across participants. **h,i**, As in **g** for the vertical cursor velocity (**h**) and cursor speed (**i**).

agreement with our results. I.e, in execution condition, the speed decoder patterns in SM1 were larger than the velocity decoder patterns.

The decoder patterns (Figures 2g-i; top and middle rows) showed also common activity in execution and observation condition. In both conditions, superior parietal and parieto-occipital areas were predictive. Their activation in both conditions is in agreement with findings of our previous study [15] and also fMRI studies on executed and observed reaching movements [28, 29].

CONCLUSION AND FUTURE WORK

In this work we have shown that non-invasive MEG signals simultaneously carry information about velocity and speed of executed and observed tracking movements. Linear, single-lag decoders extracted more information originating in contra-lateral SM1 during executed tracking movements. Whereas superior parietal and parito-occipital areas were informative in executed and observed tracking movements.

Despite the encouraging results presented here, further research is imperative. It needs to be shown that a combined decoding of velocity and speed indeed improves the SNR of the predicted end-points offline and subsequently online. Moreover, studies with humans who lost control of their upper-limb will have to demonstrate whether non-invasive, decoding of imagined body kinematics has the potential to improve their quality of life.

ACKNOWLEDGEMENTS

The authors acknowledge Catarina Lopes Dias, Joana Pereira and Lea Hehenberger for their valuable comments. This work has received funding from the European Research Council (ERC) via the consolidator grant 681231 ‘Feel Your Reach’, from Graz University of Technology via its short-time research abroad scholarship, from the Japan Society for the Promotion of Science (JSPS) program KAKENHI (26282165), from the National Institute of Information and Communications Technology (NICT), from the Council for Science, Tech-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-19

nology and Innovation (Cabinet Office, Government of Japan) program ImPACT, and from a research grant form the Ministry of Internal Affairs (Government of Japan).

REFERENCES

[1] Collinger JL, Wodlinger B, Downey JE, *et al.* High-performance neuroprosthetic control by an individual with tetraplegia. The Lancet. 2013;381(9866):557–564.
[2] Ajiboye AB, Willett FR, Young DR, *et al.* Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: A proof-of-concept demonstration. The Lancet. 2017;389(10081):1821–1830.
[3] Lebedev MA, Nicolelis MAL. Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation. Physiological Reviews. 2017;97(2):767–837.
[4] Branco MP, Boer LM de, Ramsey NF, *et al.* Encoding of kinetic and kinematic movement parameters in the sensorimotor cortex: A brain-computer interface perspective. European Journal of Neuroscience. 2019.
[5] Jerbi K, Vidal JR, Mattout J, *et al.* Inferring hand movement kinematics from MEG, EEG and intracranial EEG: From brain-machine interfaces to motor rehabilitation. Irbm. 2011;32(1):8–18.
[6] Robinson N, Vinod AP. Noninvasive Brain–Computer Interface: Decoding Arm Movement Kinematics and Motor Control. IEEE Systems, Man, & Cybernetics Magazine. 2016;2(October):4–16.
[7] Fukuma R, Yanagisawa T, Saitoh Y, *et al.* Real-time control of a neuroprosthetic hand by magnetoencephalographic signals from paralysed patients. Scientific Reports. 2016;6(1):21781.
[8] Jerbi K, Lachaux JP, N'Diaye K, *et al.* Coherent neural representation of hand speed in humans revealed by MEG imaging. PNAS. 2007;104(18):7676–7681.
[9] Bourguignon M, Jousmäki V, Op de Beeck M, *et al.* Neuronal network coherent with hand kinematics during fast repetitive hand movements. NeuroImage. 2012;59(2):1684–1691.
[10] Marty B, Bourguignon M, Jousmäki V, *et al.* Cortical kinematic processing of executed and observed goal-directed hand actions. NeuroImage. 2015;119:221–228.
[11] Bradberry TJ, Gentili RJ, Contreras-Vidal JL. Reconstructing Three-Dimensional Hand Movements from Noninvasive Electroencephalographic Signals. Journal of Neuroscience. 2010;30(9):3432–3437.
[12] Waldert S, Preissl H, Demandt E, *et al.* Hand movement direction decoded from MEG and EEG. Journal of Neuroscience. 2008;28(4):1000–1008.
[13] Bundy DT, Pahwa M, Szrama N, *et al.* Decoding three-dimensional reaching movements using electrocorticographic signals in humans. Journal of neural engineering. 2016;13(2):026021.
[14] Hammer J, Pistohl T, Fischer J, *et al.* Predominance of Movement Speed over Direction in Neuronal Population Signals of Motor Cortex: Intracranial EEG Data and A Simple Explanatory Model. Cerebral Cortex. 2016;26(6):2863–2881.
[15] Kobler RJ, Sburlea AI, Müller-Putz GR. Tuning characteristics of low-frequency EEG to positions and velocities in visuomotor and oculomotor tracking tasks. Scientific Reports. 2018;8(1):17713.
[16] Kobler RJ, Sburlea AI, Müller-Putz GR. A comparison of ocular artifact removal methods for block design based electroencephalography experiments. In: Proc. of the 7th Graz BCI Conference. 2017, 236–241.
[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] Tadel F, Baillet S, Mosher JC, *et al.* Brainstorm: A user-friendly application for MEG/EEG analysis. Computational Intelligence and Neuroscience. 2011;2011:8.
[19] Ofner P, Müller-Putz GR. Using a noninvasive decoding method to classify rhythmic movement imaginations of the arm in two planes. IEEE Trans. Bio. Eng.. 2015;62(3):972–981.
[20] Haufe S, Meinecke F, Görgen K, *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage. 2014;87:96–110.
[21] Fonov V, Evans AC, Botteron K, *et al.* Unbiased average age-appropriate atlases for pediatric studies. NeuroImage. 2011;54(1):313–327.
[22] Gramfort A, Papadopoulo T, Olivi E, *et al.* OpenMEEG: Opensource software for quasistatic bioelectromagnetics. BioMedical Engineering Online. 2010;9.
[23] Pascual-Marqui RD. Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. Methods & Findings in Experimental & Clinical Pharmacology. 2002;24:1–16.
[24] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. B. 1995;57(1):289–300.
[25] Bradberry TJ, Rong F, Contreras-Vidal JL. Decoding center-out hand velocity from MEG signals during visuomotor adaptation. NeuroImage. 2009;47(4):1691–1700.
[26] Miall RC, Wolpert DM. Forward models for physiological motor control. Neural Networks. 1996;9(8):1265–1279.
[27] Inoue Y, Mao H, Suway SB, *et al.* Decoding arm speed during reaching. Nature Communications. 2018;9(1):5243.
[28] Filimon F, Nelson JD, Huang RS, *et al.* Multiple Parietal Reach Regions in Humans: Cortical Representations for Visual and Proprioceptive Feedback during On-Line Reaching. Journal of Neuroscience. 2009;29(9):2961–2971.
[29] Magri C, Fabbri S, Caramazza A, *et al.* Directional tuning for eye and arm movements in overlapping regions in human posterior parietal cortex. NeuroImage. 2019;(in press).

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-20

TRAINING EFFECTS OF A TACTILE BCI FOR WHEELCHAIR CONTROL

M. Eidel[1], V. Klemm[1], A. Kübler[1]

[1]Institute of Psychology, University of Würzburg, Würzburg, Germany

E-mail: Matthias.Eidel@uni-wuerzburg.de

ABSTRACT: The tactile modality for brain-computer interfaces (BCIs) is rarely used because of its limitations in speed and accuracy. However, non-visual BCI paradigms are of great interest as potentially feasible options for patients with limited gaze control. For the present study, 5 participants navigated a virtual wheelchair through a 3D apartment using a 4-class tactile BCI in 5 identical sessions.
Mean P300 amplitude, mean difference between target and non-target (MD), and information transfer rate (ITR) were calculated for all sessions. Descriptively, amplitudes at Fz increased with training, MD or other electrode positions showed no obvious changes. Mean ITR during the fifth session was more than twice as high (8.7 bits/min) as compared to the first (4.0 bits/min).
Our preliminary results suggest that a previous study by Herweg and colleagues [1] can be replicated with respect to a training effect. However, we could not yet achieve ITRs as high as in [1]. We are currently increasing our sample size to 15 participants.

INTRODUCTION

Brain-Computer Interfaces allow for computer-mediated communication and interaction with the environment. Often based on electroencephalography (EEG), BCI systems record brain activity which is then interpreted by machine learning algorithms in order to classify the users' intentions. Notably, BCIs rely on brain activity and do not require any voluntary muscular movement.
This independence from muscular activity is what makes BCI a promising tool for patients with severe paralysis, for example as a consequence of Amyotrophic Lateral Sclerosis (ALS) or after brain injury [2,3]. Recently, some patients and their caretakers have been using BCIs independently at home for a prolonged time [4,5].
Still, many BCI systems use visual feedback and stimulation and thus, rely on the patients' ability to control eye movements [6] and may face usability issues once gaze control becomes limited [7], as in the case of later stages of ALS. Because of this limitation, alternatives to vision dependent BCIs have become a focus of current research. Specifically, auditory and tactile event-related potential (ERP)-based BCIs have been developed and shown to be feasible options [1,8–10]. Among these, an auditory BCI using animal sounds was tested with motor-impaired patients [10] and a tactile BCI was tested within a healthy, elderly sample [1]. In both cases, participants were invited to 5 sessions and showed major improvements in BCI performance or its physiological correlates. This putative training effect of prolonged BCI use is one of the main points of interest of the present study also designed to replicate previous results [1].
Hypotheses: We expected that prolonged training with our BCI system would lead to increased ERP amplitude, MD and ITR.

MATERIALS AND METHODS

Participants: N=6 healthy participants (5 female) were recruited. Participant 3 was excluded due to scheduling issues. All reported normal or corrected to normal vision and were BCI-naïve. All participants received a monetary reimbursement of € 7.50 per hour and gave informed consent to the procedure which was approved by the ethical review board of the Institute of Psychology at the University of Würzburg, Germany.

Stimulation: Tactile stimulation was applied via a BCI2000-controlled tactor device (C2 tactors; Engineering Acoustic Inc., Casselberry, USA) at right and left thigh, belly and neck. Tactors were adjusted until they were perceived as equally strong at all positions. During the experiment, the tactors were activated separately and pseudorandomized with equal probabilities (25%), and with a frequency of 250 Hz each. Stimulus duration was 220 ms, interstimulus interval 400 ms.
Participants were sitting in a chair in front of a monitor showing the virtual environment for wheelchair navigation and instructed to keep their eyes open and to avoid excessive blinking and keep their facial muscles relaxed.

EEG recording: EEG was recorded (512 Hz) with 12 passive Ag/AgCl electrodes and amplified using a g.USBamp (g.tec Engineering GmbH, Graz, Austria). Electrode positions were Fz, FC1, FC2, C3, Cz, C4, P3, Pz, P4, O1, Oz, and O2, with ground and reference electrodes at the right and left mastoids, respectively. Impedance was kept below 5 kΩ. Online filtering was performed using a band pass filter between 0.1 and 60 Hz and a notch filter between 48 and 52 Hz.

Procedure: To investigate training effects, participants attended five sessions on separate days, with no more than one week in between sessions.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-20

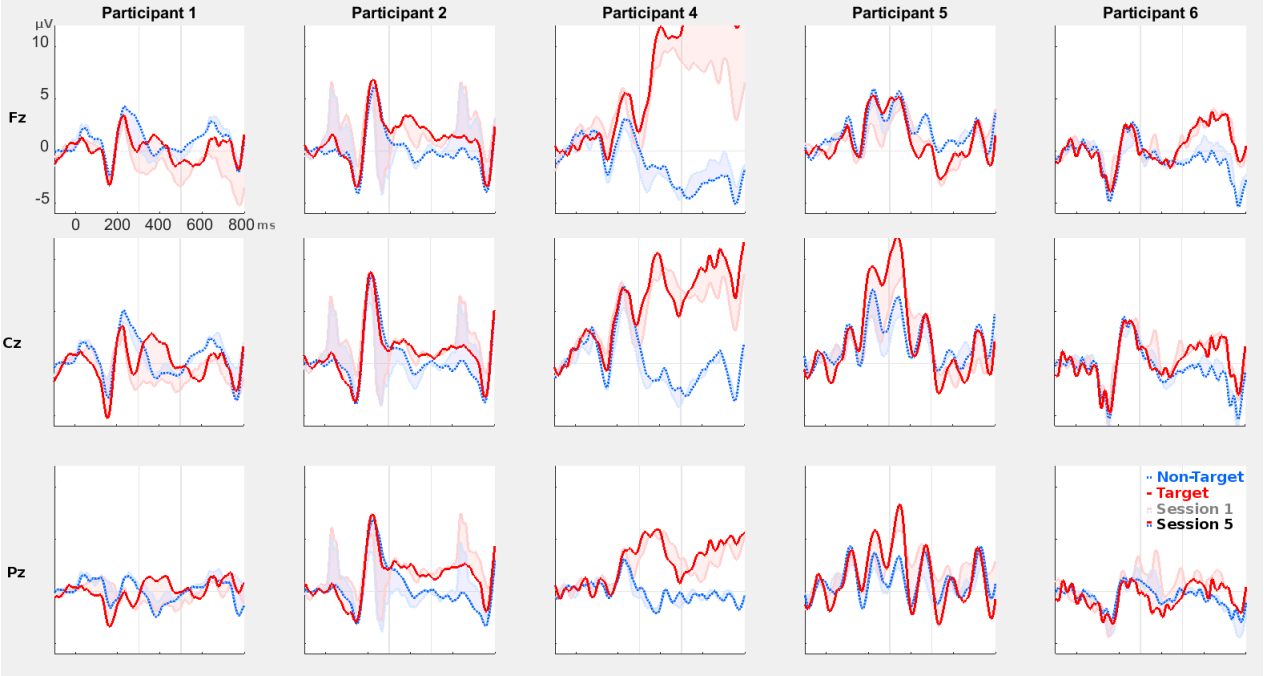


Figure 1: Post-stimulus epochs on sessions 1 (light colours) and 5 (bold colours) at positions Fz, Cz, Pz.
Target stimuli (red) appear to have elicited a positive deflection in the P300 range. Participant 6 appeared to have a delayed ERP deflection after about 450ms. Data was averaged by participant and additionally lowpass-filtered (20Hz, for visualization only). Positivity up.

During three calibration runs at the beginning of each session, participants had to concentrate on each of the body positions several times, resulting in a total of 240 target and 720 non-target epochs. These data served to train a linear classifier which was then used for two free navigation runs. Here, participants had to navigate a wheelchair through a virtual 3D apartment along three checkpoints by selecting a direction and then focusing on the corresponding body position (i.e. left knee to make a left turn). After the first run, starting and end points of the course were switched, so that participants had to navigate back to the original starting point. One successful run required at least 14 commands, however, erroneous or misleading commands were also executed and had then to be corrected. Thus, a maximum number of 22 commands, roughly corresponding to the number necessary when assuming the minimum accuracy of 70% for sufficient control [11], was allowed before the run was terminated.

To preclude possible ceiling effects like in an earlier study from Herweg and colleagues [1], which always used 8 stimulus repetitions for one selection, the number of repetitions was kept at a minimum for each participant and for every session.

We estimated the number necessary to reach 100% classification accuracy via visual analysis and predictions from the classifier algorithm. This stimulus number was determined for every participant and every session and then used for the free navigation task.

After each session, to assess workload, the NASA TLX

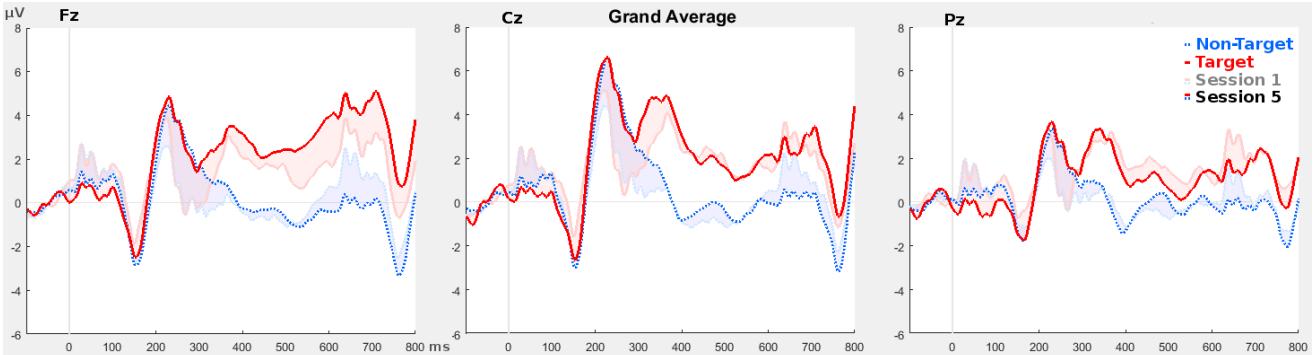


Figure 2: Grand averages of post-stimulus epochs at Fz, Cz and Pz positions. Target stimuli elicited P300 ERPs at all electrode positions. Descriptively, the ERP amplitudes have increased in magnitude from session 1 (light colours) to session 5 (bold colours) at position Fz. Overall, both amplitudes and MD appear strongest at Fz and Cz, but ERPs are still visible at Pz. Data was averaged over all participants. Positivity up.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-20

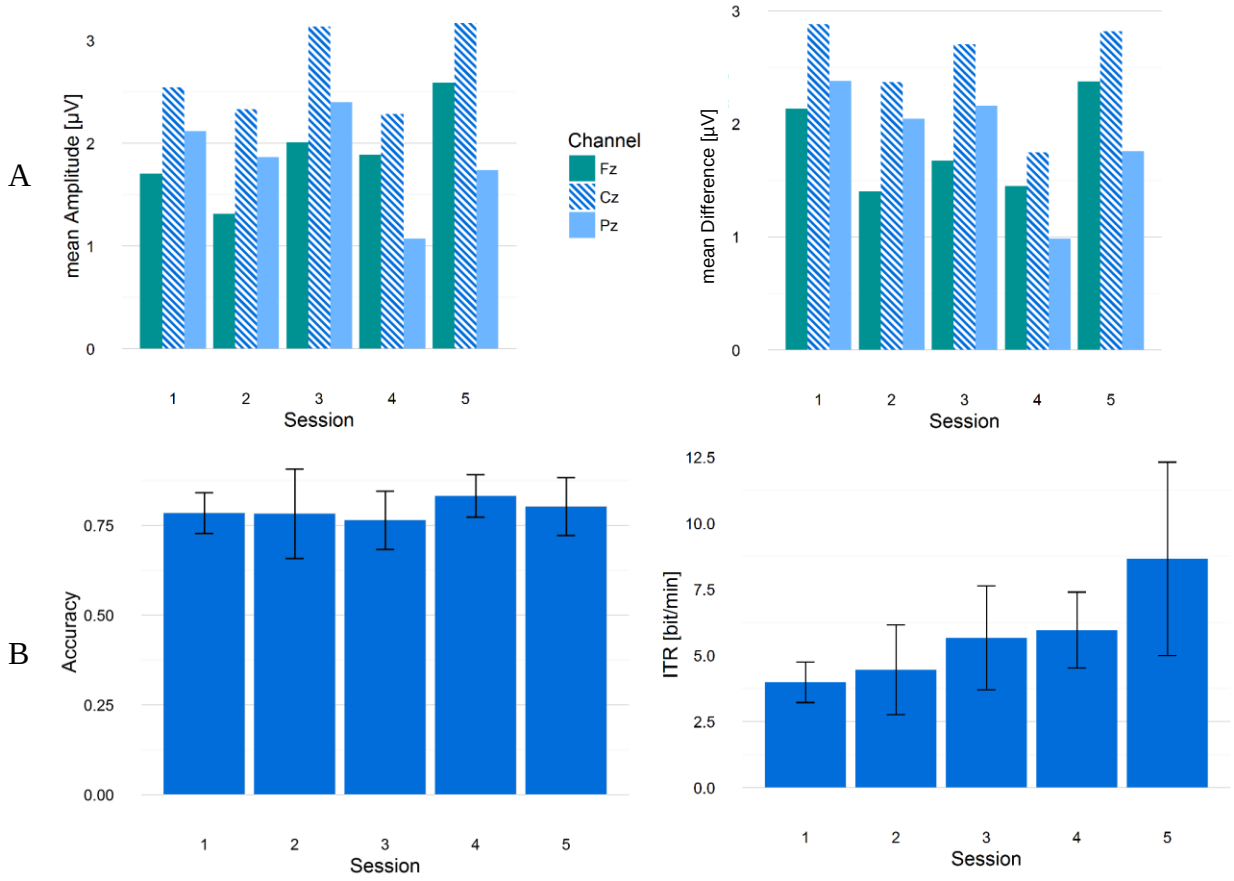


Figure 3: Bar plots of dependent variables over the course of the 5 training sessions. Data was averaged over all participants (Error bars represent SE).
A) Physiological measures, amplitude and MD, split by electrode position. Visual analysis revealed no obvious effects.
B): BCI performance measures. Online accuracy does not appear to increase with training, while ITR does appear to increase.

[12], and after the first and last session the adjusted QUEST [13,14], to assess satisfaction with the device, were filled out by the participants (Results not shown here).

EEG processing: EEG data was band pass filtered between 0.1 and 30 Hz and divided into segments of 800 ms post-stimulus, plus another leading 100 ms that was used for baseline-correction. Segments containing values exceeding ±150 µV were excluded as artifacts. Target and non-target epochs were grouped and then averaged separately. This process was performed with MATLAB© (v2013b) using functions provided by BCI2000 [15] and EEGLab [16]. Using a step-wise linear discriminant analysis as implemented in the BCI2000 package, we built new classification models for each session.

Data analysis: The use of BCI accuracy (i.e. the percentage of correct classifications) as the sole comparative measure of performance is not sufficient when the number of stimuli differs between groups or time points. Thus, we also used the ITR as a dependent variable. This ITR, given in bits per minute, is used to calculate the amount of information transferred during a given time. The number of bits (B) can be calculated with equation 1 using the accuracy (P) as well as the number of all possible selections (N = 4).

$$B = \log_2 N + \mathrm{P}\log_2 P + (1-\mathrm{P})\log_2\frac{(1-\mathrm{P})}{(\mathrm{N}-1)} \quad (1)$$

The ITR is then calculated by dividing the number of bits by the time necessary for the selection. It is thus directly dependent on the number of stimuli repetitions, but also takes accuracy into account.
Additionally, we extracted physiological features (mean target amplitudes and mean difference between target and non-target) from the EEG data from a time window ranged 300-500ms post-stimulus.
Due to the small sample size, results will be reported descriptively.

RESULTS

Wheelchair navigation through the virtual apartment was achieved with an average online accuracy of 78.5% (sd=14.8).
Target and non-target epochs from the three calibration runs at the beginning of each session are presented in

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-20

Figure 1. Visual inspection revealed a positive deflection in the P300 range at positions Fz, Cz and Pz (most pronounced at Fz) in some (1,2,4) subjects. Participant 6 had a positive deflection at a latency of around 450 ms.
Grand averages of these epochs are shown in Figure 2. Here, visual analysis again reveals a putative P300 deflection which was most pronounced at Fz and Cz and that appeared to increase with training (Session 1 vs. Session 5) at Fz.
MD at Fz increased from 2.14 to 2.37 µV. At Cz, values remained more or less stable (2.88 to 2.81 µV), whereas at Pz, values decreased (2.38 to 1.76 µV). Mean amplitude at Fz increased from 1.70 to 2.59 µV, at Cz from 2.54 to 3.17 µV but decreased at Pz (2.12 to 1.74 µV).
Accuracies (P=0.79; sd=0.15) from the navigation task were used in conjunction with the respective number of stimulus repetitions to calculate the individual participant's ITRs. Averaged online accuracies and ITRs are shown in Figure 3. There appears to be a strong trend of increasing ITRs, but not accuracies, with session numbers. Average ITR was 4.0 and 8.7 bits/min on sessions 1 and 5, respectively.

DISCUSSION

Five participants navigated a wheelchair through a virtual apartment with a reasonable level of BCI control. With a mean accuracy of 78.5%, participants mostly exceeded 70%, a number which is considered the low threshold for efficient BCI control [11]. Stimulus repetitions for each session were kept deliberately low to preclude any ceiling effects as experienced in [1], so we expect accuracies to be higher when increasing the number of stimuli used for a selection, albeit at the cost of speed.
Descriptively, mean ITR increased substantially over the 5 training sessions, with session 5 resulting in more than double the ITR as compared to session 1. Overall, this shows that participants achieved the same accuracy in less time (i.e. with fewer stimulus repetitions). We speculate that our hypothesis that ITR would increase with training might be confirmed. Since stimulus repetitions were individually adjusted, average accuracies remained consistent over the sessions. However, their comparatively [1] low values indicate that we set the number of stimuli too low.
The two physiological variables, mean amplitude and MD, did not reveal an overt training effect on the individual level, although a trend may be seen, specifically at position Fz. Our hypothesis about increasing physiological measures is thus far not supported.
The apparent training effect (as seen in ITR increase) seems not to be reflected in the physiological measures. This may be because traditional analysis struggles to accommodate for deviations in individual participants' ERP pattern, e.g. in cases of unusually high latencies or inversed polarities. This illustrates the need of machine learning approaches that automatically detect spacial and temporal features on an individual level and explains why an effect of session number is visible on machine learning-dependent variables such as ITR, but not in static measures such as amplitudes derived from a fixed time and position.

CONCLUSION

We have again shown that wheelchair control with our tactile BCI paradigm is feasible and that training effects (as measured via ITR over the five sessions) appear to be present. We will continue to recruit more subjects for the present study to allow for statistical analysis and to investigate whether the results from the study of Herweg and colleagues [1] are replicable. So far, all our dependent variables remained below the values achieved in [1].

REFERENCES

[1] Herweg, A., Gutzeit, J., Kleih, S., and Kübler, A. (2016) Wheelchair control by elderly participants in a virtual environment with a brain-computer interface (BCI) and tactile stimulation. *Biological Psychology*.
[2] Vidal, J.J. (1973) Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*. 2 (1), 157–180.
[3] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., and Vaughan, T.M. (2002) Brain–computer interfaces for communication and control. *Clinical Neurophysiology*. 113 (6), 767–791.
[4] Holz, E.M., Botrel, L., and Kübler, A. (2015) Independent home use of Brain Painting improves quality of life of two artists in the locked-in state diagnosed with amyotrophic lateral sclerosis. *Brain-Computer Interfaces*. 2 (2–3), 117–134.
[5] Botrel, L., Holz, E.M., and Kübler, A. (2017) Using Brain Painting at Home for 5 Years: Stability of the P300 During Prolonged BCI Usage by Two End-Users with ALS. in: D.D. Schmorrow, C.M. Fidopiastis (Eds.), Augment. Cogn. Enhancing Cogn. Behav. Complex Hum. Environ., Springer International Publishing, Champp. 282–292.
[6] Birbaumer, N. and Cohen, L.G. (2007) Brain–computer interfaces: communication and restoration of movement in paralysis. *The Journal of Physiology*. 579 (3), 621–636.
[7] Brunner, P., Joshi, S., Briskin, S., Wolpaw, J.R., Bischof, H., and Schalk, G. (2010) Does the "P300" speller depend on eye gaze? *Journal of Neural Engineering*. 7 (5), 56013.
[8] Kaufmann, T., Herweg, A., and Kübler, A. (2014) Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials. *Journal of NeuroEngineering and Rehabilitation*. 11 (1), 7.
[9] Simon, N., Käthner, I., Ruf, C.A., Pasqualotto,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-20

E., Kübler, A., and Halder, S. (2015) An auditory multiclass brain-computer interface with natural stimuli: Usability evaluation with healthy participants and a motor impaired end user. *Frontiers in Human Neuroscience*. 8.

[10] Halder, S., Käthner, I., and Kübler, A. (2016) Training leads to increased auditory brain–computer interface performance of end-users with motor impairments. *Clinical Neurophysiology*. 127 (2), 1288–1296.

[11] Kübler, A., Neumann, N., Kaiser, J., Kotchoubey, B., Hinterberger, T., and Birbaumer, N.P. (2001) Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication. *Archives of Physical Medicine and Rehabilitation*. 82 (11), 1533–1539.

[12] Hart, S.G. and Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. in: P.A. Hancock, N. Meshkati (Eds.), Adv. Psychol., North-Holland, pp. 139–183.

[13] Demers, L., Weiss-Lambrou, R., and Ska, B. (2002) The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): An overview and recent progress. *Technology and Disability*. 14 (3), 101–105.

[14] Zickler, C., Riccio, A., Leotta, F., Hillian-Tress, S., Halder, S., Holz, E., et al. (2011) A brain-computer interface as input channel for a standard assistive technology software. *Clinical EEG and Neuroscience*. 42 (4), 236–244.

[15] Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., and Wolpaw, J.R. (2004) BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*. 51 (6), 1034–1043.

[16] Delorme, A. and Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. 134 (1), 9–21.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-21

"WHEN DOES IT WORK ?" : AN EXPLORATORY ANALYSIS OF TRANSFER LEARNING FOR BCI

Pedro L. C. Rodrigues[1], Marco Congedo[1], Christian Jutten[1]

[1]GIPSA-lab, CNRS, University Grenoble Alpes, Grenoble Institute of Technology, Grenoble, France

E-mail: pedro.rodrigues@gipsa-lab.fr

ABSTRACT: Transfer Learning is a critical topic of research in the BCI field. Its goal is to reuse data gathered in a previous session (*source* session) in order to reduce, or completely bypass, calibration in a new session (*target* session). Although many methods have been proposed to tackle this problem, little is known about what characteristics of the datasets should be taken into account in order to ensure good performance. In this paper, we perform an exploratory analysis to study the influence of some simple descriptors of the *source* and *target* datasets over the classification scores obtained with Transfer Learning. We observe that the discriminability of the data points in the *target* session plays an important role in determining how well the Transfer Learning will work, as opposed to that of the *source* session, which has no statistically significant role in most cases.

INTRODUCTION

Reducing calibration time is an important challenge in Brain-Computer Interface (BCI) research [1]. Several Transfer Learning (TL) approaches have been proposed in the literature for doing so and most are based on the idea of reusing data from a previous recording session of a subject (the *source* session) to classify the data of a new session from the same subject or a different one (the *target* session). However, reusing data directly from previous sessions in general yields poor results. This comes from the fact that the statistical distributions of data from different sessions (same subject or not) are rarely the same [1].

A typical TL approach in BCI is to transform the data points from both the *source* and *target* datasets so that the discrepancy between their statistical distributions is reduced [2, 3]. In this paper, we match the statistics of the *source* and *target* datasets via the recently proposed *Riemannian Procrustes Analysis* (RPA) [2], a method that adapts the classical Procrustes analysis to a Riemannian geometry framework.

It is well known that, although any pair of *source–target* subjects can go through a Transfer Learning procedure, some pairs of subjects yield better results in classification than others. Our main goal in this paper is to investigate some factors that might explain this variability and how one could try to predict beforehand (i.e., before doing any matching of the datasets or classifying the data points) the "compatibility" between two datasets.

Our exploratory analysis relies on the estimation of linear models and the study of statistical significance of the coefficients estimated for those models. We use as explanatory factors the intra-scores for the *source* and *target* subjects (cross-validated classification score using the subject's dataset as training and testing dataset), and the Maximum Mean Discrepancy (MMD [4]) between the two datasets, which is a common measure of discrepancy between statistical distributions. We observe that the intra-score for the *target* subject plays an important role in determining how well the Transfer Learning will work, as opposed to the intra-scores of the *source* subjects, which plays no statistically significant role in most cases. We also observe that before doing any transformation on the data points of the *source* and *target* datasets, the MMD between their statistical distributions plays a statistically significant role over the performance of the Transfer Learning. However, once the RPA is applied, the MMD between the datasets becomes very small and no longer carries statistical information to describe the variability of the cross-subject scores. This confirms the relevance of the RPA method.

MATERIALS AND METHODS

This section begins with a formal definition of the Transfer Learning problem. Then, we give a brief introduction to concepts of Riemannian geometry and describe the RPA method. Finally, we present the statistical tools used in our exploratory analysis of Transfer Learning as well as the dataset chosen for our investigations.

Transfer Learning: We formulate the problem of Transfer Learning by first defining two datasets, the *source* ($\mathcal{S}$) and the *target* ($\mathcal{T}$) dataset. They are comprised of couples

$$\begin{aligned} \mathcal{S} &= \left\{ (C_i^{\mathcal{S}},\ y_i^{\mathcal{S}}) \text{ for } i = 1, \ldots, K_{\mathcal{S}} \right\}, \\ \mathcal{T} &= \left\{ (C_i^{\mathcal{T}},\ y_i^{\mathcal{T}}) \text{ for } i = 1, \ldots, K_{\mathcal{T}} \right\}, \end{aligned} \tag{1}$$

with $C_i^{\mathcal{S}}$ and $C_i^{\mathcal{T}} \in \mathbb{R}^{n \times n}$ being data points, and $y_i^{\mathcal{S}}$ and $y_i^{\mathcal{T}} \in \{1, \ldots, L\}$ their corresponding class labels; $K_{\mathcal{S}}$ and $K_{\mathcal{T}}$ are the number of trials in the *source* and *target* sessions respectively. In this paper, the data points in $\mathcal{S}$ and $\mathcal{T}$ are not Euclidean feature vectors as is usu-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-21

ally done, but symmetric positive definite (SPD) matrices, which are used to parametrize the statistics of EEG multivariate time series [5].

Transfer Learning concerns the case when the statistical distributions $\mu_{\mathcal{S}}$ and $\mu_{\mathcal{T}}$, describing the *source* and *target* datasets respectively, are different. In this context, one might want to train a classifier h using the information in $\mathcal{S}$ and apply it to data points in $\mathcal{T}$ (or vice-versa).

A common approach is to define a transformation for the data points in the *target* dataset so that their new statistical distribution is the same as that of the *source* dataset. To do so, most algorithms define an optimization procedure that tries to minimize some notion of distance between the statistical distributions of $\mathcal{S}$ and $\mathcal{T}$, such as the MMD. In [6], a theoretical analysis of the Transfer Learning problem has shown that methods reducing the distance between statistical distributions are mathematically well justified, since they reduce the upper bounds of the classification error of h in $\mathcal{T}$.

Riemannian geometry of SPD matrices: We denote by $X_k \in \mathbb{R}^{n\times T}$ the recording of T samples on n electrodes of the k^{th} trial of a zero-mean time series and y_k the class associated to X_k. The spatial covariance matrix C_k associated to X_k is an $n \times n$ matrix estimated as usual by

$$C_k = \frac{1}{T-1} X_k X_k^T \,. \tag{2}$$

Covariance matrices are symmetric positive definite (SPD) and form a manifold $\mathcal{P}(n)$. When associated to a metric, one can define fundamental geometric notions in $\mathcal{P}(n)$, such as geodesics (shortest curve joining two points), distance between two points (length of the geodesic connecting them), the center of mass of a set of points, etc. We endow $\mathcal{P}(n)$ with the affine-invariance Riemannian metric, which induces the distance [7]

$$\delta_R^2(C_i, C_j) = \| \log(C_i^{-1/2} C_j C_i^{-1/2}) \|_F^2 \,, \tag{3}$$

for $C_i, C_j \in \mathcal{P}(n)$. This distance is more natural for the $\mathcal{P}(n)$ manifold as compared to the Euclidean distance and has been instrumental in several BCI classification algorithms developed in recent years [5]. The geometric mean M according to distance (3) of a set of covariance matrices $\{C_1, \ldots, C_K\}$ is defined as [5]

$$M = \underset{X\in\mathcal{P}(n)}{\operatorname{argmin}} \sum_{k=1}^{K} \delta_R^2(X, C_k) \,, \tag{4}$$

where the cost function in (4) is the dispersion of the set of matrices around a matrix X. The above definitions suffice for the intents of this paper. The interested reader will find a thorough treatment of the subject in the monography of R. Bhatia [7] and its applications to BCI in [5].

Riemannian Procrustes Analysis: In this paper, we use RPA [2] for transforming data points in $\mathcal{T}$ so that their new distribution is as close as possible to $\mu_{\mathcal{S}}$. RPA works by considering the distributions of data points in $\mathcal{S}$ and $\mathcal{T}$ as shapes in a high-dimensional space. It performs rigid geometric operations over the ensemble of data points, such as translation, stretching and rotation, to make their shapes as similar as possible (see Figure 1 for a visual representation of these operations). The transformations in RPA are done respecting the intrinsic geometry of $\mathcal{P}(n)$, which is where the data points of $\mathcal{S}$ and $\mathcal{T}$ are defined. The steps involved in this procedure are summarized as follows :

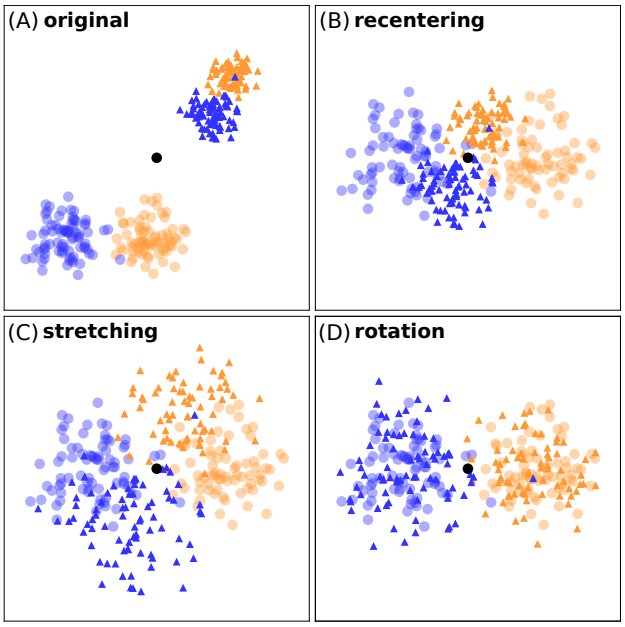


Figure 1: Schematic representation in a 2D Euclidean space of the rigid transformations involved in the classical Procrustes Analysis. The black dot represents the origin of the space and circles in blue and yellow are data points from two classes in a *source* dataset. The triangles are data points of a *target* dataset. The transformations in RPA are analogous to those in Euclidean space but applied to data points in the SPD manifold $\mathcal{P}(n)$.

1. Estimate the **geometric means** of $\mathcal{S}$ and $\mathcal{T}$, denoting them by $M_{\mathcal{S}}$ and $M_{\mathcal{T}}$, respectively, and the dispersions around the mean of each dataset, denoted by $d_{\mathcal{S}}$ and $d_{\mathcal{T}}$.
2. **Re-center** the data points in $\mathcal{S}$ and $\mathcal{T}$ by doing

$$C_i^{\mathcal{S}(\text{rct})} = M_{\mathcal{S}}^{-1/2} C_i^{\mathcal{S}} M_{\mathcal{S}}^{-1/2} \,, \tag{5}$$

$$C_i^{\mathcal{T}(\text{rct})} = M_{\mathcal{T}}^{-1/2} C_i^{\mathcal{T}} M_{\mathcal{T}}^{-1/2} \,, \tag{6}$$

   and forming new datasets

$$\mathcal{S}^{(\text{rct})} = \{C_i^{\mathcal{S}(\text{rct})}\} \quad \text{and} \quad \mathcal{T}^{(\text{rct})} = \{C_i^{\mathcal{T}(\text{rct})}\} \,.$$

   Note that the geometric mean of these two new datasets is the Identity matrix.
3. **Stretch** the dispersion around the mean of the data points in $\mathcal{T}^{(\text{rct})}$ so that it matches the dispersion around the mean of $\mathcal{S}^{(\text{rct})}$ by creating new data points

$$C_i^{\mathcal{T}(\text{str})} = \left(C_i^{\mathcal{T}(\text{rct})}\right)^s \,, \tag{7}$$

   where we require $s \in \mathbb{R}$ to verify

$$s^2 = d_{\mathcal{S}}/d_{\mathcal{T}} \,. \tag{8}$$

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-21

4. The last step consists in **rotating** the matrices from $\mathcal{T}^{(\text{str})}$ around the origin and matching the orientation of its point cloud with that of $\mathcal{S}^{(\text{rct})}$. We have then

$$C_i^{\mathcal{T}(\text{rot})} = U^T \, C_i^{\mathcal{T}(\text{str})} \, U \, , \tag{9}$$

where U is determined via an optimization procedure that minimizes the distance between the class means of each dataset after the rotation. Note that this step is a semi-supervised one, since it requires knowledge of at least a few labels of the *target* dataset for estimating its class means (see [2] for more details).

Transfer Learning classification: In this paper, whenever we want to do a classification task with data points that live in $\mathcal{P}(n)$, we use the Minimum-Distance to Mean classifier (MDM) [1, 5, 8], which is a generalization of the nearest-centroid classifier to the space of SPD matrices. It works by first estimating the geometric mean of the elements of each class in the training dataset (the class means). Then, it assigns to each unlabeled data point the class of the nearest class mean according to the δ_R distance. In the context of transfer learning, we will always consider the *source* dataset $\mathcal{S}$ as the training dataset and the *target* dataset $\mathcal{T}$ as the testing dataset. The classification score is simply the average accuracy of the classifier. In the following analysis, we will consider the results of Transfer Learning classification on three different cases :

DCT : the *source* and *target* datasets are used **d**ire**ct**ly as training and testing datasets, that is, without any transformation.

RCT : the *source* and *target* datasets are both **r**e-**c**en**t**ered and then used as training and testing dataset.

RPA : the *source* and *target* datasets go through the full RPA procedure (re-centering + stretching + rotating) and are then used as training and testing datasets.

For this analysis we will assume that all labels in the *target* dataset are available for the estimation of the class means. In fact, our intent here is not to evaluate the performance of the RPA method in a realistic situation when only a few labels from the *target* dataset are available (this has been done in [2]), but rather to understand how its performance might be influenced by other factors.

Seriation procedure: Given a dataset, all cross-subject TL scores are summarized in a matrix $S^{(m)}$, where the $S_{ij}^{(m)}$ element contains the accuracy of the classification with method $m \in \{\mathbf{DCT}, \mathbf{RCT}, \mathbf{RPA}\}$ using subject i as *target* and subject j as *source*. We use a tool from combinatorial data analysis named *seriation* [9] to rearrange the lines and columns of $S^{(m)}$ in order to make relevant patterns emerge. The rows and columns of $S^{(m)}$ are sorted in decreasing order of their marginals. The output of this procedure is a new representation where the pairs of *source-target* subjects with the best accuracy are located at the top-left region of the matrix, while the worst pairs are at the bottom-right region.

Statistical analysis procedure: Our quantitative analysis is based on the estimation of linear regression models to describe the variability on the values of $S_{ij}^{(m)}$ as defined in the previous subsection. We estimate a different linear model $\mathcal{L}_i^{(m)}$ for each *target* subject i and method m. We do this because the cross-subject scores for two *target* subjects with the same *source* subject are statistically dependent, which would undermine the estimation of a linear model mixing all scores from all *source-target* pairs. Moreover, the results after the **RPA** method are related to those for the **RCT** one, since the latter includes the former as a processing step.
We define the linear model $\mathcal{L}_i^{(m)}$ for *target* subject i on method m as :

$$S_{ij}^{(m)} = \beta_{1,i}^{(m)} S_i + \beta_{2,i}^{(m)} S_j + \beta_{3,i}^{(m)} \eta_{ij}^{(m)} + \epsilon_i^{(m)} \, , \tag{10}$$

where

- S_i (S_j) is the intra classification score of *target* (*source*) subject i (j), obtained via cross-validation with training and testing datasets coming from the same subject. Note that since each model $\mathcal{L}_i^{(m)}$ is estimated for one fixed *target* subject i, S_i is a constant in (10) and acts as a scaling for the intercept; thus, it is not considered as an independent variable in the statistical analysis.
- Factor $\eta_{ij}^{(m)}$ is the MMD between datasets $\mathcal{S}$ and $\mathcal{T}$ after the operations of method m, defined as [4]

$$\begin{aligned} \text{MMD}(\mathcal{S}, \mathcal{T}) &= \frac{1}{K_{\mathcal{S}}^2} \sum_{i,j}^{K_{\mathcal{S}}} k(C_i^{\mathcal{S}}, C_j^{\mathcal{S}}) \\ &+ \frac{1}{K_{\mathcal{T}}^2} \sum_{i,j}^{K_{\mathcal{T}}} k(C_i^{\mathcal{T}}, C_j^{\mathcal{T}}) \\ &- \frac{2}{K_{\mathcal{S}} K_{\mathcal{T}}} \sum_{i,j}^{K_S, K_T} k(C_i^{\mathcal{S}}, C_j^{\mathcal{T}}) \, , \end{aligned} \tag{11}$$

where

$$k(P, Q) = \exp\left(-\frac{\delta_R^2(P, Q)}{2\sigma^2} \right) , \tag{12}$$

for $P, Q \in \mathcal{P}(n)$ and σ is taken as the median value of all pairwise distances of elements in $\mathcal{S}$ and $\mathcal{T}$. $K_{\mathcal{S}}$ and $K_{\mathcal{T}}$ are defined in (1).

- The variable $\epsilon_i^{(m)}$ stands for all residual factors that are not explained by the linear regression model.

Once the linear models are all estimated, we perform a set of hypothesis tests for each *target* subject i. The goal is to assess the statistical significance of the coefficients of each model. The first kind of test is a F-test for the omnibus null hypothesis :

$$\begin{aligned} &\mathcal{H}_0 \, : && \beta_{2,i}^{(m)} = \beta_{3,i}^{(m)} = 0 \, , \\ &\mathcal{H}_1 \, : && \beta_{k,i}^{(m)} \neq 0 \text{ for at least one } k \text{ in } \{2, 3\} \, . \end{aligned} \tag{13}$$

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-21

This is a standard test used for inspecting whether the set of independent variables of a linear regression model, S_j and δ_{ij} in (10), is statistically significant for explaining at least part of the variability of the dependent variable, $S_{ij}^{(m)}$ in (10). When the null hypothesis is rejected, we say that there is enough statistical evidence for considering that the slope of at least one of the independent variables is different than zero. In this case, we perform t-tests for checking which explanatory variable in $\mathcal{L}_i^{(m)}$ is statistically significant. We have :

$$\begin{aligned} \mathcal{H}_0 &: \quad \beta_{\ell,i}^{(m)} = 0 \,, \\ \mathcal{H}_1 &: \quad \beta_{\ell,i}^{(m)} \neq 0 \,, \end{aligned} \tag{14}$$

for $\ell \in \{2, 3\}$. When the null hypothesis of (14) is rejected for $\beta_{\ell,i}^{(m)}$, we say that there is statistical evidence for considering it different than zero and so the independent variable related to it contributes for explaining the dependent variable $S_{ij}^{(m)}$.
The statistical procedure described above yields two sets of p-values for each method $m \in \{\mathbf{DCT}, \mathbf{RCT}, \mathbf{RPA}\}$. The first set contains the p-values for each F-test on each *target* subject i, whereas the second set gathers the p-values of the t-tests. The results presented in the next section are based on the analysis of these sets of p-values and how they are distributed along different *source* subjects for each method.

Dataset: We carried out our analysis on a publicly available dataset [10] which we will refer as *Cho2017* from now on. The dataset contains recordings of subjects performing BCI trials following a Motor Imagery (MI) paradigm with 64 EEG electrodes (sampling frequency 512 Hz) from 52 subjects, each one performing 200 trials (100 of each class). We filtered the EEG signals in the 8-30 Hz band and each trial was considered as a segment from 0.5 to 2.5 seconds after the trial onset. We estimated the spatial covariance matrices using (2). Not all subjects in *Cho2017* display data which can be well discriminated, so we kept only those with intra-score in terms of AUC (Area Under the ROC-curve) above chance level; this criterion retains 40 subjects out of the 52 in total.

RESULTS AND DISCUSSION

In this section, we present the results of our analysis of TL via RPA on the *Cho2017* dataset. We begin with a qualitative analysis of the output of the seriation procedure applied to the cross-subject scores. Then, we study the correlation of each factor defined in (10) with the cross-subject scores. Finally, we analyse the results of the statistical hypothesis tests for the linear models $\mathcal{L}_i$ and discuss the patterns observed for the whole dataset (from now on, we will indicate the superscript specifying the method m only when necessary).

Cross-subject classification accuracy: Figure 2 shows the output of the seriation procedure on the cross-subject TL scores for the *Cho2017* dataset on the three classification methods: **DCT**, **RCT**, and **RPA**. We observe that with **RCT** and **RPA** there are more pairs of subjects with high values of cross-subject classification than with **DCT**. In particular, we note that for **RCT** and even more for **RPA** there are many *target* subjects for which the classification accuracy is high for almost all possible *source* subjects. To investigate the possible explanations for this behavior, we compute the *Spearman* correlation between the average cross-subject score for each target (i.e., the average value along the rows of matrix S) and the intra-subject accuracy of the corresponding *target* subject. For the **RPA** method, we obtain a correlation of 0.58 ($p < 10^{-3}$), for **RCT** it is 0.44 ($p < 10^{-2}$) and for **DCT** is 0.45 ($p < 10^{-2}$). We interpret these results as : subjects that are "good" for classifying their own data can better receive information from other *source* subjects. We also provide a quantitative analysis of the results. Figure 3 portrays the histograms of all cross-subject Transfer Learning scores S_{ij} (rows and columns confounded) for each method. Their means are reported in Table 1. These results show that the transformations over the *source* and *target* datasets do improve the cross-subject classification scores on the average for **RCT** and greatly do so for **RPA**.

Table 1: Average values of the cross-subject Transfer Learning scores and the MMD distance between *source* and *target* datasets for each method.

Method	$(\eta_{ij})_{\text{avg}}$	$(S_{ij})_{\text{avg}}$
DCT	0.63	0.53
RCT	0.01	0.58
RPA	0.01	0.76

Changes in MMD after each RPA step: We evaluate how the MMD between each pair of *source–target* subjects changes after the re-centering step and the full RPA procedure. Table 1 gives the average values of the MMD distances for each method and shows that there is a clear decrease after each transformation. This result is not surprising, since each step of the RPA procedure was conceived exactly to make the distributions of $\mathcal{S}$ and $\mathcal{T}$ closer in some sense and the MMD allows for a quantitative assessement of it.

Study of the linear models $\mathcal{L}_i$: After exploring the grand averages of the cross-subject TL scores and how they relate to a few explanatory factors, we analyse the linear models $\mathcal{L}_i$ defined in (10) and estimated on each *target* subject i for the three methods of interest : **DCT**, **RCT**, and **RPA**.
We first plot the p-values of the F-test for each model sorted in ascending order. Notice that, under the omnibus null hypothesis for all *target* subjects, the p-values follow an uniform distribution, thus, when sorted they will lie on a straight line. The leftmost plot in Figure 4 shows that for almost all subjects the variability of the cross-subject performance is well explained by the linear model estimated for the **DCT** method and, in a lesser extent, for the **RCT** method. For **RPA** the p-values lie all very close to

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-21

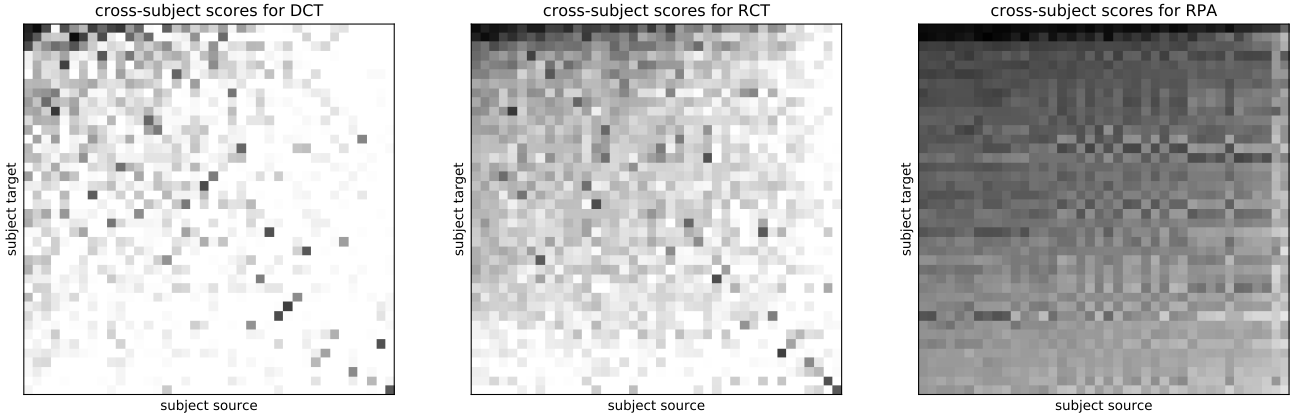


Figure 2: Accuracies of the cross-subject classification for three different Transfer Learning procedures on the *Cho2017* database. The rows and columns of each subplot were reordered using the *seriation* procedure explained in the text. The grayscale varies from white (accuracy 0.5) to black (accuracy 1.0).

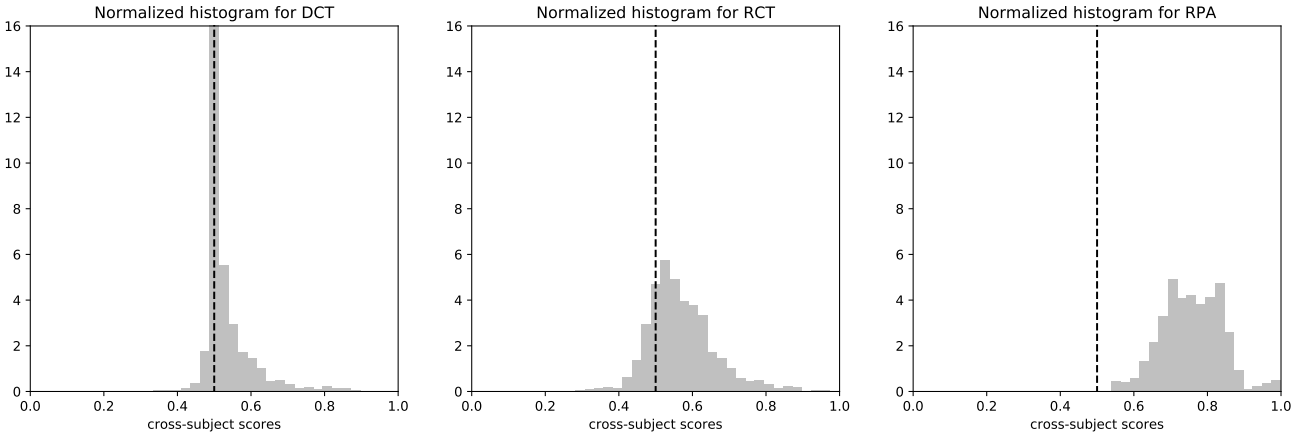


Figure 3: Normalized histograms of the cross-subject Transfer Learning scores for the three methods described in the text. The vertical dashed line indicates chance level.

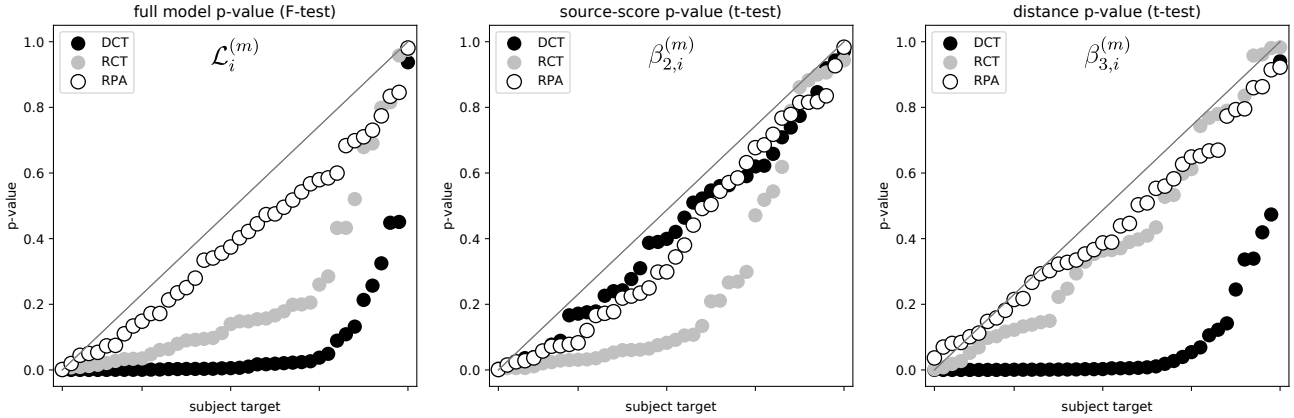


Figure 4: p-values of different statistical tests over the linear models $\mathcal{L}_i$ (each one associated to a *target* subject i). Each circle represents the p-value of a given test on a given *target* subject and the x-axis has been rearranged so that all the p-values are in increasing order. The leftmost plot represents the results of the F-test of the full linear model $\mathcal{L}_i$, whereas the center plot illustrates the p-values for the t-test of the coefficient $\beta_{2,i}$ in $\mathcal{L}_i$ (related to the intra-score of the *source* subject), and the rightmost plot displays the p-values for the t-test on the coefficient $\beta_{3,i}$ in $\mathcal{L}_i$ (related to the MMD between the *source* and *target* datasets).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-21

a 45 degree line. It is worth remembering that the statistical significance of the coefficient for the intercept, and, therefore, the influence of the intra-score S_i on describing the values of S_{ij}, is not assessed via the F-test. This is why we have calculated the *Spearman* correlation between the row-averaged S_{ij} and the S_i in the previous sub-section.

The distribution of the p-values in the center plot of Figure 4 shows that $\beta_{2,i}$ has no statistical significance in the linear model $\mathcal{L}_i$ for any of the *target* subjects in the **DCT** and **RPA** methods. However, for **RCT** it does seem to play a role for some *target* subjects. What we can conclude from these observations is that **RPA** is able to make the cross-subject TL score independent of the choice of *source* subject (at least in terms of its intra score). As a consequence, it makes it easier to find "good" source subjects for each *target* subject, as it was already observed during our qualitative analysis of Figure 2.

Finally, the rightmost plot in Figure 4 shows that the MMD between *source* and *target* datasets plays a role in describing the cross-subject Transfer Learning scores only for the **DCT** method. This result is comforting, since it brings evidence to the fact that the operations in the RPA procedure are capable of factoring out most of the differences between the statistical distributions of $\mathcal{S}$ and $\mathcal{T}$. As a consequence, we may say that any further improvement that one might want to do on the Transfer Learning procedure should take into account other aspects of the mismatch between datasets besides the MMD between them.

CONCLUSION

In this paper, we have investigated the influence of different factors on the variability of cross-subject Transfer Learning scores in the case when all *source* and *target* labels are known. Our goal has been to assess whether some basic explanatory variables, such as the intra-score of the *source* and *target* subjects, play any role for determining the scores obtained in the cross-subject classification. A simple, and yet important, application of this study is being able to predict beforehand (i.e., before doing all transformations and then classifying the trials) which *source* subject would be the most appropriate for doing classification on a given *target* subject.

We have observed that the discriminability of the trials of the *target* subjects plays a fundamental role in determining how the cross-subject Transfer Learning will perform. On the other hand, the influence of the intra-scores for the *source* subjects have proven to be rather limited, as seen by the lack of statistical significance for the values of $\beta_{2,i}$ in most cases. We have also observed that the influence of the MMD between $\mathcal{S}$ and $\mathcal{T}$ is not statistically significant after using RPA to match the two datasets. What we can conclude from this is that the RPA procedure is capable of factoring out most of the influence of the discrepancy between statistical distributions of the *source* and *target* datasets.

It is our opinion that in order to devise new and more powerful strategies for Transfer Learning it is imperative to investigate which factors determine its success. The present study is a little step in this direction. Future work may include the search for richer models for describing the variability of cross-subject TL scores. One approach would be to consider non-linear relations between the explanatory variables as well as adding new factors related to other features of the *source* and *target* subjects.

REFERENCES

[1] F Lotte, L Bougrain, A Cichocki, *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. 031 005, 2018.

[2] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for brain-computer interfaces", *IEEE Trans. on Biomedical Engineering*, pp. 1–1, 2018.

[3] N. T. H. Gayraud, A. Rakotomamonjy, and M. Clerc, "Optimal Transport Applied to Transfer Learning For P300 Detection", in *7th Graz Conference*, Graz, Austria, 2017, p. 6.

[4] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test", *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 723–773, Mar. 2012.

[5] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review", *Brain-Computer Interfaces*, pp. 1–20, 2017.

[6] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains", *Machine Learning*, vol. 79, no. 1-2, pp. 151–175, 2009.

[7] R. Bhatia, *Positive Definite Matrices*. Princeton university press, 2009.

[8] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry", *IEEE transactions on biomedical engineering*, vol. 59, no. 4, pp. 920–928, 2012.

[9] I. Liiv, "Seriation and matrix reordering methods: An historical overview", *Stat. Anal. Data Min.*, vol. 3, no. 2, pp. 70–91, Apr. 2010.

[10] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "EEG datasets for motor imagery brain–computer interface", *GigaScience*, vol. 6, no. 7, pp. 1–8, 2017.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-22

EEG SOURCE ANALYSIS OF VISUAL MOTION IMAGERY FOR APPLICATION TO BRAIN-COMPUTER INTERFACE

K. Koizumi[1], K. Ueda[1], N. Tateyama[1], M. Nakao[1]

[1] Department of Mechanical Engineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

E-mail: koizumi@hnl.t.u-tokyo.ac.jp

ABSTRACT: In the field of brain-computer interface (BCI) research, building more intuitive and goal-directed BCI systems is a key to breakthrough. "Visual motion imagery," which enables visual imaging of the operation of a target is one possible method for application to BCI. As a foothold study for applying visual motion imagery to BCI, we examined EEG source activity during visual motion imagery via comparison with visual motion perception. Source activities induced in beta band activity in the right posterior cingulate, left precuneus and left cuneus, and in low gamma activity in the right inferior parietal lobule and supramarginal gyrus were significantly higher in visual motion imagery than in visual motion perception. These findings identify promising brain areas for feature extraction for the development of BCI using visual motion imagery.

INTRODUCTION

Brain-computer interface (BCI) technology enables direct communication between the brain and a computer. When the aim is to control devices such as a computer cursor and wheelchair with electroencephalography (EEG)-based BCI, most studies have used P300 evoked potential, several steady-state visual evoked potentials (SSVEPs), and sensory-motor imagery. These conventional approaches are well established, but technological problems remain. To obtain a P300 or SSVEP signal, users must constantly concentrate on watching an external stimulus presented on a monitor. On the other hand, sensory-motor imagery requires a long training time and, because brain signals related to the user's body movement are used to operate devices, multiclass control is difficult. In the BCI development, it is also important to search for new approaches that are more goal-directed and intuitive for users compared to conventional methods.

Visual motion imagery, which involves visualization of a target operation is one approach used in recent studies [1,2]. Understanding brain source activity is a key for practical application of any new BCI approach, and this understanding contributes not only to the field of BCI but also to the field of neuroscience. However, few studies investigated brain activity and mechanism during visual motion imagery using EEG.

In this study, we examined source localization during visual motion imagery using exact low-resolution brain electromagnetic tomography (eLORETA) and compared findings with visual motion perception**.** Our aim was to clarify differences in brain activity between imagery and perception. Details regarding the experiment are provided in the MATERIALS AND METHODS section; Statistical analysis results of the differences in source activities between imagery and perception are detailed in the RESULTS section.

MATERIALS AND METHODS

Participants: A total of 18 healthy male subjects participated in the experiments. Two participants were removed from the final analysis: one due to a technical error during recording, and one due to excessive noise in the EEG data. Fifteen of the 16 remaining participants were right-handed; one was left-handed. The average age of participants was 22.9 years (standard deviation, 2.2, range, 20-29 years). All participants had normal or corrected-to-normal vision. The experimental procedure was approved by the Research Ethics Committee of the Graduate School of Engineering, the University of Tokyo, Tokyo, Japan (approval number; KE17-31). All participants provided written informed consent prior to their participation in this study.

Experimental equipment: EEG recordings were performed in a shielded room, where participants were seated in a comfortable chair approximately 90 cm from a 76.2 cm screen (MultiSync LCD-PA302W, NEC Corp) with an effective display area of 641 × 401 mm. Participants were instructed to relax and refrain from excessive body or head movements. They were also instructed to fix their gaze on the middle of the monitor.

Experimental Procedure: During imagery, participants visually imagined the movement of a drone in three planes (up/down, left/right, and forward/backward) while watching a drone hovering at the center position of three-dimensional space presented on the monitor. To reduce individual differences in imagination and perception, immediately before each imagination, participants watched a movie in which the same drone moved in one direction (Fig. 1). The three-dimensional space was displayed at a 32° visual angle in width and 20° in height at the center of the monitor. Initially, the drone subtended 6.0° in width by 2.5° in height, hovering at the center position. It then moved in

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-22

one direction for 4 s (up/down at 0.83°/s, left/right at 1.53°/s, forward/backward at 0.51°/s). When the drone moved forward or backward, its size was gradually reduced to 60% or increased to 140% at a constant speed. The movie was presented for 4 s (perception time), the fixation point was presented for 1 s, then the hovering-drone image was presented for 4 s (imagery time) (Fig.2). Participants were instructed to reproduce the same movement they had observed during the perception time in the imagery time of each trial. Six trials defined one set, and the order of the six directions in each set was random. The entire task consisted of ten sets.

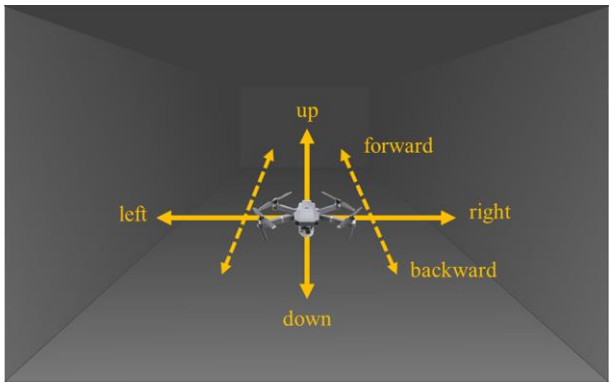


Figure 1: Illustration of the movie shown to participants

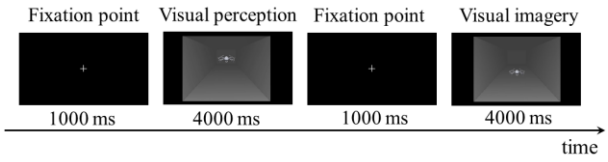


Figure 2: Experimental paradigm

Data acquisition: EEG signals were continuously recorded at a sampling rate of 1 kHz using an EEG system (EEG-1200, Nihon Kohden Corp., Tokyo, Japan). Electrodes were located at 65 positions according to the extended 10-20 system (Fig. 3). Two additional electrodes were fixed on the participant's earlobes (A1, A2) as references.

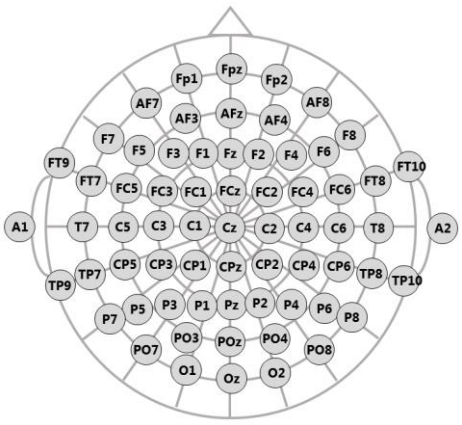


Figure 3: The 65 electrode locations of the extended 10-20 system and two electrodes (A1, A2) on the earlobes

Signal preprocessing: Data were analyzed offline using MATLAB® 2016b and EEGLAB (Delorme & Makeig, 2004) version 14.1.b. Recorded EEG signals were filtered between 1 Hz and 120 Hz using a 3300th-order finite impulse response filter, and powerline fluctuations at 50 Hz were removed using the Cleanline EEGLAB plug-in Electrooculographic artifacts due to blinks or eye movements and electromyographic (EMG) artifacts were removed using the Automatic Subspace Reconstruction (ASR) method implemented in the 'clean_rawdata' plugin of EEGLAB [3]. The threshold of ASR was set at 15 standard deviations with all the other parameters turned off. Single-trial epochs were extracted from −500 to 4000 ms relative to stimulus onset. Baseline was set as the mean signal 500 ms before onset of an perception or imagery period.

EEG source localization: The eLORETA was used to compute the intracortical distribution of current source density from multichannel head-surface EEG data [4]. Validation of LORETA for localization agreement with multimodal imaging techniques is reported in several published studies [5-7]. Further, previous research reported that eLORETA performs well in terms of less localization error and visibility compared to other low-resolution techniques (sLORETA) [8]. The head model and electrode coordinates were based on the Montreal Neurologic Institute average MRI brain (MRI152). Solution space was restricted to the cortical gray matter (6239 voxels at 5×5×5 mm spatial resolution). eLORETA functional images were computed for each subject in each of the following 6 frequency bands: delta (0.5-3.5Hz); theta (4-7.5Hz); alpha (8-12.5Hz); beta (13-30Hz); low gamma (30.5-60Hz); high gamma (60.5-120Hz). Considering differences in brain activity among participants, subject-wise normalization was performed before statistical analysis. This normalization used total power over all frequency bands and over all voxels from a participant to unity.

Statistical analysis: We used movement in six directions to investigate imagined direction classification for future BCI applications. However, our aim was to clarify differences in brain activity between imagery and perception. Hence, we did not distinguish state by direction. Thus, for source localization, we used imagery time as the period that participants visualized movement of the drone regardless of the imagined movement direction. Similarly, we used perception time as the period that participants watched the movement of the drone regardless of the movement direction. The differences of cortical source localization between conditions were assessed in each frequency band with voxel-by-voxel paired *t*-tests of log-transformed power. Randomized statistical nonparametric mapping (SnPM) was applied to determine critical probability threshold values for observed *t*-values with correction for multiple comparisons over all voxels and all frequencies [9]. A total number of 5000 permutations were used to determine the significance for each randomization test.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-22

RESULTS

Fig. 4 shows differences in brain activity in each frequency band between imagery and perception. The color becomes redder as brain activity during imagery is higher than during perception. The color becomes bluer as the brain activity during perception is higher than during imagery. In the delta and theta band, activity in the frontal lobe was relatively higher during perception, and activity over a large area of the parietal and occipital lobes was relatively higher during imagery (Fig. 4 (a), (b)). In the alpha band, activity in the left frontal lobe and parietal lobe, mainly on postcentral gyrus and inferior parietal lobule (Broadmann area 2, 3, 40), was relatively higher during visual perception, and activity in the left occipital lobes, mainly the cuneus and precuneus (Broadmann area 18, 23, 31), was relatively higher during imagery (Fig. 4 (c)). In the beta band, the activity in the right limbic lobe and left occipital lobe, mainly on the right posterior cingulate, the left cuneus and the left precuneus (Broadmann area 23, 7, 31), were relatively higher during imagery (Fig. 4 (d)). In the low gamma band, activity in the right parietal lobe and right temporal lobe, mainly in the inferior parietal lobule, superior temporal gyrus and supramarginal gyrus (Broadmann area 40, 39,) were relatively higher during imagery (Fig. 4 (e)). In the high gamma band, activity in the right frontal lobe (Broadmann area 8, 9), was relatively higher during visual perception and activity on the right parietal lobe and right temporal lobe, mainly on the inferior parietal lobule, the superior temporal gyrus, the precuneus and the angular gyrus (Broadmann area 40, 39, 7). were relatively higher during imagery (Fig. 4 (f)).
Two-tailed *t*-tests showed significant differences between imagery and perception conditions in the beta and low gamma band ($p < 0.05$ for *t*-values above 5.22, corrected for multiple comparisons). Current density in the beta band for imagery was significantly higher in the right posterior cingulate (Broadmann area 23, 30), in the left precuneus (Broadmann area 7, 31) and in the left cuneus (Broadmann area 7) compared to perception (Fig. 5(a), Tab. 1). In the low gamma band, activities in the right inferior parietal lobule and right supramarginal gyrus (Broadmann area 40), were significantly higher for imagery compared perception (Fig. 5(b), Tab. 1).

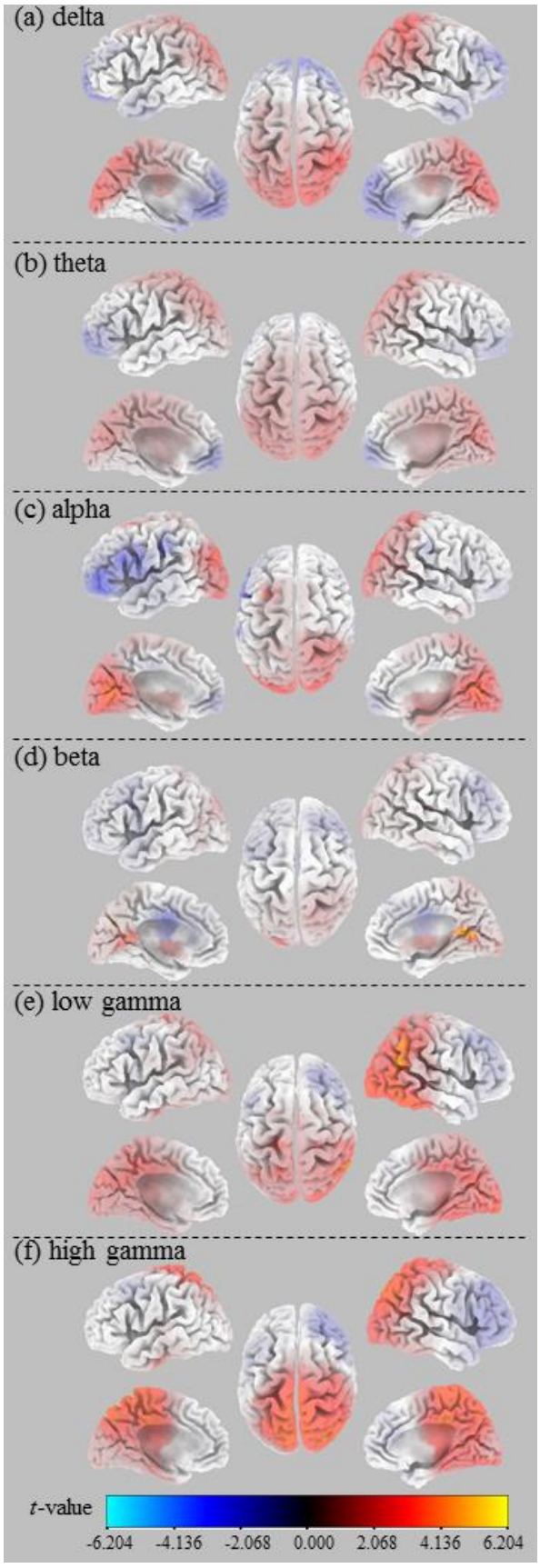


Figure 4: The difference of brain activity between imagery and perception in each frequency band ((a) delta, (b)theta, (c) alpha, (d) beta, (e) low gamma, (f) high gamma).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-22

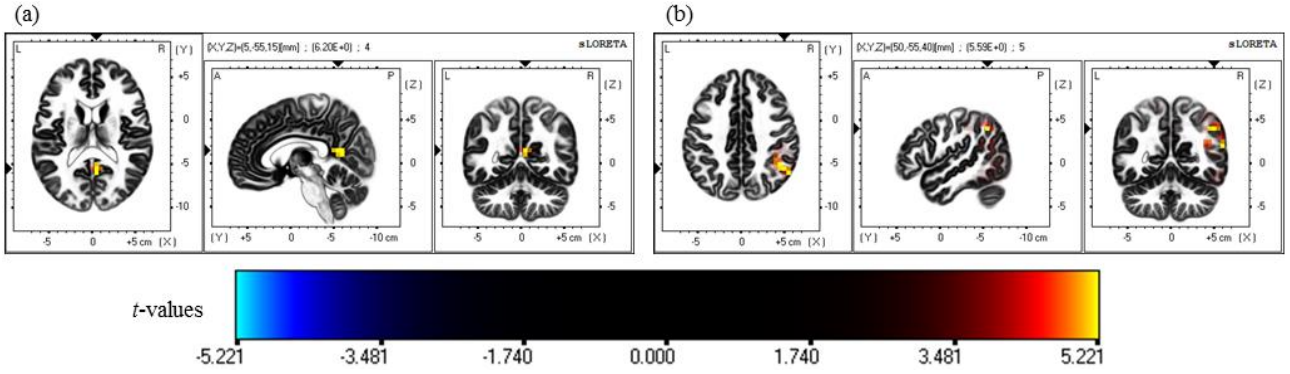


Figure 5: Significant differences in source current density between imagery and visual perception (red-coded and deep blue-coded for $p < 0.10$, $| t | > 4.75$; yellow-coded and light blue-coded for $p<0.05$, $| t | > 5.22$, corrected for multiple comparison). (a) Differences in the beta frequency band. (b) Differences in the low gamma frequency band. The color becomes yellower as brain activity in imagery is significantly higher than perception, and the color becomes bluer as the brain activity in perception is significantly higher than in imagery.

Table 1: Brain areas activated during imagery that are significantly higher than during perception

Brain Area	Broadmann area	MNI coordinate (mm)			Frequency band	*t*-value
		X	Y	Z		
Limbic Lobe, Posterior Cingulate	23	5	-55	15	Beta	6.20
	30	5	-60	10	Beta	5.49
	23	5	-60	15	Beta	5.29
Parietal Lobe, Precuneus	31	-15	-70	25	Beta	6.00
	7	-15	-70	30	Beta	5.74
Occipital Lobe, Precuneus	31	-15	-65	25	Beta	5.82
Occipital Lobe, Cuneus	7	-15	-75	30	Beta	5.68
Limbic Lobe, Precuneus	31	-15	-60	25	Beta	5.56
Parietal Lobe, Inferior Parietal Lobule	40	50	-55	40	Low gamma	5.59
	40	55	-60	40	Low gamma	5.25
	40	45	-50	40	Low gamma	5.22
Temporal Lobe, Supramarginal Gyrus	40	60	-55	20	Low gamma	5.30

t-value is the value of the statistical comparison with $p < 0.05$ for *t*-value above 5.22.

DISCUSSION

In the present study, we used eLORETA to identify brain regions where activation is associated with imagery via comparison to perception. Source activities induced in the beta band in the right posterior cingulate, left precuneus and left cuneus, and low gamma activity induced in the right inferior parietal lobule, and right supramarginal gyrus was significantly higher for imagery than perception.

There was no significant difference between perception and imagery in the alpha band where the signal-to-noise ratio is rather high. We speculate that the visual attention to the imagined movement of the drone induced the decrease in the alpha band activity in the same degree as the perceptual movement.

The precuneus is activated in several visual imagery studies using PET or functional MRI [10,11]. In this study, we used EEG for data acquisition and were able to observe the activity of precuneus induced by visual imagery. The precuneus (Broadman area 7,31) is known to have intimate interconnection with the posterior cingulate (Broadman area 23,31) and selective connections with the inferior parietal lobule and intraparietal sulcus. The latter is known to be important in visuospatial information processing [12]. Further, the ventro-dorsal stream, a processing pathway for visual information that reaches the primary visual cortex, proceeds to the inferior parietal lobule and is related to position and movement analysis of an object and to conscious of the object [13]. During perception, a participant is only watching movement of the drone, but during imagery, it is necessary to recall movement of the drone and produce a virtual visual image in the brain. We speculate that this difference in visuospatial information processing load strengthens the brain connection we describe above, and consequently, some brain areas, such as the posterior cingulate and inferior parietal lobule display significantly increased activity during imagery. Previous studies have reported that posterior cingulate is related to memory recall [14,15]. We used EEG and were able to observe the activity of posterior cingulate by visual imagery. From this result, we speculate that the participants were recalling the movie they watched

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-22

immediately before the imagery period.

CONCLUSION

We studied EEG source activity of visual motion imagery via comparison with visual motion perception. We found brain activity particular to visual motion imagery in the posterior cingulate, precuneus, cuneus, inferior parietal lobule, and supramarginal gyrus. Based on these results, we will analyze functional connectivity between brain regions related to visuospatial information processing in the future work. In addition, visual motion imagery may be a useful new approach to BCI. Therefore, we hope to investigate the possibility of classification of imagined directions focusing on the brain areas where increased activity was observed during visual motion imagery.

REFERENCES

[1] Sousa T, Amaral C, Andrade J, Pires G, Nunes UJ, Castelo-Branco M. Pure visual imagery as a potential approach to achieve three classes of control for implementation of BCI in non-motor disorders. J. Neural Eng. 2017;14(4): 046026.

[2] Koizumi K, Ueda K, Nakao M. Development of a Cognitive Brain-Machine Interface Based on a Visual Imagery Method, in Proc. EMBC'18, Honolulu, HI, USA, 2018, 1062-1065

[3] Mullen T, Kothe C, Chi YM, Ojeda A, Kerth T, Makeig S, ... & Jung TP. Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG, in Proc. EMBC'13, Osaka, Japan, 2013, 2184-2187

[4] Pascual-Marqui, RD, Lehmann D, Koukkou M, Kochi K, Anderer P, Saletu B, ..., Biscay-Lirio R. Assessing interactions in the brain with exact low-resolution electromagnetic tomography. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2011;369(1952):3768-3784.

[5] Mulert C, Jäger L, Schmitt R, Bussfeld P, Pogarell O, Möller HJ, Juckel G, Hegerl U. Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. Neuroimage. 2004;22(1): 83–94

[6] Worrell GA, Lagerlund TD, Sharbrough FW, Brinkmann BH, Busacker NE, Cicora KM, O'Brien TJ. Localization of the epileptic focus by Low-Resolution Electromagnetic Tomography in patients with a lesion demonstrated by MRI. Brain Topography. 2000;12(4): 273–282

[7] Zumsteg D, Wennberg RA, Treyer V, Buck A, Wieser HG. H215O or 13NH3 PET and electromagnetic tomography (LORETA) during partial status epilepticus. Neurology. 2005;65(10): 1657–1660

[8] Jatoi MA, Kamel N, Malik AS, Faye I. EEG based brain source localization comparison of sLORETA and eLORETA. Australas. Phys. Eng. Sci. Med. 2014;37(4): 713-721.

[9] Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human Brain Mapping. 2002;15(1): 1-25

[10] Mellet E, Petit L, Mazoyer B, Denis M, Tzourio N. Reopening the mental imagery debate: lessons from functional anatomy. Neuroimage. 1998; 8(2): 129–139.

[11] Ishai A, Haxby JV, Ungerleider LG. Visual imagery of famous faces: effects of memory and attention revealed by fMRI. Neuroimage. 2002; 17(4): 1729–1741.

[12] Cavanna AE, Trimble MR. The precuneus : a review of its functional anatomy and behavioral correlates. Brain. 2006;129(3): 564-583

[13] Gallese V. The "conscious" dorsal stream: embodied simulation and its role in space and action conscious awareness. Psyche. 2007;13(1): 1-20

[14] Maddock RJ, Garrett AS, Buonocore MH. Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. Neuroscience. 2001;104(3): 667-676.

[15] Sutherland RJ, Hoesing JM. Posterior cingulate cortex and spatial memory: A microlimnology analysis. In Neurobiology of cingulate cortex and limbic thalamus. Birkhäuser, Boston, MA 1993, pp. 461-477

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-23

AN ALGORITHM FOR DETECTION OF MULTIPLE BLINKS OF SINGLE AND BOTH EYES FROM EOG SIGNAL

I. Rejer[1], Ł. Cieszyński[1]

[1] Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Szczecin, Poland

E-mail: irejer@wi.zut.edu.pl

ABSTRACT: The paper presents an algorithm for blinks detection from an EOG (electrooculographic) signal. The algorithm is based on the analysis of time waveforms recorded from Fp1 and Fp2 and is capable to work as well in a fix trial length mode as in a free user mode. The paper covers the description of the algorithm and its verification via an experiment conducted with ten healthy subjects. During the experiment the recognition of nine blinking schemes was tested: single, double and triple blinks with left, right, and both eyes. The recognition rate calculated over nine blinking schemes and ten subjects was equal to 92%. When analyzing the results across different blinking schemes we found that: 1) it was much easier to recognize left-right blinks (96%) than both eyes blinks (83%); 2) the detection accuracy dropped with the increasing number of blinks (98% - single, 91% - double, 87% - triple blinks) but it was sufficiently high even for triple blinks.

INTRODUCTION

Eye blinks have all the features needed to establish a successful communication channel. They can be precisely characterized for an individual, easily modulated or translated to express the intention, and can be detected and tracked consistently and reliably [1]. Of course, they need some motor control from the user but if such control is possible, they constitute a fairly stable channel. Comparing to control signals used in nowadays BCIs (Brain-Computer Interfaces), the eye blinks provide more comfortable communication that SSVEP-BCI (Steady State Visually Evoked BCIs) and P300-BCI (BCI based on P300 component). First of all, they do not need any external device that would provide stimulations evoking the required brain response. That also means that nor flickering light or highlighting objects are constantly present at a user visual field. Moreover, since the different styles of eye blinks are evoked only at a user will, the user can decide when he wants to switch the interface from an 'off' to 'on' state. On the other hand, comparing EB-interface (Eye Blink) to MI-BCI (Motor Imagery BCI), the first one is much more stable and does not require any user training. At this moment both types of interfaces provide a comparable number of control states (usually no more than 2-4) but this can be easily changed with the algorithm described further in this paper.

The papers on eye blink detection usually describe recognition systems providing a rather limited number of control states, often only 1-4 states, [1-3]. This paper aims to propose a recognition algorithm that significantly increases this number. The proposed algorithm enables the recognition of succeeding blinks of left, right and both eyes. Since the algorithm allows to detect different number of blinks, the number of control states provided by the interface using this algorithm is adaptable. It starts from two states, corresponding to single blinks of left and right eye but can easily be extended to three, six, nine or more by using double, triple, and more blinks. Of course, since a single simultaneous blink with both eyes is a spontaneous physiological activity, it is usually excluded from the set of control states.

The algorithm is based on the EOG time waveforms recorded from two prefrontal contralateral channels. We do not use the EOG signal recorded from the standard positions, it is from the electrodes located around the eyes because we want to ensure the maximum comfort of the user using the EB-interface. In the experiment testing the algorithm accuracy, all the electrodes were incorporated into a forehead band and the users found it a more convenient solution than when the electrodes were placed around the eyes.

The paper covers both, the detailed description of the algorithm and its experimental verification. The experiment described in the paper was conducted with ten healthy subjects whose task was to blink according to the blinking schemes presented in the computer screen. During the experiment the recognition of nine blinking schemes was tested: single, double and triple blinks with left, right, and both eyes.

RELATED WORK

The eye blinks related signals are predominant in the frontal and prefrontal sites on the scalp. Peaks appearing in the signal waveform recorded from these areas inform that an eye open (negative peak) or eye close (positive peak) event took place. Hence, an eye blink event can be detected when the positive peak followed by a negative deflection in a signal waveform is formed. The algorithms on eye blink detection directly or indirectly utilize these features. For example, in [4] the kurtosis, a statistical coefficient that describes the relative flatness of the data distribution, was used to detect the eye blink

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-23

episodes. The kurtosis is a quite good indicator for eye blink episodes because it is positive for peaked data distribution (typical for eye blink) and negative for flat distribution (typical for noise). As the authors report in the paper, they were able to detect single, double and triple eye blinks only with kurtosis coefficient calculated from the data recorded with one bipolar EEG channel. However, the speed of the interface presented in the paper was somewhat limited because with these three choices one character per minute was selected from a 27-letters keyboard. The information on the accuracy of the proposed solution was not provided.

The kurtosis concept was also used in [5], where the kurtosis coefficient was applied in the RBF classifier. The classifiers inputs (calculated from a one-second window) were: maximum amplitude, minimum amplitude, kurtosis of the present window, kurtosis of the previous window, and kurtosis of the next window. The authors reported 76% accuracy over the test data. A neural classifier was also used in [3], where four control states corresponded to left wink, right wink, single blink, and double blink were applied to control the wheelchair movements. The classifier's inputs were DWT (Discrete Wavelet Transform) coefficients extracted from F7 and F8 channels and its sensitivity and specificity were equal to 80% and 75%, respectively.

Much more algorithms on eye blink detection and recognition can be found in the papers on eye blink artifacts correction [6,7]. However, these algorithms usually either deal with the signal in an offline mode, it is after the recording was completed, or correct the parts of the signal of the given characteristics. While the former often need long segments of signal acquired from a dense matrix of channels to work correctly (e.g. algorithms based on sources detection or trends recognition), the latter are not sensitive enough to distinguish between succeeding blinks.

All the papers mentioned so far extracted information about the blinking pattern directly from the EOG signal. However, this is not the only possibility. Another approach is based on visual detection of gaze direction and blinking episodes. For example, in [8] a 4-steps detection system based on methods for image recognition was proposed. The system first detects a face region (with Haar-like features), then extracts and eye region (via analyzing geometrical dependencies), and finally detects and classifies the eye-blinks as spontaneous (shorter than 200ms) or voluntary (longer than 200ms). Voluntary blinks are further classified as long or short and hence the algorithm provides two control signals that can be used for selecting letters from a virtual keyboard or controlling cursor movements on the screen. The authors reported the results of the experiment with 49 subjects where the average time of entering a single sign was equal to 16.8 s before the training session and 11.7 s after the training session and the accuracy of detecting control eye-blinks was equal to 99%.

RECOGNITION ALGORITHM

The EOG activity is a result of the difference in charge between the cornea and the bottom of the eye. If this activity is recorded from the forehead electrodes, it is positive when the eyeballs go up and negative when they go down. Because of the so-called Bell effect (eyeballs go in the opposite direction to the eyelids) the eyes closure induces positive potential and eyes opening – negative potential.

The eye-blink episode builds upon the phenomenon described above and hence has a very specific time pattern (Fig. 1a). It starts from a rapid and narrow positive peak that is followed by a slightly wider but also very deep negative trough. Exactly after the blink episode, two more components are formed, first – a significantly lower and wider positive component and second - a very small and wide negative come-back. When two or more blinks are followed one after another, the blinking scheme remains the same (one sharp positive peak followed by a negative trough) for each blink and the two after-blinking components at the end (Fig. 1b).

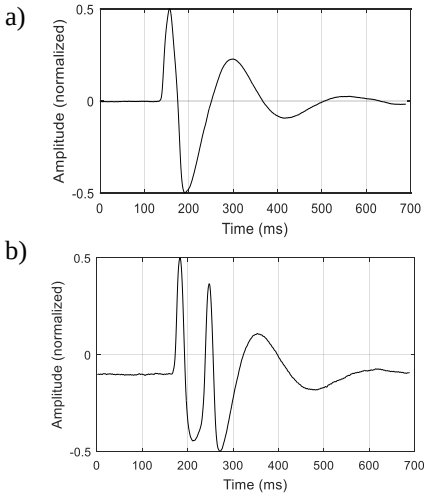


Figure 1: EOG signal from Fp1 during left eye blink event; a) single blink; b) double blink; non-filtered single-trial signals.

The height and width of the peak formed in the time course of EOG signal recorded from a single channel (preferably located at Fpz position) are the two characteristics that can be used to take a decision whether the blinking episode occurred or not and to distinguish between one, two or more consecutive blinks of one/both eyes. However, the algorithm based exclusively on counting the number of blinks provides a highly limited number of possible control states. This number can be increased (by a factor of three) if not only the number of blinks but also a blinking eye is detected from the signal. Of course, since the blinking scheme is the same for both eyes, it would be extremally difficult to distinguish the blinking eye from the signal recorded from an electrode located between both eyes (Fig. 2). This task is much easier when instead of one, two EOG channels are used.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-23

Figure 3 presents the EOG signal recorded from two prefrontal electrodes (Fp1 and Fp2) during a double blink of the left eye. As it can be noticed in this figure the involuntarily blink of the non-intended right eye is characterized by a much smaller amplitude at Fp2 (solid line) than the intended blink of the left eye (dashed line).

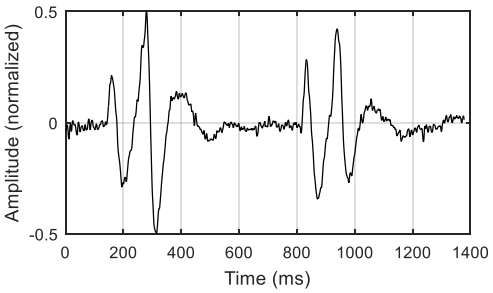


Figure 2: EOG signal from Fpz during double left eye blink (signal between 0-700ms) and double right eye blink (signal between 700-1400ms); non-filtered single-trial signal.

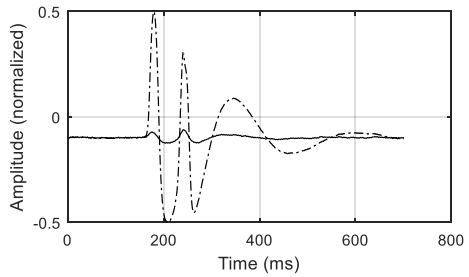


Figure 3: EOG signal from Fp1 (dashed line) and Fp2 (solid line) during double left eye blink; non-filtered single trial signal.

The last figure from this section (Fig. 4) presents the waveforms recorded from Fp1 and Fp2 during simultaneous blinks of both eyes. As it can be noticed this time peaks of both waveforms have the amplitude significantly greater than the rest of the signal. Usually, the amplitude of both peaks is not equal (because of small shifts in electrodes montage and also because of physiological differences between eyes), however, it is always well beyond the ongoing EEG activity.

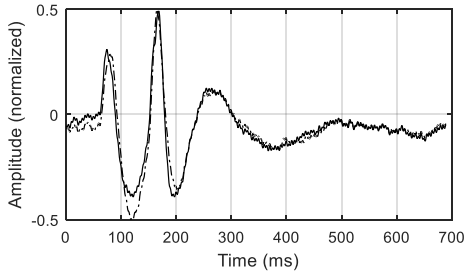


Figure 4: EOG signal from Fp1 (dashed line) and Fp2 (solid line) during double both eyes blink; non-filtered single-trial signal.

The algorithm that we propose to use for recognizing different eye blinking patterns from the EOG signal starts from the calibration session. The calibration session is necessary not only because of the subjects' differences in blink characteristics but also because these characteristics can be quite different even for the same subject in two succeeding experiments. Of course, the core of the blinking pattern does not change, the eye-close event evokes the positive peak and the eye-open event – a negative peak. However, the amplitude of the peaks in both channels can differ significantly. This what evokes this change are small shifts in the electrodes montage. Even slight differences in the electrodes' localization on the skull can highly suppress or enhance the peaks formed for left, right or both eyes. The task of the calibration session is to prevent the influence of these shifts on the algorithm accuracy. Therefore, the calibration session should be performed anytime when the user starts his/her work with the interface applying the proposed algorithm.

During the calibration session, the user is asked to blink once and twice with each eye individually and with both eyes together. Between both series, the user is asked not to blink at all. Hence, the calibration session is divided into seven segments, following one after another:

- S1, S2, S3 - single blink with left, right, and both eyes, respectively;
- S4 – no blinking period;
- S5, S6, S7 – double blink with left, right, and both eyes, respectively.

Assuming that three seconds are allocated to each condition, the whole calibration session lasts no more than 30 s. The signal acquired during the calibration session is used to calculate six parameters that are later used as thresholds in the recognition process. The description of parameters is presented in Tab. 1. All six parameters are calculated as the mean value from the two corresponding segments of the calibration session mentioned in the table.

Table 1: Parameters calculated from the calibration session

Name	Description	Segment
T1a_c	Hight of Fp1 peak corresponding to the left eye blink	S1, S5
T1b_c	Hight of Fp2 peak corresponding to the right eye blink	S2, S6
T1c_c	Hight of Fp1 peak corresponding to both eyes blink	S3, S7
T1d_c	Hight of Fp2 peak corresponding to both eyes blink	S3, S7
T2_c	Distance between peak and the consecutive zero crossing point (in samples)	S1, S2
T3_c	Distance between two succeeding peaks (in samples)	S5, S6

The parameters calculated over the calibration session are not directly used in the recognition process but first are modified to make them far less restrictive. The

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-23

modification rules are as follows:

- all four *T1_c* parameters are modified by a factor 0.5 (*T1=0.5T1_c)*;
- *T2_c* parameter is modified by a factor of 2.0 (*T2=2T2_c*);
- *T3* parameter is modified by a factor 2.0 (*T3=2T3*_c).

The algorithm is very flexible to the levels of these modification factors. It works correctly even if they are changed significantly. So far, we tested the algorithm for the levels: 0.4-0.6 (for T1s), 1.5-3 (for T2), and 1.5-3 (for T3).

The core part of the algorithm is composed of nine steps:

1. Find the first sample exceeding *min(T1a, T1c)* in Fp1 channel and/or *min(T1b, T1d)* in Fp2 channel.
2. Record the following samples (starting from the sample found in 1 until the signal crosses zero in both channels.
3. Find *max(Fp1)* and *max(Fp2)* and the distance between the sample of maximum value and the last sample from the recorded signal (*D1* – distance calculated over *Fp1, D2* - distance calculated over *Fp2*).
4. Decide on the eye (*left, right, both*) and the channel (*Fp1, Fp2*) that will be used in the process of searching for succeeding peaks. The decision on the eye (*left, right*, or *both*) and channel (*chosenChannel*) is taken according to one of two rules. Rule 1 is triggered when *T1a+T1b > T1c+T1d* and is defined as follows:
   a) if max(Fp1)>T1a & D1<T2 -> *left, Fp1, T1=T1a*
   b) if max(Fp2)>T1b & D2<T2 -> *right, Fp2, T1=T1b*
   c) *both*: if max(Fp1)>T1c & D1<T2 & max(Fp2)>T1d & D2<T2
      - if max(Fp1)>=max(Fp2) -> *Fp1, T1=T1c*
      - if max(Fp1)<max(Fp2) -> *Fp2, T1=T1d*

   Rule 2 is triggered when *T1a+T1b < T1c+T1d.* This rule is almost identical to Rule 1. The only difference is that condition *c* is tested before conditions *a* and *b*. The split into two rules is necessary because there are two blinking schemes across the subjects. While most subjects blink stronger when using both eyes simultaneously, there are also subjects who blink stronger when only one eye is used.
5. If none of the conditions given in 4 is fulfilled, discard all the recorded samples and start from 1. In the other case set *n=1* (*n* – number of peaks found in the signal), discard all the recorded samples and start looking for the next peak in the chosen channel (go to 6).
6. Find the first sample exceeding *T1* and record the following samples (starting from the this 'exceeding' sample) until the signal crosses zero.
7. Find *max(chosenChannel)* and the distance between the sample of maximum value and the last sample from the recorded signal (*D*).
8. If *max(chosenChannel)>=T1 & D<=T2* set *n=n+1*, discard all the recorded samples and start looking for the next peak in the chosen channel (go to 6).
9. If rule given in 8 is not fulfilled or the length of the recorded signal exceeds *T3* (in a free user mode) or exceeds the trial length (in a fixed trial length mode) return the number of peaks *n* and the chosen eye (*left, right*, or *both*) and start from 1.

MATERIALS AND METHODS

To test the algorithm described in the previous section, an experiment with ten healthy subjects (seven men and three women) was performed. All the subjects were right-handed, had normal vision and did not report any mental disorders. The experiment was conducted according to the Helsinki declaration on proper treatments of human subjects. Written consent was obtained from the subject before the experiment.

The detailed scheme of the experiment with one subject was as follows. The subject was placed in a comfortable chair and the electrodes were applied on his head. In front of the subject in an approximate distance of one meter an LCD monitor was located. A short sound signal announced the start of the experiment. Two seconds later the calibration session started. This session was performed two times: 1st - to make the subject familiar with the sequence of segments; 2nd - to calculate the algorithm parameters.

During the calibration session a sequence of pictures informing the subject of what was his/her task in the given segment was displayed (Fig. 5). Each picture was present on the screen for three seconds.

Figure 5: One of the pictures displayed during the calibration session (single blink of the left eye).

When the first calibration session ended, a picture presenting the EOG signal recorded during subject actions was displayed on the screen (Fig. 6). The experimenter discussed the subject's mistakes (wrong eye or wrong number of blinks) and explained once again what was the subject's task in each segment. Right then the real calibration session started.

Ten seconds after the calibration session the actual experiment started. During the whole experiment, a picture with nine texts informing the subject about his/her task was displayed on the screen (Fig. 7). Every three seconds a frame was displayed around one of the nine texts. The subject's task was to blink once, twice or three times with left, right or both eyes just as was indicated by the text surrounded by the frame. The experiment consisted of 36 trials, four trials per each blinking scheme.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-23

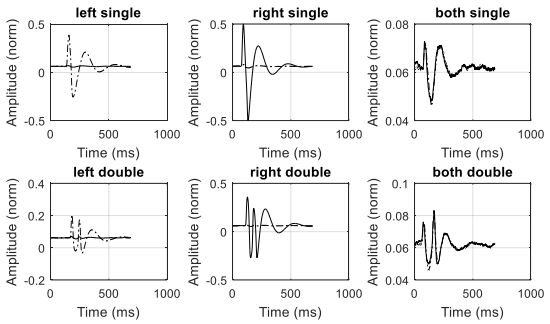


Figure 6: EOG signals recorded during the first calibration session with one of the subjects (Fp1 - dashed line, Fp2 - solid line).

EOG data was recorded from two monopolar channels at a sampling frequency of 250 Hz. Four passive gold cup electrodes were used in the experiments. Two of them were attached to the subject's skull at Fp1, and Fp2 positions according to the International 10-20 system [9]. The reference and ground electrodes were located at the left and right mastoid, respectively. The impedance of the electrodes was kept below 5 kΩ. The EOG signal was acquired with OpenBCI amplifier and recorded with OpenVibe Software [10]. The whole experiment, covering the preparation part, electrodes application, 2 calibration sessions, the actual experiment, and the after-cleaning lasted about 20 minutes with one subject.

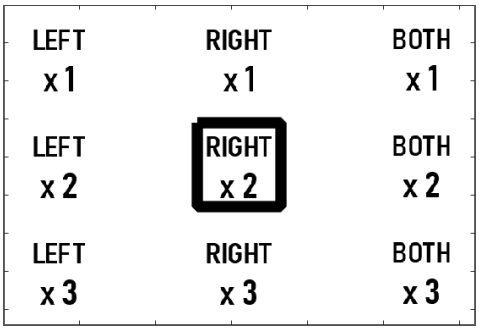


Figure 7: The tasks' panel.

RESULTS AND DISCUSSION

The classification accuracy calculated from all 36 trials for each subject is presented in Fig. 8 and the last column of Tab. 3. As it can be noticed, the accuracy was quite high for most of the subjects. For S1, S7, S8, and S9 it was even higher than 97%. However, there were also some subjects (most of all S6 and S2) that achieved much lower accuracy (S6 - 78%, S2 – 83%). Of course, having in mind that this accuracy was obtained for 9-classes classification, it was still a very good result. To find out whether these differences in the classification precision were caused by a systematic algorithm error or rather a subject specificity, we decomposed the results into 6 schemes. The three first schemes corresponded to a different blinking eye (left, right or both), the next three schemes - to a different number of blinks (1x - single, 2x - double, or 3x - triple). Each scheme covered 12 of the trials from our 36-trial experiment.

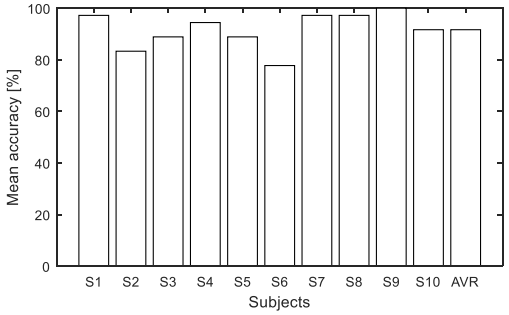


Figure 8: Classification accuracy for each subject calculated over all 9 blinking schemes.

Figures 9-10 and Tab. 2 present classification accuracy obtained for each subject under different blinking eye schemes (Fig. 9) and different number of blinks (Fig. 10). The first conclusion that should be made here is that regardless of the number of blinks, it was much easier to recognize left-right blink than both eyes blink. While the single eye recognition rate (averaged over different number of blinks) was equal to almost 96% (for both: right eye and left eye), it was significantly smaller (83%) for both eyes (according to paired sample *t*-student test with p=0.05).

The second observation that can be deduced from Tab. 2 and Fig. 10 is that the accuracy dropped with the increasing number of blinks from 98% (for single blinks) to 91% (for double blinks) and 87% (for triple blinks). While this drop was significant when comparing single to triple blinks (according to paired sample *t*-student test with p=0.05), it was incidental for single and double or double and triple blinks.

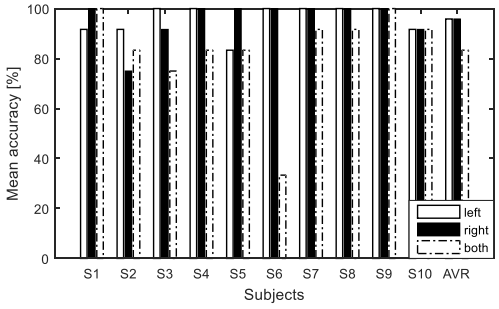


Figure 9: Classification accuracy for each subject calculated separately for left, right, and both eyes condition.

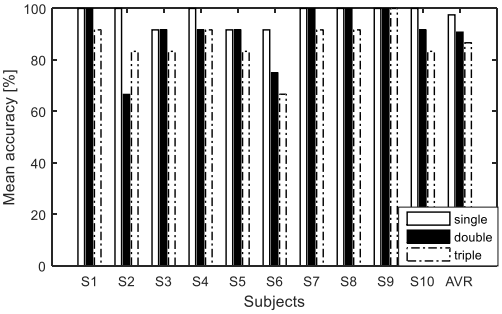


Figure 10: Classification accuracy for each subject calculated separately for single, double, and triple blinks.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-23

Table 2: Classification accuracy for each subject across analyzed number of blinks and blinking eyes (1x, 2x, 3x – single, double, and triple blinks; L, R, B – left, right, and both eyes).

	L	**R**	**B**	**1x**	**2x**	**3x**	**All**
S1	91.7	100.0	100.0	100.0	100.0	91.7	**97.2**
S2	91.7	75.0	83.3	100.0	66.7	83.3	**83.3**
S3	100.0	91.7	75.0	91.7	91.7	83.3	**88.9**
S4	100.0	100.0	83.3	100.0	91.7	91.7	**94.4**
S5	83.3	100.0	83.3	91.7	91.7	83.3	**88.9**
S6	100.0	100.0	33.3	91.7	75.0	66.7	**77.8**
S7	100.0	100.0	91.7	100.0	100.0	91.7	**97.2**
S8	100.0	100.0	91.7	100.0	100.0	91.7	**97.2**
S9	100.0	100.0	100.0	100.0	100.0	100.0	**100.0**
S10	91.7	91.7	91.7	100.0	91.7	83.3	**91.7**
Mean	**95.8**	**95.8**	**83.3**	**97.5**	**90.8**	**86.7**	**91.7**

Finally, we aggregated the results obtained from different subjects to find out which of our 9 classification schemes was correctly recognized in most cases. The aggregated results are presented in Tab. 3. As it can be noticed in the table the best detection rate was obtained for single right eye blink (100%), then for single and double left eye blink (98%), and next for double right and single both eyes blink (95%). If instead of the nine blinking schemes, only the first four were used (excluding single blink with both eyes as spontaneous physiological activity), the average accuracy would reach exactly 97.5%.

Table 3: The mean classification accuracy for each of the nine blinking schemes; results aggregated over all subjects.

	Left	Right	Both	Mean
1 x	**97.50**	**100.00**	95.00	97.50
2 x	**97.50**	**95.00**	80.00	90.83
3 x	92.50	92.50	75.00	86.67
Mean	95.83	95.83	83.33	

CONCLUSION

As it was shown in the paper, the proposed detection algorithm can recognize a quite high number of control states with a high detection rate, without any external stimulations and with use of a very short (30 s) calibration. Therefore, for the users that can still control their eyelids muscles, it can provide a good alternative to the brain-computer interfaces.

The main outcomes from the experiment described in the paper are: 1) the overall detection rate calculated over nine blinking schemes and ten subjects was equal to 92%; 2) the detection rate calculated over four best blinking schemes was equal to 98%; 3) it was much easier to recognize left-right blinks (96% - right eye and 96% - left eye) than both eyes blinks (83%); 4) the detection accuracy dropped with the increasing number of blinks (98% - single, 91% - double, 87% - triple blinks).

We believe that the algorithm described in this paper might be further improved by enhancing the recognition of both eyes' blinks. Moreover, to formulate more valid conclusions we plan to conduct experiments with more subjects and a greater number of blinks.

REFERENCES

[1] Saeid S, Chambers JA. EEG signal processing, ch.1, pp. 1-18, John Wiley and Sons Ltd (2007)

[2] Li Y, He S, Huang Q, Gu Z, Yu ZL. A EOG-based switch and its application for "start/stop" control of a wheelchair. Neurocomputing, 2018, 275, pp. 1350-1357

[3] Ahmed KS. Wheelchair movement control VIA human eye blinks. American Journal of Biomedical Engineering, 2011, 1(1), pp. 55-58

[4] Brijil C, Rajesh S, Rameshwar J. Virtual keyboard BCI using Eye blinks in EEG. In: 2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2010. p. 466-470

[5] Rihana S, Damien P, Moujaess T. Efficient eye blink detection system using RBF classifier. In: Biomedical Circuits and Systems Conference (BioCAS), IEEE, 2012. pp. 360-363

[6] Di Flumeri G, Aricó P, Borghini G, Colosimo A, Babiloni F. A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016. pp. 3187-3190

[7] Wang G, Teng C, Li K, Zhang Z, Yan X. The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition. IEEE journal of biomedical and health informatics, 2016, 20(5), pp. 1301-1308.

[8] Królak A, Strumiłło P. Eye-blink detection system for human–computer interaction. Universal Access in the Information Society, 2012, 11(4), pp. 409-419

[9] Jasper HH, The ten-twenty electrode system of the international federation in electroencephalography and clinical neurophysiology, EEG Journal, 1958, Vol. 10, 371–375

[10] Renard Y, Lotte F, Gibert G, Congedo M, Maby E, Delannoy V. et all. 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, Vol. 19, No 1

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-24

THE EFFECT OF HIGH AND LOW FREQUENCIES IN C-VEP BCI

M. Borhanazad[1], J. Thielen[1], J. Farquhar[1], P. Desain[1]

[1]Radboud University, Donders Centre for Cognition, Nijmegen, the Netherlands

E-mail: ms.borhanazad@gmail.com

ABSTRACT

Broadband code modulated visual evoked potential (BB-VEP, c-VEP) is the basis of one of the fastest brain-computer interface (BCI) paradigms. Unlike other systems, like those based on steady-state visual evoked potential (SSVEP, f-VEP), the stimulus specificity of c-VEP has not been thoroughly studied yet. One of the important stimulus characteristics that can influence both performance and user comfort is the frequency (the bit clock or frame rate). In this study, we evaluated the effect of stimuli presented at various frame rates (40, 60, 90 and 120 Hz) on c-VEP using LED lights. Accuracy and ITR were used to assess the performance and a questionnaire was used to evaluate the visual comfort. No significant differences in the performance of different frequencies were found, so comfort can be the main factor in the design decision. However, there is a trend for the frame rates of 40 and 90 Hz to yield a higher accuracy as compared to 60 and 120 Hz.

INTRODUCTION

One of the well-known brain computer interface (BCI) applications is a speller, in which the user can select a letter and type it on the screen. Different paradigms are used to develop such BCI spellers, in which the frequency-modulated visual evoked potential (f-VEP) and the code-modulated visual evoked potential (c-VEP) have gained increasing attention due to their high information transfer rate (ITR) [1].

In an f-VEP-based BCI, the user is presented with visual stimuli flickering at different frequencies. The attended stimulus generates steady-state visual evoked potential (SSVEP) responses over the occipital brain areas at the same (or harmonics) frequency of the stimulus [2]. A wide range of frequencies (between 4 and 90 Hz) can be used to elicit SSVEPs [2, 3]. The lower frequency range (<20 Hz) induce higher amplitude SSVEP responses compared to the higher frequency stimuli. However, the higher frequency stimuli are visually more comfortable and cause less visual fatigue [4]. Many research groups have investigated the effect of different frequencies on the SSVEPs response [5] and on the BCI performance [6]. For example, to compare the accuracies from high and low frequencies, Won and colleagues designed an experiment using 30 LED lights flickering at different frequencies. In their studies, they achieved higher classification accuracy using high-frequency stimuli (26-34.7 Hz) compare to the low-frequency stimuli (6-14.7 Hz) [6, 7]. Unlike the f-VEP, the stimulation frequency of c-VEP has not been thoroughly studied yet.

In c-VEP, a set of pseudo-random bit-sequences (codes) is used to modulate intensity of the visual stimuli. Though many types of codes are used in BCI, the stimulation codes are often a single m-sequence with a different phases, or are constructed as a Gold code family generated with a special pair of m-sequences. They can be run at different stimulation frequencies, we will call them frame rates, even though we use LED's and no video frame is involved. In most of the c-VEP BCI studies, due to limitations in display refresh rates, 60 and 120 Hz are commonly used in the experimental setups. However, using the latest monitors with a higher refresh rate, recent studies investigated the effect of higher stimulation frequencies on c-VEP [8, 9]. In the study by Gembke and colleagues, they compared the results of high-frequency stimulation rates (120 and 200 Hz) with the traditional 60 Hz and found a lower but comparable result in terms of accuracy and ITR for 200 Hz compared to 60 and 120 Hz. Başaklar and colleagues, examined stimulus presentation rates of 60, 120, 240 Hz and revealed that 120 Hz is better than 240 Hz for higher number of classes. Nevertheless, to understand the effects of different stimulation frequencies and their boundaries on c-VEP systems more research is needed.

In this study, we used modulated Gold codes to investigate the effects of low and high stimulation frequencies (40, 60, 90 and 120 Hz) in terms of performance and visual comfort. We used LEDs instead of a monitor to be more flexible in presenting the stimulation frequencies. In addition, in order to make the BCI experiment more practical and pleasant for the participants, water-based EEG electrodes were used.

MATERIALS AND METHODS

Participants: 10 subjects (age: 31.2 ± 11.71, average ± standard deviation, 5 male) with normal or corrected-to-normal vision took part in the experiment. The subjects agreed voluntarily to participate in the experiment. The study was approved by and conducted in accordance with the guidelines of the Ethical Committee of the Faculty of Social Sciences at the Radboud University. Prior to the experiment, all subjects read and signed a written

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-24

informed consent.

Experimental design: The visual stimuli were presented to the subjects using eight RGB LED lights. Each LED was surrounded by a plastic cylinder to create a circular target (4 cm diameter, 1 cm inter-target distance). Then, all LEDs were covered using a light diffusion sheet to prevent uneven light distribution. Subsequently, eight small transparent sheets with a picture of an animal were placed at the center of each light as shown in Figure 1(A). The LEDs were controlled by an ARDUINO UNO micro-controller to send the eight flashing patterns of the modulated Gold codes (See further, m-Gold codes). During the experiment, subjects were sitting on a chair in front of the LED box with a distance of 35 to 50 centimeters. The exact distance was chosen by each participant for a comfortable view of the LED box. The experiment consisted of four runs, one for each frame rate stimuli (i.e., 40, 60, 90, 120 Hz). Each run consisted of 10 trials of 4.2 seconds at that specific frame rate with an inter-trial interval of one second. In each trial, one LED was cued randomly as a target. After the cue, all LEDs started flashing their own code at the specific frame rate, for 4.2 seconds. The time course of one trial is shown in Figure 1(B). The subjects were instructed to focus their attention to the cued target and gaze at its center during the stimulation. The stimulation frequencies were presented to the subjects in a random order. To prevent visual fatigue, the subjects could take a short break between each run.

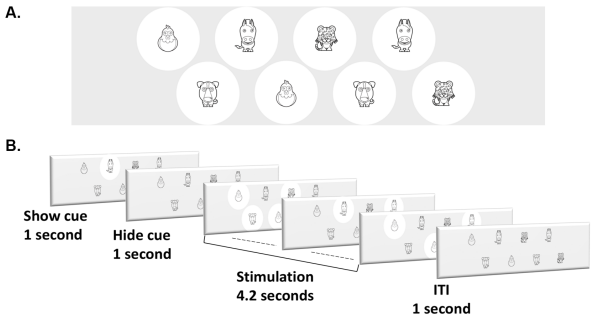


Figure 1: A. Layout design of the LED box, all are blinking at the same frame rate, but with different bit sequences; B. Time-course of one trial during the experiment. ITI: Inter trial interval.

m-Gold codes: Gold codes are a set of pseudo-random bit sequences with minimum cross-correlation and optimal auto-correlation [10]. A set of Gold codes is formed by XORing a preferred pair of m-sequences of the same length with a time-lag. Each time-lag yields a new member of a Gold code family. An m-sequence is generated by a special type of linear feedback shift register (LFSR) that has a longest non-repeating sequence for a given tap sequence. Although an m-sequences has better auto-correlation properties than a Gold code, a set of m-sequences may have large and unpredictable cross-correlation values.
In this study, we used a set of Gold codes created with a 6-bit linear feedback shift register and feedback taps at positions 6,5,2,1 and at 6,1 (See [10, 11] for the generation process). In our study, the Gold codes were modulated with a double bit clock to force a transition in each bit. These modulated codes (m-Gold codes) retained the good correlation properties while restricting the run-length distribution, as a code can be considered as a series of short (i.e., '10' or '100') and long (i.e., '110' or '1100') on-off runs, which represent short and long flashes, respectively (See Figure 2). The family consisted of 65 codes, each with a 126 bits length. In this study, we used the first eight sequences of these m-Gold codes to send to the eight LED lights.

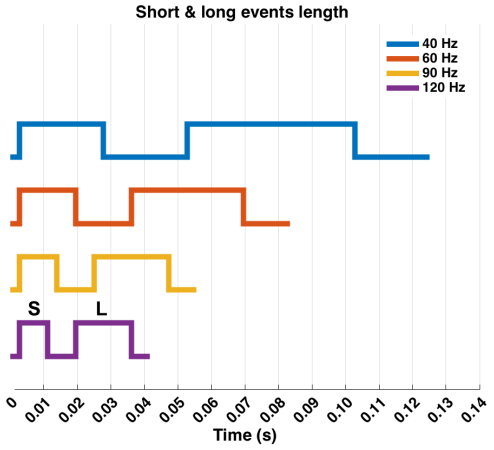


Figure 2: Pulse duration for 4 bits of the first sequence of the modulated Gold codes in different frequencies. S: short event; L: Long event.

Recording: EEG was recorded using the Twente Medical Systems International (TMSi) Porti system with 32 EEG channels and a common average reference at a sampling rate of 1000 Hz. The EEG data was collected using seven water-based electrodes located over the parietal and occipital cortex at CP5, P3, Pz, POz, Oz, P4, and CP6 according to the international system of EEG measurement [12]. The ground (GND) electrode was located over the left temporal site (T7).

Preprocessing: The raw signals were down sampled to 360 Hz. To be able to create suitable roll-on roll-off spectral sensitivity the down sampled signals were high-pass filtered at 2 Hz using a 2nd order Butterworth filter, and low-pass filtered at 50 Hz using a 4th order Chebyshev type II filter with 50 dB stop-band attenuation. The stimuli presentation and the preprocessing were implemented online using the BrainStream software [www.brainstream.nu].

Comfort scores: To assess the visual comfort at each frame rate, subjects filled out a four-level satisfaction questionnaire, (with levels: very uncomfortable = 0,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-24

uncomfortable = 3, comfortable = 7, very comfortable = 10). Participants rated the questionnaire based on their subjective experience after each run of the experiment. Thereafter, we calculated the mean score over all subjects for each frame rate.

Analysis: In the c-VEP, template matching is applied to identify the attended target. To create a template, we used a generative model called re-convolution (see [13] for details). Re-convolution is a method that models the brain response to a sequence of events (here, m-Gold codes) as the linear summation of the responses to the individual events (i.e., short and long flashes, see Figure 2). The re-convolution method used in [13] has been implemented in two sequential steps. First, it generates transient responses by performing deconvolution and subsequently learns the spatial distribution by means of Canonical Correlation Analysis (CCA). Here, we integrated re-convolution in the CCA to simultaneously derive both temporal as well as spatial dynamics of the responses [14]. Therefore, we defined the CCA as follows:

$$W,R = \underset{W,R}{argmax} \frac{W^\top X \cdot M^\top R}{\sqrt{W^\top XX^\top W \cdot R^\top MM^\top R}} \quad (1)$$

Where, $X \in \mathbb{R}^{k \cdot m,c}$ is the concatenation of k single-trials of m samples and c channels, and $m \in \mathbb{R}^{k \cdot m,c}$ is the concatenation of k structure matrices of m samples and c channels. The structure matrix, M, is a design matrix created for each event type (i.e short and a long flash), which in the first column, lists a 1 whenever the event type occurred, and zero elsewhere, and in each subsequent column the ones shifts down a row to the length of the transients (See Fig.5 in [13]). Therefore, M contains two structure matrices, $M = [M_s, M_l]$, with M_s the structure matrix listing a 1 when the short event happened, and M_l listing 1 whenever the long event happened. Thus, the CCA derives a spatial filter $W \in \mathbb{R}^{c,1}$ and a temporal filter $R \in \mathbb{R}^{l,1}$ with the responses to the short (R_s) and long (R_l) events, $R = [R_s, R_l]$. The CCA does that by optimizing the spatial and temporal filter in such a way that the correlation between spatially filtered data ($W^\top X$) and the predicted responses ($R^\top M$) is maximized. Thereafter, a template matching can be performed to classify a new single-trial using W and R as follows:

$$y = \underset{i}{argmax}\{corr(W^\top x, R^\top M_i)\} \quad (2)$$

in which, M_i is the structure matrix for the ith class and $T_i = R^\top M_i$ is the predicted responses for stimulus i.

Performance evaluation: To estimate the system performance, we measured the accuracy of the classifier, the corresponding information transfer rate (ITR) [15], as well as (correct) symbols per minute (SPM) [16], which includes the time needed for a backspace to correct misspelled symbols. The accuracy was measured using 10-fold cross-validation. The maximum correlation between the spatially filtered data and the generated templates was used as an indicator of the predicted class. We measured the ITR (in bits/minute) and the SPM (in symbols/minute) calculated as follows:

$$ITR = (\log_2 N + P\log_2 P + (1-P)\log_2 \frac{1-P}{N-1}) * (\frac{60}{T}) \quad (3)$$

$$SPM = (P - (1-P)) * \frac{60}{T} \quad (4)$$

Where N is the number of classes, P is the detection accuracy of targets and T is the time needed to convey each symbol, including both trial time and inter-trial time (in seconds). The maximum ITR value during the stimulation time was determined for each subject at each bit rate, to report as ITR.

Statistics: Since accuracy rates and comfort scores were not normally distributed, a non-parametric Friedman test was performed. This test was used to compare the result of each pair of frame rate stimuli. The significant threshold was set to 0.05.

RESULTS

Comfort scores: Figure 3 shows the score provided by volunteers about the level of visual comfort regarding different frame rate stimuli. As can be seen, the higher frequencies tend to be more comfortable than the lower frequencies. None of the subjects reported "(very) uncomfortable" for the frame rate of 120 Hz and in contrary, no subjects reported "very comfortable" for the frame rate of 40 Hz. Interestingly, some subjects reported a difficulty to attend the 120 Hz stimulation because the flicker was less perceptible compared to the other frequencies.
To determine any significant differences in the comfort scores due to different frame rate stimuli, we performed non-parametric Friedman test. The results from Friedman test revealed a significant main effect of frame rate ($x^2(3) = 17.077$, $p = 0.001$) on comfort scores. Subsequent post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction, resulting in a significance level set at $p < 0.008$. Interestingly, the results showed only a statistically significant difference between the frame rates of 40 and 120 Hz (Z = -2.762, p = 0.006). No significant differences were found between other pairs of the frame rate stimuli (40 vs. 60 Hz (Z = -1.801, p = 0.072); 40 vs. 90 Hz (Z = -2.585, p = 0.010); 60 vs. 90 Hz (Z = -1.633, p = 0.102); 60 vs. 120 Hz (Z = -2.264, p = 0.024); 90 vs. 120 Hz (Z = -1.511, p = 0.131)).

Performance evaluation: In addition to classification accuracies, Figure 4 also shows the corresponding ITRs (in bits/min) and the SPM (in sym/min) for the different frame rate stimuli. In terms of accuracy, the frame rates of 40 and 90 Hz gained higher accuracy and SPM compared to the 60 and 120 Hz. Looking at the results of the ITR, the frame rate of 40 Hz leads to the highest ITR of

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-24

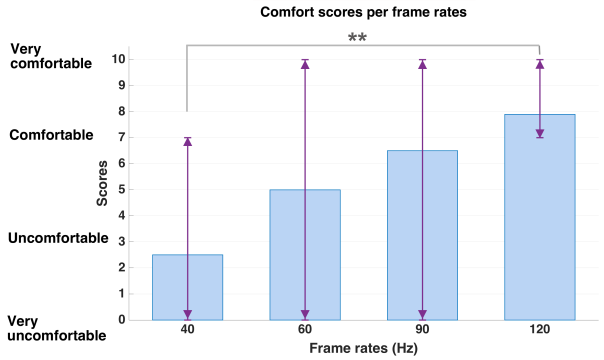


Figure 3: Scores of visual comforts provided by subjects for different frame rate stimuli. The error bars represent the maximum and minimum of the scores. ** $p < 0.01$.

53.69 ± 7.33 (mean ± SE, in bits/min) and followed by the frame rate of 60 Hz (50.41 ± 6.75) and 90 Hz (50.03 ± 6.52). In comparison, the frame rate of 120 Hz gained the lowest ITR (37.53 ± 6.89). However, the corresponding Friedman test did not reveal a significant difference between any pair of stimulation frequencies, neither in accuracy ($x^2(3) = 5.76$, $p = 0.12$) nor ITR ($x^2(3) = 6.45$, $p = 0.09$).

To inspect the evolving classification performance over time, we calculated the accuracy for each frame rate, starting from 0.5 second with time intervals of 0.1 second. The average results are depicted in Figure 5. As can be seen in this figure, the lower frequencies (40 and 60 Hz) have a tendency for better performance for short time intervals below 2 seconds as compared to the higher frequencies (90 and 120 Hz). The performances then converge close to ceiling performance at 4 second without significant difference. The frame rate of 120 Hz has a tendency to the lowest accuracy.

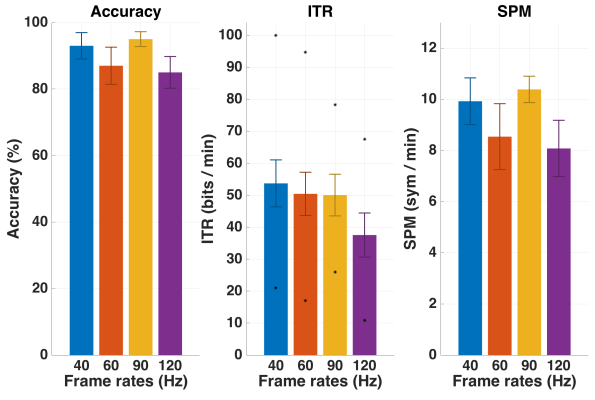


Figure 4: Average accuracy, ITR and SPM over subjects for different frame rate stimuli. The error bars indicate the standard errors. The starts show the maximum and the minimum ITR over subjects. Note: the ITR is calculated based on the maximum value during the stimulation time at each frame rate with an inter-trial interval of one second.

DISCUSSION AND CONCLUSION

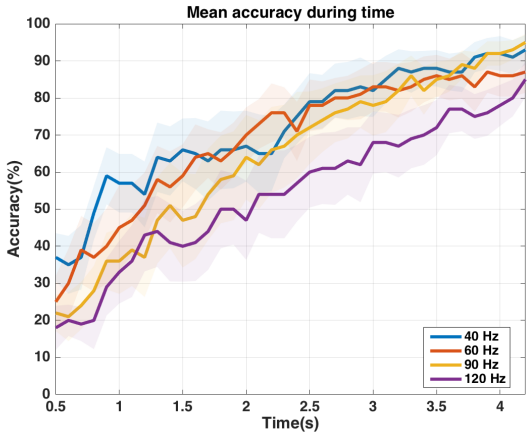


Figure 5: Average classification accuracy with respect to time for all frame rate stimuli. The shaded bars show standard errors.

In this study, we investigated the effect of high and low stimulation frequencies (40, 60, 90, 120 Hz) on c-VEP in terms of accuracy and ITR. A family of modulated Gold codes was used to control the visual stimuli on eight LED lights. The responses to these m-Gold codes were predicted by re-convolution and used as templates for a template matching classifier. To evaluate the visual comfort, a questionnaire was filled in by the subjects after each run of the experiment.

Comfort scores: The results from the questionnaire showed that most participants favor the higher stimulation frequencies especially 120 Hz while 40 Hz was irritating for most of the participants. This can be due to the fact that for high frame rates the stimuli are less perceived as flicker and this may be less tiring and more comfortable.

Performance evaluation: In terms of performance, there was a tendency for 40 and 90 Hz to yield higher accuracy compared to 60 and 120 Hz and for 90 Hz to achieve the highest accuracy and SPM. If we assume this gain will prove to be robust, 90 Hz should be considered as a good frame rate for c-VEP where traditional 60 and 120 Hz has been used. It is a happy coincidence that in Virtual Reality goggles 90 HZ is used often as frame rate. Previous studies have shown that the average accuracy from 120 Hz monitor refresh rate was slightly higher than 60 Hz [8, 9], however, our results don't substantiate that. Although the results about visual comfort showed the highest score for 120 Hz, some subjects reported a difficulty to attend the 120 Hz. This can be explained by the fact that the higher frequencies are less perceivable compared to the lower frequencies which can make it difficult for subjects to attend.

Our results showed no significant differences in the performance (accuracy and ITR) of different stimulation frequencies. This may be due to the effect of two competing mechanisms: higher frame rates yield more events per time unit thus in principle can support higher transfer rates, but the responses to fast events become smaller thus leading to a weaker discrimination. The fact that the frame rate does not affect performance much allows for

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-24

choosing 120 Hz, the most comfortable one without loss of performance. However there was a tendency for 90 Hz to be better than 120 Hz in terms of accuracy and ITR that may turn out to make up for somewhat less comfort.

ACKNOWLEDGEMENT

We would like to thank Philip van den Broek for technical assistance, Jos Wittebrood at Mindaffect company for helping in the setup, Tsvetomira Tsoneva for her help and insight and the participants for sharing their time.

REFERENCES

[1] Gao Shangkai, Wang Yijun, Gao Xiaorong, Hong Bo. Visual and auditory brain–computer interfaces. IEEE Transactions on Biomedical Engineering. 2014;61(5):1436–1447.
[2] Regan David. Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine. 1989.
[3] Herrmann Christoph S. Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. Experimental Brain Research. 2001;137(3-4):346–353.
[4] Diez Pablo F, Mut Vicente A, Perona Enrique M Avila, Leber Eric Laciar. Asynchronous BCI control using high-frequency SSVEP. Journal of Neuroengineering and Rehabilitation. 2011;8(1):39.
[5] Berumen G, Tsoneva T. Steady state visual evoked potentials at the boundaries of visual perception. Proceedings of the 7th Graz Brain-Computer Interface Conference (2017).
[6] Won Dong-Ok, Hwang Han-Jeong, Dähne Sven, Müller Klaus-Robert, Lee Seong-Whan. Effect of higher frequency on the classification of steady-state visual evoked potentials. Journal of Neural Engineering. 2015;13(1):016014.
[7] Won Dong-Ok, Zhang Hai Hong, Guan Cuntai, Lee Seong-Whan. A BCI speller based on SSVEP using high frequency stimuli design. In: Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on. IEEE. 2014, 1068–1071.
[8] Gembler Felix, Stawicki Piotr, Rezeika Aya, Saboor Abdul, Benda Mihaly, Volosyak Ivan. Effects of Monitor Refresh Rates on c-VEP BCIs. In: International Workshop on Symbiotic Interaction. Springer. 2017, 53–62.
[9] Başaklar Toygun, İder Yusuf Ziya, Tuncel Yiğit. Effects of High Stimulus Presentation Rate on c-VEP based BCIs. In: 2018 Medical Technologies National Congress (TIPTEKNO). IEEE. 2018, 1–4.
[10] Gold Robert. Optimal binary sequences for spread spectrum multiplexing (Corresp.) IEEE Transactions on Information Theory. 1967;13(4):619–621.
[11] Golomb Solomon W et al. Shift register sequences. Aegean Park Press, 1967.
[12] Oostenveld Robert, Praamstra Peter. The five percent electrode system for high-resolution EEG and ERP measurements. Clinical Neurophysiology. 2001;112(4):713–719.
[13] Thielen Jordy, Broek Philip van den, Farquhar Jason, Desain Peter. Broad-Band visually evoked potentials: re(con)volution in brain-computer interfacing. PloS One. 2015;10(7):e0133797.
[14] Thielen J, Marsman P, Farquhar J, Desain P. Re(con)volution: Accurate response prediction for broad-band evoked potentials-based brain computer interfaces. In: Brain-Computer Interface Research, 2017, 35–42.
[15] Wolpaw Jonathan R, Ramoser Herbert, McFarland Dennis J, Pfurtscheller Gert. EEG-based communication: improved accuracy by response verification. IEEE Transactions on Rehabilitation Engineering. 1998;6(3):326–333.
[16] Schreuder Martijn, Höhne Johannes, Treder Matthias, Blankertz Benjamin, Tangermann Michael. Performance optimization of ERP-based BCIs using dynamic stopping. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE. 2011, 4580–4583.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-25

ADAPTIVE PARAMETER SETTING IN A CODE MODULATED VISUAL EVOKED POTENTIALS BCI

F. Turi[1], M. Clerc[1]

[1] Athena Project-Team, Inria Sophia Antipolis-Méditerranée, Université Côte d'Azur, France

E-mail: federica.turi@inria.fr

ABSTRACT: Code-modulated visual evoked potentials (c-VEPs) BCI are designed for high-speed communication. The setting of stimulus parameters is fundamental for this type of BCI, because stimulus parameters have an influence on the performance of the system. In this work we design a c-VEP BCI for word spelling, in which it is possible to find the optimal stimulus presentation rate per each subject thanks to an adaptive setting parameter phase. This phase takes place at the beginning of each session and allows to define the stimulus parameters that are used during the spelling phase. The different stimuli are modulated by a binary m-sequence circular-shifted by a different time lag and a template matching method is applied for the target detection. We acquired data from 4 subjects in two sessions. The results obtained for the offline spelling show the variability between subjects and therefore the importance of subject-dependent adaptation of c-VEP BCI.

INTRODUCTION

Among the BCIs based on electroencephalographic signals (EEG), a VEP BCI allows for spelling from a keyboard of flashing characters, by identifying the target character, which the user is gazing at. Depending on the specific stimulus modulation design used, current VEP based BCIs can be distinguished into BCI systems using frequency modulated VEP (f-VEP), time modulated VEP (t-VEP) and BCI systems using pseudo-random code-modulated VEP (c-VEP) [1]. In a c-VEP BCI, all characters flash according to a predefined pseudo-random sequence, as a m-sequence [2], circular-shifted by a character-dependent time lag. For a given character, the m-sequence evokes a VEP in the EEG of the subject [3], which can be used as a template. This template is obtained during a calibration phase at the beginning of each session. A c-VEP BCI can potentially achieve a very high-speed communication level, reaching an average information transfer rate (ITR) of 108 ± 12 bit/min [3]. The stimulus modulation is crucial to build a high performance c-VEP BCI. Many studies investigate the effect of stimulus specificity on the target flashing modulation, applying different pseudo-random sequences [4], with different bit length sequences [5]. Isaksen et al. proved that among different types of code none provided a superior performance, showing that the "optimal-code" depends on the subject [4]. Wei et al. [5] explored different stimuli layout parameters such as the size, color and proximity of the stimuli, different length sequences and different lags between adjacent stimuli, providing the best set of parameters to increase the performance in a multi-target c-VEP BCI. Aminaka et al. propose a green-blue stimulus compared with the classical black-white [6], showing that the chromatic green–blue stimulus can give high result of accuracy, but not always better than the black-white color combination. Nazamfar et al. [7] explored different color stimulation: black and white, red and green and blue and yellow, stimulation sequence with three different bit lengths of 31, 63 and 127 bits but also three different bit presentation rates of 30, 60 and 110 bps. They showed that it is possible to find a compromise between high decision rate and subject comfort, using a m-sequence of 63-bit, a presentation rate of 60 bps and red-green for stimulus color. The stimulus presentation rate is another important parameter that can influence the system performance. The most common value used for coding sequence is 60 Hz. Wittevrongel et al. [8] proposed a study in which they compare the traditional flashing pattern frequency of 60 Hz to a faster one at 120 Hz. Applying a novel decoding algorithm based on spatio-temporal beamforming, they showed that with a faster stimulation is possible to increase the performance of the system, reaching a ITR of 172.87 bits/min. Analyzing these studies is clear that it is not possible to define an universal optimal stimulus parameter setting suitable for each BCI user. In order to obtain a BCI system with high speed communication that respects the subject's comfort is necessary to develop a system adaptable to the subject. In our study we developed a subject-dependent system with four different stimulus presentation rates of 15 Hz, 20 Hz, 30 Hz and 60 Hz. The objective is to find the optimal presentation frequency to obtain a pleasant stimulus per each subject. We demonstrate that the decreasing of the stimulation frequency does not imply the decreasing of the system performance showing the importance of the BCI-user adaptation.

MATERIALS AND METHODS

Adaptive parameter setting phase:
In a traditional c-VEP BCI with a refresh rate of 60 Hz and target encoded by binary sequence, each element

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-25

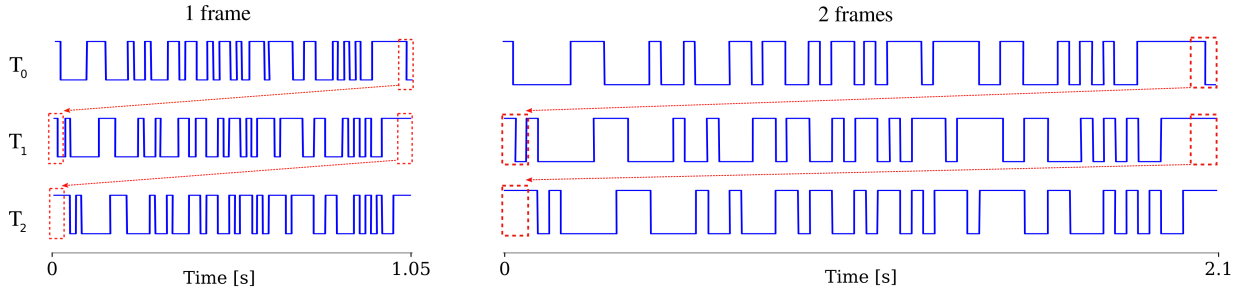


Figure 1: Illustration of the circular-shift process for the stimulus sequence for the first 3 targets. On the left, the stimulus sequences have a frame rate of 60 Hz (1 frame) and on the right, a frame rate of 30 Hz (2 frames). The figure shows the stimulation sequence for the target T_0, for the target T_1, circularly shifted with a time lag τ_s with respect to T_0 and for the target T_2, circularly shifted with a time lag τ_s with respect to T_1. The red dash boxes indicate the time lag τ_s corresponding to the # frames, listed in Tab. 1.

Table 1: Stimulus parameters set for each two consecutive targets in the adaptive parameter setting phase. The time lag $\tau_s = (2\ bits/60\ Hz) \cdot \#frames$. The length of one sequence t_s is computed as $t_s = (63\ bits/60\ Hz) \cdot \#frames$. The flashing duration of each target is $t_k = t_s \cdot \#stimulus\ cycles$

target1-target2	# frames	# stimulus cycles	τ_s	t_s	t_k
A - D	1	10	0.033s	1.05 s	10.50 s
A - P	2	8	0.067s	2.10 s	16.80 s
T - I	3	5	0.10s	3.15 s	15.75 s
V - E	4	4	0.13s	4.20 s	16.80 s

of the sequence is flashed on the screen for a time $t_b = 16.67\ ms$, corresponding to the duration of one frame. We developed a system in which the targets are encoded by a 63-bit m-sequence, but each element of this sequence can be flashed for 1, 2, 3 or 4 frames. On a 60-Hz screen, the characters on the virtual keyboard will thus flash at 60 Hz, 30 Hz, 20 Hz or 15 Hz. In this way it is possible to flash the target faster or slower and to find the stimulus pattern most adapted to each subject, while maintaining a system with high performance. Our protocol thus starts with an adaptive parameter setting phase. In this phase the subject has to focus on the targets of a word of eight characters, shown in Fig. 2, so the word can be divided in four pairs of targets and each pair follows the setting reported in Tab. 1. The flashing rate is thus changed every two targets. We compare the correlation between all the VEP responses recorded for each stimulus cycle of the two targets flashed with the same frame value per each channel c, to detect at which stimulus setting the evoked response of the subject is most prominent. Let ρ_1^c be the averaged correlation of the VEP responses over the N stimulus cycles of the target 1, ρ_2^c the averaged correlation of the VEP responses over the N stimulus cycles of the target 2 and ρ_{12}^c the averaged cross-correlation between the N stimulus cycles of the VEP response of the target 1 and the VEP response of the target 2. Based on the fact that when the user is gazing at a target, the specific VEP recorded in the EEG should be the same for each stimulus cycle of the same target, the expectation is a high value for ρ_1^c and ρ_2^c and a low value for ρ_{12}^c. The score λ^c evaluates the difference between the autocorrelation ρ_1^c and the cross-correlation ρ_{12}^c per channel for each parameters setting. The λ^c score is computed following equation (1), where std_1^c is the standard deviation of the VEP responses over the N stimulus cycles of the target 1 and std_{12}^c is the standard deviation of the VEP responses over the N stimulus cycles of the target 1 and target 2.

$$\lambda^c = \frac{\rho_1^c - \rho_{12}^c}{std_1^c + std_{12}^c} \quad (1)$$

At the end of the data acquisition of this adaptive phase the score λ per each set of parameters is averaged over the three best channels, for which ρ_1^c is the highest. The largest score λ is chosen to select the best stimulation number of frame.

Experimental setup:
The BCI software consists of OpenViBE [9] for signal acquisition, and a custom keyboard-display control software which we developed in C++. This software is run on a Windows 7 computer with an Intel(R) Xeon(R) processor. Two different LCD monitors are used: one (DELL U2711) is used during the acquisition to monitor the EEG signal quality, and the other (DELL 2709W) is set at 60 Hz with a resolution of 1920x1080 with a NVIDIA Quadro FX 580 graphic card and it is used for the stimuli presentation on a virtual keyboard. The keyboard, displayed in Fig. 2, is a 4x8 matrix containing 32 characters: letters sorted alphabetically from A to Z followed by backspace, symbols "?", "!", "." and numbers 1 and 2. Each character is placed in a circle with a dark grey background. Below the matrix of targets a text field shows the characters of the word that the subject has to gaze at. Each character of the virtual keyboard flashes according a binary sequence composed of 0 and 1. If the bit in the corresponding binary stimulation sequence is 1 the character flickers in light grey, if it is 0 in black, as illustrated in Fig. 2. The color combination light grey/black was chosen instead of white/black, in order to make the contrast more comfortable for the subjects. The stimuli are

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-25

synchronized with the refresh rate of the monitor and a trigger signal is provided by a TCP network connection to synchronize the stimulus presentation and the EEG data recordings.

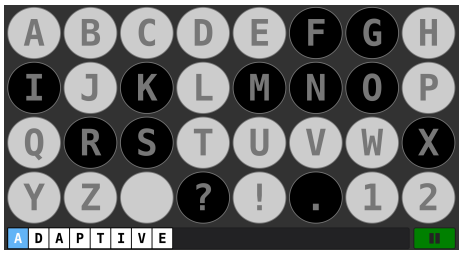


Figure 2: Virtual keyboard. During the stimulation the target is highlighted in blue below the keyboard.

Participants and data acquisition:
Four healthy volunteers (one male, average age 35, std 5 years) participated in the c-VEP BCI experiment. The experiment took place in our premises at Inria and was approved by the Operational Committee for the assessment of Legal and Ethical risks of the institute. All subjects had normal or correct to normal vision and did not suffer from epilepsy or other nervous diseases. Each subject took part in two identical sessions, half a week apart one from the other, and all of them completed the whole experiment. All the subjects were c-VEP BCI-naive participants. During the experiment, each subject was seated in a comfortable armchair 100 cm away from the computer monitor placed in a quiet room. During each session the EEG signal of the subject was recorded from ANT-Waveguard cap and a Refa8 amplifier (512 Hz sampling rate). To set the impedance between the electrodes and the subject's skin below 10 KΩ, a conductive gel was applied to the ground (FPz) and to the 12 electrodes placed in positions *Fz*, *F3*, *F4*, *Cz*, *C3*, *C4*, *Pz*, *P7*, *P8*, *Oz*, *O1*, *O2*. Each session consisted in 2 phases: the adaptive setting phase and a second phase in which the subject focuses on imposed characters, in which each target flashes according to the set of parameters found during the adaptive phase. To avoid the effect of fatigue on the experiment, the subject was allowed to take a rest of 5 minutes between the two phases. Each session lasted around 45-60 minutes, including the time for the experiment preparation (positioning of the cap and conductive gel injection) and the time for the data acquisition. During the adaptive setting phase a word of eight characters is displayed on the screen, below the keyboard. The characters are flashed following a 63-bit m-sequence and its time shifted version by 2 bits [3], an example for the circular shift of the stimulus sequence can be seen in Fig. 1. The stimulus parameters, in terms of number of frames and number stimulus cycles, follow the set of stimuli parameters reported in Tab. 1. Each target, before starting the stimulation, is highlighted in blue in the virtual keyboard to indicate to the subject the target position in the virtual keyboard. This adaptive setting phase lasts around 2 minutes. At the end of this phase the data collected is processed, as explained in the subsection *Adaptive parameter setting phase*, and the output of the processing gives the best set of parameters per subject for that session. Moreover the subject was asked for which value of frame rate the visual stimulus was more comfortable. Then, during the second phase, the subject has to focus his/her attention on the characters of a specific word written below the keyboard. In this phase each target is highlighted in blue before starting to flash and then all the characters flash following the set of best parameters computed during the first phase. Each subject has to spell five different words (42 characters in total and 10 targets), with a pause of one minute between each word.

Offline processing:
The EEG data $\mathbf{X}$ collected with N stimulus cycles on c channels, are bandpass filtered between 4 and 22 Hz with a Butterworth filter of order 4. The canonical correlation analysis (CCA) [10] is applied as spatial filter to improve the signal-to-noise ratio of the EEG signal. The objective of CCA is to find the two transformations W_X and W_S which maximize the correlation between the raw EEG-data $\mathbf{X}$ and the desired VEP waveform $\mathbf{S}$ [11].

$$CCA(X,S) = \max_{W_X W_S} \frac{W_X^T \mathbf{X}\mathbf{S}^T W_S}{\sqrt{W_X^T \mathbf{X}\mathbf{X}^T W_X} \cdot \sqrt{W_S^T \mathbf{S}\mathbf{S}^T W_S}} \quad (2)$$

To compute $\mathbf{S}$ we average over the number of stimulus cycles N the responses recorded for the first character of the first word that the subject has to gaze during the second phase, we consider only the three best channels c_b selected during the adaptive setting phase and then replicate N times the signal to obtain $\mathbf{S}$ [11], in this way $\mathbf{X}$ and $\mathbf{S}$ have the same number of stimulus cycles N. Then W_X is multiplied with $\mathbf{X}$ to compute the spatially filtered signal x. For target identification the method of template matching [3] is used. The reference template T_0 is calculated by averaging the signal x of the first character of the first word over the N stimulus cycles. The templates of all others targets T_k ($k = 0, ..., 31$) are generated by circularly shifting the template T_0. The duration t_s of the template and the time lag τ_s between two consecutive targets depends on the number of frames set for each subject, as listed in Tab. 1. To detect the attended target we compute a cumulative correlation per target. We segment the spatially filtered signal in epochs starting at the 'start' trigger, sent at the beginning of each stimulus cycles, and lasting the length of a stimulation sequence, specific to each subject (see Tab. 3), in order to obtain the epochs x_n, with $n = 0, ..., N$. To set the N stimulus cycles, the cumulative correlation between the stimulus repetitions is computed and an arbitrary threshold is fixed at 0.8, considering the normalized correlation of each subject. Finally we compute the cumulative correlation coefficient ρ_k between each template T_k and the epochs x_n, following the equation 3.

$$\rho_k = \sum_{n=0}^{N} \frac{T_k \cdot x_n}{\sqrt{T_k \cdot T_k} \cdot \sqrt{x_n \cdot x_n}} \quad (3)$$

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-25

Table 2: Average λ score with standard deviation and the frame rate preferred by the subject during the adaptive parameter setting phase. In bold the maximum λ score corresponding to the frame rate selected at the end of the adaptive parameter setting phase.

	subject	1 frame	2 frames	3 frames	4 frames	subject preference
Session1	S1	0.25 ± 0.05	$\mathbf{0.86 \pm 0.08}$	0.69 ± 0.09	0.45 ± 0.03	no preference
	S2	1.23 ± 0.12	$\mathbf{1.83 \pm 0.32}$	1.60 ± 0.18	1.53 ± 0.08	2 frames
	S3	0.43 ± 0.14	$\mathbf{0.71 \pm 0.06}$	0.51 ± 0.03	0.34 ± 0.05	2 frames
	S4	0.51 ± 0.20	0.61 ± 0.08	$\mathbf{0.93 \pm 0.15}$	0.72 ± 0.00	3 frames
Session2	S1	0.32 ± 0.05	$\mathbf{0.88 \pm 0.08}$	0.80 ± 0.09	0.36 ± 0.3	no preference
	S2	0.56 ± 0.20	1.72 ± 0.13	0.91 ± 0.12	$\mathbf{2.41 \pm 0.24}$	4 frames
	S3	0.20 ± 0.04	0.84 ± 0.05	0.54 ± 0.17	$\mathbf{0.91 \pm 0.08}$	4 frames
	S4	0.63 ± 0.09	$\mathbf{0.76 \pm 0.05}$	0.56 ± 0.08	0.38 ± 0.05	2 frames

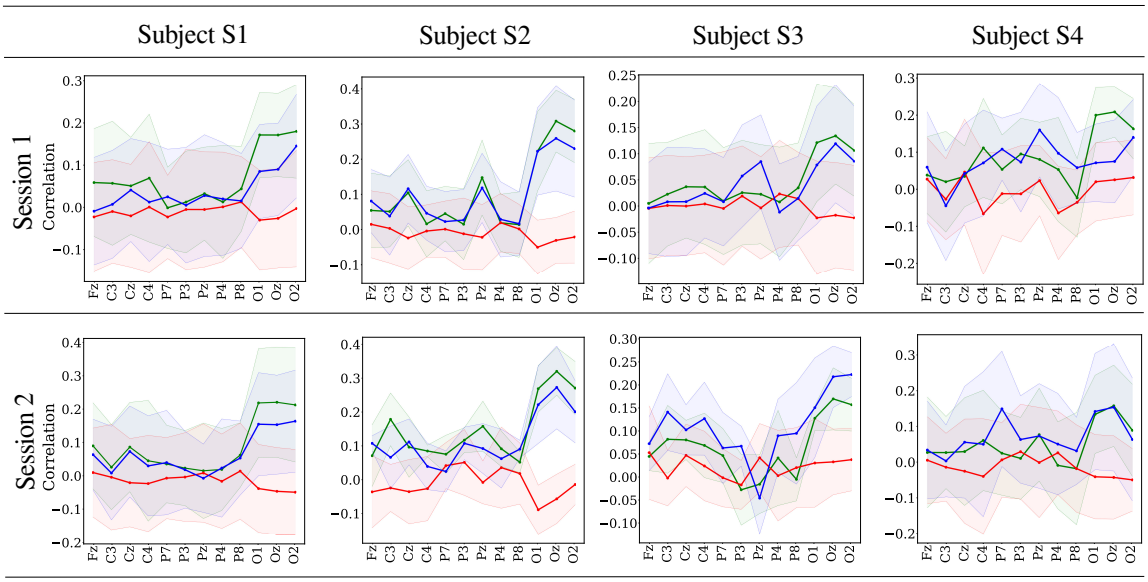


Figure 3: Correlation curves of the optimal number of frames, listed in Tab. 3, obtained for each subject at the end of session 1 and session 2. The average auto-correlation per channel ρ_1 in green and the average auto-correlation ρ_2 per channel in blue. The average cross-correlation $\rho 12$ is represented in red.

The target k with the largest coefficient ρ_k is detected as the attended target k_a. The offline spelling accuracy of each word is computed to evaluate the performance of the BCI system. Basically, if the detected target k_a corresponds to the target k at which the subject is gazing at, then the right character is detected. The number of correctly detected characters is computed for each word that the subject has to spell during the second phase. Finally the accuracy per word is computed as the proportion of correctly detected characters.

RESULTS

Tab. 2 lists the λ scores obtained per subject and session as well as the frequency rate preferred by the subject. We can observe that only for the subject S1, who did not express a preference, the optimal frame rate is unchanged between the two sessions. For the other subjects the optimal number of frames changes between one session and the other. For the subject S2 the largest λ obtained at the end of the first session is 1.83 ± 0.32 corresponding to 2 frames. If we compare this result with respect to the results obtained at the end of the second session, we can notice that for the same number of frames the λ value is 1.72 ± 0.13, but the largest λ value is reached for 4 frames, with a value equal to 2.41 ± 0.24. The same trend can be observed for the other subjects, except subject S4. Fig. 3 shows the auto-correlation and cross-correlation curves obtained for each subject for the best frame rate selected at the end of the adaptive setting parameter phase. The parameters obtained at the end of the adaptive setting phase and set for the second phase are listed in Tab. 3. In Fig. 4 we show the averaged offline spelling accuracy, over the five words, of each subject for each session. The accuracy of subject S2 reaches a good level, $80\% \pm 15\%$ in the second session and over $74\% \pm 12\%$ in the first one. We can notice that for three subjects there is an increase of performance during the second session, that proves the performance improvement of the subject over several sessions. The accuracy of the subject S3 achieves $43\% \pm 18\%$ during the second session, compared to 11% during the first session. Instead, for the subject S4, the performance during the first session, reaching $56\% \pm 12\%$ of accuracy, is better than the one of the second session, with an accuracy of $33\% \pm 8\%$.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-25

Table 3: Summary of the adaptive stimuli parameter setting. Number of frames, number of stimulus cycles and stimulus duration of the target (t_k) per subject and per session.

	subject	# frames	# stimulus cycles	t_k		subject	# frames	# stimulus cycles	t_k
Session 1	S1	2	7	14.7 s	Session 2	S1	2	7	14.7 s
	S2	2	7	14.7 s		S2	4	3	12.6 s
	S3	2	6	12.6 s		S3	4	3	12.6 s
	S4	3	4	12.6 s		S4	2	6	12.6 s

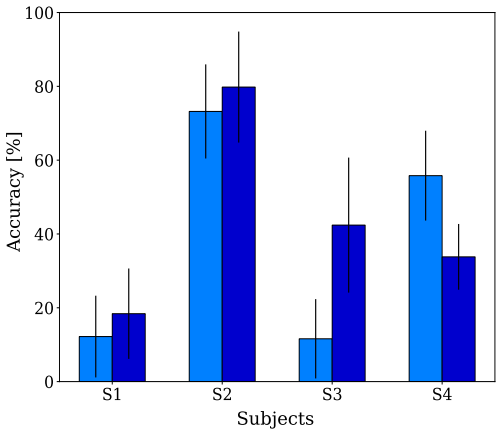


Figure 4: The box plots report the averaged accuracy over five words obtained in the offline spelling for each subject and each session. The session 1 is represented in light blue, session 2 in blue. The error bar represents the standard deviation.

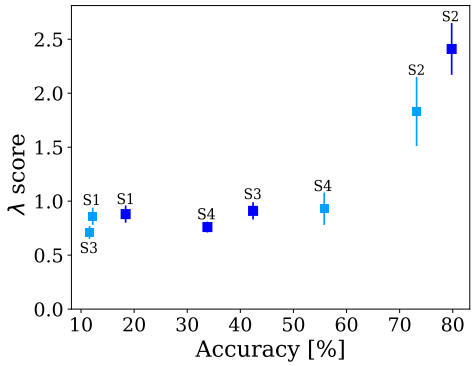


Figure 5: The best λ scores, in bold in Tab. 2, of each subject with respect to the offline spelling accuracy, showed in Fig. 4. The session 1 is represented in light blue, session 2 in blue. The error bar represents the standard deviation. The λ score increases according to the accuracy.

DISCUSSION

In our study we explored the influence of frequency stimulation on the VEP responses by the development of an adaptive setting phase. At the end of this adaptive phase we compute the λ score that can be considered as a performance estimator for a c-VEP BCI system. Indeed if we compare the largest λ value with respect to the accuracy values reached in the offline spelling, shown in Fig. 5, we can notice that a largest λ value corresponds to higher accuracy value per subject. The obtained results show an important variability, both inter-subject and intra-subject. Indeed one of the challenges of using BCIs over extended periods of time is the variation of the user's performance from a session to another. There are many factors that can influence the BCI user's performance between different sessions, for example distraction, visual fatigue, loss of concentration, motivation [12]. All these factors should be considered to build an high performance BCI. Moreover we demonstrate that decreasing the flashing pattern frequency, the length of a stimulus cycle increases, but it can be compensated by setting a lower number of stimulus cycles. As illustrated in Tab. 3, the flashing duration of each target is not longer even when the frame rate is longer than one frame, which is the most common value of frame rate used in c-VEP systems. This means that the stimulus duration at different frame rates does not impact the performance of the system. Our method does not require a long calibration phase as, for example, other c-VEP BCI in which the number of stimulus cycles is around 200 for the reference target [3, 5] in the calibration phase. During the spelling phase some systems obtain high value of spelling accuracy with only 2 stimulus cycles [3] and others with 40 stimulus cycles [5]. In these systems they reached very high level of spelling accuracy, but in our case we aim to develop a method that can reach high value of accuracy, as we obtain for the subject S2, replacing the traditional calibration with a shorter adaptive setting phase and finding a compromise on the number of stimulus cycles during the spelling phase. The objective is to obtain a system in which is possible to define the more pleasant stimulus for each subject, in order to increase the performance of the system. Tab. 2 proves that the most comfortable frame rate expressed by the subject is also the one for which the performance is the highest. This can explain why a subject that performs well for a frequency rate performed less well for another frequency rate during the same session, demonstrating the need for an adaptive system. It would certainly be interesting to understand why some subjects reach a higher accuracy value and others do not, achieving a mean accuracy lower with respect to other c-VEP BCI systems [3, 11]. Among many reasons that can explain this difference, there is the quality of the signal recorded during the experiments. Fig. 6 shows the spatially filtered responses for participants with the worst and the best performance in session 2, in term of accuracy (subject S1 and subject S2 respectively). Observing the VEP responses in Fig. 6 is evident that there is a repeatability of the VEP response for the subject S2, for

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-25

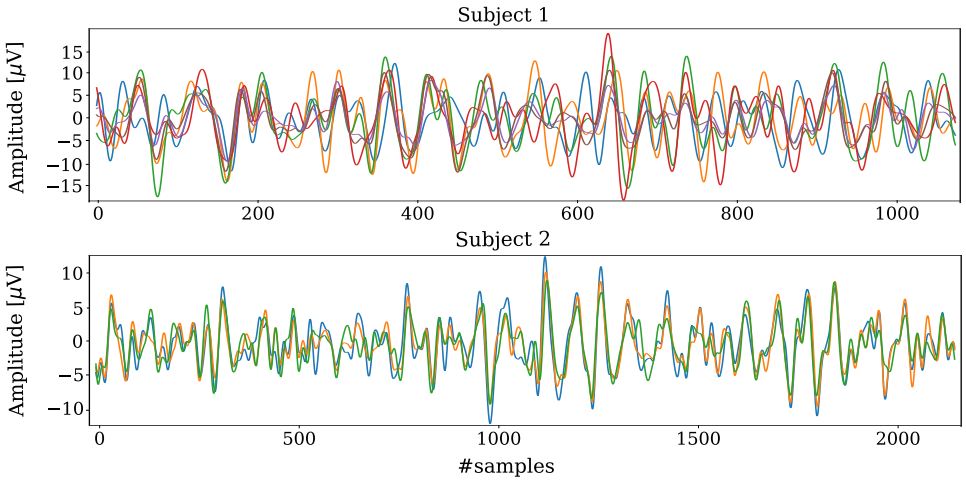


Figure 6: The spatially filtered VEP responses recorded at each stimulus cycle overlapped for target k, acquired during session 2 for subject S1 and subject S2, respectively the subjects that perform worst and best.

whom the accuracy is around 80% and a large variability on the responses for the subject S1, who obtains an accuracy of around 20%. This lack of repeatability can be due to the inexperience of the subject or also by external noise. Future work will be focused on the inclusion of other variables in the adaptive setting parameter phase, such as the selection of the stimulus sequence, color stimulus and development of an online early stopping method to find the optimal number of stimulus cycles per subject. Finally, to further improve our system, the next step will be to identify the disturbance factors and find methods to remove them in the VEP responses, by different applications of spatial filters, or by modelling the VEP response and the external disturbances and noise.

CONCLUSION

In a c-VEP BCI, the stimulus strategy applied to flash the target is fundamental to obtain a high performing BCI. In this work we parametrized the stimulus modulation with four different stimulus presentation rates. We developed an experimental protocol that deploys a preliminary phase to define the optimal setting of stimulation frequency to tackle the problem of variability inter-subject and intra-subject. Optimally, these results would help with the design of a subject-dependent c-VEP BCI with high communication performance.

ACKNOWLEDGEMENT

We thank Romain Lacroix for the precious contribution in the development of our BCI software.

REFERENCES

[1] 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).

[2] SW Golomb. Shift Register Sequences. Aegean Park Press, Laguna Hill, 1982.

[3] Bin G, Gao X, Wang Y, Li Y, Hong B, Gao S. A high-speed BCI based on code modulation VEP. Journal of neural engineering. 2011;8(2):025015.

[4] Isaksen JL, Mohebbi A, Puthusserypady S. Optimal pseudorandom sequence selection for online c-VEP based BCI control applications. PloS one. 2017;12(9):e0184785.

[5] Wei Q, Feng S, Lu Z. Stimulus specificity of brain-computer interfaces based on code modulation visual evoked potentials. PloS one. 2016;11(5):e0156416.

[6] Aminaka D, Makino S, Rutkowski TM. Chromatic and high-frequency cVEP-based BCI paradigm. In: 2015 37th EMBC. IEEE. 2015, 1906–1909.

[7] Nezamfar H, Salehi SSM, Erdogmus D. Stimuli with opponent colors and higher bit rate enable higher accuracy for C-VEP BCI. In: 2015 SPMB. IEEE. 2015, 1–6.

[8] Wittevrongel B, Van Wolputte E, Van Hulle MM. Code-modulated visual evoked potentials using fast stimulus presentation and spatiotemporal beamformer decoding. Scientific reports. 2017;7(1):15037.

[9] Renard Y, Lotte F, Gibert G, Congedo M, Maby E, Delannoy V, 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.

[10] Spüler M, Walter A, Rosenstiel W, Bogdan M. Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2014;22(6):1097–1103.

[11] Spüler M, Rosenstiel W, Bogdan M. Online adaptation of a c-VEP brain-computer interface (BCI) based on error-related potentials and unsupervised learning. PloS one. 2012;7(12):e51077.

[12] Kleih SC, Kübler A. Psychological factors influencing brain-computer interface (BCI) performance. In: SMC, 2015 IEEE Intern. Conf. 2015, 3192–3196.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-26

CAN SUGGESTIVE HYPNOSIS BE USED TO IMPROVE BCI PERFORMANCE?

S. Rimbert[1], O. Avilov[2], P. Adam[3], and L. Bougrain[1]

[1]Université de Lorraine, CNRS, Inria, LORIA, Nancy, F-54000, France
[2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[3]University Hospital of Strasbourg, Hemodialysis Department, Strasbourg, F-67000, France

E-mail: sebastien.rimbert@inria.fr

ABSTRACT:
Improving the user performance is one of the major issues in the Brain-Computer Interface (BCI) field. In this study, the following question is asked: can hypnosis enhance the performance of motor imagery based BCI? Indeed, hypnotic inductions using Ericksonian suggestions can make the user feel simultaneously more relaxed and more focused on the mental task, and therefore could be used to increase performance in BCI. The aim of this study was to evaluate the impact of an hypnotic condition on the ERD/ERS patterns during a kinesthetic motor imagery (KMI) task. To investigate this issue, 19 right-handed healthy subjects performed a KMI of the right hand during two randomized sessions: in normal and hypnotic conditions. We analyzed the event-related (De)-synchronization (ERD/ERS) over the motor cortex and compared the BCI accuracy during both conditions. Our results suggest that the state of hypnosis reduces the ERD phase in the sensorimotor bands, assuming a change in the activation of the motor cortex during the hypnotized state and thus a worst detection of the kinesthetic motor imagery.

INTRODUCTION

Brain-computer interfaces (BCI) allow end users to interact with a system using modulations of brain activities which are partially observable in electroencephalographic (EEG) signals [1]. Originally, BCIs were developed for medical and therapeutic applications, mainly to compensate people with severe motor disabilities. Since several years, BCI progress has been focused essentially to rehabilitate people with severe neuromuscular disorders who represent a very large number of potential users in the coming years [2, 3].

Most of these BCIs are based on the modulation of sensorimotor rhythms during a Kinesthetic Motor Imagery (KMI) task. A KMI is a mental representation which can be described as the ability to imagine performing a movement without executing it, by imagining haptic sensations felt during the real movement [4]. Several studies highlighted the positive effects of KMI practice for people with neurological disorders [5, 6]. These sensorimotor rhythms are characterized i) before and during an imagined movement, by a gradual decrease of power in -mainly- the mu/alpha (7-13 Hz) and beta (15-30 Hz) bands and ii) after the end of the motor imagery, by an increase of power in the beta band. These modulations are respectively known as Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS) or post-movement beta rebound [7, 8].

A KMI is a highly complex task to be performed by users [9] which makes BCI performance very heterogeneous between studies [10, 11]. Some authors claimed that approximately 15% to 30% of users are not able to gain control of a BCI, a phenomenon sometimes called BCI illiteracy [12, 13]. In this way, Ang and Guan showed that 40% of stroke patients can not achieve the critical BCI accuracy level of 70% and highlighted the critical issue of developing methods for patients to accomplish KMI tasks [14].

In order to improve BCI performance and allow a better control of KMI, it has been suggested that some of altered states of consciousness such as mind-body awareness training (MBAT) [15] or hypnosis [16] can be used. Hypnosis is particularly interesting because several conditions may be induced for patients. For example, hypnotic suggestions can make the subject feel more relaxed and at the same time more focused [17]. Hypnosis is definable as an altered state of attention, receptivity and concentration during which the hypnotized person is captured by a suggestion made by the hypnotist [18, 19]. Although this procedure is already used to reduce pain, manage stress or manage emotional problems [20], we can imagine using it to have a direct impact on the BCI performance, especially by making easier the KMI task execution.

Despite some studies have already shown the interest of MBAT for the BCI domain [21], according to our knowledge, no studies investigated the effect of hypnosis on the EEG signals over the motor cortex and the subsequent impact for BCI performance. Interestingly, Muller et al. showed activation differences between hypnotic and normal states in fMRI for the motor imagery task and suggested that hypnosis enhanced the motor control circuit engaged in the motor task by modulating the gating function of the thalamus [22]. Unfortunately, no studies have yet confirmed their findings and reveal the effect of hyp-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-26

nosis on the EEG signal of the motor cortex during motor tasks.

In this study, the following question is asked: can hypnosis improve MI-based BCI performance? To investigate this issue, 19 right-handed healthy subjects performed a KMI during 2 randomized sessions: in normal and hypnotic conditions. We analyzed the Event-related (De)-synchronization (ERD/ERS) and compared the BCI accuracy for both conditions. Our results suggest that the state of hypnosis reducees the ERD phase in the beta frequency band and thus a worst detection of the KMI task.

MATERIALS AND METHODS

Participants: 19 right-handed healthy subjects (10 females; from 18 to 30 years-old with a mean and a standard deviation of the age of 26.1 years ± 2.8) were recruited for this experiment. This study follows the statements of the WMA declaration of Helsinki on ethical principles for medical research involving human subjects [23]. In addition, participants signed an informed consent which has been approved by the ethical committee of Inria (the COERLE). The subjects had no medical history and no experience with hypnosis which could have influenced the task.

Experimental Tasks: The aim of this study was to evaluate the impact of an hypnotic condition on the ERD/ERS patterns during a KMI task. The motor imagery task was to imagine performing a simple closing movement of the right hand without executing it [24, 25].

Experimental Environment: The experiment took place in a confined room and was conducted by two investigators. The first one assumed the technological process of the experiment. The second investigator was a professional hypnotherapist responsible for instructing the subjects under both normal and hypnotic conditions. During the experiment, the subjects were sitting comfortably on a chair, with their right arm on an armrest, their hand resting on an ergonomic pillow, without any muscular tension.

Experimental Design: The protocol contained 2 sessions: the normal condition and the hypnotic condition. The sessions were performed in a counterbalanced order determined by computerized randomization. For each session, one run of 8 minutes were performed. During a run, the subject had to perform the KMI task during 4 seconds, approximately once every ten seconds. A low frequency beep indicated when the subject had to execute the KMI task. A high frequency beep indicated the end of the task. Breaks of a few minutes were planned between sessions to prevent fatigue of the subject. Before the experiment, the KMI were previously described to the subject and the subject trained to master them without feedback. At the end of each session, participants completed the same questionnaire based on a Lickert scale to assess the possible impact of hypnosis on several criteria (estimated time of the experiment, body perception, memory, detachment, stress, fatigue and motivation).

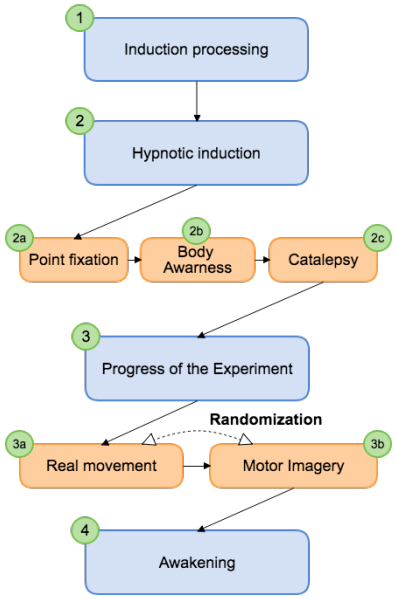


Figure 1: Description of the induction of the hypnotic condition. The Kinesthetic motor imagery was performed after the hypnotic induction composed by a (i) point of fixation, (ii) a body awareness and (iii) a catalepsy with an arm levitation

Hypnotic procedure: The induction procedure was performed by a professional hypnotist. For each subject, a standard script based on Ericksonian hypnosis was used [18]. The Ericksonian hypnosis is characterized by an indirect suggestion by using metaphors. The hypnotic condition was subdivided into several stages: (1) the induction process, (2) the suggestion, (3) the progress of the experiment (KMI executions) and (4) the awakening (Fig. 1).

Physiological Recordings: EEG signals were recorded through the OpenViBE platform with a commercial REFA amplifier developed by TMS International. The EEG cap was fitted with 32 electrodes re-referenced with respect to the common average reference across all channels over the extended international 10-20 system positions. The selected electrodes ware FC_5, FC_3, FC_1,FC_z, FC_2, FC_4, FC_6, C_5,C_3, C_1, C_z, C_2, C_4, C_6, CP_5, CP_3, CP_1, CP_z, CP_2, CP_4, CP_6, P_3, P_1, P_z, P_2, P_4, PO_3, PO_z, PO_4, O_1, O_z, O_2. These sites ware localized around the primary motor cortex, the motor cortex, the somatosensory cortex and the occipital cortex, which allowed us to observe the physiological changes due to the MI task. An external electromyogram (EMG) electrode was added in order to measure the muscle activity. Impedance was kept below 5 kΩ for all electrodes to ensure that background noise in the acquired signal was low.

Time-Frequency Analysis: To analyze the differences between both conditions (normal condition vs hypnotic condition), we performed an event-related spectral perturbation (ERSP) between 8-35 Hz. We used a 256 point sliding fast Fourier transform (FFT) window with a padratio of 4 and we computed the mean ERSP from 1 s before the task to 7 s after the task (Figure 2). ERSP visualizes event-related changes in the averaged power

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-26

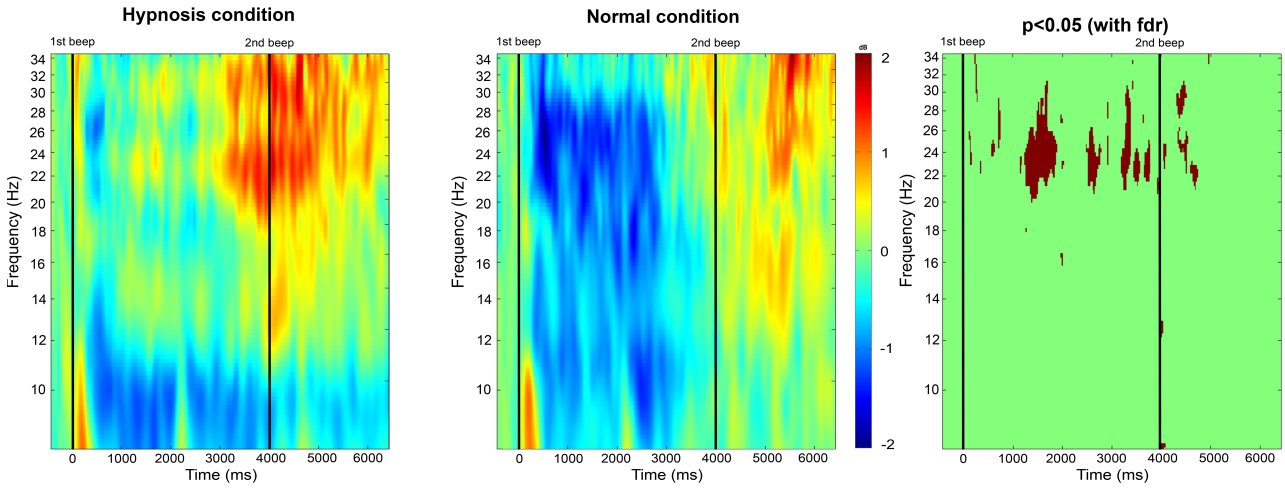


Figure 2: event-related spectral grand average perturbation (ERSP) for MI during normal and hypnosis conditions for electrode C_3. Black lines indicate when the motor task started and finished. A red color corresponds to a strong ERS and a blue one to a strong ERD. Significant difference (p<0.05) with a False Discovery Rate (FDR) correction are shown in the right part of the figure.

spectrum relative to a 2s-baseline interval took 1 s before each trial. A permutation test for a significant level of 0.05 with a False Discovery Rate (FDR) correction using the EEGLAB toolbox was done to validate differences in terms of time-frequency of this ERSPs [26].

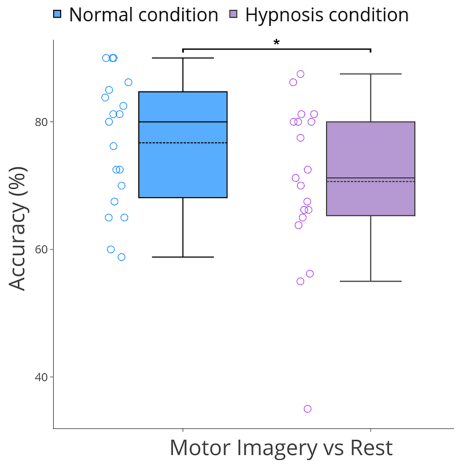


Figure 3: Boxplots showing the distribution of average classification accuracies (N=19) for MI in normal and hypnosis conditions.

Feature extraction: We used the efficient Common Spatial Pattern (CSP) algorithm to extract motor imagery features from EEG signals [27] via the MNE python library [28]. Pairs of spatial filters have been generated and applied to decompose multi-channel signals into a set of uncorrelated components that simultaneously maximize the variance of one class, while minimizing the variance of the other class [29]. The CSP algorithm has been applied on the whole set of channels which includes electrodes covering the sensorimotor cortex. The number of resulting filters equals the number of EEG channels but for dimensional reduction purpose only the first two pairs (i.e., four filters) were selected.

Once selected filters W were obtained from band-pass EEG signals X, a CSP feature vector f can be computed as follows [30]:

$$f = log(WXX^{\top}W^{\top}) = log(WCW^{\top}) = log(var(WX)) \quad (1)$$

Raw signals were filtered by a 4th order Butterworth band-pass filter using three different frequency ranges (7-13 Hz, 15-30 Hz and 8-30 Hz) to increase the ERD and ERS observations. For each frequency range, trials were cropped from the filtered signals to train two paired CSP filters in time range from 0 till 3 s to maximize differences in variance between the two conditions. Finally we aggregated each 4 features obtained separately for the three frequency ranges into a 12-feature vector.

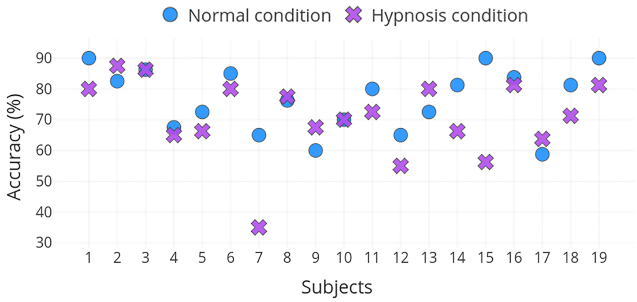


Figure 4: Individual accuracy for motor imagery in normal and hypnosis conditions

Classification: A shrinkage linear discriminant analysis (sLDA) [31] was performed using the Python library scikit-learn [32]. A leave-one-out validation was used to better estimate the accuracy on new trials. A leave-one-out process is K-fold cross-validation with K equal to n, the number of examples in the set. Thus the classifier is trained on all examples except one and a prediction is made for that example [33]. The final accuracy is the

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-26

sum over all models of correct predictions divide by the total number of examples. We chose to apply a paired t-test (two-sided) to show the significant difference between accuracies obtained for each condition (Figure 3, p-value < 0.05).

RESULTS

Time-frequency analysis: Time-frequency maps display the signal power evolution and are useful to establish the frequency and time windows in which event-related spectral perturbation (ERSP) appears (Figure 2).
During the KMI in normal condition, an ERD appears 400 ms after the first beep and is sustained for 3000 ms in the mu + beta frequency band (8-30 Hz). Note in particular that this desynchronization is predominant in the beta band (18-30 Hz). After the second beep, which represents the end of the KMI, an ERS appears to reach its maximum at 6000 ms.
When a KMI is done in hypnotic condition, an ERD is observable in the lower mu band (8-10 Hz) but not in the high mu (10-13 Hz) neither in the beta frequency bands. Interestingly, an ERS is present before the end of the KMI (at 3500 ms) and is maintained up to 6000 ms.
Comparing both conditions, the ERD disappearance during the hypnotic condition is statistically significant at p<0.05 (with a FDR correction) in the beta band.

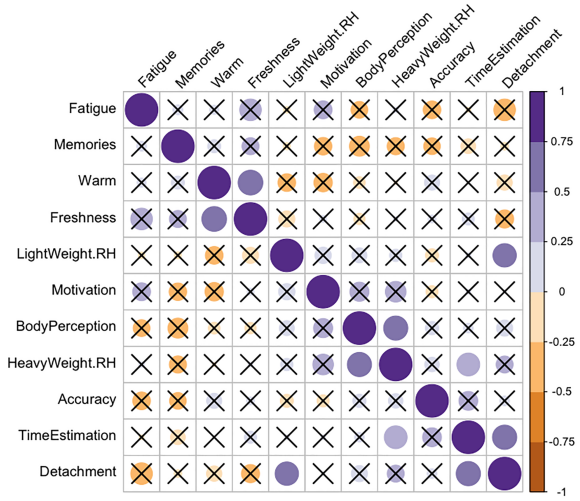


Figure 5: Correlogram representing correlations among estimated time of the experiment, body perception, memory, detachment, stress, fatigue and motivation for normal and hypnotic conditions and accuracy difference between the normal and the hypnotic conditions. Correlations with a p-value > 0.05 are considered as not significant and are crossed out. Positive correlations are displayed in purple and negative ones are in orange. The intensity of the color and the size of the circles are proportional to the correlation coefficients.

Classification:
In order to check if the hypnotic condition could be useful to improve the BCI performance, we decided to compare the accuracy for both conditions. The classification between KMI vs Rest was performed under normal and hypnotic conditions. For both conditions, each trial was segmented into a motor task period and a resting period lasting 3 seconds. Figure 3 shows that BCI performance is significantly better under normal condition compared to hypnotic condition ($76.7 \pm 10.2\%$ VS. $70.6 \pm 12.7\%$). This result is statistically significant at $p < 0.05$. The individual results also show the same tendency, except for 5 subjects: 2, 8, 9, 13, 17 (Figure 4).

Correlogram:
Subjects completed a questionnaire with several items to assess variations between both conditions: fatigue, memories of the experiment, warmness or freshness during the experiment, motivation, time estimation of the task, detachment, change in their body perception, heavyweight or lightweight of their right hand. Figure 5 shows the correlations between these items, to which we have added the *accuracy* representing the BCI performance calculated above. There are some significant correlations ($p < 0.05$) between some items: Detachment and Time estimation, Detachment and Lightweight of the right hand (RH), Freshness and Warm, Heavyweight RH and Body perception, Heavyweight RH and Time estimation. However, no correlation has been established between the variation of the different feelings during hypnosis and BCI performance.

DISCUSSION

In this discussion section, we consider (i) some study parameters, (ii) the hypotheses that could explain why hypnotic state may impact the ERD and (iii) the suggested contribution of such information to the BCI domain.

Study design:
This study was initially conducted on 23 voluntary subjects but only 19 of them were included in this article. Indeed, 4 subjects had anomalously low classification rates (less than 60%) and a time-frequency map without visible ERD and ERS. The subjects who participated were previously selected by the hypnotist as being moderately hypnotizable. We consider that the selection was well carried out since we obtain significantly different results in terms of time estimation, which has already been confirmed by another study [34]. The study was carried out with 32 electrodes, placed only at the pre-frontal, central and occipital area and it would be necessary to confirm the results with a full-head montage. Finally, we reported accuracies using aggregated features from three different frequency ranges (7-13 Hz, 15-30 Hz and 8-30 Hz). Indeed, although not statistically significant, from our experience this method overpass result obtained using features from one frequency range only. Nevertheless, the results between the normal condition and the hypnotic condition were unchanged with a conventional method.

Absence of ERD during the hypnotic state:
Interestingly, the hypnotic suggestion allows to induce multiple states (e.g. be relaxed while focusing on the feelings usually perceived during a real movement), which appears to be particularly relevant to the KMI task [17,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-26

34]. Moreover, after the experiment, a majority of subjects report a more easy way of imagining movement compared to the normal condition. Despite these positive feedbacks, our results showed that hypnosis involves a disappearance of the ERD during the KMI task (Figure 2).

The absence of ERD during hypnosis raises questions. Indeed, Neuper et al. interpret the ERD during the task as being related to a motor cortex activation and the ERS as a deactivation [35]. In this way, the absence of ERD during the KMI task would be due to a weaker activation of the motor cortex, as a consequence of the hypnotic state. This deactivation of the motor cortex could be due to the increase in automation during hypnosis [36]. Further studies had already shown that hypnotic suggestion can modulate the motor cortex activity [37]. However, Muller et al. highlighted that hypnosis could enhance the motor control circuit engaged in motor imagery by modulating the gating function of the thalamus [22], which is inconsistent with our results.

Use of hypnosis for BCIs:

ERD disappears with hypnosis can explain why BCI performance has been impacted. Indeed, the BCI classification is often based on the ERD phase because it is the most robust pattern according to the experimental conditions [8], but also to gain time to provide the end-users with the necessary feedback. If BCI performance was enhanced by the hypnotic state, as it was shown with the MBAT [15], the improvement factors would have been investigated and transferred to the BCI users. For example, designing a self-hypnosis session for BCI illiteracy might be possible. Unfortunately, our average (Figure 3) and individual (Figure 4) results indicate the opposite, and this prompts us to not recommend using hypnosis for BCI applications.

CONCLUSION

This work clearly showed that the hypnotic state has an impact on the EEG through the motor cortex, especially during a kinesthetic motor imagery task. Our results suggest that this altered state tend to remove the ERD during the mental task and implied a negative impact on BCI performance. This prompts us to not recommend using hypnosis for BCI applications.

REFERENCES

[1] Jonathan Wolpaw and Elizabeth Winter Wolpaw, eds. *Brain-Computer Interfaces: Principles and Practice*. Oxford university press, 2012.

[2] B He, S Gao, H Yuan, and J.R Wolpaw. “Brain–computer interfaces”. In: *In Neural Engineering: Second Edition*. Springer US, 2013, pp. 87–151.

[3] Ioulietta Lazarou, Spiros Nikolopoulos, Panagiotis C. Petrantonakis, Ioannis Kompatsiaris, and Magda Tsolaki. “EEG-Based Brain–Computer Interfaces for Communication and Rehabilitation of People with Motor Impairment: A Novel Approach of the 21st Century”. In: *Frontiers in Human Neuroscience* 12 (2018), p. 14.

[4] Aymeric Guillot, Christian Collet, Vo An Nguyen, Francine Malouin, Carol Richards, and Julien Doyon. “Brain activity during visual versus kinesthetic imagery: an FMRI study.” In: *Hum Brain Mapp* 30.7 (2009), pp. 2157–2172.

[5] K Liu, C Chan, T Lee, and C Hui-Chan. “Mental imagery for promoting relearning for people after stroke: a randomized controlled trial.” In: *Arch Phys Med Rehabil* 9 (2004), pp. 1403–8.

[6] Floriana Pichiorri et al. “Brain–computer interface boosts motor imagery practice during stroke recovery”. In: *Annals of Neurology* 77.5 (2015), pp. 851–865.

[7] G Pfurtscheller. “Induced oscillations in the alpha band: functional meaning”. In: *Epilepsia* 44.12 (2003), pp. 2–8.

[8] Bjorg Elisabeth Kilavik, Manuel Zaepffel, Andrea Brovelli, William A MacKay, and Alexa Riehle. “The ups and downs of beta oscillations in sensorimotor cortex.” In: *Exp Neurol* 245 (2013), pp. 15–26.

[9] C Jeunet, C Vi, D Spelmezan, B N’Kaoua, F Lotte, and S Subramanian. “Continuous tactile feedback for motor-imagery based brain-computer interaction in a multitasking context”. In: *Human-Computer Interaction* (2015), pp. 488–505.

[10] M Clerc, E Dauce, and J Mattout. “Brain-computer interfaces. Vol 1, foundations and methods”. In: London ISTE Ltd Hoboken, 2016. Chap. Adaptive Methods in Machine Learning, pp. 209–211.

[11] M Ahn and S Jun. “Performance variation in motor imagery brain–computer interface: A brief review”. In: *Journal of Neuroscience Methods* 243.1 (2015), pp. 103–110.

[12] B Allison and C Neuper. “Brain-computer interfaces: applying our minds to human-computer interaction”. In: Elsevier, 2010. Chap. Could anyone use a BCI?, pp. 35–54.

[13] Claudia Sannelli, Carmen Vidaurre, Klaus-Robert Müller, and Benjamin Blankertz. “A large scale screening study with a SMR-based BCI: Categorization of BCI users and differences in their SMR activity”. In: *PLOS ONE* 14 (Jan. 2019), e0207351.

[14] Kai Keng Ang and Cuntai Guan. “Brain–Computer Interface for Neurorehabilitation of Upper Limb After Stroke”. In: *Proceedings of the IEEE* 103 (2015), pp. 944–953.

Published by Verlag der TU Graz
Graz University of Technology

[15] Lee-Fan Tan, Zoltan Dienes, Ashok Jansari, and Sing-Yau Goh. “Effect of mindfulness meditation on brain–computer interface performance”. In: *Consciousness and Cognition* 23 (2014), pp. 12 –21.

[16] Maryam Alimardani and Kazuo Hiraki. “Development of a Real-Time Brain-Computer Interface for Interactive Robot Therapy: An Exploration of EEG and EMG Features during Hypnosis”. In: *International Journal of Computer, Electrical, Automation, Control and Information Engineering* 11.2 (2017), pp. 187 –195.

[17] Charles Simpkins and Annellen M. Simpkins. “Mindfulness and Hypnosis: The Power of Suggestion to Transform Experience by M. Yapko”. In: *American Journal of Clinical Hypnosis* 55 (Oct. 2012), pp. 199–200.

[18] MH Erickson. “Hypnosis in painful terminal illness”. In: *American Journal of Clinic Hypnosis* 56.2 (1958), pp. 67–71.

[19] JA Malrewicz, J Godin, and H Milton. “Erickson : de l’hypnose clinique à la psychotérapie stratégique”. In: ed. by Paris ESF. 986.

[20] Z Elahi, R Boostani, and A Nasrabadi. “Estimation of hypnosis susceptibility based on electroencephalogram signal features”. In: *Scientia Iranica* 20.3 (2013), pp. 730–737.

[21] Kaitlin Cassady, Albert You, Alexander Doud, and Bin He. “The impact of mind-body awareness training on the early learning of a brain-computer interface”. In: *TECHNOLOGY* 02 (Sept. 2014), pp. 254–260.

[22] K Muller, K Bacht, D Prochnow, S Schramm, and R J.Seitz. “Activation of thalamus in motor imagery results from gating by hypnosis”. In: *NeuroImage* 66 (2013), pp. 361–367.

[23] Association World Medical. “World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.” In: *J Postgrad Med* 48.3 (2002). KIE: KIE Bib: human experimentation, pp. 206–208.

[24] K Nakayashiki, M Saeki, Y Takata, Y Hayashi, and T Kondo. “Modulation of event-related desynchronization during kinematic and kinetic hand movements”. In: *J Neuroeng Rehab* 11.90 (2014).

[25] Pietro Avanzini, Maddalena Fabbri-Destro, Riccardo Dalla Volta, Elena Daprati, Giacomo Rizzolatti, and Gaetano Cantalupo. “The dynamics of sensorimotor cortical oscillations during the observation of hand movements: an EEG study”. In: *PLoS One* 7.5 (2012), e37534.

[26] Y Benjamini and Y Hochberg. “Controlling the false discovery rate: a practical and powerful approach to multiple testing”. In: *Journal of the Royal Statistical Society* 57.1 (1995), pp. 289–300.

[27] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanaba, and K.R. Müller. “Optimizing Spatial Filters for Robust EEG Single-Trial Analysis”. In: *IEEE Signal processing magazine* (2008).

[28] Alexandre Gramfort et al. “MEG and EEG data analysis with MNE-Python”. In: *Frontiers in Neuroscience* 7 (2013), p. 267.

[29] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. “Optimal spatial filtering of single trial EEG during imagined hand movement”. In: *IEEE Transactions on Rehabilitation Engineering* 8.4 (2000), pp. 441–446.

[30] Fabien Lotte. “A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces”. In: *Guide to Brain-Computer Music Interfacing*. Ed. by Eduardo Reck Miranda and Julien Caste. Springer, 2014.

[31] Jerome H. Friedman. “Regularized Discriminant Analysis”. In: *Journal of the American Statistical Association* 84.405 (1989), pp. 165–175.

[32] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[33] “Leave-One-Out Cross-Validation”. In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 600–601.

[34] Brigitte Konradt, Salim Deeb, and O Berndt Scholz. “Motor imagery in hypnosis: accuracy and duration of motor imagery in waking and hypnotic states.” In: *The International journal of clinical and experimental hypnosis* 53 2 (2005), pp. 148–69.

[35] Christa Neuper, Michael Wörtz, and Gert Pfurtscheller. “ERD/ERS patterns reflecting sensorimotor activation and deactivation”. In: *Event-Related Dynamics of Brain Oscillations*. Ed. by Christa Neuper and Wolfgang Klimesch. Vol. 159. Progress in Brain Research. Elsevier, 2006, pp. 211 –222.

[36] Catherine Déry, Natasha K.J. Campbell, Michael Lifshitz, and Amir Raz. “Suggestion overrides automatic audiovisual integration”. In: *Consciousness and Cognition* 24 (2014), pp. 33 –37.

[37] Yudai Takarada and Daichi Nozaki. “Hypnotic suggestion alters the state of the motor cortex”. In: *Neuroscience Research* 85 (2014), pp. 28 –32.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-27

BRAIN-COMPUTER INTERFACE COMMUNICATION FOR A LOCKED IN CHILD WITH EPILEPTIC ENCEPHALOPATHY

E. Zewdie[1,2,3], Z. Jadavji[1,2,3], D. Kaketsis[1], A. Kirton[1,2,3]

[1] Calgary Pediatric Stroke Program, University of Calgary, AB, Canada
[2] Hotchkiss Brain Institute, University of Calgary, Calgary AB, Canada
[3] Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada

E-mail: ephrem.zewdie@ahs.ca

ABSTRACT: There are many technologies being developed to assist individuals with severe disability. Devices based on human machine interface have been used to restore or replace lost movement and communication. Unfortunately, these technologies have not yet been extensively explored in children with severe disability. This paper describes a case study of a 16 year old patient in a locked-in state with virtually no communication. She is unable to move her body and is non-verbal. BCI potential was assessed using the mindBEAGLE system which utilizes auditory and vibro-tactile modalities to evoke a response. Long-term monitoring on clinical EEG characterized the neurophysiology of the patient including nearly continuous generalized discharges. The best classification accuracy for auditory and vibrotactile BCI was 40%. Higher accuracy (54%) was achieved using the motor imagery modality. Clinical neurophysiology measures can inform or to some extend predict the success of BCI performance of different modalities contributing to the user-centered design in BCI development.

INTRODUCTION

Being able to communicate is an essential right and need of every human being. Imagine being aware of your environment, seeing, hearing and feeling everything around you, but being unable to move or speak to express yourself. Such Lock-in Syndrome (LIS) describes patients who are conscious and aware but severely paralyzed with complete immobility and loss of verbal communication. It can be caused by acute injuries (e.g., brainstem stroke, which is the most frequent cause of LIS) to chronic causes (e.g., amyotrophic lateral sclerosis; ALS)[1] that render the motor system non-functional[2].

LIS can be divided into incomplete (some voluntary movement in addition to eye movement), classic (preserved vertical eye movement and blinking) or complete (no voluntary movement). LIS involves loss of all voluntary muscle control other than sometimes restricted lateral or vertical eye movements. As a result, most communication attempts utilize eye movement [3] for communication control. For example, eye-gaze computers can be controlled by the user's eye movement and be interpreted by a second person given the user has good control of eye movements. This kind of communication requires the help of a second person who needs to be willing and capable of following time-consuming procedures.

Brain computer interface (BCI) systems based on sensorimotor rhythms and evoked potentials have been developed and used to provide such quadriplegic patients with potential options for communication and control independent of any movement input[4]. BCIs have been used by LIS patients to communicate[5,6] and even used to navigate a computer[7]. BCIs are also used to control devices that have permitted patients with spinal cord injury to regain movement[8] and control a wheelchair[9]. Virtually all of this progress has occurred in adults.

EEG based BCI can use different modalities to evoke a response, such as event-related potentials (ERP). The most common ones are visual, auditory and vibro-tactile. Investigation of different modalities of BCI in a single LIS patient found that the vibro-tactile modality was more effective for communication than visual and auditory modalities[10]. Similarly, Guger et al. [11] reported that 2/3 LIS patients reached a classification of 100% using vibro-tactile modality.

Such non-invasive BCI systems typically rely on EEG. However, some patients with acquired brain injuries and other conditions that may benefit from BCI may also have abnormal EEG. Epileptic encephalopathies may include frequent pathological generalized discharges that are not seizures but may impair cognition and awareness[12]. It is unknown if available EEG-based BI paradigms can still work in the presence of such overlying EEG patterns.

The aim of this paper is to report a case study of exploring different BCI modalities and associated event related potentials (ERP) and event-related desynchronization(ERD) used to establish communication in a child with LIS and epileptic

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-27

encephalopathy and explore how this clinical neurophysiology influences BCI performance.

MATERIALS AND METHODS

Participant

This is a single case study of a 16 year old female in complete locked-in state following post-operative brainstem hemorrhage 3 years prior to the study date. Shortly after a complicated operation to remove a large brainstem tumor, she incurred a life-threatening hemorrhage resulting in central brainstem herniation and bilateral injury to pontomedullary regions. She suffered a cardiac arrest and secondary hypoxic-ischemic injury of the deep grey matter (thalamus and putamen) without evidence of cortical or other areas of injury.

Although there were no clinical seizures, several EEG were completed in the first 3 months following the injury. These demonstrated intermittent slowing but no epileptiform discharges and age-appropriate posterior-dominant rhythms were often present.

At 13 years old, the child was clinically left in a state of complete quadriplegia aside from some very delayed small finger movements on the right side on command, no reliable eye movements but grossly intact visual fields to confrontation, grossly intact hearing based on isolated responses to sounds and possible responses to commands based on parental interpretation.
She required tracheostomy but not mechanical ventilation and was entirely gastrostomy dependent for nutrition.
After years of trying different communication systems including eye gaze devices without success, the patient's family approached the University of Calgary Pediatric Brain Computer Interface Laboratory. Three years after her injury she and her family were provided the opportunity to attempt several different modalities of BCI.

BCI Assessment

EEG voltage signals were acquired from the scalp using the g.USBamp EEG system (g.tec medical engineering GmbH, Austria), with either 8 or 16 active gel-based electrodes located according to the preset international 10-20 system. The system aligned the event EEG with event marker, band-pass filtered (0.1-30 Hz) and digitized the signal at 256 Hz and wirelessly transmitted the signal to a laptop for processing. A linear discriminant analysis was applied.

A commercially available EEG-based BCI consciousness assessment and communication system called mindBEAGLE (g.tec medical engineering GmbH, Austria) was used to perform the following tasks:

Auditory ERP: This paradigm was based on the auditory evoked P300 which involved identification of a target auditory stimulus when they are presented randomly mixed with non-target stimuli (oddball paradigm). The auditory paradigm consisted of frequent low tones (87.5%) and infrequent high tones (12.5%). The patient was asked to close her eyes to minimize distractions, and to silently count the number of times she heard the infrequent, target stimulus. Classification accuracy reported by the mindBEAGLE was used as an outcome.
Two sessions, two trials per session, were completed with two days.
Vibrotactile ERP with two tactors (VT2): two vibro-tactile stimulators were placed on each of the patient's left and right wrists. The patient was instructed to count the rare (12.5%) stimuli presented to one wrist to elicited P300. Classification accuracy reported by the mindBEAGLE was used as an outcome. Three trials (two trials on the first session and one trial on the second session) were completed with two days.

Vibrotactile ERP with three tactors (VT3): in addition to the two vibro-tactile stimulators that were placed on each of patient's left and right wrists, a third tactor was placed on the patients left lower leg as a distractor. Classification accuracy reported by the mindBEAGLE was used as an outcome. Three trials (two trials on the first session and one trial on the second session) were completed with two days.

Motor Imagery (MI): this modality utilizes sensorimotor rhythm modulation using hand motor imagery. The patient was instructed to imagine moving her right or left hand chosen randomly for 4s. The instructions came in random interval 0.5-2s to avoid adaptation. In total, the patient completed two runs of 30 imagined movements per run for each hand. Classification accuracy reported by the mindBEAGLE was used as an outcome. Three trials were completed on a single session.
Auditory, vibrotactile and MI tests were conducted with 15 min break between each paradigm. When the patient is tired, indicated by closing her eyes and not opening them for long time, the session was ended. For all tasks, data is extracted from -100 to 600 ms, around the simulation. The data is baseline corrected and averaged. A paired t-test was applied to measure significance difference between targets and non-targets, where the green shaded area indicates $p<0.05$[13].

For auditory and vibrotactile trials, the data is classified with a linear discriminant analysis (LDA) to distinguish target from non-target stimuli. This results in a classification accuracy ranging from 0 to 100 %. For MI, the data is used to train a common-spatial pattern (CSP) algorithm that weights each electrode based on discrimination accuracy. The data from the window 3-5 seconds is used to train the CSP algorithm followed by estimating a variance of a 1.5 sec window which is used to train LDA classifier. This results in a classification

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-27

accuracy ranging from 0 to 100 % with a chance accuracy level of 50 % (2-classes are discriminated)[13].

Studies took place in the recently established Pediatric BCI Laboratory at the Alberta Children's Hospital. Protocols have been approved by the institutional research ethics board.

Clinical Neurophysiology

Ambulatory EEG Recording: Ambulatory Trex EEG was recorded for 21 hours using 19 electrodes connected to Xltek brain monitor (Natus Neuro, Middleton, USA). The sampling frequency for the EEG recording was 256Hz.
Brainstem Auditory Evoked Potential (BAEP): BAEPs were evoked bilaterally using left and right monoaural rarefaction mixed frequency clinic stimulation at 11.1Hz and were recorded with surface electrodes (Cz-A1/A2).

Somatosensory Evoked Potential (SSEP): The SSEPS were obtained by electrical stimulation of the left and right median nerves at the wrist (10.5mA, 2.5Hz) and posterior tibial nerve at the ankle (13.2mA, 2.9Hz).

*Visual Evoked Potential (VEP)**:*** VEPs were obtained by full-field monocular LED stimulation of the left and right eyes.

RESULTS

The patient and her parents attended 3 Sessions with each lasting 90-120 minutes. Tolerability appeared favorable though feedback from the patient was obviously limited. Fatigue was suggested by parental report at variable timing during the sessions. State changes were otherwise challenging to determine based solely on clinical observation.

Clinical Neurophysiology

Ambulatory EEG Recording: The majority of the ambulatory EEG recording (~90%) consisted of generalized periodic discharges (GPD) at approximately 1Hz frequency which increased during sleep states (figure 1). These discharges often assumed a triphasic wave appearance.

BAEP: There were well defined and reproducible waves I, III and V on the left with no discernible peaks on the right. Absolute latency of peak II was normal, but the interpeak I-III latency was slightly over 2.5 SD above the mean. The interpeak III-V latency was normal, with a delayed absolute latency of peak V.

SSEP: A robust twitch was reported during the study for both upper limbs. Reproducible peaks were recorded at the erb's point and the corticomedullary junction with normal latencies (left Erbs 8.6ms, right Erbs 9.7). However, no discernible peaks were obtained cortically for the upper extremity. A robust twitch was noted in the right lower extremity, with a reproducible peak at the popliteal fossa (latency 8.9ms), however, no cortical peak was seen. No twitch was reports with simulation of the left lower extremity, even with the stimulation intensities over 20mA. No reproducible peaks were obtained at the popliteal fossa or cortically.

VEP: P1 waves recorded from Oz, O1 and O2 were present. Normal conduction latencies (90ms) were recorded in the visual pathways bilaterally.

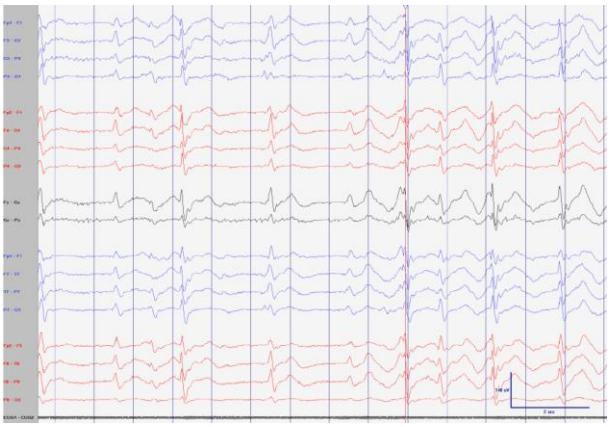


Figure 1: Generalized Periodic Discharge (GPD) in the shape of triphasic waves

BCI Assessment

Auditory P300: Both the target and non-target stimulus produced an average P300 with peak amplitude 8.75 µV, 7.68µV, 8.90µV and 8.77µV at CP1, CP2z , CPz and Pz, respectively (Figure 2).. The median accuracy for the best trial on the first day was 10%. On the second day at the same electrode locations, the P300 responses were 16.5uV and 10.8uV, respectively with the median accuracy improved to 40%.

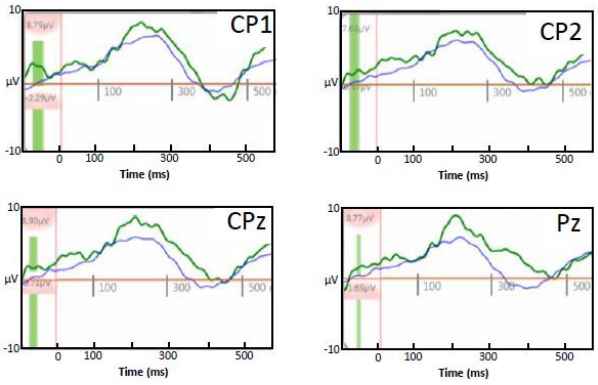


Figure 2: Auditory Evoked Potentials. P300 waveforms from selected electrodes (CP1, CP2, CPz, Pz) evoked by target (green) and non-target (blue) auditory stimulation. The vertical red line indicates when the cue

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-27

is present to the patient. The green shades indicate the time when the target and non-target waves are significantly different.

VT2 P300: The first VT 2 trial resulted in 0% median accuracy. The second and third VT2 trials resulted in 35% and 40% median accuracy, respectively. For the trial with 40% accuracy, the peak amplitude were 3.33µV, 9.35µV, 8.77µV and 5.48µV FCz, CP1, Cz, and C4, respectively, in response to target vibro-tactile stimulus (Figure 3).

VT3 P300: All VT 3 trials resulted 0% median accuracy.

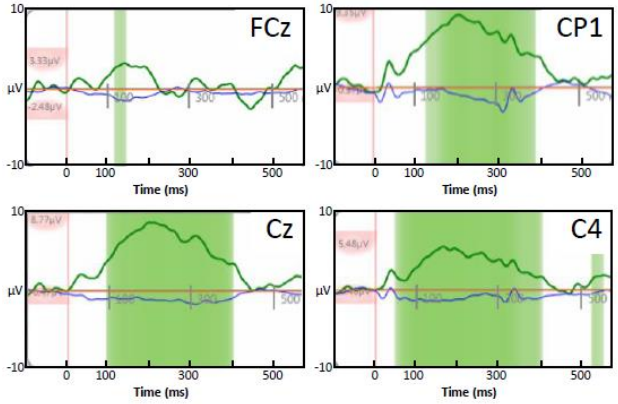


Figure 3: Vibro-tactile Evoked Potentials (VT2). P300 waveforms from selected electrodes (FCz, CP1, Cz, C4) evoked by target (green) and non-target (blue) vibro-tactile stimulation. The vertical red line indicates when the cue is present to the patient. The green shades indicate the time when the target and non-target waves are significantly different.

Motor Imagery: The first trial resulted in a median accuracy of 46.3%, which improved to 54.3% accuracy in after two trials (Figure 4).

DISCUSSION

While earlier results have shown that LIS patients can use BCI based on motor imagery[14], auditory P300 [15] and vibro-tactile P300[16], this case study extends the potential use of such BCI systems for a child with LIS. With this patient, communication using any modality was not possible due to the low classification accuracy. This in part may have related to comorbid epileptic encephalopathy, suggesting that baseline clinical neurophysiology may be important in advancing BCI applications in such clinical populations.

Our unsatisfactory BCI performance may partly be explained by the clinical neurophysiology that was observed. The presence of EEG-patterns consistent with epileptic encephalopathy may indicate more severe global cerebral dysfunction.

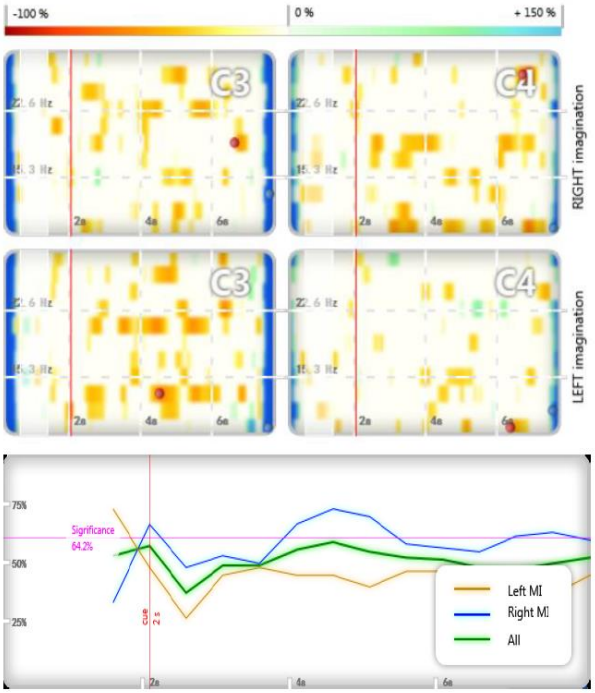


Figure 4: Motor Imagery (MI) evoked event related desynchronization and accuracy. The top 4 graphs show EEG desynchronization map for the right (top) and left (bottom) imagination. The vertical red line indicates when the cue is present to the patient. The bottom plot shows accuracy for left MI (yellow), right MI (blue) and all combined (green). The horizontal purple line indicates 64.2% accuracy which is the threshold for significance.

However, many other pediatric conditions that feature such frequent, generalized discharges are not uniformly associated with severe impairment of cognition or consciousness. For example, epileptic encephalopathies like continuous spike-and-wave discharges in sleep (CSWS), electrical status epilepticus I sleep (ESES) and Landau-Kleffner syndrome (LKS) can be seen in fully conscious, ambulatory children who might appear normal aside from slow effects on learning and development over time. In other words, young brains can sometimes be highly functional “beneath” such grossly abnormal EEG recordings. Such patterns are also potentially treatable with anticonvulsant medications or steroids[17,18], suggesting a potentially modifiable factor that could improve BCI performance in such subjects.

The BAEP test also indicated abnormal auditory activation which may have affected BCI performance based on auditory evoked potentials. An earphone with identical right and left output may be used to overcome unbalanced auditory evoked potential.

The abnormal SSEP of upper and lower extremities and lack of cortical responses suggest abnormal conduction above the cervicomodullary junction, consistent with the known injuries in this patient. The lack of twitches

Published by Verlag der TU Graz
Graz University of Technology

also indicates peripheral nerve conduction defect. As a result, BCI performance based on both VT2 and VT3 was likely limited due to interruption of the signal at cervicomodullary junction. This suggests SSEP results can inform better placement of vibrators for BCI based on vibro-tactile stimulation.

The above results emphasize the importance of user-centered design in BCI development. While some studies find that the visual modality to be superior to other[19], Kaufmann and colleagues[10] found out that vibro-tactile based BCI resulted in the highest classification accuracy. Furthermore, the accuracies achieved in healthy participants do not necessarily transfer too LIS patients[10].

Unique ethical consideration must be taken into account when working with adolescent patients lacking autonomy and communication. As in the case of complete locked-in syndrome, consent to participate is determined by the child's parent or guardian. Under such sensitive circumstance, it is important to maintain balance between offering hope for improved communication and realistic expectations among families.

CONCLUSION

This case study explored different BCI modalities that were used to establish communication in a child with LIS and epileptic encephalopathy. Our unsatisfactory assessment of BCI performance was partly explained by the clinical neurophysiology that was observe. Most patients have one or more clinical neurophysiology done following their injury. The data from these tests should be used to refine the classifier used in BCI as well as to choose the best BCI modality for better BCI performance.

REFERENCES

[1] Lulé D, Zickler C, Häcker S, Bruno MA, Demertzi A, Pellas F, et al. Life can be worth living in locked-in syndrome. Prog Brain Res 2009;177:339–51. doi:10.1016/S0079-6123(09)17723-3.

[2] Posner JB, Saper CB, Schiff N, Plum F. Plum and Posner's Diagnosis of Stupor and Coma. Fourth edition. Oxford ; New York: Oxford University Press; 2007.

[3] Laureys S, Pellas F, Van Eeckhout P, Ghorbel S, Schnakers C, Perrin F, et al. The locked-in syndrome : what is it like to be conscious but paralyzed and voiceless? Prog Brain Res 2005;150:495–511. doi:10.1016/S0079-6123(05)50034-7.

[4] Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, et al. Brain-computer interface technology: a review of the first international meeting. IEEE Trans Rehabil Eng 2000;8:164–73.

[5] Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, et al. A spelling device for the paralysed. Nature 1999;398:297–8. doi:10.1038/18581.

[6] Nijboer F, Sellers EW, Mellinger J, Jordan MA, Matuz T, Furdea A, et al. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. Clin Neurophysiol 2008;119:1909–16. doi:10.1016/j.clinph.2008.03.034.

[7] Karim AA, Hinterberger T, Richter J, Mellinger J, Neumann N, Flor H, et al. Neural internet: Web surfing with brain potentials for the completely paralyzed. Neurorehabil Neural Repair 2006;20:508–15. doi:10.1177/1545968306290661.

[8] Pfurtscheller G, Müller GR, Pfurtscheller J, Gerner HJ, Rupp R. 'Thought'--control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. Neurosci Lett 2003;351:33–6.

[9] A Brain-Actuated Wheelchair: Asynchronous and Non-Invasive Brain-Computer Interfaces for Continuous Control of Robots | Request PDF. ResearchGate n.d. doi:http://dx.doi.org/10.1016/j.clinph.2008.06.001.

[10] Kaufmann T, Holz EM, Kübler A. Comparison of tactile, auditory, and visual modality for brain-computer interface use: a case study with a patient in the locked-in state. Front Neurosci 2013;7:129. doi:10.3389/fnins.2013.00129.

[11] Guger C, Spataro R, Allison BZ, Heilinger A, Ortner R, Cho W, et al. Complete Locked-in and Locked-in Patients: Command Following Assessment and Communication with Vibro-Tactile P300 and Motor Imagery Brain-Computer Interface Tools. Front Neurosci 2017;11:251. doi:10.3389/fnins.2017.00251.

[12] Van Bogaert P. Epileptic encephalopathy with continuous spike-waves during slow-wave sleep including Landau-Kleffner syndrome. Handb Clin Neurol 2013;111:635–40. doi:10.1016/B978-0-444-52891-9.00066-X.

[13] Guger C, Allison B, Spataro R, Bella VL, Kammerhofer A, Guttmann F, et al. MindBEAGLE — A new system for the assessment and communication with patients with disorders of consciousness and complete locked-in syndrom. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017, p. 3008–13. doi:10.1109/SMC.2017.8123086.

[14] McFarland DJ, Vaughan TM. BCI in practice. Prog Brain Res 2016;228:389–404. doi:10.1016/bs.pbr.2016.06.005.

[15] Risetti M, Formisano R, Toppi J, Quitadamo LR,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-27

Bianchi L, Astolfi L, et al. On ERPs detection in disorders of consciousness rehabilitation. Front Hum Neurosci 2013;7:775. doi:10.3389/fnhum.2013.00775.

[16] Lugo ZR, Rodriguez J, Lechner A, Ortner R, Gantner IS, Laureys S, et al. A vibrotactile p300-based brain-computer interface for consciousness detection and communication. Clin EEG Neurosci 2014;45:14–21. doi:10.1177/1550059413505533.

[17] Buzatu M, Bulteau C, Altuzarra C, Dulac O, Van Bogaert P. Corticosteroids as treatment of epileptic syndromes with continuous spike-waves during slow-wave sleep. Epilepsia 2009;50 Suppl 7:68–72. doi:10.1111/j.1528-1167.2009.02224.x.

[18] Shinnar S, Kang H, Berg AT, Goldensohn ES, Hauser WA, Moshé SL. EEG abnormalities in children with a first unprovoked seizure. Epilepsia 1994;35:471–6.

[19] Riccio A, Mattia D, Simione L, Olivetti M, Cincotti F. Eye-gaze independent EEG-based brain-computer interfaces for communication. J Neural Eng 2012;9:045001. doi:10.1088/1741-2560/9/4/045001.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-28

PEDIATRIC BRAIN-COMPUTER INTERFACE COMPETENCY: A PILOT STUDY

Dion Kelly MBT, Ephrem Zewdie PhD, Adam Kirton, MD

Alberta Children's Hospital Research Institute, Hotchkiss Brain Institute,
Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

E-mail: adam.kirton@ahs.ca

ABSTRACT: Imagine having no functional control over your body—your cognition is unaffected, but you cannot speak or move. Unfortunately, this is the case for many children with severe neurological disabilities, including quadriplegic cerebral palsy. Brain-computer interfaces (BCIs) represent a potential solution for such individuals. BCIs have been widely studied in adults with neurological disabilities, however there is limited research into their use in children. We have therefore established a clinical pilot pediatric BCI program. In order to gauge competency in patients, we aimed here to establish the ability of normal children to use non-invasive BCI systems. Nine healthy children completed tasks on five established EEG-based BCI systems. Performance was assessed by accuracy scores for each paradigm. Our results demonstrate that children can effectively operate a BCI system with accuracies above 60%—comparable to what's been reported in adults. Establishing performance norms for typically developing children across specific BCI systems will inform program development and provide comparisons for children with disabilities.

INTRODUCTION

Cerebral palsy (CP) is the leading cause of lifelong disability, affecting 17 million people worldwide [1]. It is estimated that 1 in 500 babies are born with CP each year, which translates to over 60,000 affected Canadians [2], [3]. A non-progressive neurologic condition characterized by motor impairments, CP is the result of an injury to the brain during early development. [4], [5]. CP may be induced by complications during birth such as hypoxic-ischemic encephalopathy, stroke, prematurity, infections, brain malformations, or an accumulation of bilirubin [4], [5]. Depending on the severity of motor impairment and the limbs affected, CP can be classified into different subtypes: monoplegia describes mild impairments affecting only one limb; diplegia describes impairments of the lower body; hemiplegia describes impairments on one side of the body; and quadriplegia describes the most severe type of CP, which affects all four limbs [6]. Quadriplegic CP often results in devastating loss of all voluntary movement and verbal communication. A substantial proportion of such individuals have preserved cognitive function and may be intellectually normal but trapped in a body that cannot move.

Perinatal stroke is a well understood human model for CP. Perinatal stroke is defined as "a focal disruption" of blood flow to the brain, which occurs between 20 weeks of fetal life up to three weeks after birth [7]. The outcomes for perinatal stroke survivors are usually poor, with neurodevelopmental deficits occurring in 75% of patients and often resulting in a lifetime of disability [8]. It is the leading cause of hemiparetic CP, which is characterized by unilateral motor deficits, as well as higher risks for other developmental disabilities [9], [10]. With no strategies for prevention, recovery mechanisms for children with CP involve supportive care and rehabilitation strategies. When perinatal stroke occurs bilaterally, it can result in quadriplegic CP much like the more global brain injuries described above.

Therapies to improve motor outcomes for children affected by perinatal stroke are limited. Conventional rehabilitation strategies lack supporting evidence and numerous provided interventions are ineffective [11]. In randomized clinical trials, constraint-induced movement therapy (CIMT) has been shown to be effective in improving motor control in children with congenital hemiparesis [11]. However, this intervention excludes individuals with severe CP because they do not meet the minimum criteria for motor function necessary for participation [12]. Recently, neuromodulation therapies, such as transcranial direct current stimulation (tDCS) and repetitive transcranial magnetic stimulation (rTMS) have been combined with CIMT and have shown additive effects on improvement of motor function in children with hemiparetic CP [13], [14]. Recent evidence from adult stroke suggests that BCI can provide new avenues for motor rehabilitation [15], [16]. This concept has not yet been attempted in children with hemiparesis where limited understanding of how children can perform on BCI systems is a barrier to progress.

Children with severe CP are unable to benefit from the combination therapies noted above due to their extensive motor impairments, which prevent them from moving their limbs and, in many cases, speaking. A 2012 review found that 1 in 3 children who have CP cannot walk, and 1 in 4 children with CP cannot talk [17].

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-28

In contrast to most children with perinatal stroke and hemiparetic CP, those with severe bilateral neurological disabilities such as quadriplegic CP, have limited or no motor function and are fully dependent on others for all activities of daily living [18]. Even more unfortunate is that in some children, the cortex is spared and cognitive function is partially or entirely preserved [18]. With virtually no way to communicate or interact with their environments, such cognitively aware children are literally trapped inside a body that does not work. Consequently, they are deprived of their fundamental human rights, including the right to freedom of expression.

Brain-computer interfaces (BCIs) represent a promising solution by which such children might be able to better communicate, interact with their environment, and better realize their own independence [19]. BCIs can allow for real-time mental control of an external device, such as a computer cursor or a communication device. This can be achieved via non-invasive sampling of brain signals with electroencephalography (EEG) [20]. Such mental control signals can be produced in response to external stimuli (evoked potentials and P300 event-related potentials) or they can be generated internally by imagined movements (sensorimotor rhythms) [20], [21]. While there has been an explosion in recent BCI development, the majority of studies have been in adults with a relative neglect of pediatric populations [21], [22]. BCIs represent a potential breakthrough for children with severe neurological disabilities, with the potential to greatly increase their independence and enhance their quality of life [23].

A recent study demonstrated that children can control simple EEG-based BCI systems with minimal training and performance comparable to adults [24]. Clinical research is required to fill a fundamental gap that currently exists between technology development and the pediatric patients and families who could benefit from these new tools [25]. We have therefore initiated a clinical pediatric BCI program to provide severely disabled children with unique opportunities to try a suite of BCI systems to explore possible new avenues for their independence. A major limitation in advancing this program is that we do not yet know how well typically developing children can perform such tasks.

Therefore, the main goal of this study was to establish baseline levels of performance for typically developing children across a suite of non-invasive BCI systems and tasks. This will provide the foundational knowledge necessary for future applications of this technology in children with severe neurological disabilities that have no motor ability. We hypothesized that most typically developing children can achieve basic BCI competency on all modern systems comparable to that seen in adults.

MATERIALS AND METHODS

This was an open-label pilot study of typically developing, school-aged children. Participants were recruited from the community. Nine healthy volunteers aged 11-17 years (2 female) have participated at the time of this report (Tab. 1). Studies took place in the recently established Pediatric BCI Laboratory at the Alberta Children's Hospital. Protocols have been approved by the institutional research ethics board.

Table 1. Details of the participants in this study.

Participant (P)	Sex	Age (years)
1	M	14
2	M	11
3	M	16
4	F	16
5	M	17
6	F	15
7	M	12
8	M	16
9	M	14

Modern, commercially available EEG-based BCI systems were employed (mindBEAGLE and intendiX systems from g.tec medical engineering GmbH, Austria). EEG voltage signals were acquired from the scalp using the g.USBamp EEG system (g.tec medical engineering GmbH, Austria) with either 8 or 16 active gel-based electrodes located according to the international 10-20 system. Electrode locations differed based on which BCI system was being tested. For the mindBEAGLE system, signals were sampled from electrodes at positions FC3, FCz, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1 CPz, CP2, CP4 and Pz. For the intendiX system, signals were sampled from electrodes positioned at Fz, Cz, P3, Pz, P4, PO7, Oz, and PO8. Additionally, a ground electrode was positioned at AFz and a reference electrode was placed on the ear lobe. Data were sampled at 256 Hz and signal features were extracted using band-pass filtering between 0.1-30 Hz for the mindBEAGLE and 0.5-30 Hz for the intendiX. The digitized data was then transmitted to a laptop via universal serial bus (USB) for processing using a linear discriminant analysis (LDA).

Five BCI paradigms well established in adults were studied: sensorimotor rhythms (SMRs), and auditory, visual, and sensory P300 event-related potentials (ERPs) (two- and three-tactor vibro-tactile stimulation). Each paradigm consisted of tasks with predefined goals. For the mindBEAGLE system, the first run of each paradigm was to train a classifier to recognize the unique brain signals of the individual participant. This was completed using a generic classifier calculated from an online mixed dataset of multiple participants [26]. For the intendiX system, a classifier was trained by spelling the word "LUKAS" at 15 flashes (the number of times each row and column flashed) [27]. The classification was done using LDA, as was developed by Guger et al. (2012) by extracting individual ERPs following each trial [28].

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-28

Task 1: Communication board spelling using visual P300 event-related potentials: In the visual P300 paradigm, participants performed a spelling task using the intendiX row-column speller at various difficulty levels (flash rates). Faces of adult celebrities were presented in colour as the stimuli for the flashing characters of the spelling board [29], [30]. Participants were seated in front of a laptop and were instructed to remain still and relaxed. Once a classifier was trained, the participant started by spelling the 3-letter word "cat" at a range of 8 (min.) to 10 (max.) flashes. The participant was instructed to pay attention to and to silently count the number of times the target letter flashed while disregarding other flashes, until a letter was selected and displayed on the screen. If a letter other than the target letter was chosen, the participant was instructed to move on to the next letter in the word. This trial was counted as a failed attempt at the specified flash rate and the participant was instructed to try again. If the participant was successful in spelling the word at this difficulty level, they advanced to the next level, which involved spelling the same word with less time (decreased number of flashes). Participants had three chances to spell the word before it was established that they were unable to perform the task at the difficulty level. If the participant was unsuccessful in spelling the word at the starting maximum of 10 flashes after three chances, the flash number was increased by two after each unsuccessful trial, up to a maximum of fourteen flashes. If the participant remained unsuccessful at fourteen flashes, a new classifier was generated and the participant retried. Accuracy was averaged over each trial per specified number of flashes. For example, if the participant correctly selected 1 out of 3 letters in the word 'cat', they received an accuracy score of 33% for that trial and were instructed to try again at the same flash rate. Subsequently, if they correctly selected 2 out of 3 letters, they received an accuracy score of 67% and were instructed to try again at the same flash rate. If, on their third try, the participant correctly selected all of the letters of the word, they received an accuracy score of 100% for that trial and advanced to the next level with a lower flash rate. The accuracy scores for the three trials at the same specified flash rate were then averaged into one score for that flash rate. Performance was determined by the lowest number flashes in which the participant was successfully able to spell the word within three trials.

Task 2: Identification of a target stimulus in the oddball paradigm using 3-tactor vibro-tactile P300 event-related potentials (VTP3): Vibro-tactile stimulators were fixed on each wrist, with an additional distractor stimulator fixed to the lower right leg. Using the mindBEAGLE system (g.tec medical engineering GmbH, Austria), participants were asked to identify target stimuli during the oddball paradigm. Non-target, distractor stimuli on the leg were presented frequently (75%), while target stimuli on the left and right wrists occurred infrequently (12.5% for each wrist). Participants were asked to close their eyes to minimize distractions, and to silently count the number of times they felt the infrequent, target stimulus in order to evoke a P300 response. The primary outcome was classification accuracy, as generated by the mindBEAGLE software by comparing target and non-target signals generated by the system with those generated by the participant.

Task 3: Identification of a target stimulus in the oddball paradigm using 2-tactor vibro-tactile P300 event-related potentials (VTP2): Vibro-tactile stimulators were fixed on each wrist. Using the mindBEAGLE system, participants were asked to identify target stimuli during the oddball paradigm. Non-target stimuli were presented frequently (87.5%) while target stimuli were presented infrequently (12.5%). Participants were asked to close their eyes to minimize distractions, and to silently count the number of times they felt the infrequent, target stimulus in order to evoke a P300 response. The primary outcome was classification accuracy, which was calculated by the mindBEAGLE software.

Task 4: Identification of a target stimulus in the oddball paradigm using auditory event-related potentials (AEP): In the auditory P300 paradigm, performance was assessed on the mindBEAGLE system and involved the identification of a target stimulus by selectively paying attention to deviant, target stimuli among frequent, non-target stimuli (oddball paradigm). The auditory paradigm consisted of frequent low tones (87.5%) and infrequent high tones (12.5%) and lasted approximately 7.3 minutes. Participants were asked to close their eyes to minimize distractions, and to silently count the number of times they heard the infrequent, target stimulus. The primary outcome was classification accuracy, which was calculated by the mindBEAGLE software.

Task 5: Sensorimotor rhythm modulation using hand motor imagery: In the SMR paradigm, participants were assessed on the mindBEAGLE system. They were instructed to imagine left- and right-hand movements (opening and closing of each hand) for 8 seconds, with 2 seconds between trials. One run consisted of 60 imagined movements in randomized order, lasting approximately 9 minutes. The primary outcome was a classification accuracy score ranging from 0 to 100%. This score was generated by the mindBEAGLE software via cross validation of the classifier data with the testing data.

RESULTS

All nine participants completed one or more sessions successfully. There were no serious adverse events. Procedures were generally well tolerated. One participant expressed that auditory stimuli in the AEP paradigm induced feelings of sleepiness. Sessions averaged 40-60 minutes with the longest being 90 minutes. No participant ended the session early and all were agreeable

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-28

to participating again.

Visual P300 paradigm: Figure 1 demonstrates the performance of each of 5 participants, as indicated by average accuracy score at various number of flashes. The average accuracy scores ranged from 33% to 100%. The general trend of the results show that performance decreased as the number of flashes decreased and, consequently, the difficulty level increased. Three out of the five participants achieved accuracy scores of 100% on the first two difficulty levels (8-10 and 6-8 flashes) (P1, P3, P9), however their accuracy scores decreased by as much as 45% at a flash rate between 4-6 flashes. The lowest number of flashes achieved by any participant was 2-4 flashes, with the highest accuracy score at this flash rate being 50%. Participant 2 (P2) was unable to spell the word after three tries at 8-10 flashes, therefore the flash rate was increased by two up to 12-14 flashes. Correspondingly, a linear increase in average accuracy of 33% was observed each time flash rate was increased, up to 100% at 12-14 flashes.

VTP3 paradigm: Five participants were assessed in the VTP3 paradigm (Fig. 2). Accuracy scores ranged from 25% to 100%, with a median accuracy of 70% and an average accuracy of 66%. Three out of the five participants demonstrated improvements in accuracy scores between their first and last run (P2, P7, P8). The accuracy score of participant 5 did not change between the first and last run (remained at 90%) however, once a new classifier was generated for the participant (following run 1) an improvement of 30% was observed between the second and third run and the last run. A decrease in performance was observed in participant 4, however this participant performed only two runs.

VTP2 paradigm: Four participants were assessed in the VTP2 paradigm over two runs (Fig. 2). Accuracy scores ranged from 65% to 100%, with a median accuracy of 100% and an average accuracy of 94%. All four participants achieved 100% in the first run using the generic classifier. In the second run, one participant achieved 100% using their own classifier (P1), while participants 2 and 8 also achieved high accuracy scores of 90% and 95%, respectively. The lowest accuracy score observed in this paradigm was 65% (P4).

AEP paradigm: Two participants were assessed in the AEP paradigm (Fig. 2). Participant 5 completed three runs and was able to achieve an accuracy score of 100% in all runs. Participant 8 completed one run using the generic classifier and achieved an accuracy score of 100%.

SMR paradigm: One participant was assessed in the SMR paradigm in one run using the generic classifier (Fig. 2). Participant 8 was able to achieve an accuracy score of 61%.

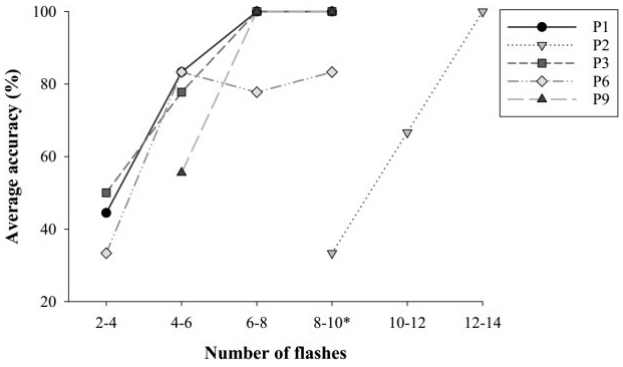


Figure 1. Average accuracy of five participants performing a spelling task using visual P300 event-related potentials on the intendiX row-column speller system. *indicates the starting number of flashes.

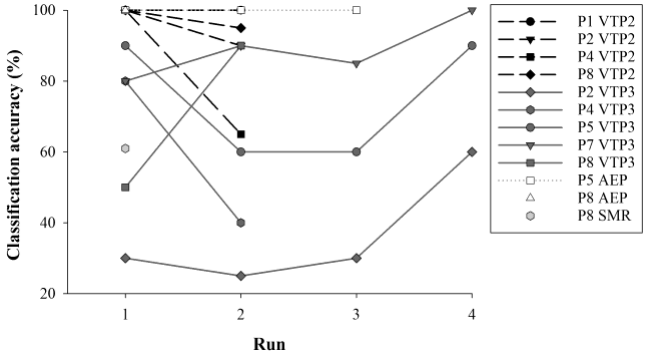


Figure 2. Classification accuracies for six participants who completed various runs in VTP2, VTP3, AEP, and SMR paradigms using the mindBEAGLE system.

DISCUSSION

This pilot study investigated performance on multiple noninvasive modern BCI systems in a pediatric population. Our results suggest that many children can operate such systems with minimal training and favorable tolerability. Performance appears within the ranges established for adults though additional testing on larger samples is obviously required. Establishing pediatric-specific performance standards will facilitate the development of BCI training programs for children with severe neurological impairments.

Our findings included children successfully completing a spelling task using the intendiX system with an average accuracy of at least 80% within a maximum number of 14 flashes. This is comparable to performance in adults using the same system [31]**.** Most participants were able to achieve an average accuracy of at least 80% within a maximum of 10 flashes, except for one participant. Interestingly, this individual was also the youngest to be assessed in this paradigm and was at least three years younger than the other participants performing the same task. Such communication systems may be particularly valuable for clinical populations such as severe quadriplegic cerebral palsy where affected children are

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-28

often non-verbal but possess the visual fields and eye movements required to use such systems.

Across the four paradigms assessed using the mindBEAGLE system, performance was more wide ranging across our modest sample. Abilities to operate were suggested in most subjects though accuracies ranged from 25% to 100%. The average accuracy was 66% for the VTP3 paradigm, 94% for VTP2, and 100% for AEP. The sample size of this pilot study is too small to reach any conclusions regarding relative performance across the different tasks. However, our observed performance on the VTP2 and AEP paradigms does appear to be on par with that seen in adults [32]. Average performance in the VTP3 paradigm in adults has been reported to be 88% [32], [33]. Although we did not clearly observe comparable performance in this pediatric population, it is important to recognize that our performance results were all above the recommended threshold of 60% suggested by g.tec medical engineering [33]. The recommended threshold for the SMR paradigm is 61%, which was achieved by the sole participant in their first and only run. This accuracy score is comparable to and better than some scores achieved by healthy adults in multiple runs [32], [33]. All participants also achieved at least 60% accuracy in at least one run using the mindBEAGLE system.

CONCLUSION

We report pilot data suggesting that many children can competently operate modern BCI systems with accuracy scores comparable to established thresholds necessary for determining conscious awareness, device control, and augmented communication. Additionally, we have demonstrated that at least some children can rapidly achieve BCI competency comparable to that observed in adults. The comparability in performance between children and adults has important implications for future BCI research in both populations where results from progressing adult research may be applicable to children while lessons learned in the more plastic brains of children may also advance the field of BCI. If nothing else, this work provides a useful starting point for screening programs and the development of BCI training programs for children with severe neurological disabilities.

In future work, we intend to investigate performance on additional BCI systems across multiple runs and we will evaluate factors affecting performance, including age and gender. Additionally, once we have established a complete baseline level of performance across these systems in typically developing children, we can begin to modify the systems and tailor tasks specifically for use by disabled pediatric populations with no motor ability. This will provide such children with increased independence and will enhance their quality of life.

REFERENCES

[1] M. Oskoui, F. Coutinho, J. Dykeman, N. Jetté, and T. Pringsheim, “An update on the prevalence of cerebral palsy: a systematic review and meta-analysis,” *Dev. Med. Child Neurol.*, vol. 55, no. 6, pp. 509–519, Jun. 2013.

[2] N. Colledge, “A Guide to Cerebral Palsy,” The Ontario Federation for Cerebral Palsy, 2011.

[3] C. Cans, J. De-la-Cruz, and M.-A. Mermet, “Epidemiology of cerebral palsy,” *Paediatr. Child Health*, vol. 18, no. 9, pp. 393–398, Sep. 2008.

[4] I. Daly *et al.*, “Exploration of the neural correlates of cerebral palsy for sensorimotor BCI control,” *Front. Neuroengineering*, vol. 7, p. 20, 2014.

[5] K. Krigger, “Cerebral Palsy: An Overview,” *Am. Fam. Physician*, vol. 73, no. 1, pp. 91–100, Jan. 2006.

[6] P. Minocha, S. Sitaraman, and P. Sachdeva, “Clinical Spectrum, Comorbidities, and Risk Factor Profile of Cerebral Palsy Children: A Prospective Study,” *J. Pediatr. Neurosci.*, vol. 12, no. 1, pp. 15–18, 2017.

[7] T. N. Raju, K. B. Nelson, D. Ferriero, and J. K. Lynch, “Ischemic perinatal stroke: summary of a workshop sponsored by the National Institute of Child Health and Human Development and the National Institute of Neurological Disorders and Stroke,” *Pediatrics*, vol. 120, pp. 609–616, Sep. 2007.

[8] A. Kirton, “Modeling developmental plasticity after perinatal stroke: defining central therapeutic targets in cerebral palsy,” *Pediatr Neurol*, vol. 48, no. 2, pp. 81–94, Feb. 2013.

[9] A. Kirton, “Life After Perinatal Stroke,” *Stroke*, vol. 44, no. 11, pp. 3265–3271, Oct. 2013.

[10] D. G. Whitney *et al.*, “Age trajectories of musculoskeletal morbidities in adults with cerebral palsy,” *Bone*, vol. 114, pp. 285–291, Sep. 2018.

[11] I. Novak *et al.*, “A systematic review of interventions for children with cerebral palsy: state of the evidence,” *Dev. Med. Child Neurol.*, vol. 55, no. 10, pp. 885–910, Oct. 2013.

[12] E. Taub, G. Uswatte, and R. Pidikiti, “Constraint-induced Movement Therapy: A new family of techniques with broad application to physical rehabilitation - A clinical review,” *J. Rehabil. Res. Dev.*, vol. 36, no. 3, pp. 237–251, 1999.

[13] B. Gillick *et al.*, “Transcranial direct current stimulation and constraint-induced therapy in cerebral palsy: A randomized, blinded, sham-controlled clinical trial,” *Eur. J. Paediatr. Neurol.*, vol. 22, no. 3, pp. 358–368, May 2018.

[14] A. Kirton *et al.*, “Brain stimulation and constraint for perinatal stroke hemiparesis: The PLASTIC CHAMPS Trial,” *Neurology*, vol. 86, no. 18, pp. 1659–1667, May 2016.

[15] M. A. Cervera *et al.*, “Brain-computer interfaces

Published by Verlag der TU Graz
Graz University of Technology

for post-stroke motor rehabilitation: a meta-analysis," *Ann. Clin. Transl. Neurol.*, vol. 5, no. 5, pp. 651–663, May 2018.

[16] 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.

[17] I. Novak, M. Hines, S. Goldsmith, and R. Barclay, "Clinical Prognostic Messages From a Systematic Review on Cerebral Palsy," *Pediatrics*, vol. 130, no. 5, pp. e1285–e1312, Nov. 2012.

[18] M.-A. Bruno *et al.*, "Locked-In Syndrome in Children: Report of Five Cases and Review of the Literature," *Pediatr. Neurol.*, vol. 41, no. 4, pp. 237–246, Oct. 2009.

[19] J. N. Mak and J. R. Wolpaw, "Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 187–199, 2009.

[20] R. A. Ramadan and A. V. Vasilakos, "Brain computer interface: control signals review," *Neurocomputing*, vol. 223, pp. 26–44, Feb. 2017.

[21] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain-computer interfaces for communication and rehabilitation," *Nat. Rev. Neurol.*, vol. 12, no. 9, pp. 513–525, Sep. 2016.

[22] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *Lancet Neurol.*, vol. 7, no. 11, pp. 1032–1043, Nov. 2008.

[23] E. Mikołajewska and D. Mikołajewski, "The prospects of brain — computer interface applications in children," *Cent. Eur. J. Med.*, vol. 9, no. 1, pp. 74–79, Feb. 2014.

[24] J. Zhang, Z. Jadavji, E. Zewdie, and A. Kirton, "Evaluating If Children Can Use Simple Brain Computer Interfaces," *Front. Hum. Neurosci.*, vol. 13, p. 24, 2019.

[25] S. Letourneau, E. Zewdie, L. Burkholder, J. Andersen, and A. Kirton, "Clinician Awareness of Brain Computer Interfaces and Eligible Pediatric Populations," in *ANNALS OF NEUROLOGY*, 2018, vol. 84, pp. S416–S416.

[26] C. Guger *et al.*, "MindBEAGLE — A new system for the assessment and communication with patients with disorders of consciousness and complete locked-in syndrom," 2017, pp. 3008–3013.

[27] G. Krausz, R. Ortner, and E. Opisso, "Accuracy of a Brain Computer Interface (P300 Spelling Device) used by People with Motor Impairments," *J. CyberTherapy Rehabil.*, vol. 4, no. 2, pp. 148–151, 2011.

[28] C. Guger, G. Krausz, B. Z. Allison, and G. Edlinger, "Comparison of Dry and Gel Based Electrodes for P300 Brain–Computer Interfaces," *Front. Neurosci.*, vol. 6, 2012.

[29] T. Kaufmann, S. M. Schulz, C. Grünzinger, and A. Kübler, "Flashing characters with famous faces improves ERP-based brain–computer interface performance," *J. Neural Eng.*, vol. 8, no. 5, p. 056016, Sep. 2011.

[30] C. Guger, R. Ortner, S. Dimov, and B. Allison, "A comparison of face speller approaches for P300 BCIs," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 004809–004812.

[31] C. Guger *et al.*, "How many people are able to control a P300-based brain–computer interface (BCI)?," *Neurosci. Lett.*, vol. 462, no. 1, pp. 94–98, Sep. 2009.

[32] A. Schnurer, A. Espinosa, and C. Guger, "P125. mindBEAGLE: A BCI for communication and assessment of consciousness for patients with disorders of consciousness - ScienceDirect," *Clin. Neurophysiol.*, vol. 126, no. 8, pp. e152–e153, Aug. 2015.

[33] B. Z. Allison, W. Cho, R. Ortner, A. Heilinger, G. Edlinger, and C. Guger, "Validation of a Brain-Computer Interface (BCI) System Designed for Patients with Disorders of Consciousness (DOC): Regular and Sham Testing with Healthy Participants," in *Augmented Cognition. Enhancing Cognition and Behavior in Complex Human Environments*, 2017, pp. 253–265.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-29

UTRECHT NEUROPROSTHESIS SYSTEM: NEW FEATURES TO ACCOMMODATE USER NEEDS

B.H. Van der Vijgh, M.A.Van den Boom, M.P. Branco, S. Leinders, Z.V. Freudenburg, E.G.M. Pels, M.J. Vansteensel, N.F. Ramsey and E.J. Aarnoutse

UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, Utrecht, the Netherlands

E-mail: b.h.vandervijgh@umcutrecht.nl

ABSTRACT: Individuals with locked-in syndrome can benefit from Brain-Computer Interfaces (BCIs) as an alternative assistive technology for communication. The Utrecht NeuroProsthesis (UNP) is a fully implanted ECoG based BCI that provides the user with independent control of a computer using intentional brain signals. In order to avoid technology abandonment and to stimulate home use, a user-centered approach to design and development of the system is essential. Here we show accommodation of several of the needs expressed by users of the UNP system, including new features that provide the user with control over the system during the night and which increase training efficacy.

INTRODUCTION

The Utrecht NeuroProsthesis (UNP; www.neuroprosthesis.eu) has been recently presented as an augmentative and alternative communication system for individuals with severe paralysis and communication problems [1]. This system is fully implanted and is specifically designed for autonomous home use. It makes use of two electrocorticography (ECoG) strips, one placed over the hand area of the sensorimotor cortex and the other placed over the dorsolateral pre-frontal cortex (DLPFC). The selection of these locations is based on previous work showing that neuroelectrical activity can be generated by patients with quadriplegia by attempting hand movements [2], and on work showing the feasibility of accurate BCI control by modulation of gamma-power of the DLPFC in epilepsy patients [3]. The activity of the sensorimotor area is used for home use by both users as it yields the most consistent BCI control [1]. By attempting to move the hand, the user generates a decodable brain signal and can control the UNP with the resulting 'brain click', which selects the currently highlighted item on the screen, analogous to a mouse click.

We have previously reported on the signal processing pipeline and presented the graphical user interface [4]. In short, the signal is processed by six consecutive filters that smooth and normalize the signal and convert it to a brain click. Simultaneously, the user can activate an 'escape sequence', which results in a specific screen being shown immediately, with the option to alert a caregiver [5]. The brain clicks and escape sequences are produced by attempting to move the hand with different durations: short and long respectively.

To avoid technology abandonment and to improve home-use, the system needs to be usable in an easy and intuitive manner. To this end software was developed within a quality system for medical software that provides the user with a personalized user menu, training applications, a spelling application, possibility to alert a caregiver with sound, and stand-by functions.

Here, we give an update on the development of the UNP software and describe the novel features of the system that have been employed in two users implanted with the UNP.

MATERIALS AND METHODS

The UNP was implanted in two locked-in individuals, one suffering from late-stage Amyotrophic Lateral Sclerosis (ALS), referred to as UNP1 here, and another individual who suffered a brainstem stroke, referred to as UNP4. Subdural ECoG strips with 4 electrodes each (Resume II®, Medtronic, 4 mm electrode diameter, 1 cm inter-electrode distance, off-label use) were placed over the sensorimotor hand region and over the DLPFC. Electrode strips are connected subcutaneously to an amplifier and transmitter device (Activa® PC+S, Medtronic, off-label use), subcutaneously placed under the left clavicula. This device amplifies and digitalizes the signals, both in the power and time domain, from pre-selected bipolar electrode pairs. Signals from sensorimotor and DLPFC strips can be used to control the UNP menu.

User requirements have led to the development of new features in the UNP system. In this paper we address two new features: the *Sleep Mode,* and the *Continuous Click Task* (CCT). The Sleep Mode was developed on the request by one of the UNP users. It allows the user to autonomously wake the system up during the night and alert the caregiver. In this way the user can get help when needed during the night. The CCT was developed to allow users to more efficiently practice generating brain clicks.

The Sleep Mode can be accessed from the menu by both the user and caretaker, by means of a brain click or tap on the screen, respectively. When the Sleep Mode is activated, the screen turns black and the

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-29

brightness is decreased to not disturb sleep. A grey lock symbol indicates the system is in standby. The caretaker can exit the Sleepmode by tapping the screen, which brings up the main menu. The user can wake the system by activating an escape sequence, which requires a longer duration of attempted movement than the regular escape setting used during the day. This prolonged duration is used to prevent unintended waking of the system due to spontaneous brain activity occurring during sleep. When the user awakes the system, an auditory notification is played, which informs the caretaker that assistance is required. This notification continues to play until the caretaker taps the screen, after which three options are displayed: to return to the Sleep Mode, go to the main menu, or start the spelling application (Tobii Dynavox Communicator 5, referred to as 'Typer' in the menu), which allows the user to spell words on the screen. Based on the judgment of the caretaker an option is selected, allowing the user to either return to sleep, give further instructions through the spelling application, or access the main menu for any other desired actions. As a safety measure, in case the caregiver forgets to click one of the three options, the user can still activate the night mode escape, which will again sound the auditory notification, until a caregiver taps the screen. This flow of activity is depicted in Figure 1.

The CCT is a modification of the previously reported click task [1] and was implemented with three options, aiming to make practicing more efficient using different approaches. In the 'basic' mode, the user can practice his or her timing of creating brain clicks. In the 'help' mode, the system can aid the user with creating clicks during initial training, and in the 'dynamic' mode, the system can optimize the parameters of the task in real time. In all modi the interface is the same, to provide the user with a consistent view.

Through this interface the user is shown a row of eight molehills, with a mole sitting on top of one. Aim of the task is to select the molehill with the mole on top. To this end a selection box highlights the molehills

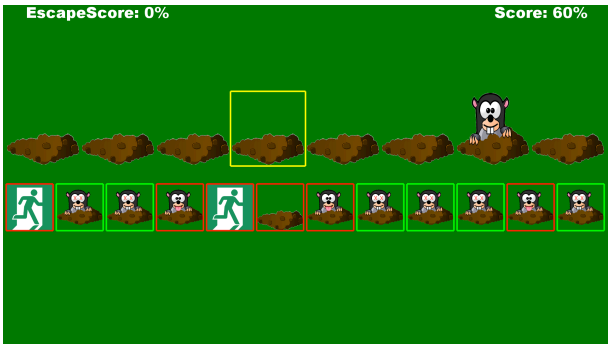


Figure 2: Interface of the CCT, showing the row of eight molehills. In this screenshot the mole sits on the seventh molehill and the selection box highlights the fourth molehill. Below this the feedback icons are shown. The scores for escape sequences and brain clicks are shown in the top corners.

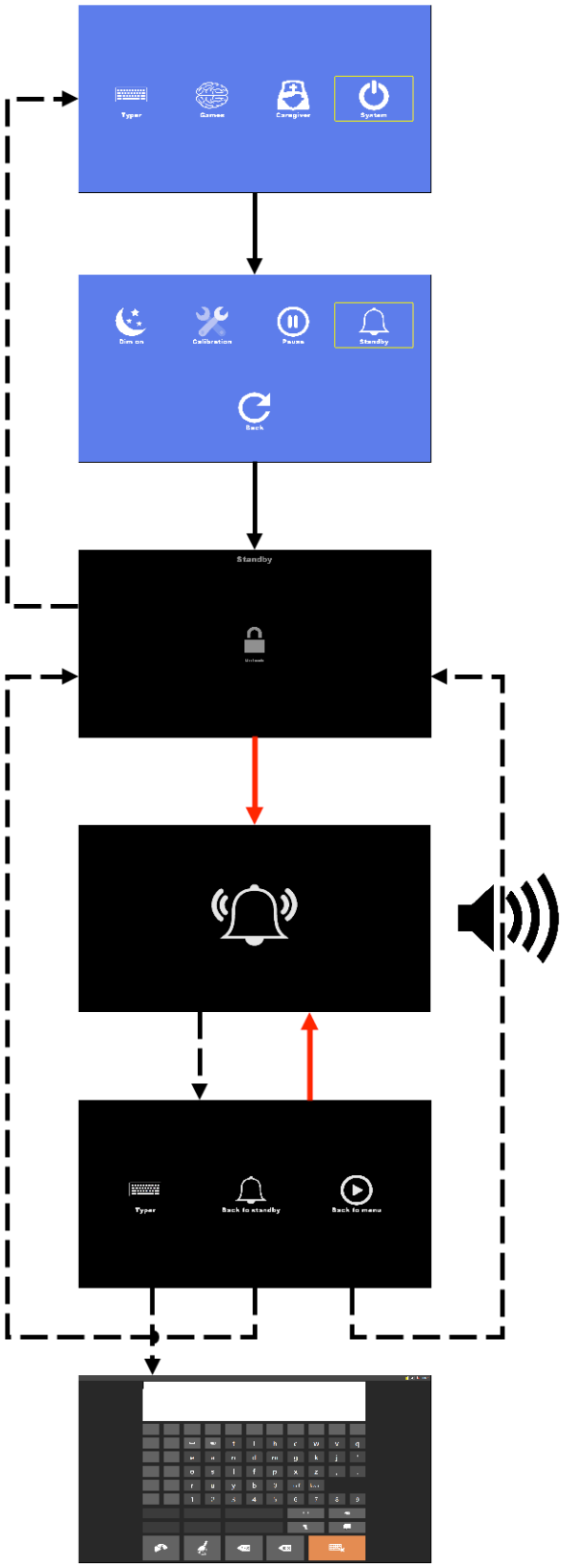


Figure 1: Schematic overview of the Sleep Mode. Solid black lines depict transitions between screens that both the user or caretaker can perform, red solid lines are transitions that only the user can perform, and dashed lines depict transitions only the caretaker can perform.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-29

sequentially, starting from the left and moving to the right. After the rightmost hill is highlighted, the selection box returns to the leftmost hill and the process repeats. The user can at any time create a brain click to select the currently highlighted molehill. If this molehill has the mole on top, one point is awarded and the accuracy is displayed at the top-right corner of the screen. Below the row of molehills a green-boxed icon (a mole with X's in his eyes) indicating a 'hit' (true positive) is added to a row of feedback icons. Conversely, if the highlighted molehill with the mole on top is not selected by the user, a 'miss' (false negative) is registered, adding a red-boxed icon of a mole sticking its tongue out to the feedback icons. To facilitate efficient practice, the mole changes position after a hit or miss. If a molehill is selected that does not have the mole on top, a red-boxed icon (an empty molehill), indicating a false positive, is added to the feedback icons. Escape sequences can also be trained using this task: in this case an icon of an escape sign is displayed above the row with molehills, cueing the user to produce an escape sequence. Based on whether an escape sequence is successfully produced, points are awarded and an escape sign is added to the feedback icons, with the colour of the border indicating whether the escape sequence was successfully created: green for successful and red for unsuccessful. These points are kept separately from the points awarded for hits with moles. Moles and escape cues can be presented within the same task run, or only moles or only escape cues can be presented. This process is depicted in Figure 2.

In the basic modus the CCT works as described above. In the help modus the CCT can provide additional help to the user: in case of a false positive or negative the system can instead assign the input as a true positive or true negative and give feedback accordingly. The amount of help can be varied: from adjusting a high percentage of false classes to a few, and eventually to no help. Rationale here is to gradually increase the control the user has, allowing the user to gradually adapt and train to create brain clicks, an approach that has been used successfully in other BCI systems [6]. In addition, this approach helps in keeping the user motivated during training, in order to further facilitate efficient training.

The dynamic mode allows for online optimisation of the parameters underlying user performance. These parameters pertain to either the feedback given to the user, such as the speed with which the molehills are highlighted (scan rate), or to the construction of a brain click, such as how long the movement should be attempted to create a brain click, or to the calibration of the signal, such as the z-scoring performed on the signal (cf. [4]). Based on the feedback (e.g. false positives or negatives) these parameters are continuously optimised online, converging to settings that allow the user to produce correct brain clicks and escape sequences as efficiently as possible. This modus can be used to make training more efficient by both reducing the time needed to optimise parameters required for optimal training, as well as allowing users to improve their performance by training with parameters that continuously alter towards increasingly effective performance. In addition, this modus allows us to monitor optimal parameter settings over time to see if (temporary) changes are useful, for example when the user experiences a period of time in which he or she is feeling more tired than usual, or is experiencing other temporary conditions that can alter the signal. To facilitate this, the CCT can be started by the user at home and the resulting optimal parameters are stored to be reviewed by the research team, allowing changes to be made to the standard parameters if necessary.

RESULTS

The Sleep Mode requires different settings with respect to the duration the control signal has to exceed the threshold, and with respect to the value of the threshold during sleep, compared to the wake situation. We are currently in the process of optimising the parameters to the characteristics of the brain signal during sleep to accommodate use of the Sleep Mode without spurious alerting of the caretaker.

The CCT has been used by both users for 73 runs in total, with UNP1 using an average run time of 80 seconds and UNP4 an average time of 150 seconds. Of these runs, the basic mode has been used by both users, for a total of 46 runs, in order to practice timing. The help mode has been used 11 times by UNP4 and is perceived as more fun. Based on this feedback, the CCT was used during specific sessions to make training more rewarding. The dynamic mode has been used 16 times in total, by both users. With UNP1 we used this mode with the aim of reducing the scan rate. This resulted in a decrease of the used scan rate from 2.4s to 1.8s over the course of 7 training runs. UNP1 now uses the resulting 1.8s scan rate for home use and reports being pleased with the task and the resulting optimised scan rate.

DISCUSSION & CONCLUSION

User input and specifications are crucial for the development of home use assistive technology. In the last two years we have designed and developed new features of the UNP system based on the request and feedback provided by the users implanted with this BCI. We have reported on two new features, which provide the user with control of the system during the night and improve the accuracy of BCI control. These features are especially important for users in a nearly complete locked-in state, likely providing the only reliable communication channel once no volitional residual movement is left.

Preliminary results are promising: users report finding the CCT pleasant, and based on the use of the dynamic mode of the CCT, one user now uses a faster scan rate for spelling, allowing for faster communication. For the Sleep Mode no results are yet available.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-29

The UNP system is continuously being updated to integrate user requests. There are a number of features that are expected to be implemented in the future, allowing, for example, faster menu navigation, and increased speed of spelling, both of which are features that users of BCI's have reported to be desirable [7].

ACKNOWLEDGEMENTS

This work has been funded by the ERC-Advanced 'iConnect' project (grant ERC-Adv 320708) and the Dutch Technology Foundation STW (grant UGT7685). We thank the UNP participants for their testing and feedback on the software and we acknowledge the work by Stavrina Devetzoglou-Toliou and Nikos Vardalakis in programming several of the features of the UNP system.

REFERENCES

[1] Vansteensel, M. J., Pels, E. G. M., Bleichner, M. G., Branco, M. P., Denison, T., Freudenburg, Z. V., Gosselaar, P., Leinders, S., Ottens, T., Van den Boom, M., Van Rijen, P., Aarnoutse, E., Ramsey, N. F. (2016). Fully Implanted Brain–Computer Interface in a Locked-In Patient with ALS. New England Journal of Medicine, 375(21), 2060–2066.

[2] Bouton, C. E., Shaikhouni, A., Annetta, N. V., Bockbrader, M. A., Friedenberg, D. A., Nielson, D. M., ... & Morgan, A. G. (2016). Restoring cortical control of functional movement in a human with quadriplegia. Nature, 533(7602), 247.

[3] Vansteensel, M. J., Hermes, D., Aarnoutse, E. J., Bleichner, M. G., Schalk, G., van Rijen, P. C., ... & Ramsey, N. F. (2010). Brain–computer interfacing based on cognitive control. Annals of neurology, 67(6), 809-816.

[4] M.A. van den Boom, M. Vermaas, E.J. Aarnoutse, S. Leinders, E. G. M. Pels, Z.V. Freudenburg, M.P. Branco, M.J. Vansteensel, N.F. Ramsey (2017). Utrecht neuroprosthesis: from brain signal to Independent control. Proceedings of the 7th Graz Brain-Computer Interface Conference 2017, DOI: 10.3217/978-3-85125-533-1-90.

[5] Leinders, S., Pels, E.G.M., Vansteensel, M.J., Branco, M.P., Freudenburg, Z.V., van den Boom, M.A., Vermaas, M., Aarnoutse, E.J., & Ramsey, N.F. Using A One-Dimensional Control Signal For Two Different Output Commands In An Implanted BCI. Proceedings of the 7th Graz Brain-Computer Interface Conference 2017, 50, 270-273. DOI: 10.3217/978-3-85125-533-1-50

[6] Wang, W., Collinger, J. L., Degenhart, A. D., Tyler-Kabara, E. C., Schwartz, A. B., Moran, D. W., ... & Kelly, J. W. (2013). An electrocorticographic brain interface in an individual with tetraplegia. PloS one, 8(2), e55344.

[7] Peters, B., Bieker, G., Heckman, S. M., Huggins, J. E., Wolf, C., Zeitlin, D., & Fried-Oken, M. (2015). Brain-computer interface users speak up: the Virtual Users' Forum at the 2013 International Brain-Computer Interface meeting. Archives of physical medicine and rehabilitation, 96(3), S33-S37.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-30

SYSTEMATIC REVIEW OF THE STATE-OF-THE-ART IN BRAIN-COMPUTER INTERFACE ROBOTIC WALKING IN STROKE

Alexander Heilinger[1], Máire Claffey[2], Olive Lennon[2]

1 g.tec medical engineering GmbH, Sierninstraße 14, 4521 Schiedlberg, Austria
2 University College of Dublin, School of Public Health, Physiotherapy and Sports Science, Belfield, Dublin 4, D14 YH57, Ireland

E-mail: heilinger@gtec.at

ABSTRACT: In stroke rehabilitation intelligent robots are required that can interpret walking intention accurately. This systematic review, guided by PRISMA and registered with PROSPERO (CRD42018112252), identifies current state-of-the-art in brain-computer interface (BCI) robotic walking in stroke using electroencephalogram (EEG). A detailed search strategy was applied across the following databases: PubMed (1949-2018), EMBASE (1947-2018), Web of Science (1945-2018), COMPENDEX (1967-2018), CINAHL (1982-2018), SPORTDiscus (1985-2018), ScienceDirect (1997-2018), Cochrane Library (1974-2018). Eligibility for inclusion was considered independently by two reviewers. From a total of 38895 publications, 2 studies comprising N=48 chronic stroke patients met inclusion criteria and rated moderate to strong in quality. Different exoskeleton devices were employed in studies (Ekso^{TM}; H2 exoskeleton). No study closed the BCI loop. Longitudinal training was noted to increase classification accuracy. Improved frontoparietal connectivity in the affected side was observed during robotic gait training. BCI robotic gait training after stroke is not reported in the literature to date but shows promise with adequate training for classification.

INTRODUCTION

Many patients with stroke experience a restriction in their mobility. Impairment of gait effects the functional ambulation capacity, balance, walking velocity, cadence, stride length and muscle activation pattern[1].
With conventional rehabilitation and gait training, 22% of all stroke patients do not regain the ability to walk[2]. Intensive gait training with or without body weight support (BWS), while associated with short term gains in walking speed and endurance does not increase the likelihood of walking independently [3]–[6]. Moreover, only patients with stroke who are able to walk benefit from such an intervention [3]–[6]. There is growing effort to increase the efficacy of gait rehabilitation for stroke patients using advanced technical devices. In the last decade, advances in robotic technologies, actuators & sensors, new materials, control algorithms and miniaturization of computers have led to the development of wearable lower-body exoskeleton robotic orthoses. There is also evidence that robotic-assisted gait training has an additional beneficial effect on functional ambulation outcomes[7], [8] and increases the likelihood of walking independently[9].
Brain-Computer Interface (BCI) technology is also a growing field in stroke rehabilitation with promising results as a therapeutic intervention in upper limb training. [10], [11]. Gains in BCI based gait training in stroke lag behind and there is a critical need for reliable BCIs that interpret user intent directly from brain signals for making context-based decisions with respect to stepping.
EEG recording in stroke, where the pathology is at brain level has been problematic when compared to other neurological pathologies such as spinal cord injury. Uncertainty exists in the literature on the best choice of EEG metric [12]–[14] and in inherent difficulties in identifying the precise role of the motor cortex during phases of the gait cycle due to difficulty capturing low artefact EEG in an ambulatory context and the proximity of both motor cortices to each other [15].
Our overarching goal in this review is to summarize existing studies employing BCI robotic walking in stroke using EEG. Thereby providing a current state-of-the art summary of this research field. Current BCI robotic walking devices, algorithms, signal processing methods and classification methods described in the literature are presented.

MATERIALS AND METHODS

Study identification and selection

A systematic literature search was conducted. As this review is related to engineering and physiotherapy / rehabilitation, an automated search in the following main databases available were undertaken to identify relevant publications: PubMed (1949-2018), EMBASE (1947-2018), Web of Science (1945-2018), COMPENDEX (1967-2018), CINAHL (1982-2018), SPORTDiscus (1985-2018), ScienceDirect (1997-2018), Cochrane Library (1974-2018). The search terms and Boolean

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-30

ID	Stroke patients	RGD	EMP	BLC
1	40	Ekso™	FPEC CSE SMI	N
2	8	H2	Neural Decoding	N

Table 1 | Table of studies. In the first row the three selected studies are shown. (1) Calabrò et al., (2) Contreras Vidal et al. 2018. RGD: Robotic gait device; EMP: EEG Measurement Protocol; BLC: BCI loop closure achieved; FPEC: Frontoparietal effective connectivity; CSE: Cortico-spinal excitability; SMI: Sensory-motor integration

operators employed are summarised in figure 1. Only publications in English were considered. The following types of study were included: Randomized control trials, cross-over or quasi randomized control studies, case-control studies, cohort studies, cross-sectional studies, case series and case reports. Reviews, opinion pieces, editorials and conference abstracts were excluded.

Data extraction

Two independent researchers reviewed the literature at title, abstract and paper stages and performed the data extraction and the results were compared afterwards. Disagreements were resolved by broad discussion and observance of the set criteria. Data extraction fields focused on the following topics: Fields relating to achievement of BCI, description of bio-signal capture, description of hardware/software devices employed, description of bio-signal recording, description of processing methodologies employed and description of bio-feedback mechanisms.

Quality assessment

Risk of bias was assessed using the effective public health practice project tool (EPHPP). Two reviewers independently rated the studies under headings of selection bias, study design, con-founders, blinding, data collection methods and withdrawal and drop-outs, a final rating of strong, moderate or weak quality.

RESULTS

The search was completed across the databases in November 2018. Figure 1 summarizes the studies identified from the search and those included at each stage of the review process. A total of 2 studies were identified following full manuscript review that met the inclusion criteria. These were rated by the *EPHPP quality assessment* as moderate to strong in quality.
Table 1 provides a summary of the key characteristics of these studies. Results are then synthesized narratively under headings of interest.

Feasibility of decoding walking from brain activity in stroke survivors

Contreras-Vidal et al. (2018)**[16]** investigated the feasibility of decoding walking from brain activity in stroke survivors during therapy by using a powered exoskeleton integrated with an EEG-based BCI in five chronic stroke patients. Classification accuracies of predicting joint angles improved with multiple training sessions and gait speed, suggesting improved neural representation for gait and the feasibility of designing an EEG-based BCI to monitor brain activity or control a rehabilitative robotic exoskeleton. EEG was recorded using a high-input impedance amplifier (referential input noise < 0.5μVrms @ 1÷20,000 Hz, referential input signal range 150-1000mVPP, input impedance >1GΩ, CMRR > 100 dB, 22bit ADC) of Brain Quick System PLUS (Micromed; Mogliano Veneto, Italy), wired to an EEG cap equipped with 21 Ag tin disk electrodes, positioned according to the international 10-20 system. The cortical activations induced by gait training were identified from EEG recordings by using Low-Resolution Brain Electromagnetic Tomography (LORETA). The brain compartment of the three-shell spherical head model used in LORETA was thereby restricted to the cortical gray matter using a resolution of 7 mm, thus obtaining 2394 voxels. The voxels were collapsed into 7 regions of interest (prefrontal, PF, supplementary motor, SMA, centroparietal, CP, and occipital, O, areas of both hemispheres) determined according to the brain model coded into Talairach space, by using MATLAB. Then, structural equation modeling technique (or path analysis) was employed to measure the effective connectivity (that assesses the causal influence that one brain area, i.e., electrode-group, exerts over another, under the assumption of a given mechanistic model) among the cortical activations induced by gait trainings.

This works builds from a previous publication (not included in this review) outlining a possible roadmap for EEG-based BCIs in lower body robotic exoskeletons using the NeuroRex System [Contreras-Vidal et al. (2013).

The second study identified for inclusion was an RCT by Calabrò et al.[17]. This study with 40 sub-acute and chronic stroke patients employed EEG to evaluate frontoparietal effective connectivity (FPEC) as a metric of neuroplasticity to identify if additional gains were made from Robotic gait training in addition to conventional rehabilitation and overground walking when compared with same duration therapy and overground training alone. The strengthening effect of robotic gait training on FPEC when compared with the control group ($r = 0.601$, $p < 0.001$) was observed. and were the most important factors that correlated with the clinical improvement following robotic gait training.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-30

EEG was recorded in this study using a high-input impedance amplifier (referential input noise < 0.5 μVrms @ 1÷20,000 Hz) of Brain Quick SystemPLUS (Micromed; Mogliano Veneto, Italy), wired to an EEG cap equipped with 21 Ag tin disk electrodes, positioned according to the international 10-20 system. An electrooculogram (EOG) was also recorded for blinking artefact detection. The recording occurred in the morning (about 11^{am}) and lasted at least 10 min, with the eyes open (fixing a point in front of the patient). The EEG end EOG were sampled at 512 Hz, filtered at 0.3-70 Hz, and referenced to linked earlobes. the cortical activations induced by gait training from the EEG recordings were identified by using Low-Resolution Brain Electromagnetic Tomography (LORETA; LORETA-KEY alpha-software)

DISCUSSION

This systematic review describes the current state of the art in robotic gait training with BCI integration in stroke. Following a scientifically rigorous systematic review identifying over 38,000 potentially relevant publications, only two studies met the criteria for inclusion by describing a clinical population of stroke, having a robotic gait training intervention and having registered EEG signals during gait training in the device. No study was identified that closed the BCI loop in robotic walking after stroke.

Many papers reviewed presented data from healthy norms during robotic gait and identified utility for populations such as stroke. However, these foundational works have not yet translated into the next phase of testing in the clinical population of interest, suggesting barriers to implementation. Inclusion of study participants with physical disability requires interdisciplinary expertise beyond engineering and barriers to EEG in clinical practice have been identified including the length of time needed to mount traditional AgCl EEG electrodes [12] and uncertainty regarding choice of EEG metric [12]–[14]. Advances in EEG hardware, recording electrodes, and software analysis methods may now overcome many of these barriers with dense--array systems with up to 256 electrodes showing improved spatial resolution [18] and newer computational methods for example, using partial least squares (PLS) regression has outperformed traditional EEG methods in capturing behavioral variation in stroke populations [19]. Renewed, interdisciplinary focus is now required in this field.

Of the two studies identified addressing robotic walking and EEG biosignals; data collection and processing were dealt with differently, limiting synthesis and future projections. Agreed standards in EEG capture and processing are required in this field to allow future meta-synthesis.

Contreras-Vidal et al. showed in their publication/s a foundation step towards EEG-based BCI controlled lower-limb rehabilitation using the NeuroRex system after stroke. They identified the feasibility of accurately decoding lower limb movements during robotic gait training after stroke, an important first step and presented a clinical neural interface roadmap for EEG-based BCIs in lower body robotic exoskeletons, identifying requirements that include: a reliable BCI system with an option of shared control between the user and the robotic device; appropriate risk assessment and a better understanding of the neural representation for the action and perception of bipedal locomotion at cortical level.fMRI-EEG.

Calabrò et al. showed utility of EEG as an outcome measure for neuroplasticity, identifying a strengthening in the frontoparietal effective connectivity and demonstrating promise for robotic gait training as an intervention after stroke to driving positive brain plasticity over and above usual physiotherapy and overground gait training.

CONCLUSION

The small number of studies available in robotic gait in stroke and EEG limit conclusions that can be drawn about BCI integration in the future. A lack of standardization around EEG data capture and processing was evident that limited data synthesis. However, the current review revealed encouraging preliminary results in the road to BCI integration in robotic gait training after stroke.

REFERENCES

[1] A. R. Den Otter, A. C. H. Geurts, Th. Mulder, and J. Duysens, “Abnormalities in the temporal patterning of lower extremity muscle activity in hemiparetic gait,” *Gait & Posture*, vol. 25, no. 3, pp. 342–352, Mar. 2007.

[2] “Irish Heart Foundation, in association with the Department of Health and ChildrenNational Audit of Stroke Care,” *Irish Heart Foundation, Dublin, Ireland (April 2008)*.

[3] J. Mehrholz, M. Pohl, and B. Elsner, “Treadmill training and body weight support for walking after stroke,” in *Cochrane Database of Systematic Reviews*, The Cochrane Collaboration, Ed. Chichester, UK: John Wiley & Sons, Ltd, 2014.

[4] R. A. States, Y. Salem, and E. Pappas, “Overground Gait Training for Individuals with Chronic Stroke: A Cochrane Systematic Review:,” *Journal of Neurologic Physical Therapy*, vol. 33, no. 4, pp. 179–186, Dec. 2009.

[5] J.-M. Belda-Lois *et al.*, “Rehabilitation of gait after stroke: a review towards a top-down approach,” *Journal of neuroengineering and rehabilitation*, vol. 8, no. 1, p. 66, 2011.

[6] M. Franceschini, S. Carda, M. Agosti, R. Antenucci, D. Malgrati, and C. Cisari, “Walking

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-30

After Stroke: What Does Treadmill Training With Body Weight Support Add to Overground Gait Training in Patients Early After Stroke?: A Single-Blind, Randomized, Controlled Trial," *Stroke*, vol. 40, no. 9, pp. 3079–3085, Sep. 2009.

[7] C. Tefertiller, B. Pharo, N. Evans, and P. Winchester, "Efficacy of rehabilitation robotics for walking training in neurological disorders: A review," *The Journal of Rehabilitation Research and Development*, vol. 48, no. 4, p. 387, 2011.

[8] I. Schwartz and Z. Meiner, "Robotic-Assisted Gait Training in Neurological Patients: Who May Benefit?," *Annals of Biomedical Engineering*, vol. 43, no. 5, pp. 1260–1269, May 2015.

[9] J. Mehrholz, C. Werner, J. Kugler, and M. Pohl, "Electromechanical-assisted training for walking after stroke.," *Cochrane Database of Systematic Reviews*, vol. 4, 2007.

[10] D. Irimia *et al.*, "recoveriX: A new BCI-based technology for persons with stroke," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 1504–1507.

[11] W. Cho *et al.*, "Paired Associative Stimulation Using Brain-Computer Interfaces for Stroke Rehabilitation: A Pilot Study," *Eur J Transl Myol*, vol. 26, no. 3, Jun. 2016.

[12] S. P. Finnigan, M. Walsh, S. E. Rose, and J. B. Chalk, "Quantitative EEG indices of sub-acute ischaemic stroke correlate with clinical outcomes," *Clinical Neurophysiology*, vol. 118, no. 11, pp. 2525–2532, Nov. 2007.

[13] R. V. A. Sheorajpanday, G. Nagels, A. J. T. M. Weeren, M. J. A. M. van Putten, and P. P. De Deyn, "Reproducibility and clinical relevance of quantitative EEG parameters in cerebral ischemia: A basic approach," *Clinical Neurophysiology*, vol. 120, no. 5, pp. 845–855, May 2009.

[14] J. Jin, B. Z. Allison, X. Wang, and C. Neuper, "A combined brain–computer interface based on P300 potentials and motion-onset visual evoked potentials," *Journal of neuroscience methods*, vol. 205, no. 2, pp. 265–276, 2012.

[15] T. Castermans, M. Duvinage, G. Cheron, and T. Dutoit, "Towards Effective Non-Invasive Brain-Computer Interfaces Dedicated to Gait Rehabilitation Systems," *Brain Sciences*, vol. 4, no. 1, pp. 1–48, Dec. 2013.

[16] J. L. Contreras-Vidal *et al.*, "Neural Decoding of Robot-Assisted Gait During Rehabilitation After Stroke:," *American Journal of Physical Medicine & Rehabilitation*, vol. 97, no. 8, pp. 541–550, Aug. 2018.

[17] R. S. Calabrò *et al.*, "Shaping neuroplasticity by using powered exoskeletons in patients with stroke: a randomized clinical trial," *Journal of NeuroEngineering and Rehabilitation*, vol. 15, no. 1, Dec. 2018.

[18] Y. Petrov, J. Nador, C. Hughes, S. Tran, O. Yavuzcetin, and S. Sridhar, "Ultra-dense EEG sampling results in two-fold increase of functional brain information," *NeuroImage*, vol. 90, pp. 140–145, Apr. 2014.

[19] J. Wu, R. Srinivasan, E. Burke Quinlan, A. Solodkin, S. L. Small, and S. C. Cramer, "Utility of EEG measures of brain function in patients with acute stroke," *Journal of Neurophysiology*, vol. 115, no. 5, pp. 2399–2405, May 2016.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-30

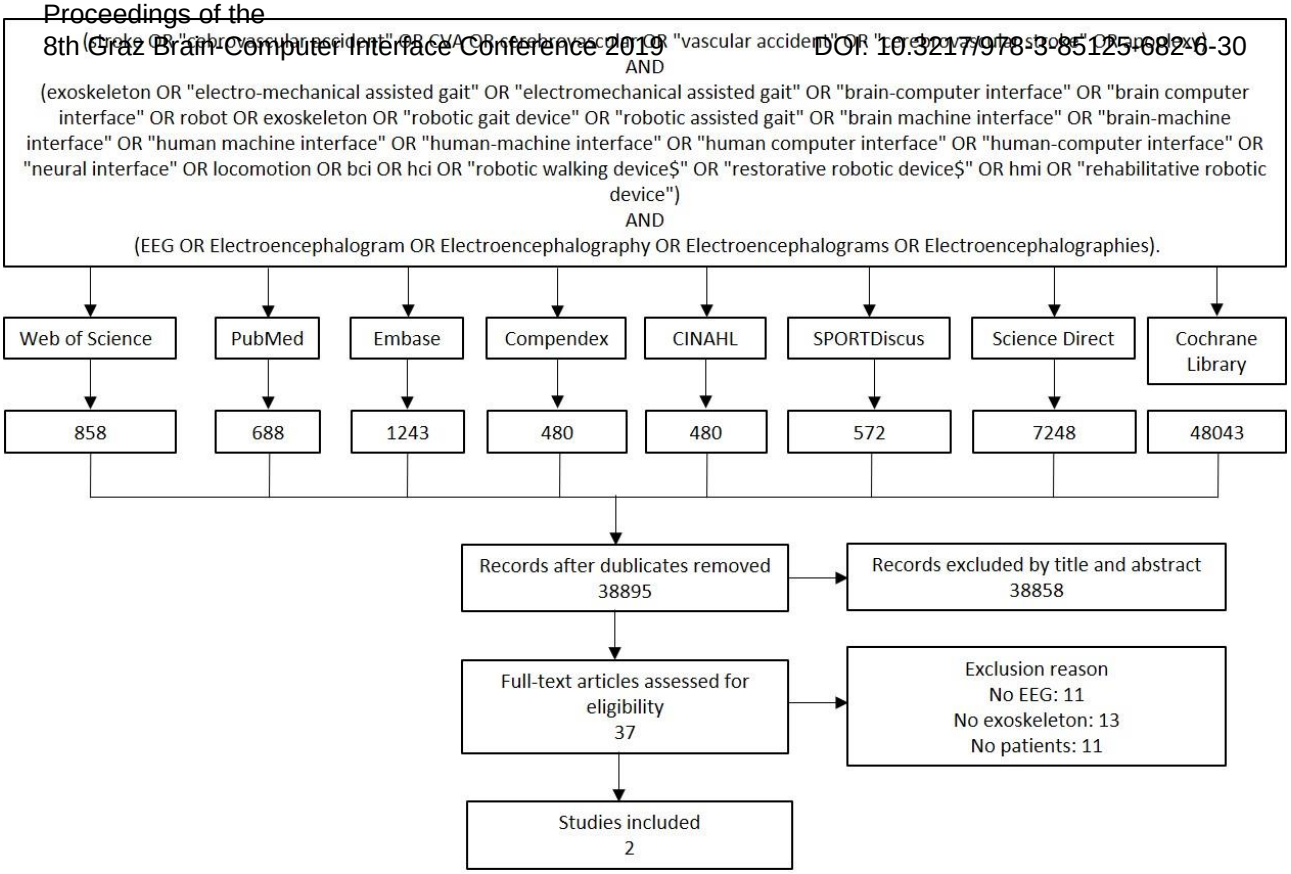


Figure 1 | Scheme of systematic review search process.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-31

A SIMULATED ENVIRONMENT FOR STUDYING PARTIAL OBSERVABILITY IN NOVEL ADAPTIVE DEEP BRAIN STIMULATION

Sebastián Castaño-Candamil[1], Mara Vaihinger[1], Michael Tangermann[1,2]

[1]Brain State Decoding Lab, Department of Computer Science,
BrainLinks-BrainTools, University of Freiburg, Germany
[2]Autonomous Intelligent Systems, Department of Computer Science,
University of Freiburg, Germany

E-mail: sebastian.castano|michael.tangermann@blbt.uni-freiburg.de

ABSTRACT: Adaptive deep brain stimulation (aDBS) can profit from data-driven approaches developed for BCIs. These aDBS systems improve upon the constant DBS in terms of efficiency and side effects amelioration by taking the ongoing brain state into consideration.

The environment controlled by aDBS is governed by partial observability, rendering classic control strategies suboptimal. In this regard, development of novel approaches is critical for improved aDBS therapy. However, early stage aDBS development is a difficult endeavor, given the lack of suitable development platforms.

In our contribution, we present a simulated environment that allows to modularly embed different surrogates of key challenges found in the aDBS problem. Specifically, we will focus on partial observability stemming from non-stationary dynamics and noisy state representations. Our simulations are used to analyze representative reinforcement learning approaches regarding their ability to cope with the partial observability.

To allow reproducibility and encourage adoption of our approach, the source code of our experiments is made available online.

INTRODUCTION

Deep brain stimulation (DBS) has been established as standard clinical treatment for movement disorders, such as Parkinson's disease (PD) and essential tremor (ET) [1, 2]. In addition, it is investigated to provide symptom relief in several neuropsychiatric diseases such as obsessive compulsive disorder (OCD) and major depressive disorder (MDP) [3, 4].

The stimulation characteristics and thus the efficiency of DBS treatment can be shaped by a number of parameters, e.g., electrode contacts used to deliver the electric stimulation pulses, the shape, width and amplitude of pulses and the frequency of these pulses. In a standard clinical setting, DBS parameters are determined by a highly trained clinician and are kept constant until the next consultation. This manual adaptation, performed a few times per year, will ideally account for initial post-surgical transient effects and long-term variations caused by disease progress and DBS-induced plasticity changes. However, such a constant DBS (cDBS) strategy can not cope with changes occurring on much shorter timescales. As a result, patients undergoing cDBS therapy are prone to acute and chronic motor- and neuropsychiatric side-effects, such as speech disorders, dysarthria, depression, emotional disinhibition, and paresthesias [5–7].

Closed loop strategies for DBS: Fortunately, closed-loop adaptive DBS (aDBS) provides a promising approach for tackling the shortcomings of cDBS strategies [8, 9]. Closed-loop aDBS systems provide stimulation as a function of symptoms and DBS-induced side-effects surrogates, extracted directly from brain signals, and termed neural markers (NMs) [10, 11]. Such NMs, however, are highly contaminated by background activity and are co-modulated by multiple brain processes; thus, providing only a partial representation of the real neural state of the patient [9]. Furthermore, despite improving upon cDBS, aDBS systems usually implement control strategies—such as threshold-based and proportional control—that neglect time dynamics.

Alternatively, more complex strategies have attempted to use latent neural dynamics as a potential source of information for improving aDBS efficacy [12–15]. While such dynamics can originate from inherent temporal brain activity, others can be explicitly associated with external factors such as medication intake, activity of daily living (ADL), and circadian rhythm. Another major source of non-stationary dynamics is the so-called DBS washout effect, describing the persistent clinical effect of DBS after stimulation withdrawal [16].

Data-driven approaches for dynamics-aware aDBS: Many of the diseases treated with DBS, as PD and MDP, are characterized by a remarkably heterogeneous phenotype [17, 18], where group studies are unable to deliver NMs and stimulation strategies that are universally suitable. In contrast, data-driven optimization of dynamics-aware control strategies offers a promising approach for obtaining effective and efficient aDBS systems. In this regard, Kumar and colleagues [19] presented a proof-of-concept *in-vitro* study using a tabular reinforcement learning (RL) strategy to control a neural network. Such

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-31

classic RL strategies assume that the underlying controlled system (*environment*, in RL literature) is *Markovian* and *fully observable*, i.e., predictions of future states of the environment depend solely on the current observation and this observation should offer a full representation of the environment's state. Given that in aDBS both assumptions are not fulfilled, classic RL strategies might deliver a sub-optimal control strategy (*policy*).

Strictly speaking, the non-stationary dynamics seen in aDBS are a consequence of partial observability: if information about non-stationary sources is included in the state representation (e.g., medication intake or ADL context), then the environment could be considered stationary. However, such information is usually not available. Consequently, we use the term *partial observability* to cover non-stationary dynamics and noisy NMs, hereafter.

Development platforms for aDBS: Working directly with patients is an expensive and strongly constrained endeavor (safety regulations), and *in-vitro* development protocols are relatively expensive and may suffer from oversimplifying assumptions regarding the structure of the underlying neural network. For these reasons, *in-silico* frameworks have been widely utilized in early aDBS development stages [20–22].

With our current contribution we introduce a novel *in-silico* approach. We adopt a modified version of an environment used in the standard RL testbench *openAI Gym*[23] and show, how partial observability properties of aDBS can be explicitly embedded into a RL task, thus making partial observability a benchmarkable challenge. Finally, we provide a comparison of state-of-the-art RL algorithms that deliver a more efficient control strategy than classic aDBS approaches under this partially observable environment.

METHODS

The core concept of RL is to learn how to control an environment alone from interactions with it. During the learning phase, the decision making *agent* continuously improves its control *policy* based on the *reward* it gains by interacting with the—potentially unknown—*environment*. In the aDBS context, the mapping between the RL components can be defined as: agent↔DBS controller, environment↔neural system, policy↔stimulation strategy, action↔apply a (parameterized) stimulation, and reward↔symptom suppression / side effects. This mapping can be formalized by defining an aDBS system as a Markov decision process, as follows.

Closed-loop aDBS as a Markov decision process: A closed-loop aDBS system can be defined as a Markov decision process $\mathscr{M} = \langle \mathscr{S}, \mathscr{A}, T, R \rangle$, where the space $\mathscr{S}$ is formed by all possible motor states, $\mathscr{A}$ is the set of possible stimulation parameters, $T : (s_k, s_{k+1}, a) \mapsto p(s_{k+1}|s_k, a_k) \in [0,1]$ is the probability distribution over brain state transitions, such that applying the stimulation parameter $a_k \in \mathscr{A}$ in the brain state $s_k \in \mathscr{S}$ at time point k leads to a new brain state $s_{k+1} \in \mathscr{S}$ at time $k+1$, and $R : (s_k, s_{k+1}) \mapsto R(s_k, s_{k+1}) \in [a,b]$ is the reward function (bounded by $\{a,b\} \in \mathbb{R}$) obtained by transitioning from s_k to s_{k+1}. In aDBS, a reward may express, e.g., the amelioration of symptoms or the suppression of DBS-induced side effects.

Partial observability in aDBS:

Noise regimes in NMs: Brain imaging techniques, such as electrocorticographic recordings or local field potential recordings from deep brain electrodes, are contaminated by measurement noise and background activity and are co-modulated by several, possibly independent, neural processes. In PD, the beta-band power of local field potentials recorded from the subthalamic nucleus (STN) is a widely used NM. However, it is not only determined by the symptom state, but is also modulated by motor preparation and execution[24, 25]—similar to cortical beta band power—. These characteristics render many extracted NMs highly noisy and thus they contribute to partial observability found in aDBS.

Washout-induced non-stationary dynamics: The washout effect observed in DBS contributes to partial observability by generating non-stationary dynamics on multiple timescales. For example in PD, DBS washout effects w.r.t. axial symptoms that influence gait or speech, can span from minutes to several hours, whereas washout w.r.t. rigor, tremor, and bradykinesia, typically lasts seconds only. In the Markov decision process defined above, a washout phase amounts to state transition distribution T that not only depends on the current stimulation, but also on the history thereof.

As some classes of control algorithms have not been designed to cope well with partial observability, it is important to investigate its effects upon the effectiveness and efficiency of aDBS control strategies. However, in *in-vivo* or *in-vitro* scenarios, it is difficult to analyze the specific impact of such dynamics individually. A surrogate environment provides the possibility to model those aDBS-specific challenges explicitly.

The flappingBird environment: We adapted the *FlappyBird* environment provided in the openAI gym platform to incorporate challenges of aDBS. We term the adapted environment the *continuous FlappyBird* (CFB). An agent's goal in this environment is to fly through horizontal tunnels that constantly pass by, as a gravity force pulls the agent downwards.Two main criteria were considered for selecting CFB as our surrogate environment: First, dimensionality of the state and action space is similar to the aDBS problem [26], and second, the model's engine provides a computationally inexpensive way of modifying the environment dynamics.

State representation: The state of the CFB environment is given by a 7-dimensional signal. It comprises 1) agent vertical position, 2) agent vertical velocity 3) and 4) bottom and top vertical position of the current tunnel, 5) agent's distance to the next tunnel and 6) and 7) bottom and top position of the next tunnel. The dimensionality of the CFB state space is similar to that of simple aDBS

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-31

setups: $\mathscr{S}$ is usually represented by a small number of power features extracted from local field potential signals [27–29]. Assuming that each of the bilaterally implanted DBS electrodes has 4 contacts, $\mathscr{S}$ results in an 8-dimensional representation.

Actions: The binary action space comprises two actions: vertical thrust and no vertical thrust. In a simplified approach to aDBS, stimulus amplitude parameter can also be defined as binary, i.e., DBS-on/off. Note that in more advanced setups, the action space might be continuous (if a continuous amplitude control is desired) or multidimensional (if other DBS parameters like stimulation frequency or stimulating contacts are considered).

Reward signal: Designing a good reward signal is difficult and problem specific. In aDBS, it requires a trade-off between at least the amelioration of PD-related symptoms and stimulation-induced side effects. In CFB, we define the reward signal as a function of the proximity with the center of a tunnel, for each time point k:

$$R(s_k, s_{k+1}) = \begin{cases} 0.1 & \text{if in } s_{k+1} \text{ agent is inside tunnel} \\ -0.4 & \text{if in } s_{k+1} \text{ agent outside tunnel} \\ -0.9 & \text{if in } s_{k+1} \text{ top or bottom is hit} \end{cases}$$

Partial observability in the CFB environment:

Noise in the state representation: Noisy state measurements are embedded in the environment by adding zero-mean Gaussian noise to each feature describing the state. The standard deviation of each noise source is defined individually per feature as $\sigma_f = \xi \cdot (l_f^u - l_f^d)$, where ξ denotes the noise level and the interval $[l_f^d, l_f^u]$ define domain of feature f. This modified version of CFB is called **CFB-Nξ** in the following.

History dependent action effect: The washout effect is simulated as a sustained aftereffect of each thrust action, and termed CFB-H. It is implemented using an action history which considers (at maximum) the last 100 thrusts. Specifically, the decaying thrust T_k^d at a time point k after a sequence of N_{thrust} thrust actions in the last 100 time steps is defined as: $T_k^d = T_{k-1}^d - \frac{T}{N_{thrust}/3}$, where T represents the thrust generated by a single *thrust* action. The constants selected here for the decaying thrust function are based on studies reporting ratios between accumulated stimulation N_{thrust} and washout duration in a range of 8:1 to 2:1 [30, 31]. The horizon is limited to a history of 100 time steps to ensure at least a limited level of controllability in the CFB-H environment.

EXPERIMENTAL SETUP

Choice of RL algorithms: Three model-free RL strategies were chosen as base algorithms, from the three main method families in RL: Value-function based, policy gradient, and actor critic. The selection of specific methods involved the following criteria: First, we considered the reported performance across multiple RL tasks in the OpenAI benchmark[1]. Second, we took into account the scientific impact of each algorithm within the RL community, as measured by the number of citations of the corresponding papers, their publication date, and number of appearances in review studies. The resulting collection of representative RL base algorithms, each using a feed-forward (FF) neural network, comprises: 1) Deep Q-Learning (DQN) with experience replay and a target network [32], an off-policy value function based method, 2) Advantage Actor-Critic (A2C), a 1-step advantage actor-critic method [33], and 3) Proximal Policy Optimization (PPO), a 1-step advantage actor-critic method focusing on an improved policy gradient estimate [34].

For comparison, we have also included a simple reactive agent, designed to resemble a threshold-based control strategy. It applies thrust whenever the agent finds itself below the center of a tunnel and stops when it is above the tunnel.

RL approaches to non-stationary MDP: By their design, the RL base algorithms DQN+FF, A2C+FF and PPO+FF can not be expected to deal well with partial observability of non-stationary origin. However, they can be equipped with the ability to consider a (potentially infinite) state (or state-action) history as proposed by [35–38]. For this reasons we have extended the DQN, A2C, and PPO models by recurrent networks implemented using either gated recurrent units (GRUs) or long short term memory (LSTM) units [39].

As a result, we could benchmark the following nine RL approaches: DQN+FF, DQN+LSTM, DQN+GRU, A2C+FF, A2C+LSTM, A2C+GRU, PPO+FF, PPO+LSTM, and PPO+GRU.

Benchmark design:

Model architectures and hyperparameter optimization: For all models, hyperparameters were optimized using the sequential model-based configuration framework introduced in [40]. The source code including all parameter details used to optimize and train our agents and the resulting architectures is provided online[2].

Training stage: A population of twenty RL agents per method was trained, each trained in an individual instance of the same RL task, but initialized with different random seeds. All agents were trained over two million interactions in the environments CFB-N and CFB-H. An Adam optimizer [41] updated a model's parameters every 64 interactions. The performance of a learning agent during training is evaluated in an independent test environment every 300 parameter updates. The performance reported corresponds to the average reward obtained by each population of agents.

Testing stage: After training, each agent population is tested in twenty randomly initialized environments during 50k interactions. The reported performance corresponds to the average reward for each agent population.

RESULTS

Performance in CFB-H environments:

[1] https://github.com/openai/baselines-results

[2] https://github.com/mVaihinger/RLAgentsFlappyBird

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-31

The training performance of the different agents in the CFB-H environments are depicted in Figure 1. Only the A2C agents consistently achieved a better performance than the reactive agents. PPO-FF achieved also a similar performance, however, it was less stable throughout training. All DQN-based agents had a much slower convergence rate and stayed considerably below the performance of the simple reactive agents.

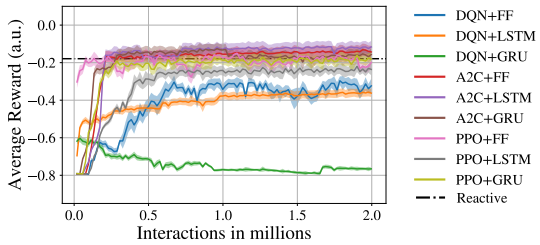


Figure 1: Time courses of training performance over two million steps in the CFB-H task.

Consistent with the training stage, Figure 2 shows that DQN agents yielded the worst performance in the test stage, whereas agents based on A2C and PPO better than the reactive agents on average. Overall, agents based on LSTM units achieved the best performance.

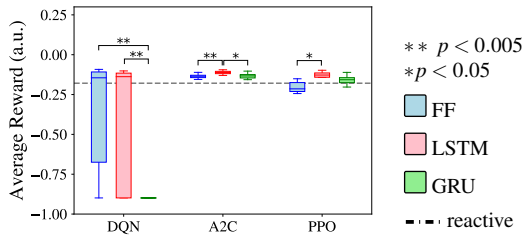


Figure 2: Boxplot of RL agent's performance in the unseen CFB-H test environments. Statistical significance was tested with the Wilcoxon ranksum using Bonferroni correction.

Performance in CFB-N environments:
Figure 3 shows the training performance of all agents in CFB-N environments. For the RL-methods, the main difference elicited by varying noise levels is the final training performance, while the convergence rate was rather unaffected by noise. Among all, A2C and PPO agents showed the greatest sample efficiency, as their performance improved the fastest in early stages of training, with A2C showing the most stable performance throughout training. While A2C-LSTM showed the best performance, even under strong noise regimes ($\xi < 0.5$), it is interesting to see, that A2C+FF could still cope quite well with the noise, while GRU-based models yielded the worst performance. These observations also hold during test stage, shown in Figure 4.

DISCUSSION

We have introduced a simulated environment to support early development stages of data-driven aDBS strategies based on RL. Specifically, the environment is designed to study approaches coping with partial observability properties. In addition, we have benchmarked representative RL methods, that have been modified to cope with the challenges faced by an aDBS system.

Suitability of considered methods for partially observable environments: The comparative analysis presented was motivated by the hypothesis that models aimed at capturing long term dependencies yield a higher end performance in partial observable environments, compared to classic approaches. We have shown that using LSTM- and GRU-based models does not consistently improve end performance compared to agents based on classic FF networks; however, LSTM-based models enable higher sample efficiency, as proved by the faster convergence of such models in early stages of training. We have also observed that the major performance difference is caused by the RL method chosen, and not by the type of recurrent units used in them.

CFB as a development environment of RL-aDBS: Although it is not possible to establish a one to one correspondence between a real aDBS environment and the CFB environment used, our framework allows to study specific characteristics of aDBS in early stages of control algorithm development, when physiological and functional constraints and clinical interpretability are not critical. A key feature of our contribution is the flexibility to explicitly embed key challenges found in aDBS in a modular fashion. In the present contribution, we have studied precise non-stationary dynamics caused by DBS washout, as well as noisy state representations. However, our framework can easily include other major sources of partial observability such as circadian rhythm variations, medication induced changes, among others. The only prerequisite is an appropriate scaling of time constants, as we have exemplified with the CFN-H environment.
In conclusion, our framework provides a cost efficient platform for early stage development of novel aDBS strategies before accessing more complicated setups, as physiologically-motivated simulations and, as a final goal, patients.

ACKNOWLEDGMENTS

This work was supported by BrainLinks-BrainTools Cluster of Excellence funded by the German Research Foundation (DFG, EXC1086), by the Federal Ministry of Education and Research (BMBF, 16SV8012), and by the state of Baden-Württemberg, and the DFG through grants bwHPC and INST 39/963-1 FUGG.

REFERENCES

[1] Baizabal-Carvallo J, Kagnoff M, Jimenez-Shahad J, Fekete R, Jankovic J. The safety and efficacy of thalamic deep brain stimulation in essential tremor: 10 years and beyond. J Neurol Neurosurg Psychiatry. 2014;85(5):567–72.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-31

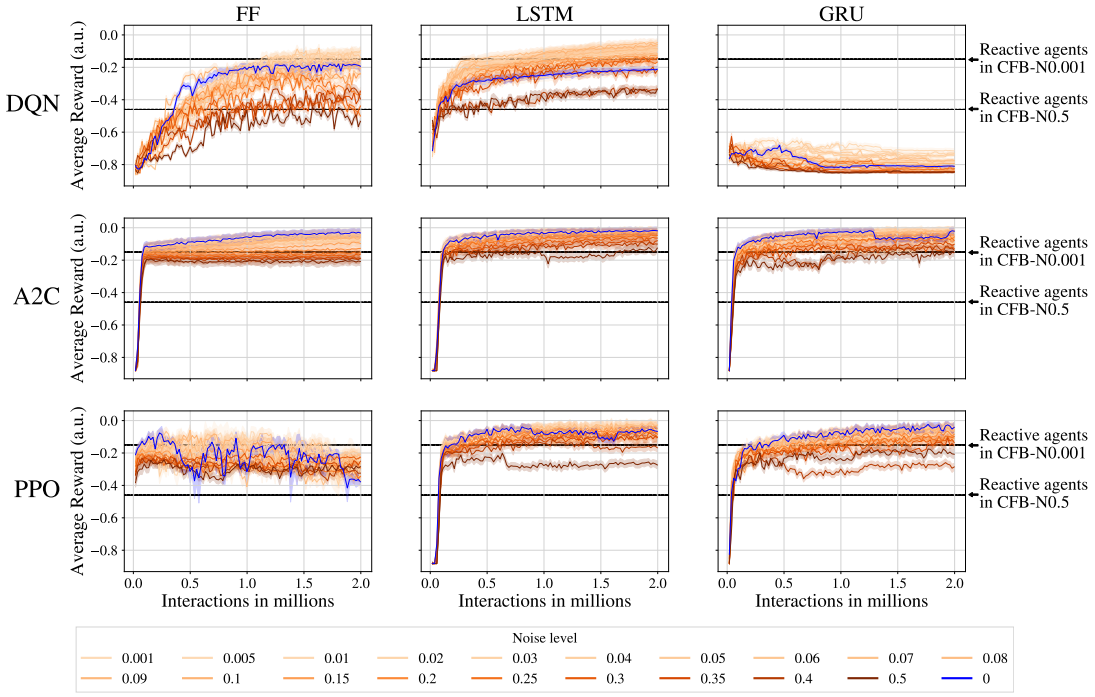


Figure 3: Time courses of the training performance over two million steps in the CFB-N task for several noise levels.

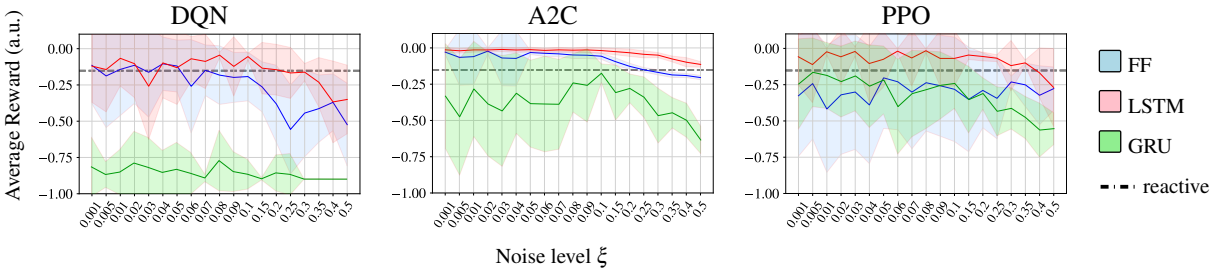


Figure 4: RL agent's performance in the unseen CFB-N test scenarios for several noise levels.

[2] Rodriguez-Oroz MC, Obeso JA, Lang AE, *et al.* Bilateral deep brain stimulation in Parkinson's disease: A multicentre study with 4 years follow-up. Brain. 2005;128(10):2240–49.

[3] Widge AS, Malone Jr DA, Dougherty DD. Closing the loop on deep brain stimulation for treatment-resistant depression. Frontiers in neuroscience. 2018;12:175.

[4] Coenen VA, Schlaepfer TE, Goll P, *et al.* The medial forebrain bundle as a target for deep brain stimulation for obsessive-compulsive disorder. CNS spectrums. 2017;22(3):282–289.

[5] Witt K, Daniels C, Volkmann J. Factors associated with neuropsychiatric side effects after STN-DBS in Parkinson's disease. Parkinsonism & related disorders. 2012;18:S168–S170.

[6] Little S, Tripoliti E, Beudel M, *et al.* Adaptive deep brain stimulation for Parkinson's disease demonstrates reduced speech side effects compared to conventional stimulation in the acute setting. Journal of Neurology, Neurosurgery & Psychiatry. 2016;87(12):1388–1389.

[7] Castrioto A, Lhommée E, Moro E, Krack P. Mood and behavioural effects of subthalamic stimulation in Parkinson's disease. The Lancet Neurology. 2014;13(3):287 –305.

[8] Kühn AA, Volkmann J. Innovations in deep brain stimulation methodology: Innovations in DBS Methodology. Movement Disorders. 2017;32(1):11–19.

[9] Little S, Brown P. What brain signals are suitable for feedback control of deep brain stimulation in Parkinson's disease?: Brain signals for control of DBS in PD. Annals of the New York Academy of Sciences. 2012;1265(1):9–24.

[10] Beudel M, Brown P. Adaptive deep brain stimulation in parkinson's disease. Parkinsonism & related disorders. 2016;22:S123–S126.

[11] Kuo C-H, White-Dzuro GA, Ko AL. Approaches to closed-loop deep brain stimulation for movement disorders. Neurosurgical focus. 2018;45(2):E2.

[12] Haddock A, Velisar A, Herron J, Bronte-Stewart H, Chizeck HJ. Model predictive control of deep brain stimulation for Parkinsonian tremor. In: Neural Engineering (NER), 2017 8th International IEEE/EMBS Conference on. 2017, 358–362.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-31

[13] Adamchic I, Hauptmann C, Barnikol UB, *et al.* Coordinated reset neuromodulation for Parkinson's disease: Proof-of-concept study. Movement disorders. 2014;29(13):1679–1684.
[14] Wang J, Nebeck S, Muralidharan A, Johnson MD, Vitek JL, Baker KB. Coordinated reset deep brain stimulation of subthalamic nucleus produces long-lasting, dose-dependent motor improvements in the 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine non-human primate model of Parkinsonism. Brain stimulation. 2016;9(4):609–617.
[15] Cagnan H, Pedrosa D, Little S, *et al.* Stimulating at the right time: Phase-specific deep brain stimulation. Brain. 2016;140(1):132–145.
[16] Cooper SE, McIntyre CC, Fernandez HH, Vitek JL. Association of deep brain stimulation washout effects with Parkinson disease duration. JAMA neurology. 2013;70(1):95–99.
[17] Thenganatt MA, Jankovic J. Parkinson disease subtypes. JAMA neurology. 2014;71(4):499–504.
[18] Widge AS, Ellard KK, Paulk AC, *et al.* Treating refractory mental illness with closed-loop brain stimulation: Progress towards a patient-specific transdiagnostic approach. Experimental neurology. 2017;287:461–472.
[19] Kumar SS, Wülfing J, Okujeni S, Boedecker J, Riedmiller M, Egert U. Autonomous optimization of targeted stimulation of neuronal networks. PLoS computational biology. 2016;12(8):e1005054.
[20] Popovych OV, Lysyansky B, Tass PA. Closed-loop deep brain stimulation by pulsatile delayed feedback with increased gap between pulse phases. Scientific reports. 2017;7(1):1033.
[21] Karamintziou SD, Custódio AL, Piallat B, *et al.* Algorithmic design of a noise-resistant and efficient closed-loop deep brain stimulation system: A computational approach. PloS one. 2017;12(2):e0171458.
[22] Karamintziou SD, Deligiannis NG, Piallat B, *et al.* Dominant efficiency of nonregular patterns of subthalamic nucleus deep brain stimulation for Parkinson's disease and obsessive-compulsive disorder in a data-driven computational model. Journal of neural engineering. 2015;13(1):016013.
[23] Brockman G, Cheung V, Pettersson L, *et al.* Openai gym. arXiv preprint arXiv:1606.01540. 2016.
[24] Kühn AA, Williams D, Kupsch A, *et al.* Event-related beta desynchronization in human subthalamic nucleus correlates with motor performance. Brain. 2004;127(4):735–746.
[25] Engel AK, Fries P. Beta-band oscillations—signalling the status quo? Current opinion in neurobiology. 2010;20(2):156–165.
[26] Castaño-Candamil S, Vaihinger M, Tangermann M. A Simulated Environment for Early Development Stages of Reinforcement Learning Algorithms for Closed-Loop Deep Brain Stimulation (under review). In: Engineering in Medicine and Biology Society (EMBC), 2019 41st Annual International Conference of the IEEE. 2019.
[27] Kühn AA, Tsui A, Aziz T, *et al.* Pathological synchronisation in the subthalamic nucleus of patients with Parkinson's disease relates to both bradykinesia and rigidity. Experimental neurology. 2009;215(2):380–387.
[28] Blumenfeld Z, Brontë-Stewart H. High Frequency Deep Brain Stimulation and Neural Rhythms in Parkinson's Disease. Neuropsychology Review. 2015;25(4):384–397.
[29] Neumann W-J, Staub-Bartelt F, Horn A, *et al.* Long term correlation of subthalamic beta band activity with motor impairment in patients with Parkinson's disease. Clinical Neurophysiology. 2017;128(11):2286–2291.
[30] Wingeier B, Tcheng T, Koop MM, Hill BC, Heit G, Bronte-Stewart HM. Intra-operative STN DBS attenuates the prominent beta rhythm in the STN in Parkinson's disease. Experimental neurology. 2006;197(1):244–251.
[31] Bronte-Stewart H, Barberini C, Koop MM, Hill BC, Henderson JM, Wingeier B. The STN beta-band profile in Parkinson's disease is stationary and shows prolonged attenuation after deep brain stimulation. Experimental neurology. 2009;215(1):20–28.
[32] Mnih V, Kavukcuoglu K, Silver D, *et al.* Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529.
[33] Mnih V, Badia AP, Mirza M, *et al.* Asynchronous methods for deep reinforcement learning. In: International conference on machine learning. 2016, 1928–1937.
[34] Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. 2017.
[35] Hausknecht M, Stone P. Deep recurrent Q-learning for partially observable MDPS. CoRR, abs/1507.06527. 2015;7(1).
[36] Schmidhuber J. Reinforcement learning in Markovian and non-Markovian environments. In: Advances in neural information processing systems. 1991, 500–506.
[37] Wierstra D, Foerster A, Peters J, Schmidhuber J. Solving deep memory POMDPs with recurrent policy gradients. In: International Conference on Artificial Neural Networks. 2007, 697–706.
[38] Heess N, Hunt JJ, Lillicrap TP, Silver D. Memory-based control with recurrent neural networks. arXiv preprint arXiv:1512.04455. 2015.
[39] Bakker B. Reinforcement learning with long short-term memory. In: Advances in neural information processing systems. 2002, 1475–1482.
[40] Hutter F, Hoos H, Leyton-Brown K. Sequential model-based optimization for general algorithm configuration. LION. 2011;5:507–523.
[41] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-32

COMBINATION OF CONNECTIVITY AND SPECTRAL FEATURES FOR MOTOR-IMAGERY BCI

Tiziana Cattai[1,2,3], Stefania Colonnese[3], Marie-Constance Corsi[1,2],Danielle S. Bassett[4,5,6,7], Gaetano Scarano[3], Fabrizio De Vico Fallani[1,2]

[1]Inria Paris, Aramis project-team, Paris, France
[2]Institut du Cerveau et de la Moelle Épinière,ICM,Inserm, U 1127, CNRS, UMR 7225,Sorbonne Université, F-75013, Paris, France
[3]Dept. of Information Engineering, Electronics and Telecommunication, Sapienza University of Rome, Rome, Italy
[4]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, 19104, USA
[5]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, 19104, USA
[6]Department of Physics & Astronomy, University of Pennsylvania, Philadelphia, PA, 19104, USA
[7]Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA, 19104, USA

E-mail: tiziana.cattai@inria.fr

ABSTRACT: In brain-computer interfaces (BCI), the detection of different mental states is a key element. In Motor Imagery (MI)-based BCIs, the considered features typically rely on the power spectral density (PSD) of brain signals, but alternative features can be explored looking for better performance. One possibility is the integration of functional connectivity (FC). These features quantify the interactions between different brain areas and they could represent a valuable tool to detect differences between two mental conditions. Here, we investigated the behavior of coherence-based FC features and PSD features, alone and in combination. For a better comparison, we characterized the network centrality of each brain area by computing the weighted node degrees from the estimated FC networks. Our findings show that in both alpha and beta frequency bands, and for almost all the subjects, the fusion of FC network indices and PSD features give better performance. This preliminary results open the way to the use for network-based approaches in BCIs.

INTRODUCTION

A brain-computer interface is a system that enables the interaction subject-external world without peripheral nerve or muscles [26]. It allows communication [17] and the control of real or virtual objects [4]. Nowadays, the performance of BCI appears to be inconsistent across subjects. Indeed, there is an non-negligible percentage of users who cannot use the interface [3], [24].
To face this problem, investigators historically searched for better brain decoders [18]. To do so, one possibility consists of working on the features extraction and selection and more specifically of looking for alternative features that could better discriminate the subjects' mental state.
Recently, a great interest was born in connectivity applied in brain-computer interfaces and in particular in motor imagery, giving promising results [22], [2]. The reason behind this choice is that brain mechanisms involved in BCI are complex and they could be better described by functional interactions changes. Indeed, connectivity describes the interaction between different brain areas that can reflect specific mechanisms such as synchronization in time-others in phase domain, or the causal interactions for instance [11]. Notably, a variation of inter hemispheric connections has been found using phase-locking value, but in other case the prominent mechanisms was the increase of information flow regarding the contralateral motor area [15]. The frequency band that appears involved in the task differs being alpha, beta or gamma and a completely changing behavior can be found in the different bands. These contradictory results motivated us to investigate more the subject.
Here, we hypothesized that features based on connectivity can provide relevant information to better discriminate the subjects' mental state. For this purpose, we first tested the feasibility of the use of connectivity-based features already identified as a feature with high discriminating potential in motor imagery task [7]. Then, we tested the combination of features based on power spectra and on connectivity to take advantage of their complementarity.

MATERIALS AND METHODS

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-32

1. Experimental Protocol: Fifteen healthy subjects (aged 27.73 ± 4.45 years, 7 women) all right handed participated to our motor imagery-based BCI protocol. The users were seated in front of a screen. When the target was up, the subjects had to image to move the right hand (e.g. grasping) [27]. When the target was down they had to remain at rest. During the tasks, we recorded the electroencephalography (EEG) activity with 74 electrodes in a 10-10 standard configuration. We collected 64 trials for motor imagery and 64 for resting state. Each trial lasted 5s. EEG signals were recorded with frequency sampling of 1 kHz and then downsampled to 250 Hz. To remove those related to eye and cardiac artifacts in the sensor space, a pre-processing step consisted in an Independent Component Analysis (ICA) [10] with Infomax algorithm [1] performed with Fieldtrip toolbox [19]. The selection of the components was performed by a visual inspection of the signals.
2. Analysis: We can schematize our recording systems in the following way. N signal samples were acquired at frequency F_s in two different mental states, S_1 and S_2. EEG samples were recorded from M electrodes for each state, obtaining T_{S1} trials in S_1 and T_{S2} trials in S_2; the same procedure was replicated for I subjects.
The signal Acquisition Stage (AS) in the two mental states S_1 and S_2 for the i-th user, t-th trial, m-th channel, n-th signal sample yields the real measurements' sets:

$$X^{(S_1)} = \{x_m^{(i,t)}[n]\},\ S_1\ \text{AS}\} \tag{1}$$

$$X^{(S_2)} = \{x_m^{(i,t)}[n]\},\ S_2\ \text{AS}\} \tag{2}$$

We reported the framework of this study (Fig. 1). The first step consisted of recording brain signals from a given subject i followed by the features extraction. Here, we considered two types of features: PSD, that reflects the local activation of the cortex, and connectivity, describing interactions of brain areas. The last step is the evaluation of compact metric (node degree) that synthetically quantifies the connectivity of each node. In the next sections, a detailed description of each step is given.
2.1 Features estimations: For both the ASs, the following power spectral estimates are computed:

$$P_{x_m}^{(i,t)}[k] = \frac{1}{L_W}\sum_{l=0}^{L_W}\left|\sum_{n=0}^{N-1} w_l[n]x_m^{(i,t)}[n]e^{-j2\pi nk/N}\right|^2 \tag{3}$$

$$P_{x_{m_1}x_{m_2}}^{(i,t)}[k] = \frac{1}{L_W}\sum_{l=0}^{L_W}\left(\sum_{n=0}^{N-1} w_l[n]x_{m_1}^{(i,t)}[n]e^{-j2\pi nk/N}\right)\cdot \left(\sum_{n=0}^{N-1} w_l[n]x_{m_2}^{(i,t)}[n]e^{-j2\pi nk/N}\right)^* \tag{4}$$

where $w_l[n]$, with $l = 0, \cdots L_W - 1$ and $n = 0, \cdots N-1$ are real windows L_W , depending on the estimation method. In the case of Welch method, a set of time orthogonal functions are used [25].
For each subject i, each trial t, each frequency bin we evaluated the two features. Specifically, Welch method

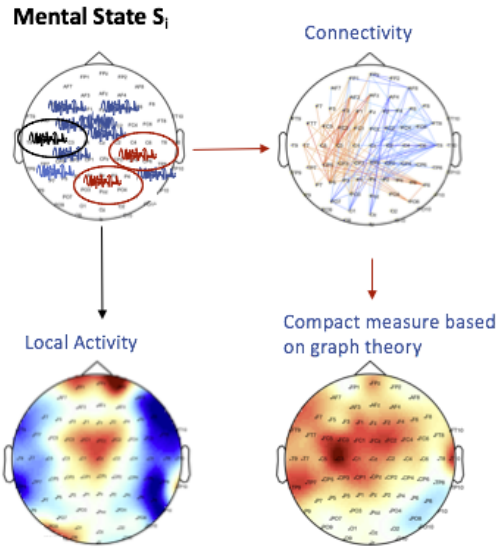


Figure 1: Schematic representation of different steps of our analysis. EEG signals are recorded from each channels. Signals vary in time and frequency domain. Since we collected all the signals, we can extract features, which can be local features (as PSD) or connectivity (as coherence). We can use compact measures based on graph theory (as weighted node degree).

is used for the evaluation of the auto and cross-spectrum. Hanning time windows characterized by a length of 1s and an overlap of 50% are used. To this aim,we narrowed the frequency set, to select features from 4 to 40 Hz in steps of 1 Hz.
The connectivity estimator that we use in this work, is the spectral coherence [6], defined as:

$$C_{m1\,m2}^{(i,t)}[k] = \frac{\left|P_{x_{m_1}x_{m_2}}^{(i,t)}[k]\right|}{\left(P_{x_{m_1}}^{(i,t)}[k]\cdot P_{x_{m_2}}^{(i,t)}[k]\right)^{1/2}} \tag{5}$$

This quantity reflects in the frequency domain the synchronization in amplitude between two signals. The advantage compared to the simpler cross-correlation is that it is possible to separate different frequency bands. This aspect is important in EEG applications where, the response changes in different frequency bands.

We defined a graph as a set of vertices (or nodes) and edges (or links)[5]. In our case, nodes represents the EEG-electrodes and the edges are the connectivity estimate, without thresholding or binarizing [11].
To extract information about the connectivity of each electrode and following the approach in [7], we defined the Coherence-based Node Degree (CND), as follows:

$$\text{CND}_m^{(i,t)}[k] = \sum_{c=0}^{M-1} C_{mc}^{(i,t)}[k] \tag{6}$$

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-32

This measure describes in compact way the connectivity of each node. In fact, it is evaluated for each node and it represents how much one node is connected to all the others [11].
To take into account the neurophysiology aspects underlying the task used here [21],[20], we decided to restrict our study on the set of electrodes located in the controlateral motor area (Fig.2). Nevertheless, the CND is evaluated taking into consideration all the possible connections including the selected nodes.

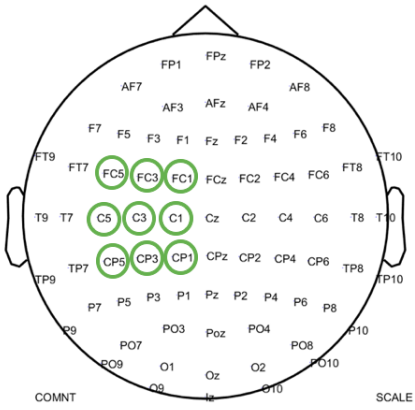


Figure 2: Representation of 74 electrodes in 10-10 standard configuration as used in our experiments. Electrodes in left central are circled in green

2.2. Features fusion: To understand if a combination of different information would better discriminate between two conditions, different approaches have been proposed in the literature [23]. Notably, the fusion of features from different modalities in motor imagery BCI has already been studied, with promising findings [9]. In particular, the fusion can be performed at the classifier level, by attributing a specific weight to the output of each classifier, or at features level. Here we used the latter with a concatenation of the trials of different features.

2.3 Classification: To classify the subjects' mental state, we used the linear discriminant analysis (LDA) method [12]. Our classifier operates on single subject, single frequency bin and single channel. More specifically, for a given subject, for the k -frequency bin and the m-channel we randomly chose the 80% of the trials for the training set. The total number of trials for the training set is 103. The remaining 25 trials, related to the other 20%, are used for the testing part. Performances of the classification were evaluated during the testing part. We repeated this operation 50 times and the performances reported here corresponds to the average in all the repetitions. For i -subject, k -bin, m -channel, we measured classification performances, in terms of accuracy, sensitivity, specificity, and area under curve (AUC). For a necessity of representation, we only reported here the most significative results in terms of accuracy, defined as follows:

$$ACC = \frac{TP+TN}{P+N} \tag{7}$$

Where TP are the true positives, TN the true negatives and P and N are positives and negatives respectively.

RESULTS

We performed the classification test as explained before and we repeated the operation for each subject, each frequency bin and each channel. For each subject, we selected the best value in terms of classification performances (e.g. accuracy) comprised in the central left area. The best accuracy value is associated with a given (electrode; frequency bin) couple. In the Fig 3, we represented the maximum accuracy obtained for each subject and each modality. In particular, we reported in a bar plot the maximum value of accuracy. In order to make the results clearer, we associated to each frequency bin, the related frequency band.
To be more precise, the frequency band that we use are: $B_{theta} = 4-7$Hz, $B_{alpha} = 8-13$Hz, $B_{beta} = 14-29$Hz and $B_{gamma} = 30-40$Hz.

In Fig. 3, when we compare the PSD with coherence, we can notice that 9 on 15 subjects reached higher accuracy with PSD, while for the remaining 6, the coherence is better. If we compare coherence with the fusion, the latter is better in 9 over 15 subjects. The last comparison is between fusion and PSD. The fusion of the two features is better in 13 over 15 subjects. The mean values of accuracy with the associated standard deviations are: 0.66$\pm$ 0.03 for PSD, 0.66$\pm$ 0.02 for coherence and 0.68 $\pm$0.02 for the fusion. The frequency bands associated with the reported results vary according to the subject, but the most selected bands are alpha and beta.
We reported a set of figures to represent the most frequently chosen channels (Fig.4) The associated plots represent the occurrences on the scalp for each explored mode, Coherence-based Node degree, PSD and fusion and for each frequency band. In each picture, the dimension of the circle is proportional to the occurrence of each electrode as the one related with the highest accuracy. C5 and C3 channels are usually the most involved in the task, have a high representation, especially in alpha and beta band. Interestingly, the occurrences associated with these two channels, in the case of coherence feature, are also high in gamma band. For sake of completeness, in Fig.5, for each subject and each feature we report the standard deviation related to the cross validation. Specifically, we evaluated the standard deviation of the accuracy across the 50 random repetitions for the best couple channel-frequency bin.

DISCUSSION AND CONCLUSION

Functional connectivity is a valuable tool to describe and

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-32

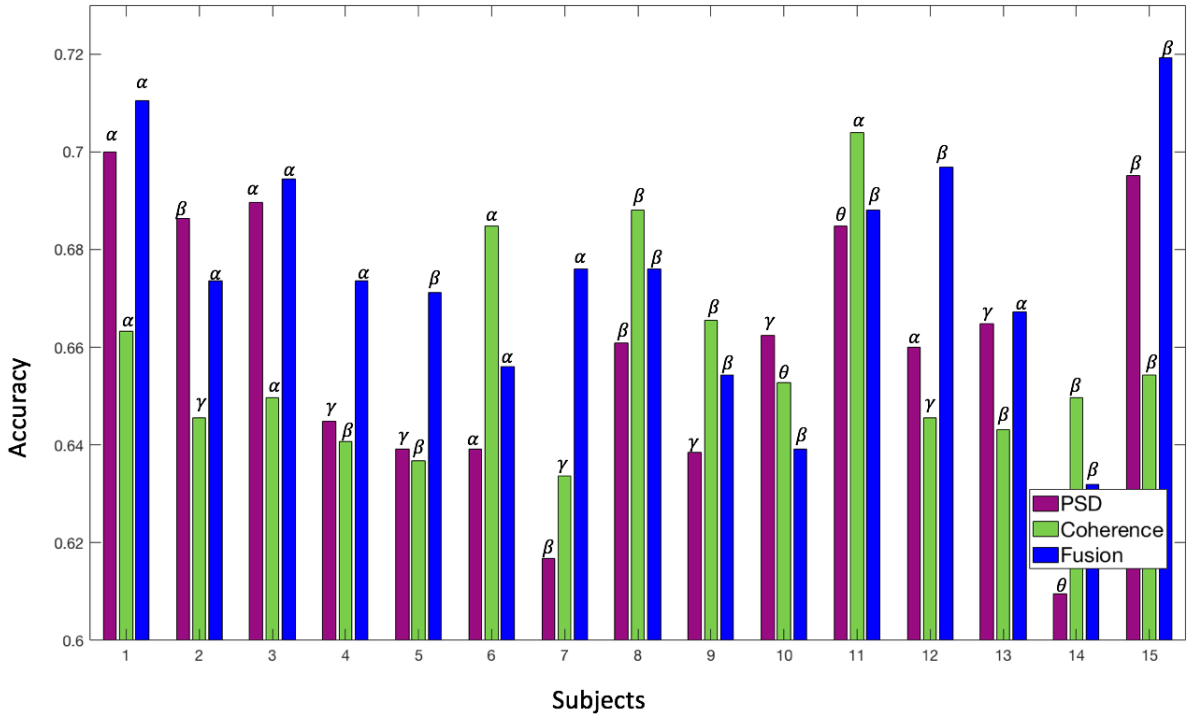


Figure 3: Bar plot with the best accuracy level for each subject. Each value is related to a specific couple frequency bin-electrode. For each subject we have three bars that refer to the three modes: PSD (in pink), Node Coherence (in green) and Fusion (in blue). On the top of each bar we report the frequency band associated to the selected bin.

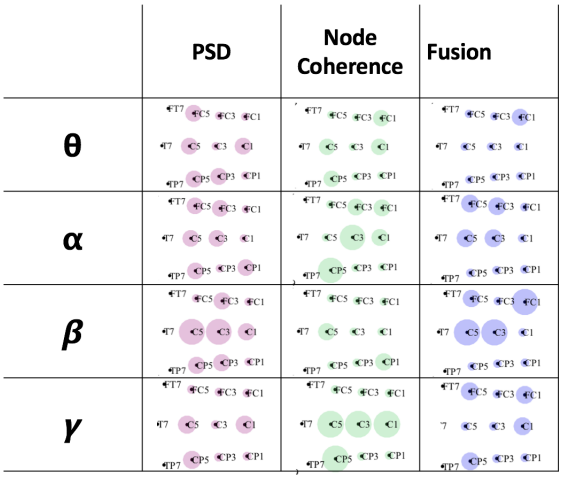


Figure 4: Table with occurrences of electrodes related to the highest accuracy. The size of the circles is proportional to the number of subjects for which that channel was the best. Each frequency band and each mode is considered separately: PSD in pink, coherence-based node degree in green and the fusion in blue

explain the organization and functionality of the brain during the tasks. For these reasons, we believe that it can be an important feature for brain computer interface applications, in which the user has to perform difficult tasks, involving different portions of the brain. In this work, we presented a mental state classifier based on connectivity features and we compared its performances with those achieved using local spectral features. Most importantly, we considered the classification results with the combination of the two features.

Our findings demonstrated that alpha and beta frequency bands are usually more involved during the task. In fact, the accuracy obtained in those bands are higher.Another interesting result is that the fusion between connectivity feature and a local one gives higher accuracy than PSD alone for the majority of the subject, for 13 over 15. This means that the integration of features reflecting different brain mechanisms can be useful for the discrimination of different mental states.

We notice that in general our accuracy values are higher than the actual chance level that in this configuration corresponds to 63% [8]. In addition, we reported a relatively high variance in the accuracy across the cross-validation that is explained by the high inter-trial variability.

From our findings, we can notice that there is a high variability between subjects concerning the best feature to use. This is reflected by the high standard deviation, 0.02 for coherence and fusion and 0.03 PSD. The inter-subjects variability is also evident in terms of frequency band of interest [16]. However, the most selected

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-32

	Subjects														
	1	2	3	3	5	6	7	8	9	10	11	12	13	14	15
PSD	0,1	0,1	0,07	0,09	0,09	0,08	0,1	0,12	0,09	0,1	0,08	0,08	0,1	0,09	0,08
Coherence	0,08	0,09	0,09	0,09	0,09	0,08	0,09	0,09	0,07	0,11	0,09	0,09	0,13	0,11	0,06
Fusion	0,08	0,11	0,08	0,1	0,09	0,1	0,08	0,08	0,08	0,08	0,09	0,1	0,09	0,08	0,1

Figure 5: Table with values of standard deviation of the accuracy associated to cross validation folds. For each subject and each feature, for the best couple channel-frequency bin, we evaluated the standard deviation related to the 50 random repetitions

bands are alpha and beta as expected because they are usually related to motor tasks. Interestingly, in the case of coherence, the gamma frequency band is intensively involved. One possible explanation is that connectivity mechanisms during motor imagery tasks happens also in high frequencies [13].

In this preliminary work, we did not put a constraint on the frequency band in the fusion case. Indeed, for each case, we selected the best frequency bin independently from all the other cases. For instance, it could be interesting to put a constraint on the frequency band associated to the fusion on the basis of the selected frequency band for the other two ways. Alternatively, it could be possible to select the same frequency for all the features. These changes could help in the interpretation of the results.

For the fusion between two different features, another possibility is to perform a fusion of the classifiers output. In this case, the advantage will be a better control of features contribution and it can be helpful for the interpretation. Indeed, for each subject it will be possible to determine which feature contributed the more to the fusion [9]

Another possible improvement would be to increase the number of selected features. Indeed, in this work, we only considered the single feature case, or the combination of one of each class to obtain the fusion. One option to integrate more features could be the ranking of the features on the training set, based for instance of statistical tests performed between conditions [14].

This preliminary work confirmed our hypothesis that connectivity features bring relevant information on brain functionality. Notably, the integration of these features with standard one could increase BCI performances and be a valuable tool to reduce the inter-subject variability.

References

[1] Anthony J Bell and Terrence J Sejnowski. “An information-maximization approach to blind separation and blind deconvolution”. In: *Neural computation* 7.6 (1995), pp. 1129–1159.

[2] Martin Billinger, Clemens Brunner, and Gernot R Müller-Putz. “Single-trial connectivity estimation for classification of motor imagery data”. In: *Journal of neural engineering* 10.4 (2013), p. 046006.

[3] Benjamin Blankertz et al. “Predicting BCI performance to study BCI illiteracy”. In: (2009).

[4] Stevo Bozinovski, Mihail Sestakov, and Liljana Bozinovska. “Using EEG alpha rhythm to control a mobile robot”. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 1988, pp. 1515–1516.

[5] Ed Bullmore and Olaf Sporns. “Complex brain networks: graph theoretical analysis of structural and functional systems”. In: *Nature Reviews Neuroscience* 10.3 (2009), p. 186.

[6] G Clifford Carter. “Coherence and time delay estimation”. In: *Proceedings of the IEEE* 75.2 (1987), pp. 236–255.

[7] Tiziana Cattai, Stefania Colonnese, Marie-Constance Corsi, Danielle S Bassett, Gaetano Scarano, and Fabrizio De Vico Fallani. “Characterization of mental states through node connectivity between brain signals”. In: *Signal Processing Conference (EUSIPCO), 2018 26th European*. IEEE. 2018, pp. 1391–1395.

[8] Etienne Combrisson and Karim Jerbi. “Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy”. In: *Journal of neuroscience methods* 250 (2015), pp. 126–136.

[9] Marie-Constance Corsi et al. “Integrating EEG and MEG Signals to Improve Motor Imagery Classification in Brain–Computer Interface”. In: *International journal of neural systems* (2018), p. 1850014.

[10] Arnaud Delorme, Terrence Sejnowski, and Scott Makeig. “Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis”. In: *Neuroimage* 34.4 (2007), pp. 1443–1449.

[11] Fabrizio De Vico Fallani, Jonas Richiardi, Mario Chavez, and Sophie Achard. “Graph analysis of functional brain networks: practical issues in translational neuroscience”. In: *Phil. Trans. R. Soc. B* 369.1653 (2014), p. 20130521.

[12] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Elsevier, 2013.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-32

[13] J Ginter Jr, KJ Blinowska, M Kamiński, PJ Durka, Gert Pfurtscheller, and Christa Neuper. "Propagation of EEG activity in the beta and gamma band during movement imagery in humans". In: *Methods of information in medicine* 44.01 (2005), pp. 106–113.

[14] Jeremy Guillon et al. "Loss of brain inter-frequency hubs in Alzheimer's disease". In: *Scientific reports* 7.1 (2017), p. 10879.

[15] Mahyar Hamedi, Sh-Hussain Salleh, and Alias Mohd Noor. "Electroencephalographic motor imagery brain connectivity analysis for BCI: a review". In: *Neural computation* 28.6 (2016), pp. 999–1041.

[16] Nuri F Ince, Ahmed H Tewfik, and Sami Arica. "Extraction subject-specific motor imagery time–frequency patterns for single trial EEG classification". In: *Computers in biology and medicine* 37.4 (2007), pp. 499–508.

[17] Jing Jin et al. "The changing face of P300 BCIs: a comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement". In: *PloS one* 7.11 (2012), e49688.

[18] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. "A review of classification algorithms for EEG-based brain–computer interfaces". In: *Journal of neural engineering* 4.2 (2007), R1.

[19] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. "FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data". In: *Computational intelligence and neuroscience* 2011 (2011), p. 1.

[20] G Pfurtscheller. "Spatiotemporal ERD/ERS patterns during voluntary movement and motor imagery". In: *Supplements to Clinical neurophysiology*. Vol. 53. Elsevier, 2000, pp. 196–198.

[21] Gert Pfurtscheller and A Aranibar. "Event-related cortical desynchronization detected by power measurements of scalp EEG". In: *Clinical Neurophysiology* 42.6 (1977), pp. 817–826.

[22] Dheeraj Rathee, Hubert Cecotti, and Girijesh Prasad. "Single-trial effective brain connectivity patterns enhance discriminability of mental imagery tasks". In: *Journal of neural engineering* 14.5 (2017), p. 056005.

[23] Dymitr Ruta and Bogdan Gabrys. "An overview of classifier fusion methods". In: *Computing and Information systems* 7.1 (2000), pp. 1–10.

[24] Carmen Vidaurre and Benjamin Blankertz. "Towards a cure for BCI illiteracy". In: *Brain topography* 23.2 (2010), pp. 194–198.

[25] Peter Welch. "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms". In: *IEEE Transactions on audio and electroacoustics* 15.2 (1967), pp. 70–73.

[26] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. "Brain–computer interfaces for communication and control". In: *Clinical neurophysiology* 113.6 (2002), pp. 767–791.

[27] Jonathan R Wolpaw, Dennis J McFarland, Theresa M Vaughan, and Gerwin Schalk. "The Wadsworth Center brain-computer interface (BCI) research and development program". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11.2 (2003), pp. 1–4.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-33

HETEROGENEOUS REAL-TIME MULTI-CHANNEL TIME-DOMAIN FEATURE EXTRACTION USING PARALLEL SUM REDUCTION ON GPU

J. Arnin, D. Kahani, H. Lakany, B.A. Conway[1]

[1] Centre of Excellence in Rehabilitation Engineering, Department of Biomedical Engineering, University of Strathclyde, Glasgow, United Kingdom

E-mail: jetsada.arnin@strath.ac.uk, b.a.conway@strath.ac.uk

ABSTRACT: Online BCI has become a fascinating field of research nowadays. One of the main challenges in this field is to reduce the latency caused by the computational complexity of the signal processing algorithms. This issue leads to difficulty in processing real-time data. Usually, a trade-off needs to be considered between the number of input samples and precision of the processing algorithms. In this paper, heterogeneous computing concept is investigated to alleviate the computational complexity occurred in real-time processing. An OpenCL was utilized to implement signal processing algorithms in parallel. Feature extraction methods including band power and statistical moments were selected to examine the power of heterogeneous computing using parallel sum reduction. As a result, varying the number of work-group sizes which is an essential parameter of parallel processing provided dissimilar computing times. Also, running at a higher sampling rate yielded a higher benchmark ratio between sequential and parallel. However, system optimization is still necessary when processing BCI in real time.

INTRODUCTION

Processing signal in real-time brain-computer interface (BCI) could usually encounter many difficulties ranging from hardware level to software level. One of the most challenging issues is the system latency [1] which may raise a major problem because this could lead to missing some important data such as an EEG component or an event under half-second. To minimize the latency by only optimizing sequential algorithms to reduce the processing time may not be enough to capture those components. The most common solution on the hardware level is to increase the speed of data transmission, buffering a sufficient amount of incoming data, and using fast processing units [2]. This may cost developer much money and a comprehensive technique.

Nowadays, high-performance computing (HPC) technology plays an important role in solving complex computational problems such as simulation, modeling, and analysis [3-4]. This technology does not only focus on developing faster hardware but the algorithms also [5]. According to the HPC, its concept is based on parallel computing for running application efficiently, reliably, and quickly [6]. To understand the concept of parallel computing, any sequential task can be split into a section which each is run separately on hardware acceleration. There is a lot of hardware acceleration available on the market that has a reasonable price such as consumer graphics processing unit (GPU) and a user-friendly field-programmable gate array (FPGA) [7]. These hardware units are programmable with their specific languages that may take much time to learn in a programming language. To resolve this issue, an open computing language (OpenCL) has been developed to overcome cross-platform programming [8]. It means that any hardware acceleration unit can be executed with one-time coding. So far, the OpenCL platform has become an industry standard for programming those hardware units [9]. In addition, certainly understanding heterogeneous computing concept which is the use of parallel processing techniques is essential and required when programming in OpenCL [10].

To demonstrate the heterogeneous computing concept for real-time signal processing, implementing in major processing steps and their bottlenecks was discussed in this paper. One of the most challenging BCI problems is running feature extraction algorithms in real time [11]. Since the number of processing channel is always much more than one or two, some complex features such as independent component analysis (ICA), autoregressive (AR) model, and discrete wavelet transform (DWT) are mostly implemented in an offline BCI [12]. However, these time-consuming features can be used in real-time processing with optimization that may limit the performance of the algorithms [13]. As an advantage of heterogeneous computing, these algorithms can be broken down into a smaller part and compute each part concurrently then concatenate to a final solution.

In this paper, we applied the commercial and open-source OpenCL technology into real-time signal processing which time-domain feature extraction methods including selective band power and statistical moments were selected to evaluate the computing performance. The system includes both EEG simulation and signal processing module. The proposed module offers up to 32 channels for real-time signal processing based on the heterogeneous computing concept. The archive EEG dataset was used to test the computing performance in real time with different sampling frequency acquired. The parallel computing time was compared to sequential processing approach as well.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-33

MATERIALS AND METHODS

Latency analysis: Regarding the cause of computational latency in real-time processing, this can be divided into four categories, i.e., signal acquisition hardware, data transmission, types of application, and processing algorithms. Table 1 shows the comparison of the latency causes in terms of delay and versatility. The comparison was on the basis of cost-effectiveness and current technology. According to Table 1, the main cause of latency in a real-time BCI is computational complexity of the signal processing algorithms. To overcome this bottleneck, using different approach such as parallel processing instead of traditionally sequential method could more reduce the latency.

Table 1: Delay and versatility in signal processing

Category	Delay	Versatility
Signal acquisition	Low	Low
Data transmission	Medium	Low
Types of application	Medium	Low
Processing algorithms	High	High

EEG dataset: The archive EEG, a collection of 32-channel data from 14 subjects (7 males, 7 females), provided by Swartz Center for Computational Neuroscience [14] was used to evaluate the performance of our parallel design. According to the data, the participants were asked to perform a go-nogo categorization task and a go-no recognition task on natural photographs displayed every 20 milliseconds. The experiment ran a total of 2500 trials on each participant. Note that as the archive data was sampled at 1000 Hz with a specific amplifier but for the purpose of full usage, the data were regenerated and rectified to appropriately match with the voltage range of the analog output device.

Simulation system: This study was developed based on Qt platform (Qt 5.12 LTS) using C++ programming language which can integrate the OpenCL and related libraries together. The simulation system consists of a signal generator and signal acquisition. The archive dataset was generated waveform through the 32-channel analog output device (NI PCIe-6738) and then fed back into the 32-channel analog input device (NI PCIe-6343) using RG58 50-Ohm coaxial cables. For the output device, each channel was generated at the sampling frequency of 1kHz according to the dataset. Note that the output resolution is 16 bits with voltage range of ±10V. For signal acquisition, the sampling rate was varied, including 128, 256 512, 1024, and 2048 Hertz.

Heterogeneous signal processing: According to the general signal processing pipeline, it is frequently processed in sequential approach. This may result in a delay when loads of processing steps are added. Using heterogeneous computing concept in this problem can decrease the latency dramatically. In this paper, the calculation part of the feature extraction on each channel was processed separately and concurrently by multiple compute units. Fig. 1 demonstrates the overall system of the heterogeneous feature extraction which each channel processes simultaneously.

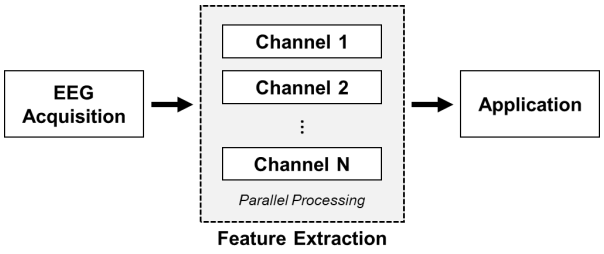


Figure 1: The overall system of heterogeneous feature extraction for real-time BCI.

OpenCL initialization: As an advantage of the cross-platform parallel programming, the graphics card is the easiest unit to be used with OpenCL. The AMD graphic cards (Radeon™ Pro WX 7100) were used to deploy the computation based on the OpenCL 2.0 which the shared virtual memory technique was introduced [15]. The shared virtual memory can reduce the latency of transferring data between the host and devices. Setting up the number of work items manually split into global and local to yield the best computing result. Fig. 2 shows an overview of the structure of the OpenCL 2.0 platform used in the study.

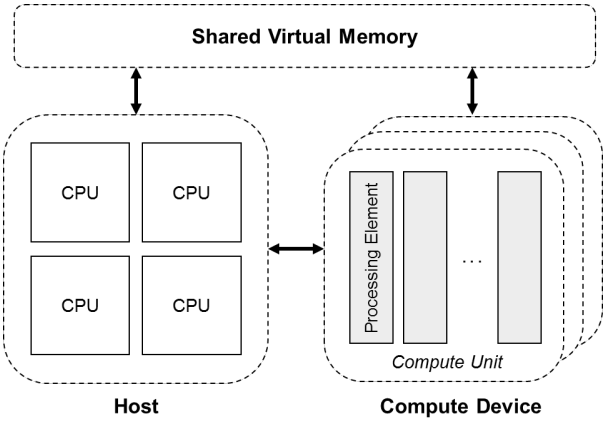


Figure 2: The overview of the structure of the OpenCL 2.0 platform.

Feature extraction: Many familiar feature extraction methods have been developed for processing BCI such as discrete Fourier transform (DFT) or power spectral density (PDS) and wavelet transform (WT) based which are based on frequency analysis. These methods are considered to be the most effective techniques for dealing with time-varying EEG signals. Regarding the time-domain analysis, the most commonly used method for EEG feature extraction is selective band power which is the average power of a signal in a specific frequency range. In this paper, the band power was used as a feature for event detection. Besides the band power, another time-domain method named statistical moments is also used to evaluate. The statistical moments are specific quantitative measurements in time domain

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-33

analysis. The general formula of the n-th order statistical moments was described in the literature [16-17] which mean, variance, skewness, and kurtosis are mostly used for feature extraction. In this paper, the first moment and second moment which are mean and variance were calculated concurrently on each channel.

Parallel sum reduction: To achieve the highest performance from the heterogeneous computing concept, as the calculation of band power and statistical moments mostly uses a summation, this can be managed by using a parallel sum reduction technique [18]. This technique maximizes the performance of compute unit by copying values from global memory into a local memory of the same work-group. Then each work-group processes its local work-item concurrently that is partitioning the whole summation into a small summation and finalize when all work-items finished their own tasks. Fig. 3 shows the concept of parallel sum reduction which introduces the use of local memory to store each element concurrently and then reduce to half by using a stride. Note that in the OpenCL 2.0 parallel sum reduction is integrated into a workgroup function so there is no need to write a nested loop to calculate the summation.

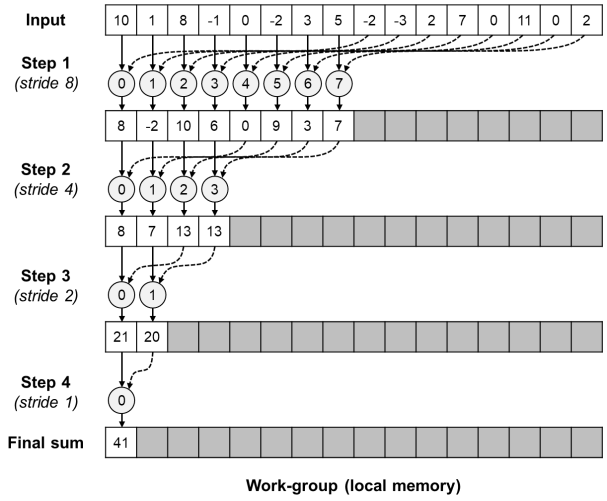


Figure 3: The concept of parallel sum reduction.

Performance improvement: In order to improve the computational speed, fine-tuning parameters for parallel sum reduction is required. As the appropriate number of input data for sum reduction should be a power of 2, we decided the local work-item at 16, 32, 64, and 128, respectively. Conversely, using sum reduction inside the work-group function of the OpenCL 2.0, the input number is not necessary to follow the power of 2.

Benchmarking: The average computation time of a full command execution running on the GPU each local work-item are recorded for 100 times and compared to the result from sequential computing. Note that the full execution starts from transferring data from host to device, processing the kernel, and transferring data back to the host. This study ran on 64-bit Window 10 OS, with 32-GB DDR4 and Intel Xeon E5-1630 v4.

RESULTS

Feature extraction methods including band power and statistical moments were examined in real-time signal processing using the OpenCL platform. The execution time and benchmarking of sequential processing and parallel processing approach were reported in this section. Fig. 4 illustrates the execution time in microsecond when the band power feature was performed. Regarding the result, the execution time was related to the number of a processed sample which higher number required more processing time. With a modification of the number of work-group sizes, a large number of work-group size provided the better performance which the computing time was reduced. In addition, Fig. 5 shows the benchmarking of sequential processing and parallel processing on the same compute device. The ratio was calculated from the execution time of the sequential approach divided by the execution parallel approach. According to the benchmarking result, the higher the sampling rate set, the higher the ratio received. Apart from the band power, using statistical moments also provided likely an identical result. Fig. 6 presents its execution time at different numbers of input. Interestingly, adjusting work-group size had an impacted on speed especially at 256Hz and 512Hz. Fig. 7 also showed the ratio as explained previously.

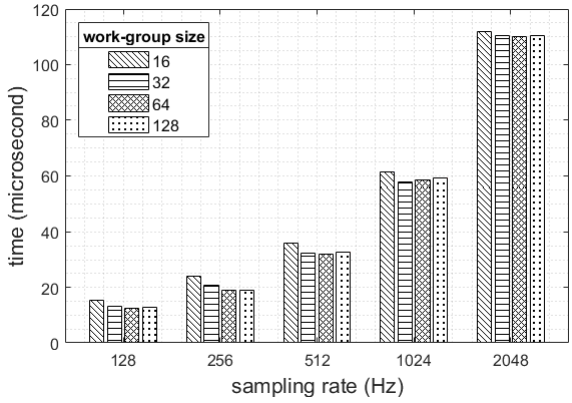


Figure 4: Execution time of band power feature.

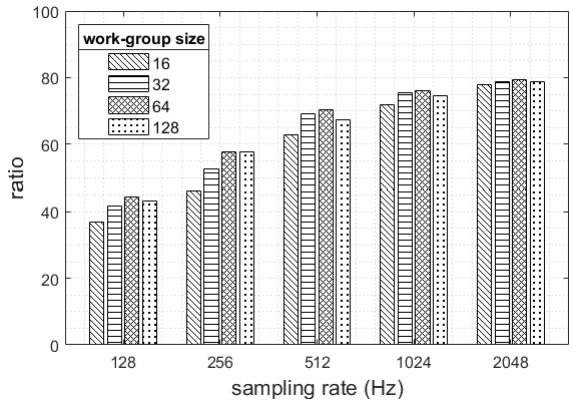


Figure 5: Benchmarking of band power feature.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-33

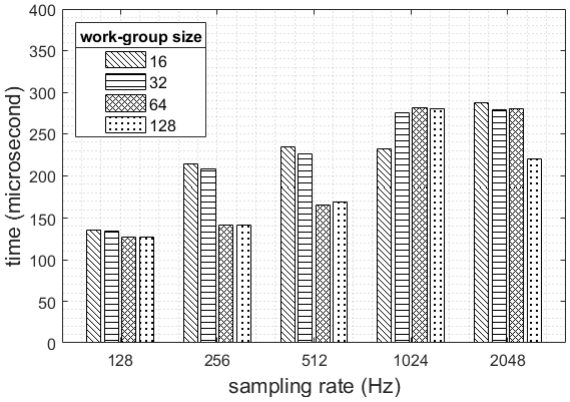


Figure 6: Execution time of statistical moments feature.

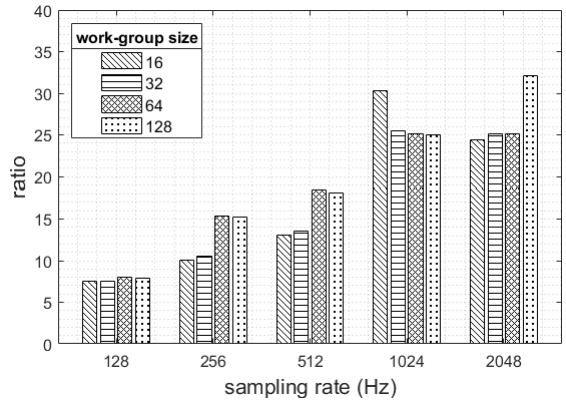


Figure 7: Benchmarking of statistical moments feature.

DISCUSSION

According to the results, the latency of feature extraction was relatively small and appropriately enough for real-time processing. For example, at 1kHz of the sampling rate, band power feature was about 0.06 milliseconds while the statistical moments method was roughly 0.25 milliseconds. Fascinatingly, running data at 2kHz with band power method achieved the execution time about 0.11 milliseconds which is much higher than running at 1kHz (nearly double time) whereas statistical moments method yielded the slightly identical result at 1kHz. It is to be observed that running on different processors and environments may yield different computing times.

Remarkably to the benchmarking results, the reduction technique provided a higher ratio when the size of the input was larger. Besides, increasing the number of work-group size from 16 to 64 provided almost the higher ratio for both features. While setting a work-group size at 128, the ratio dropped slightly. This is because it is allowed enough times for the compute unit to initiate internal parameters (hardware level) and then process data continuously and efficiently as discussed in the previous study [19]. Therefore, a tradeoff between the number of input samples and the acceptable latency in the system should be considered. Note that the GPU used in the study has the maximum work-group size of 256 work-items on each dimension.

Regarding the sum reduction, it can be applied to other time-consuming algorithms such as frequency-domain analysis like DFT and WT. These multiple computing steps can be separated into multiple kernels and directly execute from device side without any request commands from the host side as introduced in the new features of OpenCL 2.0. This technique has also been implemented into the statistical moments which there were two kernels, i.e., one for mean calculation and another for variance calculation, running concurrently using shared virtual memory. Not only running on the GPU, by using the same OpenCL program the project can be run on other systems such as FPGAs. This is expeditious and required only minor parameters adjustment.

With regard to the latency analysis, some limitations can be resolved but have to tradeoff between time- and cost-effectiveness. Furthermore, an OpenCL library for signal processing could help researchers to gain the most benefit from heterogeneous computing because in this study we have developed all steps from the beginning including setting up complicated parameters such as initializing a platform and a context.

CONCLUSION

This study presents the use of heterogeneous computing technique to implement into BCI processing. As a result, speeding up the computation by using a parallel processing scheme is possible and flexible for real-time computing. To reduce the system latency, optimization of both hardware and software should be considered when using in such real-time applications. Lastly, an OpenCL library could help researchers to reduce the developing time for the BCI applications which is the next step of our work.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support from Scottish Government Health Directorates and the Royal Thai Government scholarship. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

[1] Xu R, Jiang N, Lin C, Mrachacz-Kersting N, Dremstrup K, Farina D. Enhanced low-latency detection of motor intention from EEG for closed-loop brain-computer interface applications. IEEE Transactions on Biomedical Engineering. 2014;61(2):288-96

[2] Oweiss KG. A systems approach for data compression and latency reduction in cortically controlled brain machine interfaces. IEEE Transactions on Biomedical Engineering. 2006;53(7):1364-77

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-33

[3] Gulo CA, Sementille AC, Tavares JM. Techniques of medical image processing and analysis accelerated by high-performance computing: A systematic literature review. Journal of Real-Time Image Processing. 2017;16:1-8
[4] Xu J, Huang E, Chen CH, Lee LH. Simulation optimization: A review and exploration in the new era of cloud computing and big data. Asia-Pacific Journal of Operational Research. 2015;32(03):1550019
[5] Wang T, Kemao Q. Parallel computing in experimental mechanics and optical measurement: A review (II). Optics and Lasers in Engineering. 2018;104:181-91
[6] Pratx G, Xing L. GPU computing in medical physics: A review. Medical physics. 2011;38(5):2685-97
[7] Brodtkorb AR, Hagen TR, Sætra ML. Graphics processing unit (GPU) programming strategies and trends in GPU computing. Journal of Parallel and Distributed Computing. 2013;73(1):4-13
[8] Munshi A, Gaster B, Mattson TG, Ginsburg D. OpenCL Programming Guide, Pearson Education (2011)
[9] Stone JE, Gohara D, Shi G. OpenCL: A parallel programming standard for heterogeneous computing systems. Computing in science & engineering. 2010;12(3):66
[10] Gaster B, Howes L, Kaeli DR, Mistry P, Schaa D. Heterogeneous computing with openCL: revised openCL 1, Newnes (2012)
[11] Tangermann M, et al. Review of the BCI competition IV. Frontiers in neuroscience. 2012;6:55
[12] 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
[13] Al-Fahoum AS, Al-Fraihat AA. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. ISRN neuroscience. 2014
[14] Delorme A, Rousselet GA, Mace MJ, Fabre-Thorpe M. Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. Brain research Cognitive brain research. 2004;19(2):103-13
[15] Kaeli DR, Mistry P, Schaa D, Zhang DP. Heterogeneous Computing with OpenCL 2.0, Morgan Kaufmann (2015)
[16] Soliman SS, Hsue SZ. Signal classification using statistical moments. IEEE Transactions on Communications. 1992;40(5):908-16
[17] Hjorth B. EEG analysis based on time domain properties. Electroencephalography and clinical neurophysiology. 1970;29(3):306-10
[18] Zaki MJ, Parthasarathy S, Ogihara M, Li W. Parallel algorithms for discovery of association rules. Data mining and knowledge discovery. 1997;1(4):343-73
[19] Fang J, Varbanescu AL, Sips H. A comprehensive performance comparison of CUDA and OpenCL, in Proc. International Conference on Parallel Processing, 2011, 216-225

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-34

UMA-BCI SPELLER, A P300-BASED SPELLING TOOL

F. Velasco-Álvarez[1], S. Sancha-Ros[2], E. García-Garaluz[2], A. Fernández-Rodríguez[1], M.T. Medina-Juliá[1], R. Ron-Angevin[1]

[1] Departamento de Tecnología Electrónica, Universidad de Málaga, Málaga, Spain
[2] ENESO Tecnología de Adaptación S.L., Málaga, Spain

E-mail: rra@dte.uma.es

ABSTRACT: The brain-computer interface research group of the University of Malaga (UMA-BCI) has developed a speller application based on the well-known P300 potential which can be easily installed, configured and used. The application supports the common P300 paradigms: the Row-Column Paradigm and the Rapid Serial Visual Presentation Paradigm. The inner core of the application is implemented with a widely used and studied platform, BCI2000, which ensures its reliability and allows other researchers to apply modifications at will in order to test new features. There are many studies regarding brain-controlled spellers; however, these systems do not usually leap out of the lab because of technical and economic requirements. As a consequence, the potential end users do not benefit from these scientific advances in their daily life. The objective of this paper is to present a novel brain-controlled speller designed to be used by patients due to its versatility and ease of use.

INTRODUCTION

Most of the studies regarding Brain-Computer Interface (BCI) research are carried out in experimental environments with healthy subjects. Even when patients suffering from different kinds of paralysis participate as subjects, they usually do it only for a few sessions in order to validate the proposed BCI systems. Several authors have also studied the use of BCIs taking into account the perspective of patients, caregivers, and professionals:

- A recent study by Wolpaw et al. [1] found that, although there are more than 4000 research studies regarding BCI systems, only three cases corresponded to independent home use of a BCI for communication. Encouraged by these case reports, they carried out a research with 37 ALS patients who initially began to use a BCI system at home; when the study ended (after up to 18 months), only 7 patients kept the BCI for further use. The authors mention several reasons that explain this difference: 13 patients died during the study or abandoned it because of a rapid disease progression, 9 patients could not use it, 4 lost interest and 2 preferred another assistive communication device, among other minor reasons.
- The results of Taherian and Davies [2] showed that BCIs are not suitable for independent use outside experimental environments. For them, "the hardware needs to be configurable, comfortable and accommodate physical support needs. The training approach needs to be less cognitively demanding, motivating and support personalized mental tasks". They also cite how the reliability should improve and the need for adequate technological support.
- Liberati et al. [3] concluded that four reasons were fundamental to understand the low intensity use of BCI: "i) lack of information on BCI and its everyday applications; ii) importance of a customizable system that supports individuals throughout the various stages of the disease; iii) relationship between affectivity and technology use; and iv) importance of individuals retaining a sense of agency".

Taking into account these results, the BCI research group of the University of Malaga (UMA-BCI) decided to implement a BCI system based on a P300 speller that helped to solve some of the previously mentioned issues in order to be used by patients and caregivers at home. It should be an easy to install and use tool and it should be modular and configurable to support personalization. At the same time, the developed tool should be flexible enough to be used by researchers in the field in order to study variations and alternative paradigms. Both Row-Column Paradigm (RCP) and Rapid Serial Visual Presentation (RSVP) P300-based paradigms should be supported. The target tool should include the possibility of providing dynamic spellers, i.e., spellers that change the items in their layout depending on previous selections. One last specification of the application is that it should be free and open source.

Three options were available: i) to fully develop a custom BCI system; ii) to use a general purpose platform to implement the desired speller; and iii) to partially develop a custom BCI system based on a BCI platform. We declined the first option because it would require a lot of development and testing in order to get a reliable application. The second option would avoid these processes, but it would be limited in its functionality and it would require advanced technical skills. Finally, we decided to benefit from the extended use of platforms to use reliable signal acquisition and signal processing modules, thus using a general purpose platform as the internal core of a custom application.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-34

There exist multi-purpose platforms for BCI analysis and development for scientists to carry out their research; a complete survey on these platforms can be found in [4]. As previously mentioned, the aim of this work is to provide end users with an easy to use P300 speller that is flexible enough to be useful for researchers as well. From this point of view, BCI2000 [5] and OpenViBE [6] are the most interesting among the BCI platforms as they are both widely used, with up-to-date software releases, documentation and support. These two platforms are intended to build end user BCI applications, however, they still require technical skills in order to implement a P300 speller. As both are general-purpose platforms, with a high degree of configurability, it may be complex to parameterize them in order to obtain the desired speller. Besides, the tool proposed in the present paper is intended to be easily configurable for researchers as well, in such a way that the testing of new spelling modalities using alternative stimuli may help in their research (e.g., through the use of images or the independent customization of each cell). BCI2000 is used notably more by researchers in their works [4], so a speller based on it could be easier to use and modify by the scientific community than one based on OpenViBE. This was the main reason to choose BCI2000 as the software core of the speller implementation.

BCI2000. This platform allows a system to be implemented using three modules: *signal source*, *signal processing* and *application*. Each one of these modules can be independently configured in order to adjust them to the desired system. The user can modify many parameters in order to configure the three modules:

- The *signal source* module supports multiple data acquisition hardware that can be selected. Besides, manufacturers can create their own C++ modules to be included.
- The *signal processing* module allows the processing of the most common EEG signals (P300, steady-state visual evoked potential (SSVEP), sensorimotor rhythms (SMR), slow cortical potentials (SCP) etc.) with predefined or custom-implemented algorithms.
- The *application* module contains several basic graphical applications which use the data coming from signal processing as input.

All these parameters can be modified and saved through a tab-based GUI provided together with BCI2000 and they are stored in a parameters file with the extension *.prm*. The use of parameterization allows a high degree of customization, but at the same time it makes the application harder to configure. Among the pre-defined BCI2000 applications, the platform includes implementations of the RCP speller (*P3Speller*) and RSVP paradigm (*StimulusPresentation*). The *UMA-BCI Speller* here presented is built using these two BCI2000 presentation paradigms.

This paper is a short version of a full length manuscript sent to a scientific journal. On the date this paper is written the full length version is under review.

MATERIALS AND METHODS

The *UMA-BCI Speller* is a P300-based speller application that is built using BCI2000, that simplifies the use of it and extends its functions acting as a wrapper of BCI2000 (see Fig. 1).

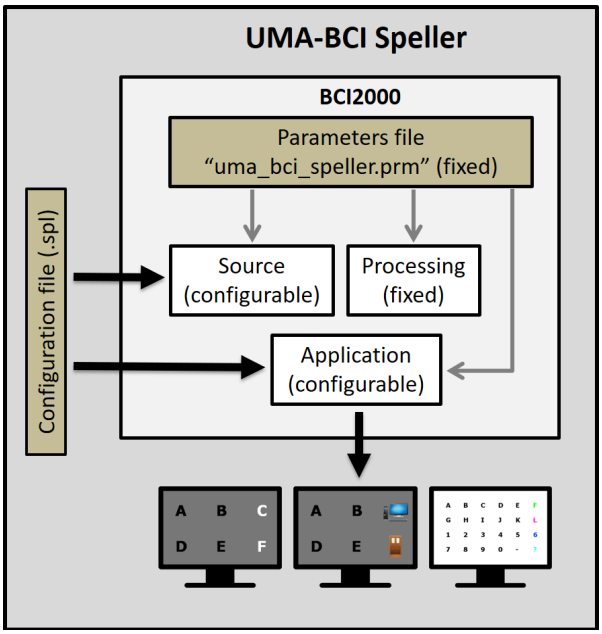


Figure 1: *UMA-BCI Speller* scheme. The application acts as a wrapper of BCI2000.

Most of BCI2000's configurability is intentionally limited because *UMA-BCI Speller*'s aim is to easily implement a P300 speller. For this reason, the BCI2000 signal processing module is fixed to *P3SignalProcessing*, as well as the majority of the application module parameters (fixed to be *P3Speller* or *StimulusPresentation* for RCP or RSVP, respectively) and the source module (which loads the common electrode positions for a P300 experiment). These predefined settings are applied to BCI2000 through a parameters file (*uma_bci_speller.prm* or *uma_bci_speller_rsvp.prm*) that is loaded automatically. Through the *UMA-BCI Speller* GUI, users can indirectly modify some BCI2000 parameters that affect the source signal selection and the number of elements of the speller matrix (see Fig. 1). Besides, users can control the visual appearance of the speller through new options outside BCI2000 that let them change colours and add images. The set of configuration parameters that users can modify is grouped in a new configuration file **.spl*. Thus, users only control a reduced set of parameters in order to use and personalize their own P300 speller keyboard. Advanced users (i.e., users who already know how to manage BCI2000) can edit the mentioned "uma_bci_speller.prm" file in order to customize the application beyond what *UMA-BCI Speller* offers by default (e.g., parameters regarding the timing of the speller).

The customization of *UMA-BCI Speller* is achieved

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-34

through the GUI configuration, which can be modified in two ways: i) by changing the values of the parameters that are related to the general appearance of the speller (number of rows and columns, space in between them, background colour...) and to all the cells equally; and ii) by particularizing the appearance and function of each individual cell. There are two menus to modify these two sets of parameters, the grid parameters and the cell parameters menus, which will be described in their respective sections below. The *UMA-BCI Speller* configuration files (with extension *.spl*) save both types of parameters.

Fig. 2 shows the top partial view of the main screen which contains a toolbar with the most common functionality: to create, open and save configurations; to start and stop experimental runs; to access the settings menu; to toggle full screen mode, RCP/RSVP paradigms and calibration/online test mode; to set the calibration text, the number of sequences, the subject name and session; and to launch the classifier tool. The image shows a calibration session with the word "text" in which the two first letters have already been calibrated; the central column is being highlighted.

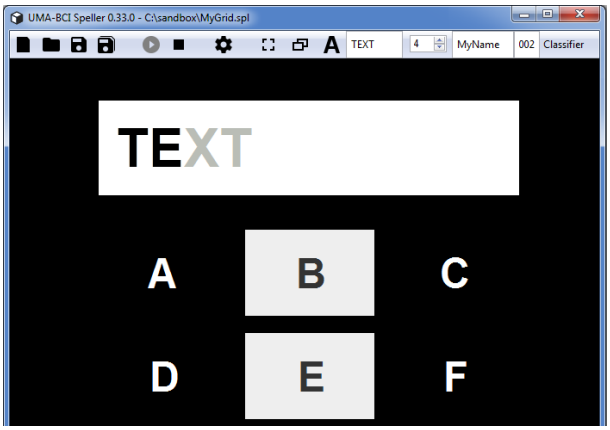


Figure 2. *UMA-BCI Speller* partial screen capture.

Grid parameters. This set of parameters (Fig. 3) is accessed by double-clicking in the idle part of the GUI (i.e., the speller background). Users can modify the number of rows and columns of the display and specify the size of the cells, as well as the empty space in between. The typing bar layout is defined here as well. There are some other parameters that allow the user to make changes on the keyboard letters' font type and size, as well as the cells' background colour for either the idle or highlighted state. Through this menu, users can fix also the size of the stimulus in the case of activating the RSVP paradigm. These parameters affect all cells equally.

The grid parameters shown in Fig. 3 configure the layout in Fig. 5b with four rows and three columns, highlighting the letters in blue.

Figure 3. Grid parameters menu.

Cell parameters. The capability to customize each individual cell beyond the general appearance of the grid parameters is one of the main aspects of the *UMA-BCI Speller*. If users want to change a single cell's appearance, they can do it by double-clicking on a particular item; a cell-specific menu will appear (Fig. 4). In this menu, two columns of parameters are present, *Idle* and *Highlight*, for the case of the cell being in idle or highlighted state, respectively. The cell text and background colour for both states can be modified; if not marked in this menu, the values defined in the general grid parameters menu will apply.

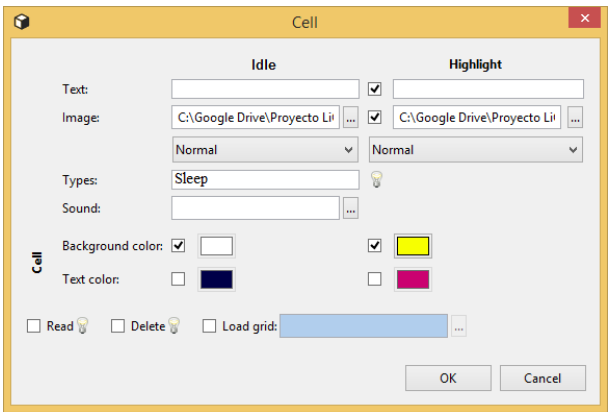


Figure 4. Cell parameters menu.

Another important feature is that images can be included here instead of text, for both idle and highlight states (field *Image*). The images presented this way will fit the space of the corresponding cell determined through the grid parameters.

In the field *Types*, users can set the text to be written in the typing bar when the cell is selected. In the common case of a speller matrix containing characters, the *Types*

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-34

text should be a single character corresponding to the one actually shown. However, in other situations the *Types* field may be a whole word or sentence; that would be the case of cells containing pictograms that could mean "water", or "turn on the light", for example. The field *Sound* allows a sound to be loaded to be heard when this cell is presented in the RSVP paradigm.

Two checkable options are present, *Read* and *Delete*. With the first one, the user indicates that when this cell is selected, the system "reads" with a synthesized voice all the characters spelled to this point. The other option is used to include a *Delete* cell that removes the last spelled character.

Finally, the *Load grid* option allows the speller elements to be changed to a previously saved grid. In this way, the elements in the speller are updated depending on previous selections, so users can navigate through several keyboards, as in [7].

In the case of using the RSVP paradigm, only the *Highlight* parameters apply, as the symbols are presented serially in their highlighted state.

The cell parameters in Fig. 4 define the appearance of the upper-right cell in Fig. 5d: there, the same image with alpha channel is used for the idle and highlighted state; however, the highlighting is achieved through the use of a yellow background. When this cell is selected, the speller writes the word "Sleep".

P300 Classifier automatic calling. The offline signal processing procedure of the calibration data is simplified in the *UMA-BCI Speller*, as it is integrated with the tool: once the training session is finished, one single action groups all the subject's EEG raw signals and performs the necessary steps to obtain the optimal classifier. Clicking on the button Classifier (see Fig. 2) causes the system to open a window to select the files with the calibration data. Once selected, an internal process uses the BCI2000 tool *P300 Classifier* to obtain the classifier parameters to be used in the free spelling session. Another window with the results of the process is then shown, which indicates if a classifier could be obtained with the supplied data and the classification accuracy corresponding to the number of sequences. This information can be used to determine the optimal number of sequences to be applied in the free spelling mode.

Settings Menu. This menu allows the user to parameterize some miscellaneous aspects of the application. One option in this menu is the signal source selection: the process for using a bio-signal amplifier in *UMA-BCI Speller* is the same as in BCI2000: a specific module needs to be provided by manufacturers, which may be a *.exe* and a *.dll* file or source code to be compiled. In both cases, once the BCI2000 system is ready to use them, the *UMA-BCI Speller* just needs the user to select the desired option in the *Source* selector.

The Settings menu also allows the user to add an additional *.prm* parameters file that directly affects any aspect of BCI2000, beyond the default settings determined by the automatically loaded parameters files (*uma_bci_speller.prm* or *uma_bci_speller_rsvp.prm*) mentioned before. The parameters loaded in this way overwrite those automatically loaded, but they do not persist for the next runs. If the users want to persistently update these parameters, the default parameters files need to be directly modified (e.g., with a text editor).

GUI configuration examples. Fig. 5 shows four variations obtained by modifying grid and cell parameters; the last column is highlighted in all the examples. Fig. 5a corresponds to a common implementation of a P300 speller. Fig. 5b is the result of applying grid parameters to increment the number of rows and columns and to change the highlighting colour to blue (for details see Fig. 3). In Fig. 5c the highlighting and cell background colours are different in the last column cells, it is achieved through the use of cell parameters. Finally, Fig. 5d shows the use of images in cell parameters: every cell contains a pictogram with white background while the highlighting consists of a changing the background colour; the upper-right cell parameters are detailed in Fig. 4.

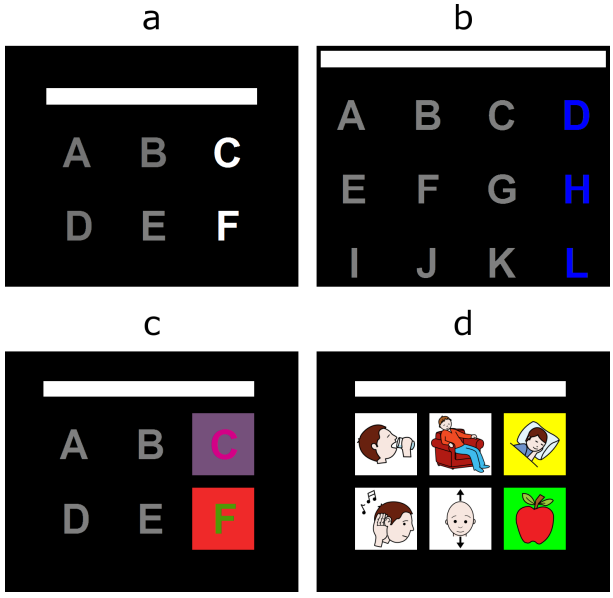


Figure 5. GUI configuration examples.

Hardware and software requirements. The *UMA-BCI Speller* application does not require any special features. As it wraps BCI2000, the whole system requirements would be those of BCI2000. The BCI2000 platform is not a heavy software, it only requires a Windows-based operating system (Windows 2000 or newer) and one of the data acquisition devices that BCI2000 supports [8]. BCI2000 can also run on Linux and OS X systems, but there is no binary distribution for these systems (users would need to compile BCI2000 before using it). Besides, only a minority of source modules support operating systems other than Windows (those modules that receive data through a TCP connection [9]).

Licensing and availability. The *UMA-BCI Speller* application is implemented using BCI2000, which is available under the GNU General Public License (GPL). As a consequence, the use and distribution of

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-34

UMA-BCI Speller is protected by the same GPL license. This means that anyone can use *UMA-BCI Speller*, that the source code is publicly available, and that anyone can develop and distribute derived software products under any terms, provided that the full source code of the derived product is made publicly available for download under the terms of the GPL.
The installation files of *UMA-BCI Speller*, the full source code and a detailed User Manual with all the information regarding the installation, configuration and use can be accessed in the following URL: https://proyectos.diana.uma.es/umabci_speller.

CONCLUSION

We have presented in this paper a BCI tool focused on a speller application: the *UMA-BCI Speller*. The aim of this tool is to provide end users with an easy to use open source P300 speller. It is based on the widely used platform BCI2000, so it takes advantage of the reliability that such a platform offers. The *UMA-BCI Speller* wraps BCI2000 in such a way that its configuration and use is more visual and easier, a fact that can enhance the use of BCI systems at end users' homes. The UMA-BCI Speller allows new features to be added in order to test different variations of the speller layout, a capability that BCI researchers can exploit. The *UMA-BCI Speller* supports two P300 stimulations: RCP and RSVP. Users can configure their speller more appropriately using characters, images or sound cues, and they can navigate through different layouts, thus opening the door to complex speller configurations.

The *UMA-BCI Speller* uses the current release of BCI2000 (v3.0.5); however recent beta versions (e.g. BCI2000 v3.6) include new features that enable a "quasi-asynchronous" control and let the system dynamically change the number of stimulus repetitions through the use of the parameters MinimumEvidence and AccumulateEvidence [10]. Upcoming versions of *UMA-BCI Speller* will be updated to include the new features of BCI2000 beta releases.

Other features to be included could be the support of variations of the RCP (e.g. the Checkerboard paradigm [11]) and RSVP (e.g. Triple RSVP [12]). Other current trends on BCI could help to reduce the calibration time, like the use of generic models or adaptative calibration techniques. It would be interesting also to adapt the interface in order to extend its functionality beyond spelling, for example, use it to browse the internet or to activate external actuators.

ACKNOWLEDGEMENTS

This work was partially supported by the Spanish Ministry of Economy and Competitiveness through the project LICOM (DPI2015-67064-R), the European Regional Development Fund (ERDF) and the University of Malaga.

REFERENCES

[1] J. R. Wolpaw et al., Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis, Neurology, p. 10.1212/WNL.0000000000005812, 2018.
[2] S. Taherian and T. C. Davies, Caregiver and special education staff perspectives of a commercial brain-computer interface as access technology: a qualitative study, Brain-Computer Interfaces, vol. 5, no. 2–3, pp. 73–87, 2018.
[3] G. Liberati et al., Developing brain-computer interfaces from a user-centered perspective: Assessing the needs of persons with amyotrophic lateral sclerosis, caregivers, and professionals, Appl. Ergon., vol. 50, pp. 139–146, 2015.
[4] C. Brunner et al., BCI Software Platforms, in Towards Practical Brain-Computer Interfaces, 2013, pp. 303–331.
[5] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, BCI2000: A general-purpose brain-computer interface (BCI) system, IEEE Trans. Biomed. Eng., vol. 51, no. 6, pp. 1034–1043, 2004.
[6] C. Arrouët, M. Congedo, J.-E. Marvie, F. Lamarche, A. Lécuyer, and B. Arnaldi, Open-ViBE: A three dimensional platform for real-time neuroscience, J. Neurother., vol. 9, no. 1, pp. 3–25, 2005.
[7] G. Pires, U. Nunes, and M. Castelo-Branco, GIBS block speller: Toward a gaze-independent P300-based BCI, Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, pp. 6360–6364, 2011.
[8] G. Schalk and J. Mellinger, A Practical Guide to Brain-Computer Interfacing with BCI2000. Springer London Dordrecht Heidelberg New York, 2010.
[9] BCI2000 Binaries - BCI2000 Wiki. [Online]. Available: https://www.bci2000.org/mediawiki/index.php/BCI2000_Binaries#Non-Windows_Operating_Systems. [Accessed: 31-Jan-2019].
[10] User Reference:P3SpellerTask - BCI2000 Wiki. [Online]. Available: https://www.bci2000.org/mediawiki/index.php/User_Reference:P3SpellerTask#MinimumEvidence. [Accessed: 31-Jan-2019].
[11] G. Townsend et al., A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns, Clin. Neurophysiol., vol. 121, no. 7, pp. 1109–1120, 2010.
[12] Z. Lin, C. Zhang, Y. Zeng, L. Tong, and B. Yan, A novel P300 BCI speller based on the Triple RSVP paradigm, Sci. Rep., vol. 8, no. 1, pp. 1–9, 2018.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-35

ONLINE DETECTION OF HAND OPEN VS PALMAR GRASP ATTEMPTS IN A PERSON WITH SPINAL CORD INJURY

P. Ofner[1], J. Pereira[1], A. Schwarz[1], G.R. Müller-Putz[1]

[1]Institute of Neural Engineering, Graz University of Technology, Graz, Austria

E-mail: patrick@ofner.science, gernot.mueller@tugraz.at

ABSTRACT: In a case study with a person with high cervical spinal cord injury, we show a first proof-of-concept on how to detect and classify different movement attempts of the same upper limb. The lesion was complete (AIS A) at level C4 and no hand function was preserved. We detected in a self-paced online setup hand open and palmar grasp with an accuracy of 68.4 % (chance level 50%).

INTRODUCTION

A brain-computer interface (BCI) can detect various intentionally modulated brain signals and use them as a control signal or as a communication channel [15]. Persons with cervical spinal cord injury (SCI) may profit from a BCI combined with functional electrical stimulation (FES) [26], a so called motor neuroprosthesis. The BCI detects the user's movement intention, and transforms the movement intention into a real movement via FES. It is a technical bypass of the lesion in the spinal cord. So far, several attempts have been shown to restore grasp function in persons with SCI using a BCI with FES [16, 22, 24]. Non-invasive BCI/FES neuroprostheses often exploit modulations of brain oscillations in the mu or beta band accompanying movement imaginations (MIs) [21]. However, these BCIs have also a downside. Despite using brain signals to control motor functions, the mental strategy used to modulate the brain signals is often not similar to the intended movement. For example, [16, 20, 22] used a repetitive foot MI or contralateral hand MI to control the hand. Furthermore, details how the MI is performed are hardly accessible in EEG brain oscillations. EEG brain oscillations rather allow to detect the state of performing repetitive MI, i.e. *that* one moves the limb but not *how* it is moved [1]. However, there is evidence that source imaging can be used to classify more complex MI [6].

Instead of brain oscillations, we exploit another brain signal called movement related cortical potentials (MRCPs) [28]. MRCPs were shown to encode, e.g. force [10] or reaching directions/targets [7, 13, 29], and were furthermore used to detect movements [18]. MRCPs could provide a non-invasive control signal which allows for a more natural neuroprosthesis control compared to EEG brain oscillations. A BCI based on MRCPs could detect movement attempts in persons with SCI like hand open, palmar grasp, pronation, supination, etc., and use these detections for neuroprosthesis control. In that context, we have shown that single movements of the same upper limb as well as grasps can be classified from MRCPs in healthy persons [19, 27]. However, a classification of self-paced attempted movements in the ongoing EEG of a person with SCI is lacking. We therefore propose in this work an MRCPs-based asynchronous online classifier, and show a proof-of-concept (without FES) of detecting hand open and palmar grasp in a person with SCI.

Table 1: ISNCSCI motor scores of the right upper limb

motor key muscles		score
elbow flexors	C5	4
wrist extensors	C6	1
elbow extensors	C7	0
finger flexors	C8	0
finger abductors	T1	0

MATERIALS AND METHODS

Participant: We recruited a right-handed male participant of age 55 with a chronic cervical SCI. The SCI has been sustained 6 years ago with a neurological level of injury of C4 and AIS A classification, i.e. complete SCI. No hand function is preserved, see Table 1 for ISNCSCI motor scores. Written informed consent was obtained.

Paradigm: We measured two sessions with the participant. He sat in his wheelchair in front of a computer screen and carried out instructions given on the computer screen. A *training paradigm* and a *test paradigm* were used to evaluate the asynchronous online classifier.

The training paradigm comprised of two types of trials: movement and rest. Movement trials were used to record hand open and palmar grasp. A movement trial started with a beep and a class cue which indicated either hand open or palmar grasp, see Figure 1. At second 2, the ready cue appeared and replaced the class cue. The ready cue was a green circle with a smaller inner white circle. 0.5 s to 1 s after the appearance of the ready cue, the green circle started to shrink within 2 s to 4 s to the size of the inner white circle. We instructed the participant to attempt the movement when the outer circle hit the inner circle. We refer to this moment as go cue. In session 1, we instructed the participant to attempt to open or grasp, and deliberately hold the position until the end

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-35

of the trial, i.e. attempt a sustained movement. In session 2, we gave the instruction not to hold the position, but to make a short single movement attempt. Two seconds after the go cue, the trial ended and the screen turned black. During the movement trials, a cross was shown in the middle of the screen to fixate the gaze. Trials were spaced by a 2 s to 3 s time interval. In rest trials, a cross was shown for 70 s, and we instructed the participant to avoid any movement during this period. We recorded 5 movement runs and 4 rest runs. A movement run comprised of 30 movement trials, and a rest run comprised of 1 rest trial. In total we recorded 150 movement trials (75 trials per movement class) and 4 rest trials. We epoched then the 70 s long rest trials at random positions so that we had 150 rest trials.

The test paradigm is shown in Figure 2. The class cue (hand open, palmar grasp, rest), a beep, and a fixation cross were presented at the trial start. After 5 s, the class cue disappeared and only the fixation cross remained on the screen for 60 s. We instructed the participant to attempt several self-paced movements of the respective class during this 60 s period, or avoid any movement if it was a rest class. The participant was instructed to wait at least 5 s between movement attempts, and to report every movement attempt 2 s later by a soft speech sound. However, due to a misunderstanding, the participant reported movement attempts immediately afterwards in session 1 (but not session 2). When the participant reported a movement attempt, the experimenter pressed a button on the computer to log the time point of the movement event. The online classifier was constantly active and showed the respective movement icon (i.e. hand open or palmar grasp) for 2s whenever a movement attempt was detected. Thus, it was a closed-loop classification as feedback was provided. We recorded 6 runs in session 1 and 5 runs in session 2. Each run comprised of 4 movement trials and 1 rest trial.

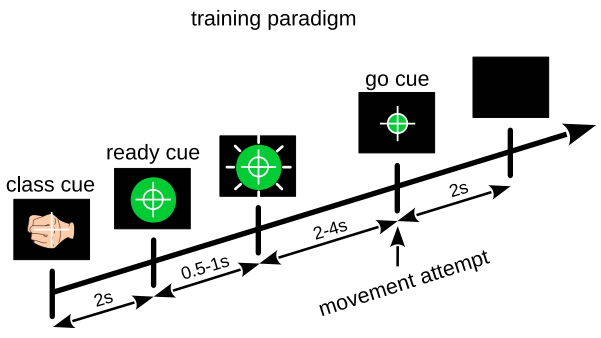


Figure 1: Training paradigm. A green filled circle shrunk at random speed. The participant attempted a hand open movement or palmar grasp when the green circle hit the inner white circle (the go cue).

Recording: We measured EEG with 61 electrodes covering frontal, central, parietal, and temporal areas. Reference was placed on the left earlobe and ground on AFF2h. Signals were sampled with 256 Hz with four 16-channel biosignal amplifiers and an active electrode system (g.tec medical engineering GmbH, Austria). A notch filter at 50 Hz and a band-pass filter with 0.01 Hz to 100 Hz (8th order Chebyshev filter) were used.

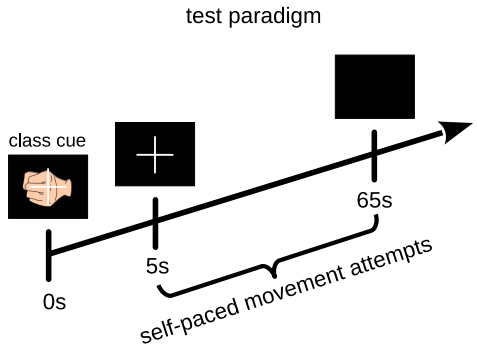


Figure 2: Test paradigm. The participant attempted several single self-paced movements.

Preprocessing: We excluded channel AFz, and re-referenced the remaining channels to a common average reference (CAR). Next, we filtered signals with a causal 4th order Butterworth filter with 0.3 Hz to 3 Hz to extract low-frequency signals. In the training paradigm, artifact contaminated trials were removed with a statistical outlier rejection method like in [19].

Classifier and Detector: We classified the EEG from the training paradigm with a multi-class shrinkage linear discriminant analysis (sLDA) [3]. The input features to the sLDA classifier were the preprocessed EEG data, which were extracted from a causal time-window of length 1.4 s (feature extraction window). To find the optimal training time point, we time-locked to the go cue and calculated offline classification accuracies with a 10-fold cross-validation. For this purpose, we shifted the right corner of the feature extraction window from 1 s to 2 s relative to the go cue, and calculated for each time point the classification accuracy. The time point with the highest classification accuracy was then found as t_{train} (t_{train} = 1.875s in training session 1, and t_{train} = 1.625s in training session 2). The output of the classifier was subjected to a softmax transformation to obtain class probabilities.

The classifier included two additional classes: a pre and a post class. MRCPs have a duration of more than 2 s, and the pre and post classes are supposed to detect the early and late phases of MRCPs (irrespective of hand open or palmar grasp). These MRCPs phases could otherwise increase the chance of detecting a wrong movement class if the MRCPs are not yet (or are no longer) fully covered by the feature extraction window during online operation. We grouped therefore the movement classes (hand open and palmar grasp) at $t_{train} - 500\,\text{ms}$ and $t_{train} + 500\,\text{ms}$ to pre and post classes, respectively. The final online classifier was then trained on hand open, palmar grasp and rest classes at t_{train}, and on the time shifted pre and post classes, yielding in fact a 5-class classifier. See Figure 3 for an explanation how the trial-averaged classifier output looks like when applied on the training paradigm and time-locked to the go cue. An attempted movement

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-35

should lead to a peak of the pre class probability, followed 500ms later by a peak of the hand open or palmar grasp class probability, and another 500ms later by a peak of the post class probability.
We build then a detector which detected attempted movements based on the time-sequences of the classifier outputs. This detector employed 3 time windows specified relative to t_0, see Figure 4. A movement was detected if the 3 peaks of the pre class, movement class, and post class fell within these 3 windows. The pre window was centered at $t_0 - 500$ ms, the movement window at t_0, and the post window at $t_0 + 500$ ms. Pre and post windows had a length of 300 ms, the movement window had a length of 100 ms. To detect a movement, the pre class probability had to be above 0.7 for at least half of the pre window (and analog for the post probability and window), and hand open class or palmar grasp class probabilities had to be above 0.9 for at least the full movement window. If all conditions were fulfilled, the movement class with the highest probability in the movement window (i.e. hand open or palmar grasp) was then eventually detected. A refractory period of 2 s followed every detection.

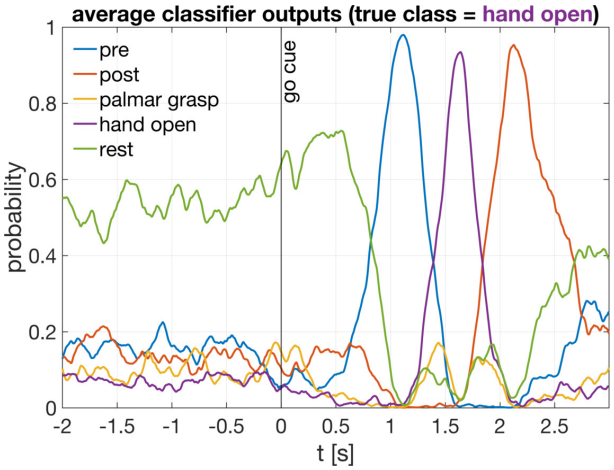


Figure 3: Trial-averaged output of the classifier for hand open trials in the training paradigm. The plot is time-locked to the go cue (0 s). One can see the peak of the hand open class, and preceding and subsequent peaks of the pre and post classes, respectively.

Detection delay: Several factors caused a delay between the onset of the movement attempt and the detection. As we do not know the true onset of the movement attempt we have to make some assumptions to estimate the delay. First, we assume that the onset of the movement attempt coincided with the go cue. Second, the average post class probability was maximal 500 ms after t_{train} and symmetric. There is a first delay between the onset of the movement attempt and the classifier training time point, found as t_{train}. This delay is due to the filter delay and the length of the feature extraction window. Furthermore, there is a second delay due to the detection logic based on pre and post classes. We can only provide a conservative estimation of this delay, i.e. the average maximum delay. We presume that the average post class

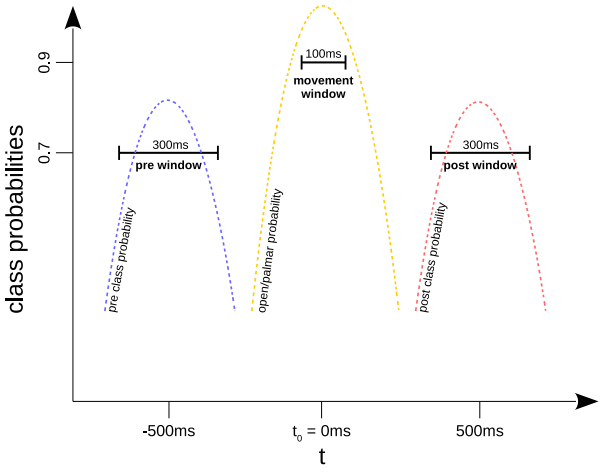


Figure 4: Illustration of the pre, movement and post windows of the detector. All windows had to be crossed for a certain amount of time by their respective probability to cause a movement detection.

probability crosses the post window probability-threshold (0.7) for exactly the length of the time-threshold (150ms, half the post window size). Shorter crossings do not cause a detection, and longer crossings cause an earlier detection. This second delay is the time from t_{train} to the point when the post class probability falls below the post window probability, which would be (under the second assumption) 500 ms + 75 ms after t_{train}. Thus, the maximum detection delay is then on average t_{train} + 500 ms + 75 ms.

Definition of the true-positive window:
We defined a true positive window to evaluate the online classifier. Every detected movement onset within the true positive window was counted as a true positive (TP), irrespective if it was a hand open or palmar grasp class. Every detection outside this window was a false positive (FP_{win}). The length of the TP window was set to 2 s which allowed for not more than one detection per window due to the refractory period of the detector. The center of the TP window was set with an offset to the logged movement onset, and accounted for the detection delay. Thus, the center should correspond to the assumed movement onset (i.e. the detection time point corrected by the detection delay). To find the offset between the TP window center and the logged movement onsets, it was necessary to compensate for the response time of the participant and the reaction time of the experimenter during reporting. As the average response and reaction times were unknown, we used a systematic approach to determine the offset. We iterated the offset from 0 s to 5 s and calculated the TP/FP_{win} ratio for each offset. The offset with the maximal TP/FP_{win} ratio was then determined as the final offset (the offset was 2.2 s in session 1, and 4.2 s in session 2).

RESULTS

We present our results separately for detection and clas-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-35

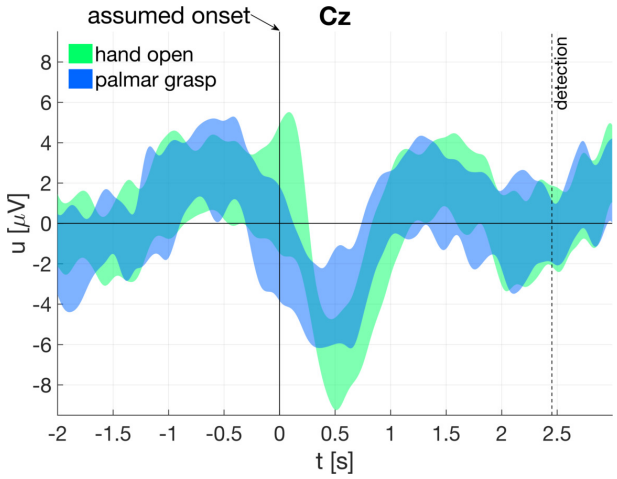


Figure 5: Electrode potentials of the test paradigm in session 1. Shown are the 95% confidence intervals at electrode Cz time locked to the assumed movement onset. The detection time point and the assumed onset are shown (i.e. the detection time point corrected by the detection delay).

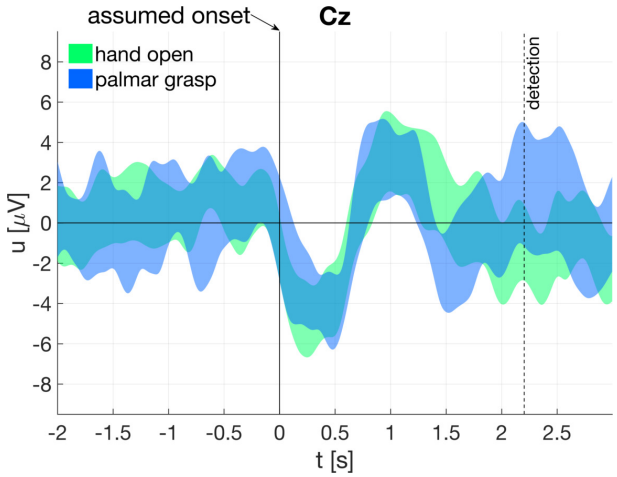


Figure 6: Electrode potentials of the test paradigm in session 2. Shown are the 95% confidence intervals at electrode Cz time locked to the assumed movement onset. The detection time point and the assumed onset are shown (i.e. the detection time point corrected by the detection delay).

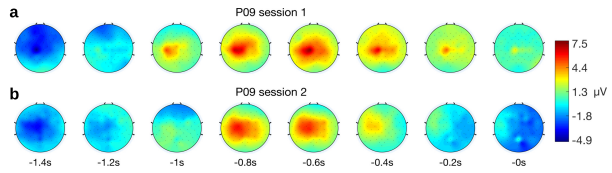


Figure 7: Test paradigm topoplots. The time lags are the input to the classifier when a movement was detected (the topoplots are time-locked to true positive movement detections, second 0 corresponds to t_{train}). **a:** Topoplots from session 1. **b:** Topoplots from session 2.

Table 2: Detection and classification results of both sessions. TPR and FP/min were calculated independent of the movement class. Accuracy was calculated on TPs.

sess.	TPR [%]	FP/min	acc. [%]	sig. level [%]
1	26.6	3.2	66.0	62.4
2	36.9	3.6	70.8	65.3

sification. Detection refers to the identification of any movement class in the ongoing EEG, regardless if it was hand open or palmar grasp. Here, the detection performance is quantified with true positive ratio (TPR) and false positives per minute (FP/min). TPs were all detections within the TP window, FPs were all detections during the 70 s rest trials (not to be confused with FP_{win}). Classification refers to the identification of the movement class itself and considers only TPs. Thus, detections outside the true positive window were ignored. The classification performance is quantified with the classification accuracy. The results of both sessions are shown in Table 2; the significance levels were determined with an adjusted wald interval [2, 17] with $\alpha = 0.05$. When averaging over both sessions, we obtained a TPR of 31.8 and 3.4 FP/min, with an accuracy of 68.4 %.

We also analyzed if the classification in the test paradigm is based on plausible brain signals. The electrode potentials on Cz are therefore shown in Figure 5 (session 1) and Figure 6 (session 2) for both movement classes. We considered only TP, and aligned to the assumed movement onset, i.e. the detection time point minus the detection delay. The detection delay was 2.5s in session 1 and 2.2s in session 2. The plots include 95% confidence intervals based on a t-distribution over trials. Both plots show around movement onset the characteristic negative deflection present in MRCPs. Noteworthy, the EEG signal in Figure 5 and Figure 6 are non-causally filtered to avoid phase distortion and ease the interpretation.

Furthermore we show in Figure 7 the topoplots of the trial-averaged time lags used as input to the classifier when a movement was detected (i.e. the feature extraction window). The topoplots indicate that the classifier used brain signals originating from lateral and central motor areas.

DISCUSSION

We show a first proof-of-concept on how to detect and classify attempted hand open and palmar grasp movements in a closed-loop. Noteworthy, the participant suffered from a complete SCI at level C4 and was not able to execute any movements with his hand.

The motivation of detecting movements online is to connect a BCI with a motor neuroprosthesis and restore movements in persons with SCI. We did another step towards that goal with an online detection in a closed-loop in a person with SCI. However, the performance is modest and at the moment not sufficient to control a motor neuroprosthesis in daily life activities. Most of all, the FP rate is not acceptable. Other movement detectors were shown to achieve better performances [10, 18]. Reasons

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-35

may be that those studies did not test participants with SCI, and furthermore relied on lower limb movements which produce larger amplitude MRCPs than upper limb movements [14]. The brain is adaptive and user training could therefore improve the performance, as seen in oscillation-based BCIs [8]. However, there is evidence that such a training is of limited benefit [11], and – if possible at all – novel training strategies need to be found.

We chose hand open and palmar grasp as movement classes because the participant was not able to execute these movements. We have shown in healthy participants that MRCPs encode other movements like pronation, supination, or different grasps [19, 27]. We expect therefore that the detection and classification is not limited to hand open and palmar grasp and generalizes also to other movements.

The externally-cued MRCPs elicited in the training paradigm should be similar to the MRCPs elicited in the self-paced test paradigm. Otherwise the classifier would haven been trained with improper data. We can expect that this is the case as MRCPs preceding self-paced and regular-cued movements are similar in shape and topography [5, 9]. Furthermore, we employed a training paradigm with no sudden appearance of any visual or auditory stimulus around the go cue. With that strategy we avoided any evoked potential related to low-level stimulus processing in the brain. However, we can not exclude that our training paradigm was contaminated with contingent negative variation (CNV) potentials [30]. CNVs can be elicited with a warning and an imperative (i.e. go) stimulus, which is also the basic structure of our training paradigm. CNVs comprise of two main components: one after the warning stimulus and one before the imperative stimulus. The latter component could form a potential complex with the Bereitschaftspotential (BP) [12]. However, with longer intervals between the warning and imperative stimulus (in the second range), the second component of the CNV becomes similar to the late BP in time course and scalp distribution [4, 23, 25], and therefore should not have strongly affected the classifier calibration.

We show that the source signals originate from the brain and are not residual movement artifacts for example. The topoplots indicate that the classifier mainly exploited a lateralized positivity for movement classification (preceded by a broad but centralized negativity). This positive deflection followed the typical negative peak of MRCPs [28] (c.f. Figure 5 and Figure 6).

The offset of the true positive window was chosen systematically to avoid any subjective influence on the classification accuracy. As the offset was chosen to maximize the TP/FP_{win} ratio of the movement trials, we may have made an overoptimistic estimation of the TPR (the FP measure is not affected as FP were counted on the rest trials). Nevertheless, the main result of this work is the classification of hand open vs palmar grasp in a closed-loop. A suboptimal choice of the true positive window would have led to a suboptimal estimation of the true classification accuracy, as the estimation would then consider more false positives and/or less true positives.

CONCLUSION

We introduce a proof-of-concept on how to detect and classify attempted hand opening and palmar grasp movements based on MRCPs. We tested our proof-of-concept on a person with a complete cervical SCI without any preserved hand function. We achieved a significant classification accuracy but also a high number of false positive detections. Thus, attempted hand movements can be non-invasively decoded from EEG, even if no hand function has been preserved. If the detection and classification performances can be improved, this may provide a control option for future neuroprostheses.

ACKNOWLEDGEMENTS

This work was partially supported by the European ICT Programme Project H2020-643955 "MoreGrasp" and the ERC Consolidator Grant "Feel Your Reach".

REFERENCES

[1] Ball Tonio, Schulze-Bonhage Andreas, Aertsen Ad, Mehring Carsten. Differential representation of arm movement direction in relation to cortical anatomy and function. Journal of Neural Engineering. 2009;6(1):016006.

[2] Billinger Martin et al. Is It Significant? Guidelines for Reporting BCI Performance. In: Towards Practical Brain-Computer Interfaces, 2012, 333–354.

[3] Blankertz Benjamin, Lemm Steven, Treder Matthias, Haufe Stefan, Müller Klaus-Robert. Single-trial analysis and classification of ERP components–a tutorial. NeuroImage. 2011;56:814–825.

[4] Cui R. Q., Egkher A., Huter D., Lang W., Lindinger G., Deecke L. High resolution spatiotemporal analysis of the contingent negative variation in simple or complex motor tasks and a non-motor task. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology. 2000;111:1847–1859.

[5] Cui Rongqing, MacKinnon Colum D. The effect of temporal accuracy constraints on movement-related potentials. Experimental Brain Research. 2009;194(3):477–488.

[6] Edelman Bradley J., Baxter Bryan, He Bin. EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. IEEE transactions on bio-medical engineering. 2016;63:4–14.

[7] Hammon Paul, Makeig Scott, Poizner Howard, Todorov Emanuel, Sa Virginia De. Predicting Reaching Targets from Human EEG. IEEE Signal Processing Magazine. 2008;25(1):69–77.

[8] He B., Baxter B., Edelman B. J., Cline C. C., Ye W. W. Noninvasive Brain-Computer Interfaces Based

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-35

on Sensorimotor Rhythms. Proceedings of the IEEE. 2015;103(6):907–925.
[9] Jankelowitz S., Colebatch J. Movement-related potentials associated with self-paced, cued and imagined arm movements. Experimental Brain Research. 2002;147(1):98–107.
[10] Jochumsen Mads, Niazi Imran Khan, Mrachacz-Kersting Natalie, Farina Dario, Dremstrup Kim. Detection and classification of movement-related cortical potentials associated with task force and speed. Journal of Neural Engineering. 2013;10(5):056015.
[11] Jochumsen Mads, Niazi Imran Khan, Nedergaard Rasmus Wiberg, Navid Muhammad Samran, Dremstrup Kim. Effect of subject training on a movement-related cortical potential-based brain-computer interface. Biomedical Signal Processing and Control. 2018;41:63–68.
[12] Kornhuber Hans Helmut, Deecke Lüder. Hirnpotentialanderungen beim Menschen vor und nach Willkurbewegungen dargestellt mit Magnetbandspeicherung und Ruckwartsanalyse. In: Pflugers Archiv-European Journal of Physiology. Vol. 281. 1. SPRINGER VERLAG 175 FIFTH AVE, NEW YORK, NY 10010. 1964, 52.
[13] Lew Eileen Y. L., Chavarriaga Ricardo, Silvoni Stefano, R. Millán José del. Single trial prediction of self-paced reaching directions from EEG signals. Frontiers in Neuroscience. 2014;8.
[14] Martínez-Expósito A., Ibáñez J., Resquín F., Pons J. L. Task Influence on Motor-Related Cortical Signals: Comparison Between Upper and Lower Limb Coordinated and Analytic Movements. In: Converging Clinical and Engineering Research on Neurorehabilitation II, 2017, vol. 15, 1139–1143.
[15] Millán J. D. R. et al. Combining Brain-Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges. Frontiers in neuroscience. 2010;4.
[16] Müller-Putz Gernot R., Scherer Reinhold, Pfurtscheller Gert, Rupp Rüdiger. EEG-based neuroprosthesis control: a step towards clinical practice. Neuroscience letters. 2005;382:169–174.
[17] Müller-Putz Gernot, Scherer Reinhold, Brunner Clemens, Leeb Robert, Pfurtscheller Gert. Better than random: a closer look on BCI results. International Journal of Bioelectromagnetism. 2008;10(1):52–55.
[18] Niazi Imran Khan, Jiang Ning, Tiberghien Olivier, Nielsen Jørgen Feldbæk, Dremstrup Kim, Farina Dario. Detection of movement intention from single-trial movement-related cortical potentials. Journal of Neural Engineering. 2011;8(6):066009.
[19] Ofner Patrick, Schwarz Andreas, Pereira Joana, Müller-Putz Gernot R. Upper limb movements can be decoded from the time-domain of low-frequency EEG. PLOS ONE. 2017;12(8):e0182578.
[20] Pfurtscheller G., Guger C., Müller G., Krausz G., Neuper C. Brain oscillations control hand orthosis in a tetraplegic. Neuroscience letters. 2000;292:211–214.
[21] Pfurtscheller G., Lopes da Silva F. H. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology. 1999;110:1842–1857.
[22] Pfurtscheller Gert, Müller Gernot R., Pfurtscheller Jörg, Gerner Hans Jürgen, Rupp Rüdiger. 'Thought'–control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. Neuroscience letters. 2003;351:33–36.
[23] Prescott J. The effects of response parameters on CNV amplitude. Biological psychology. 1986;22:107–135.
[24] Rohm Martin et al. Hybrid brain–computer interfaces and hybrid neuroprostheses for restoration of upper limb functions in individuals with high-level spinal cord injury. Artificial Intelligence in Medicine. 2013;59(2):133–142.
[25] Rohrbaugh J. W., Syndulko K., Lindsley D. B. Brain wave components of the contingent negative variation in humans. Science (New York, N.Y.) 1976;191:1055–1057.
[26] Rupp R., Rohm M., Schneiders M., Kreilinger A., Müller-Putz G. R. Functional Rehabilitation of the Paralyzed Upper Extremity After Spinal Cord Injury by Noninvasive Hybrid Neuroprostheses. Proceedings of the IEEE. 2015;103(6):954–968.
[27] Schwarz Andreas, Ofner Patrick, Pereira Joana, Sburlea Andreea Ioana, Müller-Putz Gernot R. Decoding natural reach-and-grasp actions from human EEG. Journal of Neural Engineering. 2017;15(1):016005.
[28] Shibasaki Hiroshi, Hallett Mark. What is the Bereitschaftspotential? Clinical Neurophysiology. 2006;117(11):2341–2356.
[29] Úbeda Andrés, Azorı´ n José M., Chavarriaga Ricardo, R. Millán José del. Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques. Journal of NeuroEngineering and Rehabilitation. 2017;14(1).
[30] Walter W. G., Cooper R., Aldridge V. J., Mccallum W. C., Winter A. L. Contingent Negative Variation: An Electric Sign of Sensorimotor Association and Expectancy in the Human Brain. Nature. 1964;203:380–384.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-36

WOULD MOTOR-IMAGERY BASED BCI USER TRAINING BENEFIT FROM MORE WOMEN EXPERIMENTERS?

A. Roc[1, 2], L. Pillette[1, 2], B. N'Kaoua[3], F. Lotte[1, 2]

[1]Inria Bordeaux Sud-Ouest, Talence, France
[2]LaBRI (University of Bordeaux / CNRS / Bordeaux-INP), Talence, France
[3]Handicap, Activity, Cognition, Health, University of Bordeaux, Bordeaux, France

E-mail: aline.roc@inria.fr, lea.pillette@inria.fr

ABSTRACT: Mental Imagery based Brain-Computer Interfaces (MI-BCI) are a mean to control digital technologies by performing MI tasks alone. Throughout MI-BCI use, human supervision (e.g., experimenter or caregiver) plays a central role. While providing emotional and social feedback, people present BCIs to users and ensure smooth users' progress with BCI use. Though, very little is known about the influence experimenters might have on the results obtained. Such influence is to be expected as social and emotional feedback were shown to influence MI-BCI performances. Furthermore, literature from different fields showed an experimenter effect, and specifically of their gender, on experimental outcome. We assessed the impact of the interaction between experimenter and participant gender on MI-BCI performances and progress throughout a session. Our results revealed an interaction between participants gender, experimenter gender and progress over runs. It seems to suggest that women experimenters may positively influence participants' progress compared to men experimenters.

INTRODUCTION

Brain Computer Interfaces (BCI) enable their users to interact with technologies by extrapolating their intentions from their brain activity, often measured using an electroencephalogram (EEG). In this article, we focus on Motor-Imagery based BCI (MI-BCI), for which users express their intentions by performing Motor Imagery (MI) tasks, such as imagining hands or feet movements, to induce change in their EEG signals and thereby control the BCI. MI-BCI represent promising new technologies. For example, they have proven effective for motor rehabilitation post-stroke and to interact with a variety of automated system, such as orthoses or video games [6].
Though, currently, MI-BCI do not enable a sufficient accuracy in detecting which task is performed by users [17]. Indeed, on average when trying to differentiate one MI task between two from EEG signals, a task is correctly recognized 75% of the time [1]. Users need to train to acquire MI-BCI skills. Yet, around 10 to 30% of users are unable to control MI-BCI [1]. Therefore, MI-BCI are still mainly at the development stage in research laboratories. During MI-BCI experimental protocols, experimenters play a key role [27]. For instance, they introduce the technology to the participants, provide the participants with advice regarding how they should perform the MI tasks and keep the participants motivated throughout the training. Nonetheless, social and emotional feedback were shown to have an impact on user experience, motivation and MI-BCI performances [23]. Despite the main role that experimenters have in the experimental process and the literature regarding the impact of social and emotional feedback, no studies had yet been led in MI-BCI to evaluate the influence experimenters might have on their own experimental results.

Experimenter related biases are an important concern in other fields such as ethics and business [19], social research [26] or economic research [32]. Literature from different fields states that the characteristics of the experimenter may consciously or unconsciously affect the responses, behavior and performance of the participants via direct and/or indirect interactions [26]. For example, it has been shown that an "experimenter demand effect" occurs when participants unconsciously try to fit the appropriate image reflected by the experimenter's behavior and therefore want to please and assist the experimenters in obtaining their expected results [26].

Several studies investigating experimenters' influence suggested that gender-interaction could have an impact. For example, the interplay of participant's and experimenter's genders may shape the experimenter demand effect. When participants are instructed by an opposite-sex experimenter, they seem more likely to act in ways that confirm the experimenter's hypothesis [21]. Also, men participants seem to elaborate more on autobiographical memory report with women experimenters than with men experimenters and more than women participants in general [8]. Proxemics studies, which study the amount of space that people feel necessary to set between themselves and others, provide another example of gender interaction. Men participants seem to keep a shorter distance from women than from men [29]. Interestingly, participants also prefer a larger comfort and reachability distance when facing a virtual man as compared to a virtual woman [9]. In a pain-related study, it was shown that men participants tend to report higher cold presser

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-36

pain to a man experimenter than to a woman one [14]. Studies revealed that there were no interactions in the physiological data between experimenter gender and participant gender, suggesting that men participants reporting lower pain report to women experimenters is probably due to psychosocial factors [2]. Another gender-related example would be that defensiveness is associated with greater relative left frontal activation in the presence of experimenters from the opposite-sex compared to experimenters from the same-sex [12]. Thus, participants who work with an opposite-gender or same-gender experimenter can have different neurological responses, such as differences in their EEG recordings [5]. In a recent study examining the effect of psychosocial factors —particularly sex-related effects— on a given neurofeedback training (learning to modulate sensorimotor rhythm power and theta/beta power), women participants trained by women experimenters learned significantly less than their counterparts trained by men experimenters [31].

These observations have led us to think that a gender-interaction could have an effect on MI-BCI experimental results. Yet, it has never been tested in BCI. The aim of our study was therefore to investigate if there was an influence of the experimenters' gender depending on the participants' gender on MI-BCI performances and progression (i.e., the evolution of performances).

MATERIALS & METHODS

Participants:

Fifty-nine healthy MI-BCI naïve participants (29 women; age 19-59; $\bar{X}$=29; SD=9.318) completed the study. None of them reported a history of neurological or psychiatric disorder. Experimenters who conducted the study were six scientists (3 women; age 23-37; $\bar{X}$=29.2; SD=5.60) among which two experienced in BCI experimentation (1 woman) and four beginners who were trained to perform a BCI experiment beforehand. Each experimenter was randomly assigned to 10 participants (5 women and 5 men) they had never met before the session.

Our study was conducted in accordance with the relevant guidelines for ethical research according to the Declaration of Helsinki. Both participants and experimenters gave informed consent before participating in the study. In order to avoid biased behavior, this study was conducted using a deception strategy, partially masking the purpose of the study. Participants were told that the study aimed at understanding which factors (unspecified) could influence BCI progress and/or performance. The study has been reviewed and approved by Inria's ethics committee, the COERLE.

Experimental protocol:

Each participant participated in one MI-BCI 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) six 7-minute runs during which participants had to learn to perform two MI-tasks, i.e., imagine right or left hand movements, (around 60 min, including breaks between the runs), (4) completion of post-session questionnaires (around 5 min) and (5) uninstallation and debriefing (around 10 min).

During each run, participants had to perform 40 trials (20 per MI-task, presented in a random order), each trial lasting 8s. At t = 0s, an arrow was displayed on the screen. At t = 2s, an acoustic signal announced the appearance of a red arrow, which appeared one second later (at t = 3s) and remained displayed for 1.250s.The arrow pointed in the direction of the task to be performed, namely left or right to imagine a movement of the left hand or the right hand. Finally, at t = 4.250s, a visual feedback was provided in the shape of a blue bar, the length of which varied according to the classifier output. Only positive feedback was displayed, i.e., the feedback was provided only when the instruction matched the recognized task. The feedback lasted 3.75 s and was updated at 16Hz, using a 1s sliding window. After 8 seconds of testing, the screen turned black again. The participant could then rest for a few seconds, and a new cross was then displayed on the screen, marking the beginning of the next trial.

The training protocol used was the Graz protocol [22] which is divided into two steps: (1) training of the system and (2) training of the user. The first two runs were used as calibration in order to provide examples of EEG patterns associated with each of the MI tasks to the system. During the first two runs, as the classifier was not yet trained to recognize the mental tasks being performed by the user, it could not provide a consistent feedback. In order to limit biases with the other runs, e.g., EEG changes due to different visual processing between runs, the user was provided with an equivalent sham feedback, i.e., a blue bar randomly appearing and varying in length.

We respected the following recommendations: encourage the user to perform a kinesthetic imagination [20] and leave users free to choose their mental imagery strategy [13], e.g., imagining waving at someone or playing the piano. Participants were instructed to find a strategy for each task so that the system would display the longest possible feedback bar. Instructions were written in advance so that all the participants started with the same standardized information.

Questionnaires:

We assessed personality and cognitive profile for both experimenters and participants with the 5th edition of the 16 Personality Factors (16PF5) [4], a validated psychometric questionnaire to assess different aspects of personality and cognitive profile. This questionnaire identifies 16 primary factors of personality, such as anxiety or autonomy. Participants also completed a mental rotation test measuring spatial abilities [30].

EEG Recordings & Signal Processing:

To record the EEG signals, 27 active scalp electrodes, referenced to the left ear, were used (Fz, FCz, Cz, CPz, Pz, C1, C3, C5, C2, C4, C6, F4, FC2, FC4, FC6, CP2, CP4, CP6, P4, F3, FC1, FC3, FC5, CP1, CP3, CP5, P3, 10-20 system). Electromyographic (EMG) activity of the hands

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-36

were recorded using two active electrodes situated 2.5cm below the skinfold on each wrists. Electrooculographic (EOG) activity of one eye was recorded using three active electrodes. Two, situated below and above the eye, aimed at recording vertical movements of the eye and one on the side aimed at recording horizontal movements. Physiological signals were measured using a g.USBAmp (g.tec, Austria), sampled at 256 Hz, and processed online using OpenviBE 2.1.0 [25].

To classify the two MI tasks from EEG data, we used participant-specific spectral and spatial filters. First, from the EEG signals recorded during the calibration runs, we identified a participant-specific discriminant frequency band using the heuristic algorithm proposed by Blankertz et al. in [3] (Algorithm 1 in that paper). Roughly, this algorithm selects the frequency band whose power in the sensorimotor channels maximally correlates with the class labels. Here we used channels C3 & C4 after spatial filtering with a Laplacian filter as sensorimotor channels, as recommended in [**blankertz08**]. We selected a discriminant frequency band in the interval from 5 Hz to 35 Hz, with 0.5Hz large bins. Once this discriminant frequency band identified, we filtered EEG signals in that band using a butterworth filter of order 5.

Then, we used the Common Spatial Pattern (CSP) algorithm [24] to optimize 3 pairs of spatial filters, still using the data from the two calibration runs. Such spatially filtered EEG signals should thus have a band power which is maximally different between the two MI conditions. We then computed the band power of these spatially filtered signals by squaring the EEG signals, averaging them over a 1 second sliding window (with 1/16th second between consecutive windows), and log-transforming the results. This led to 6 different features per time window, which were used as input to a Linear Discriminant Analysis (LDA) classifier [18]. As mentioned above, this LDA was calibrated on the data from the two calibration runs. These filters and classifier were then applied on the subsequent runs to provide online feedback.

Variables & Factors:

The aim was to evaluate the influence of the gender of the experimenters and participants on the MI-BCI performances of the participants over a series of 4 runs with online BCI use. Two measures were used to assess the performance of the participants.

The first is the mean classification accuracy which is traditionally used by the community. This measure represents the percentage of time windows from the feedback periods that were correctly classified. However, this metric only considers whether the classification was correct, but not the quality of this classification, i.e., it does not take into account the classifier output. Since our participants were instructed to train to obtain not only a correct classification, but also a feedback bar as long as possible, we also studied a metric considering the feedback bar length, i.e., the classifier output.

Thus, we also used the Quality-Weighted Accuracy (QWA), the standard performance metric provided in OpenViBE MI-BCI applications, which is inspired by the SensoriMotor Rhythm quality score in [7]. To compute it, we first summed the (signed) LDA classifier outputs (distance to the separating hyperplane) over all time windows during a trial feedback period. If this sum sign matched the required trial label, i.e. negative for left hand MI and positive for right hand MI, then the trial was considered as correctly classified, otherwise it was not. Finally, a run QWA was estimated as the percentage of trials considered as correctly classified using this approach.

RESULTS

Comparability of groups:

Among 59 participants, 3 outperformed the others (by more than two SDs) both in term of mean classification accuracy (Outliers $\bar{X}1$=88.94, $\bar{X}2$=90.36, $\bar{X}3$=94.51; $\bar{X}$grp=59.33%; SDgrp=12.3) and QWA (Respectively, outliers $\bar{X}1$=98.13, $\bar{X}2$=98.13, $\bar{X}3$=99.38; $\bar{X}$grp=62.78%; SDgrp=16.2). Thus, the following analyzes are based on the results of 56 participants (27 women).

Before it all, we verified if the distribution of the data collected was normal using Shapiro-Wilk tests. The variables describing the mental rotation scores (p=0.34), anxiety (p=0.06) and autonomy (p=0.14) of our participants could be considered as having a normal distribution. Though, the mean classification accuracy of the runs did not have a normal distribution ($p \leq 10^{-3}$) and neither did the QWA metrics for the different runs ($p \leq 10^{-3}$).

We also checked that groups formed by participants' gender, i.e., *"ParGender"*, and experimenters' gender, i.e., *"ExpGender"*, were comparable. We focused on mental rotation scores (MRS), anxiety or autonomy, which were shown to influence on MI-BCI performances [10]. Participants with low MRS [30], anxious or non-autonomous (both measured using the 16PF5 questionnaire [4]) were shown to have lower MI-BCI performances than the others [10, 28]. To check that groups were comparable, we ran 2-way ANOVAs with *"ExpGender*ParGender"* as independent variables and either mental rotation scores, anxiety or autonomy as dependent variable.

Results indicate that groups are comparable in terms of anxiety. Though, participants' gender influence their MRS [F(1,52)=17.47; $p \leq 10^{-3}$, η^2=0.25]. Men ($\bar{X}$men=0.072; SD=0.024) had higher MRS than women ($\bar{X}$women=0.045; SD=0.023), which is in accordance with the literature [15]. Furthermore, participants training with men or women experimenters did not have the same level of autonomy [F(1,52)=4.01; p=0.05, η^2=0.07]. Participants training with men experimenters ($\bar{X}$menExp=6.35; SD=1.74) were more autonomous than participants training with women experimenters ($\bar{X}$womenExp=5.67; SD=1.66). Therefore, we controlled for the influence of these variables in our subsequent analyses by using them as covariates in ANCOVAs (see paragraph *Checking for confounding factors*).

Participants' and experimenters' gender:

Then, we analyzed the influence of the gender of the

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-36

experimenters and participants on the MI-BCI performances of the participants over the runs, i.e., *"Run"*. To do so, we performed a 3-way repeated measures mixed ANOVAs with *"ExpGender*ParGender*Run"* as independent variables and the repeated measures of performance over the runs, i.e., mean classification accuracy or QWA, as dependent variable. Even though the normality of the data is a pre-requisite of an ANOVA, the ANOVA is considered as robust against the normality assumption and to the best of our knowledge no other non parametric test enabled to perform such analysis.

First, we performed such ANOVA using the mean classification accuracy. Results revealed no simple effect of *"Run"* [F(3,156)=1.53; p=0.22, η^2=0.03], *"ExpGender"* [F(1,52)=0.26; p=0.61, $\eta^2 \leq$0.01] and *"ParGender"* [F(1,52)=0.23; p=0.64, $\eta^2 \leq$0.01]. They revealed no interaction of *"Run*ParGender"* [F(3,156)=1.92; p=0.13, η^2=0.04], *"Run*ExpGender"* [F(3,156)=0.23; p=0.87, η^2=0.01] nor *"ParGender*ExpGender"* [F(1,52)=0.92; p=0.34, η^2=0.02]. Finally, the interaction of *"Run*ParGender*ExpGender"* was not significant either [F(3,156)=1.38; p=0.25, η^2=0.03].

Next, we performed this same analysis using QWA. Results revealed no simple effect of *"Run"* [F(3,16)=1.81; p=0.15, η^2=0.03], *"ExpGender"* [F(1,52)=0.54; p=0.47, η^2=0.01] nor *"ParGender"* [F(1,52)=0.09; p=0.76, η^2=0.01]. They also revealed no interaction of *"Run*ExpGender"* [F(3,16)=0.08; p=0.97, η^2=10^{-2}] nor *"ParGender*ExpGender"* [F(1,52)=0.60; p=0.44, η^2=0.01]. Though, the *"Run*ParGender"* interaction was significant [F(3,156)=5.98; p=0.001, η^2=0.1]. Figure 1 represents the evolution of the participants' QWA depending on their gender.

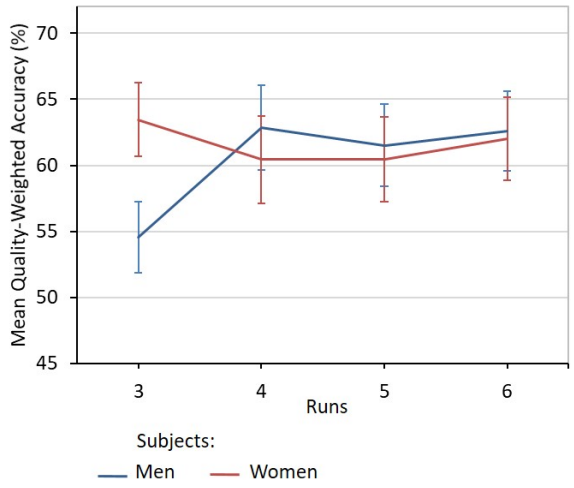


Figure 1: QWA evolution depending on participants' gender with standard errors.

Finally, a significant *"Run*ParGender*ExpGender"* interaction was found [F(3,156)=3.46; p=0.02, η^2=0.06]. Figure 2 represents the participants' QWA evolution depending on the participants' and experimenters' gender.

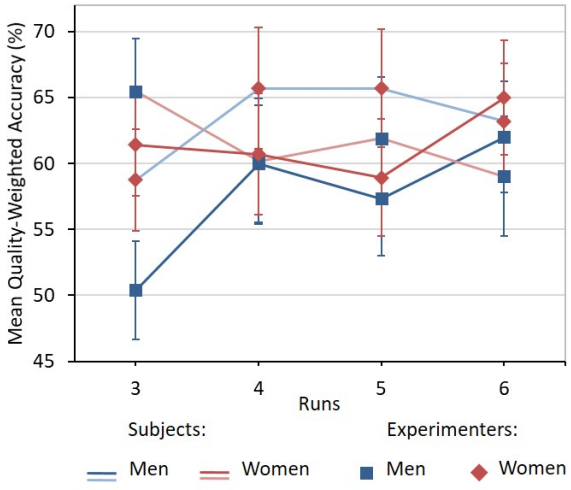


Figure 2: Participants' QWA evolution depending on the participants' and experimenters' gender with standard errors.

Checking for confounding factors:

As stated before, the groups of participants formed using the participants' and experimenters' gender had differences in terms of mental rotation scores and autonomy. Therefore, we studied the potential impact these differences could have had on our results. First, we checked if a correlation could be found between our metrics of performances and these variables. No significant correlation was found between the autonomy and the mean classification accuracy ($r = -0.11$, $p = 0.40$) nor the QWA ($r = -0.07$, $p = 0.62$). The correlations between the mental rotation score and the mean classification accuracy ($r = -0.13$, $p = 0.36$) or the QWA ($r = -0.24$, $p = 0.08$) was not significant either.

Second, we ran a 3-way repeated measures mixed ANCOVA with *"ExpGender, ParGender, Run"* as independent variables and one of the measures of performance, i.e., mean classification accuracy or QWA, as dependent variable, with the autonomy, i.e., *"Aut"*, or the mental rotation score, i.e., *"MRs"*, of the participants as covariate. When performing the analysis on the QWA we did not find any single effect or interaction of the autonomy (*"Aut"* [F(1,51)=0.26; p=0.61, η^2=10^{-2}], *"Aut*Run"* [F(3,15)=0.81; p=0.49, η^2=0.02]) or the mental rotation score (*"MRs"* [F(1,51)=1.75; p=0.19, η^2=0.03], *"MRs*Run"* [F(3,15)=1.52; p=0.21, η^2=0.03]).

When investigating the mean classification accuracy we found as well no impact of the autonomy (*"Aut"* [F(1,51)=0.44; p=0.51, η^2=10^{-2}], *"Aut*Run"* [F(3,15)=1.46; p=0.23, η^2=0.03]) or the mental rotation score (*"MRs"* [F(1,51)=1.05; p=0.31, η^2=0.02], *"MRs*Run"* [F(3,15)=1.35; p=0.26, η^2=0.03]).

DISCUSSION

We analyzed results using two metrics of performances: QWA which represented what the participants were instructed to improve during training, and the mean classification accuracy, a traditional measure of BCI perfor-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-36

mances. Initial differences in mental rotation scores and autonomy between groups did not seem to bias results.

No influence of the experimenters' and/or participants' gender on the mean accuracy performance was found. Though, we found a significantly different evolution across runs of QWA between men and women participants (see Figure 1). Women participants seemed to start the training with already good QWA, which decreased on the second run and increased again during the last run. Men participants however, started with rather low QWA and then drastically improved on the second run and then stagnated to reach slightly higher final QWA performances than women.

In addition, experimenters' gender seemed to have an influence on this previous interaction. Indeed, the evolution of the QWA appear to depend on participants' and experimenters' gender (see Figure 2). On the one hand, we found the same tendency for men participants to start with lower QWA at the beginning of the session independently of the experimenter's gender. However, men seemed to start with drastically lower QWA performances when they were training with men experimenters. They also seemed to have higher QWA performances throughout the session when they were training with women experimenters. On the other hand, women participants seemed to start with higher QWA when training with men experimenters, though their QWA performances tended to drop throughout the session. However, when training with women experimenters they seemed to have a great increase in QWA during the last run.

Interestingly enough, this result does not match those of a recently published neurofeedback study [31], in which the combination woman experimenter–woman participant appeared to hamper the training outcomes of the last, so that no learning effect was observed in this group.

CONCLUSION

We investigated the potential influence of the experimenters' gender depending on the participants' gender on MI-BCI performances and progression throughout one MI-BCI session. Six experimenters (3 men; 3 women) trained 59 participants (30 men; 29 women). The general observation emerging from this study is that women experimenters seemed to induce better QWA performance progress for both men and women participants. Men participants seemed to start with substantially lower performances when they were training with men experimenters compared to when they were training with women experimenters. Also, even though women participants started with higher performances when training with men experimenters, their performances decreased throughout the session when they overall increased when training with women experimenters.

These results naturally need to be confirmed with larger populations. Further analysis are also needed regarding other variables that might influence or provide insights on our results. This includes inter-experimenter variables (e.g., traits or teaching competence), intra-experimenters variables (e.g., appearance or states), inter and intra-participants variables (e.g., traits or motivation) and interaction related variables (e.g., quantity and quality of interaction between the participant and the experimenter). There might also be other analysis to perform based on different performance metrics reflecting user performances independently of the classifier output [16].

Further formal studies investigating the role of the BCI experimenter are needed. The need for research methods that explicit larger amounts of influencing factors (such as the experimenter) emerging from experimental protocol and context is equally important. In BCI research, the instructions (i.e., what participants are instructed to do during mental-imagery tasks) are rarely formalized, or in any case they are not taken into account and mentioned in papers. Similarly, protocols rarely evoke demonstrations [17] (i.e. showing a demonstration of a successful BCI use to the participant, together with a demonstration of feedback during (in)correctly performed mental tasks). It is common practice for studies in the BCI field not to report experimenter gender, though the literature as well as our results indicate that the influence of experimenters should be considered carefully while designing and reporting experimental protocols.

Literature suggests several solutions to limit the potential bias arising from the experimenter [19, 26]. These methods include: monitoring participant-experimenter interaction; increasing the number and diversity of data collectors; pre-testing the method and controlling expectancy; providing an extensive training for administrators/ data collectors; monitoring and standardizing the behavior of experimenters with detailed protocol and prewritten instructions for the participant; and statistically controlling for bias.

Beyond the potential bias that could arise from the experimenters' presence, the social and emotional feedback that experimenters provide could benefit MI-BCI. Indeed, the use of social feedback in BCI has been encouraged [27]. Social presence and trust relationship between the user and the experimenter are essential for maintaining training motivation, which has been shown to facilitate the BCI learning process [11].

During MI-BCI training, using a learning companion to provide the participants with social and emotional feedback have proven effective in improving the user experience [23]. An advanced conversational agent could also be used to supplement the role of the experimenter. It would represent yet another interesting method to control and/or enhance the experimenter influence. Taking experimenter-related factors into account might lead to a conjoint progress of the global BCI performance and the validity and understanding of BCI experimental results.

ACKNOWLEDGEMENTS

This work was supported by the European Research Council (grant ERC-2016-STG-714567) and the French

Published by Verlag der TU Graz
Graz University of Technology

National Research Agency (grant ANR-15-CE23-0013-01). We express our gratitude to all the experimenters and participants who took part in this study.

REFERENCES

[1] Allison BZ, Neuper C. Could anyone use a BCI? In: Brain-computer interfaces, 2010, 35–54.
[2] Aslaksen PM, Myrbakk IN, Høifødt RS, Flaten MA. The effect of experimenter gender on autonomic and subjective responses to pain stimuli. Pain. 2007;129(3):260–268.
[3] 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.
[4] Cattell RB, P. Cattell HE. Personality structure and the new fifth edition of the 16PF. Educational and Psychological Measurement. 1995;55(6):926–937.
[5] Chapman CD, Benedict C, Schiöth HB. Experimenter gender and replicability in science. Science advances. 2018;4(1):e1701427.
[6] Clerc M, Bougrain L, Lotte F. Brain-Computer Interfaces 2: Technology and Applications. John Wiley & Sons, 2016.
[7] Grosse-Wentrup M, Schölkopf B. High gamma-power predicts performance in sensorimotor-rhythm brain–computer interfaces. Journal of neural engineering. 2012;9(4):046001.
[8] Grysman A, Denney A. Gender, experimenter gender and medium of report influence the content of autobiographical memory report. Memory. 2017;25(1):132–145.
[9] Iachini T, Coello Y, Frassinetti F, Senese VP, Galante F, Ruggiero G. Peripersonal and interpersonal space in virtual and real environments: Effects of gender and age. Journal of Environmental Psychology. 2016;45:154–164.
[10] Jeunet C, N'Kaoua B, Subramanian S, Hachet M, Lotte F. Predicting Mental Imagery-Based BCI Performance from Personality, Cognitive Profile and Neurophysiological Patterns. PLoS ONE. 2015;10(12):20.
[11] Kleih SC et al. Motivation influences performance in SMR-BCI. na, 2011.
[12] Kline JP, Blackhart GC, Joiner TE. Sex, lie scales, and electrode caps: An interpersonal context for defensiveness and anterior electroencephalographic asymmetry. Personality and Individual Differences. 2002;33(3):459–478.
[13] Kober SE, Witte M, Ninaus M, Neuper C, Wood G. Learning to modulate one's own brain activity: the effect of spontaneous mental strategies. Frontiers in human neuroscience. 2013;7:695.
[14] Levine FM, De Simone LL. The effects of experimenter gender on pain report in male and female subjects. Pain. 1991;44(1):69–72.
[15] Linn MC, Petersen AC. Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child development. 1985;1479–1498.
[16] Lotte F, Jeunet C. Defining and quantifying users' mental imagery-based BCI skills: a first step. Journal of neural engineering. 2018;15(4):046030.
[17] Lotte F, Larrue F, Mühl C. Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design. Frontiers in Human Neuroscience. 2013;7:568.
[18] 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.
[19] Miyazaki AD, Taylor KA. Researcher interaction biases and business ethics research: Respondent reactions to researcher characteristics. Journal of Business Ethics. 2008;81(4):779–795.
[20] 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.
[21] Nichols AL, Maner JK. The good-subject effect: Investigating participant demand characteristics. The Journal of general psychology. 2008;135(2):151–166.
[22] Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. Proceedings of the IEEE. 2001;89(7):1123–1134.
[23] Pillette L, Jeunet C, Mansencal B, N'Kambou R, N'Kaoua B, Lotte F. Peanut: Personalised emotional agent for neurotechnology user-training. In: 7th International BCI Conference. 2017.
[24] Ramoser H, Muller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE transactions on rehabilitation engineering. 2000;8(4):441–446.
[25] 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.
[26] Rosnow R, Rosenthal R. People studying people: Artifacts and ethics in behavioral research. WH Freeman, 1997.
[27] Sexton CA. The overlooked potential for social factors to improve effectiveness of brain-computer interfaces. Frontiers in systems neuroscience. 2015;9:70.
[28] Tan LF, Dienes Z, Jansari A, Goh SY. Effect of mindfulness meditation on brain–computer interface performance. Consciousness and cognition. 2014;23:12–21.
[29] Uzzell D, Horne N. The influence of biological sex, sexuality and gender role on interpersonal distance. British Journal of Social Psychology. 2006;45(3):579–597.
[30] Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Perceptual and motor skills. 1978;47(2):599–604.
[31] Wood G, Kober SE. EEG Neurofeedback Is Under Strong Control of Psychosocial Factors. Applied psychophysiology and biofeedback. 2018;43(4):293–300.
[32] Zizzo DJ. Experimenter demand effects in economic experiments. Experimental Economics. 2010;13(1):75–98.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-37

DESIGN AND PRELIMINARY STUDY OF A NEUROFEEDBACK PROTOCOL TO SELF-REGULATE AN EEG MARKER OF DROWSINESS

Thibaut Monseigne[1,3], Fabien Lotte[1,2], Stephanie Bioulac[3,4], Pierre Philip[3,4], Jean-Arthur Micoulaud-Franchi[3,4]

[1]Inria Bordeaux - Sud-Ouest, Talence, France
[2]LaBRI - (CNRS / Univ. Bordeaux / Bordeaux INP), Talence, France
[3]SANPSY (CNRS / Univ. Bordeaux), USR 3413, Bordeaux, France
[4]Clinique du sommeil, Hôpital Pellegrin-Tripode, Bordeaux, France

E-mail: thibaut.monseigne@inria.fr

ABSTRACT : Neurofeedback (NF) consists in using electroencephalographic (EEG) measurements to guide users to perform a cognitive learning using information coming from their own brain activity, by means of a real-time sensory feedback (e.g., visual or auditory).

Many NF approaches have been studied to improve attentional abilities, notably for attention deficit hyper activity disorder. However, to our knowledge, no NF solution has been proposed to specifically reduce drowsiness. Thus, we propose an EEG-NF solution to train users to self-regulate an EEG marker of drowsiness, and evaluate it with a preliminary study.

Results with five healthy subjects showed that three of them could learn to self-regulate this EEG marker with a relatively short number of NF sessions (up to 8 sessions of 40 min). Clinical trials with sleep-deprived subjects should begin in 2019 to study possible cognitive and clinical benefits of this self-regulation. This NF solution implementation is available for free, with the OpenViBE platform, under the AGPL-3.0 license.

INTRODUCTION

Brain-computer interfaces (BCI) and Neurofeedback (NF) are two inter-related disciplines, that can and should learn from each other [12]. A substantial part of BCI research is dedicated to train computational models to classify patterns of brain activity and translate them into commands for a machine [16]. For instance, BCIs are often used to classify imagined movements of the left and right hand from brain signals [20]. NF, on the other hand, mostly focuses on teaching users to self-regulate their own brain activity to reach a desired state [23], which may possibly lead to clinical or cognitive benefits. For instance, the desired state could be an increased attention level, as measured by an electroencephalographic (EEG) marker of attention, to treat attention deficit hyperactivity disorder (ADHD) [18]. NF is thus a form of cognitive remediation guided by physiology.

One of the main clinical application of EEG-NF is ADHD rehabilitation [18, 19]. However, while several papers reported positive clinical benefits with NF, there is currently a debate in the community about the origin of these benefits [17, 26], and the level of evidence is still relatively weak [11, 12, 25]. This recently motivated the NF community to encourage further research into NF with higher research standards, e.g., by identifying open research challenges or by proposing consensus evaluation and reporting guidelines [1, 3, 9, 22]. Among the identified open challenges, exploring innovative psychiatric applications of NF, other than ADHD rehabilitation, has been encouraged [4].

To the best of our knowledge, there is no NF approach that has been studied to reduce excessive daytime sleepiness (EDS) and drowsiness so far. EDS is a common complaint among young people between 18 and 30 years of age [10]. EDS leads to increased accident risks. However, to date, there is no countermeasure to EDS based on the principles of cognitive remediation such as NF. Although EDS is linked to ADHD [2, 5], sleep disorders are rarely highlighted in ADHD research. Moreover, EDS may have a physiological target that can be more easily identified than attentional markers [7]. NF might thus be used to train subjects to regulate their EEG activity, and in particular an EEG marker of drowsiness, in order to reinforce their ability to stay awake. This could hopefully reduce their drowsiness and increasing their cognitive performance.

Therefore, in this paper, we present the design, implementation and preliminary evaluation of a complete NF protocol that is designed to train subjects to self regulate a neurophysiological target related to drowsiness. In this preliminary evaluation, we aim at validating our NF solution, i.e., at assessing whether subjects can self-regulate the identified marker of drowsiness using NF. The objective quantification of drowsiness, and how it varies with NF training, is not addressed in the present study. However, it will naturally be the primary outcome measure of the future clinical study based on our NF solution that is now validated.

This paper is organized as follows: first it details, in the materials and methods section, the different methodolog-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-37

ical, hardware and software elements used for our solution. Then, it describes the three main components of this protocol: the EEG signal processing methods, the training procedure and the learning evaluation. This section finishes by describing a preliminary evaluation of our NF solution with five healthy users. The results section then reports on the preliminary data obtained with such evaluation. The paper ends with a discussion on these results and the future clinical studies.

MATERIALS AND METHODS

Overview: Our NF system (see Figure 1) first consists in measuring and extracting a marker of drowsiness from EEG signals. Then, it consists in providing subjects with a visual feedback representing this EEG marker, in order to train them to self-regulate it. To do so, subjects are rewarded with an audio feedback if they manage to substantially reduce this EEG marker of drowsiness. To prevent artifacts from deteriorating this EEG self-regulation learning, we also need to detect them in order to adapt the feedback accordingly. In the following, we describe these various components.

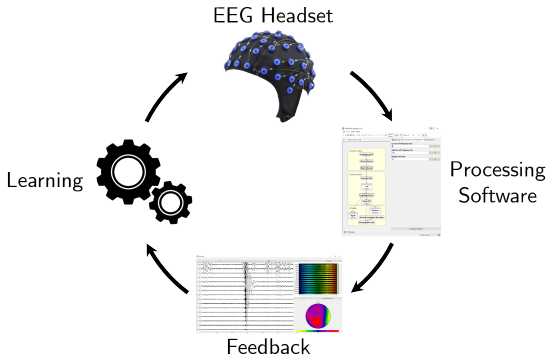


Figure 1: Our NF system: The EEG cap measures brain signals that the software analyzes to extract a marker of drowsiness. Then, a visual feedback representing this marker is provided to subjects. With feedback training, subjects should learn to self-regulate it and thus, hopefully, to reduce their drowsiness.

Hardware: We used a g.Nautilus (g.tec, Austria) wireless EEG acquisition device with 16 g.SAHARA active dry electrodes localized at Fp1, Fp3, F3, Fz, F4, T7, C3, Cz, C4, T8, P3, Pz, P4, PO7, PO8 and Oz.

Software: We used OpenViBE [21], a free and open source software platform for the acquisition, processing, classification and visualization of brain signals.

Signal Processing: As a neurophysiological target (EEG marker of drowsiness), we chose the homeostatic sleep pressure defined in [7] as the power of the spectral band θ-α (6.25-9Hz) in Cz. In that paper, it was shown that during sustained wakefulness, the θ-α spectral band in Cz was the one whose power increased the most over time.
However, this spectral band power may increase independently of the homeostatic sleep pressure, due to a general increase of EEG activity for example. In order to overcome this problem, we used a second spectral band, also related to drowsiness, to compute a ratio of spectral band power. This solution is quite common in NF [18]. Some noises and artifacts that may affect the whole signal will be attenuated (or suppressed) in a ratio of spectral band power, since we divide one band power by the other. Inspired by ADHD studies [18], we chose to divide our main target (i.e., θ-α power in Cz) by the power of the β (15-30Hz) spectral band, also in Cz. Indeed, this band is related to vigilance and its power should thus decrease with drowsiness [18].
Overall, we thus used as neurophysiological NF target $\frac{\theta\text{-}\alpha}{\beta}$, which should be related to the subject's level of drowsiness. The β band power should indeed decrease with increasing drowsiness whereas the θ-α band power should increase with increasing drowsiness. In order to give subjects a more positive impression of exercising, we defined their goal during NF training as to increase their wakefulness. Therefore, we reversed this ratio to use the $\frac{\beta}{\theta\text{-}\alpha}$ ratio as a target that subjects should learn to up-regulate, i.e., to increase.
To estimate the power of a spectral band, we first band-pass filtered the EEG signal, here in Cz, in the selected band (e.g., 15-30Hz) using a fourth order Butterworth filter [6]. Next, we used a sliding window analysis, with 1s long windows with an interval of 1/16s (0.0625s) between consecutive windows (with overlap). For each window, we squared the filtered EEG signals from that window, and then averaged them over the window duration to obtain the power of the selected spectral band. We then log-transformed the data ($x' = \log(1+x)$) to make these spectral band power values more normal-like.
In order to deal with artifacts, we eliminated the only easily identifiable ones, i.e., those leading to EEG signal amplitudes greater than 100 μV. These are typically considered as artifacts due to electrode movements or to muscular or ocular activities, resulting in larger EEG amplitudes. Epochs whose samples exceeded 100 μV were thus considered as artifacts and marked as such. These epochs were rejected and no feedback was provided to participants during them.

Training - sequencing: Regarding the duration and sequencing of the NF training, we arranged a NF session as follows (Figure 2):

- 1 calibration block of 2min.
- 6 work blocks of 5min.
- 1 transfer block of 5min.
- Each block is separated by a break of about 30s.
- Total Time $= 2+7\times(0.5+5) = 40.5$min

The calibration block (baseline) aims at measuring EEG activity "at rest", without any NF task. This was used as a reference, with measurements under similar conditions each time. During the work blocks, the subjects had to perform the various training tasks with feedback: the subject had to find a mental strategy to increase his

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-37

wakefulness level by using the provided feedback. This feedback represents the value of the band power ratio described above. The last type of block is called a transfer block and consists in performing a work block without any feedback, in order to ease the transition between NF training and everyday life (where there is no feedback).

Training - rewards: Rewards were given to subjects via a sound (audio feedback) and a score system, when their wakefulness level, as measured using the EEG marker, became high enough during 0.25s. These audios rewards were thus provided when subjects managed to make their EEG band power ratio exceeds a given threshold. On the other hand, if subjects' wakefulness level was judged too low (i.e., lower than a given threshold), the experimenter was warned, as he could see the subject's count of "negative" points increasing on a visual display. Note that such negative points were not shown to subjects, who did not receive any negative feedback when their wakefulness level was too low. In order to deal with EEG signal non-stationarities, we regularly updated our threshold values. More precisely, we updated the thresholds every two work blocks (see Figure 2).

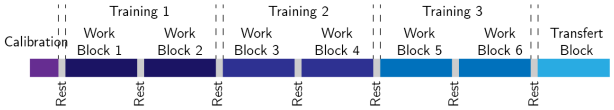


Figure 2: Session sequencing: 1 Calibration block, 3 pairs of work blocks (the dotted lines indicate the times when the threshold values were updated) and one last transfer block.

Current NF literature usually does not mention how the reward thresholds are defined, except when this is the paper central topic, see, e.g., [8]. This is however a crucial point in any NF protocol design. Indeed, such thresholds are used to define the EEG target values that subjects should reach or avoid. Here, we defined 4 thresholds:

- A theoretical minimum (S_{m}) and maximum (S_{M}), used to identify the interval in which the neurophysiological target value varies. They are used to create a standardized color range for the visual feedback.
- A threshold to be reached to receive a reward (S_{R}).
- A threshold to be avoided, not to receive negative points (S_{C}) (again, these negative points are only shown to the experimenter).

Visual inspections of our NF target empirical values showed that it was approximately normally distributed, or at least normal-like enough so we could define the thresholds according to its mean and variance. The NF target mean and variance are estimated before blocks 1, 3, 5 and 7 (Figure 2) on all the neurophysiological target values from the previous block pair (for block 3, 5 and 7), or from the calibration block (for the first work block).
By estimating the mean μ and variance σ of the NF target distribution (assuming a Gaussian distribution), we can then define our 4 thresholds as follows (see Figure 3):

$$S_{\mathrm{m}} = \mu - 3 \times \sigma \qquad S_{\mathrm{M}} = \mu + 3 \times \sigma$$
$$S_{\mathrm{C}} = \mu - 1.5 \times \sigma \qquad S_{\mathrm{R}} = \mu + 1.5 \times \sigma$$

Assuming normally distributed data, we can estimate the probability of a value to belong to a specific interval (Figure 3) :

$$\mathbb{P}(S_{\mathrm{m}} \leq x \leq S_{\mathrm{M}}) \approx 99.73\%$$
$$\mathbb{P}(S_{\mathrm{C}} \leq x \leq S_{\mathrm{R}}) \approx 86.64\%$$

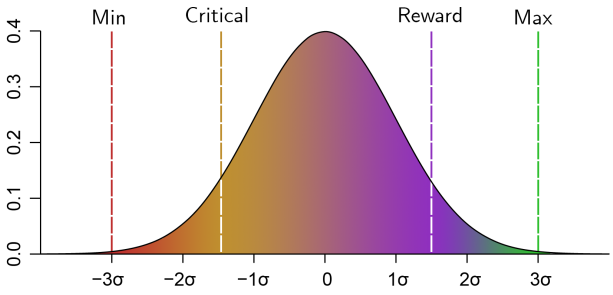


Figure 3: Distribution of thresholds (with σ the standard deviation and an average of 0 for this example).

Training - feedback: The last key element of NF training is the neurophysiological target visualization, i.e., the feedback. As a metaphor of the notion of wakefulness, we displayed as feedback a gray level that goes from black (low wakefulness) to white (high wakefulness), as can be seen in Figure 4. The gray color range was delimited by our thresholds S_{m} and S_{M}. Subjects could also see, at the bottom of the screen, the elapsed time since the beginning of the block, the time points at which they received rewards and the time points when artifacts were detected.

Learning: NF training aims at favoring learning over sessions, which requires to define the number and frequency of the training sessions. This project is a proof of concept of NF training in a small number of sessions. We thus chose to train subjects for a maximum of eight sessions. In order to help subjects to assimilate their training sessions, to reflect on their strategies and to test them in a real situation, the training frequency was defined as from 1 to 2 sessions per week. Finally, the 8 sessions had to be completed in a maximum of 6 weeks.
In addition, to facilitate EEG self-regulation learning, which is not something humans are used to do consciously, we offered subjects discussion times at the beginning of the protocol, between the work blocks and at the end of each session. The first discussion explained subjects the principle of the protocol, to demystify the NF process. Note that subjects were already provided with information on NF beforehand, but this allowed them to ask questions if necessary. This step aimed at preventing any apprehension with the EEG machine or any suspicion with the overall NF method. The discussions between work blocks were used to guide subjects to gain

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-37

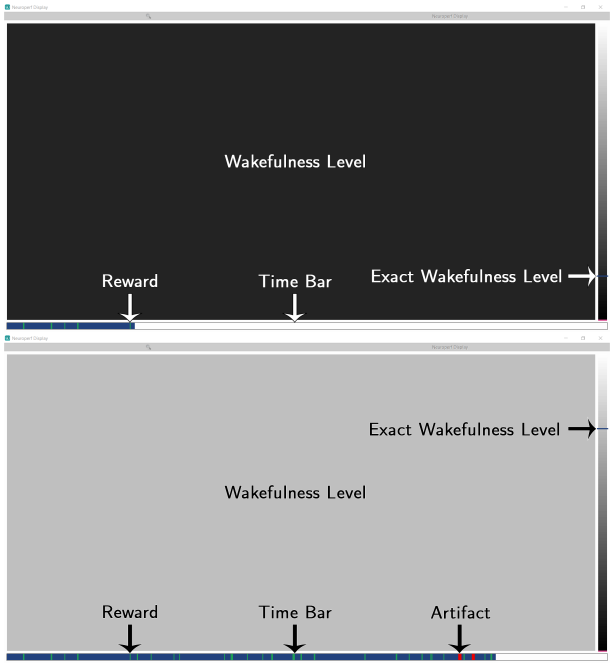


Figure 4: Work block Feedback for low (top figure) and high (bottom figure) wakefulness level estimates.

insights about the strategies they used and their effectiveness. Thus, oriented questions were used to help subjects becoming aware of what they thought and felt during the NF training. In addition, when thresholds were changed, the practitioner indicated to the subjects whether it increased or decreased. At the end of each session, subjects were shown the different curves representing the EEG signal, the $\frac{\beta}{\theta\text{-}\alpha}$ ratio, the number of rewards received and the threshold changes over time and blocks.

Preliminary Experimental Validation: As a proof of concept, in order to assess whether our NF training protocol could enable subjects to self-regulate the selected EEG marker of drowsiness, we tested it with five healthy medical students (two women and three men). They were able to perform between 4 and 8 NF sessions each, depending on their availability. Each session was held at a fixed time for each subject (at 10am for S1 and S2 and 2pm for the others) with a minimum of 2 days between two sessions.

RESULTS

Figure 5 shows the evolution of the average value of the $\frac{\beta}{\theta\text{-}\alpha}$ ratio over sessions, for each subject, while Figure 6 shows the same information averaged across all subjects. Such results suggest a positive evolution of the neurophysiological target of interest for 3 subjects (S1, S3, S4). For one subject (S2), there does not seem to be any subtantial change over time. Moreover, for this subject, the experimenter could observe that the subject was bored during the sessions, due to a lack of success with self-regulating the EEG marker. The last subject (S5) obtained inconclusive results because he obtained rewards during two sessions by playing with the artifacts detection system, using micro movements of his jaw.

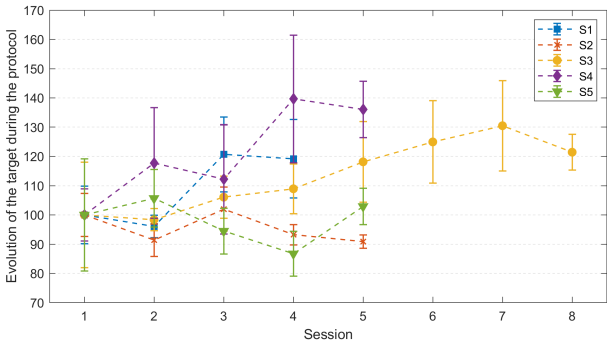


Figure 5: Evolution of the physiological target ($\frac{\beta}{\theta\text{-}\alpha}$ ratio) during the sessions. For each subject, we averaged the value of the physiological target on each session. We then normalized these values so that the first session normalized value was 100. Thus, this gave us a percentage of evolution over the sessions. The displayed standard deviations represent the variation of the ratio average value across work blocks.

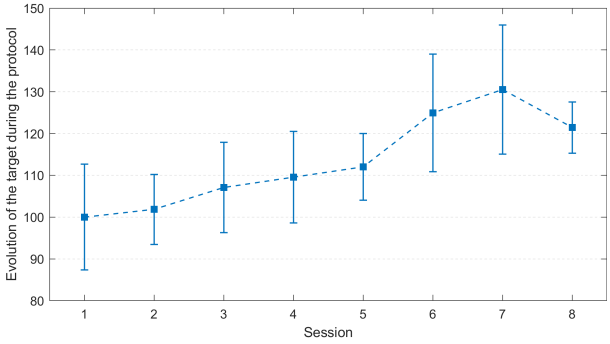


Figure 6: From Figure 5. Average across subjects, for the sessions in which they participated.

We also analyzed the change of the average target value during NF blocks as compared to its average value during the baseline of each session (see Figure 7), i.e., the relative increase of EEG marker value with respect to the baseline. To do so, we estimated the difference of the average target value during work blocks with its value during the baseline, expressed as relative percentage of the baseline value. Positive values thus indicates that the target value was higher during the work blocks than during the baseline of the same session.
Here as well, we observed that the same 3 subjects who showed a substantial increase in EEG marker absolute value across sessions, also showed a subtantial increase of EEG marker relative increase with respect to the baseline, across sessions. Interestingly enough, the other 2 subjects, who did not reveal any subtantial change of absolute target value across sessions, are the ones with the weakest intra-session relative target changes. This further confirmed that these 2 subjects did not succeed to learn to self-regulate our EEG marker of drowsiness.

Finally, we statistically analyzed these relative EEG

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-37

marker increases, to assess whether learning did occur across sessions, or whether this was due to chance. To do so, for each subject, we computed the pearson correlation between the time index of each work block and the relative EEG marker increase of that block. We obtained correlation of $\rho = 0.67$ ($p < 0.00005$), $\rho = -0.36$ ($p = 0.02$), $\rho = 0.7$ ($p < 0.000000001$), $\rho = 0.66$ ($p < 0.000005$) and $\rho = -0.13$ ($p = 0.43$), for subjects S1, S2, S3, S4 and S5 respectively. This confirmed that subjects S1, S3 and S4 significantly improved across blocks and sessions, and thus successfully learned to self-regulate the EEG marker, whereas subjects S2 and S5 did not.

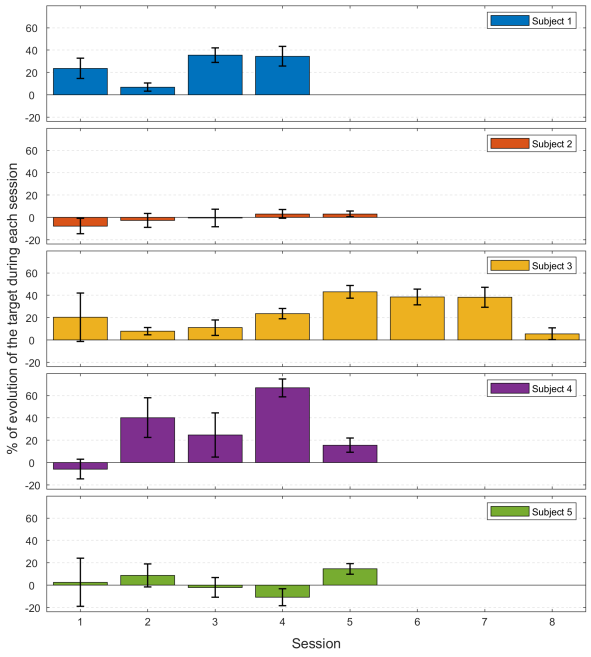


Figure 7: Average value of the physiological target ($\frac{\beta}{\theta\text{-}\alpha}$ ratio) for each session, evaluated as a relative percentage difference with the baseline average target value.

DISCUSSION

The results of the preliminary study described above on five subjects are encouraging, as they show that three of them could learn to increase the neurophysiological target across sessions. However, it is still unknown whether this successful self-regulation could lead to improved vigilance in day-to-day tasks, or whether this could reduce drowsiness for people affected by EDS. Thus, further studies are necessary. Accordingly, a clinical trial with a panel of 20 sleep-deprived subjects should start in the coming months. Such clinical trial will include clinical and objective measures of drowsiness and cognitive performances. It will aim at complementing, confirming or invalidating our results, as well as at assessing this NF training influence on subjects drowsiness and cognitive performances.

There are many open questions about the effectiveness of NF [25], as well as about how to best design relevant NF protocols. Some of these questions are actually common with BCI research. Is there a better neurophysiological target? Here, we used results from neurophysiology research to define the main target of interest according to a spectral band that was found related to drowsiness. However, it is likely that the optimal spectral band and EEG channel may be slightly different depending on the subjects. Another open question is whether we could further improve learning by exploring different feedback. For example, could the use of screen brightness in our experiment have a similar influence on performance as tactile feedback had for motor imagery BCI experiments [14]? Next, we should evaluate whether learning is effective in the long term and therefore induces neuroplasticity. Moreover, as often in NF and BCI experiments, the results between subjects vary greatly. Thus, research on identifying predictive variables of learning [13, 15] could help us to understand the cause of these variations, and thus possibly help us to reduce them.

CONCLUSION

In this article, we presented a complete NF protocol to train subjects to self-regulate an EEG marker of drowsiness, with a longer-term objective to reduce drowsiness. Starting from the neurophysiological target to be analyzed and the signal processing methods, we have defined a training program for NF that aims at guiding subjects to learn to develop strategies to self-regulate their wakefulness level. We conducted a preliminary study on five healthy subjects, that showed that three of them managed to learn to increase the selected neurophysiological target across sessions.

The work done during this study allowed us to develop a set of signal processing algorithms, features, feedback and scripts for the OpenViBE software. All sources are available[1] under AGPL-3.0 license for later use with or without modification. In addition, we did our best to report all the parameters of our experiment to favor reproducibility.

A clinical trial is planned in 2019 to validate these preliminary results, and assess possible clinical benefits.

Many of the previously mentioned NF training parameters could be optimized to further improve learning. For instance, we believe that interviews could be used to reinforce the insights that subjects have about their strategies. Pre NF training procedures, as meditation training did to improve some BCI protocols [24], could also be considered to improve this NF protocol. Finally, it would also be interesting to identify the optimal frequency, duration, and form of training and feedback that would maximize the subjects' performance, learning and clinical benefits.

REFERENCES

[1] https://github.com/tmonseigne/NEUROPERF

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-37

[1] Arns M., Heinrich H., Strehl U. Evaluation of neurofeedback in ADHD: The long and winding road. Biological Psychology. 2014;95:108–115.
[2] Arns M., Kenemans J. L. Neurofeedback in ADHD and insomnia: Vigilance stabilization through sleep spindles and circadian networks. Neuroscience & Biobehavioral Reviews. 2014;44:183–194.
[3] Arns M. et al. Neurofeedback: One of today's techniques in psychiatry? L'Encéphale. 2017;43(2):135–145.
[4] Batail J-M et al. EEG neurofeedback research: A fertile ground for psychiatry? L'Encéphale. 2019.
[5] Bioulac S. et al. Excessive daytime sleepiness in adult patients with ADHD as measured by the Maintenance of Wakefulness Test, an electrophysiologic measure. The Journal of clinical psychiatry. 2015;76(7):943–948.
[6] Butterworth S. On the theory of filter amplifiers. 1930.
[7] Cajochen C., Brunner D., Krauchi K., Graw P., Wirz-Justice A. Power density in theta/alpha frequencies of the waking EEG progressively increases during sustained wakefulness. Sleep. 1995;18(10):890–894.
[8] Dhindsa K., Gauder K. D, Marszalek K. A, Terpou B., Becker S. Progressive Thresholding: Shaping and Specificity in Automated Neurofeedback Training. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2018;26(12):2297–2305.
[9] Fovet T.s et al. On assessing neurofeedback effects: should double-blind replace neurophysiological mechanisms? Brain. 2017;140(10).
[10] Jaussent I., Morin C. M., Ivers H., Dauvilliers Y. Incidence, worsening and risk factors of daytime sleepiness in a population-based 5-year longitudinal study. Scientific Reports. 2017;7(1).
[11] Jeunet C., Glize B., McGonigal A., Batail J.-M., Micoulaud-Franchi J.-A. Using EEG-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: Theoretical background, applications and prospects. Neurophysiologie Clinique. 2018.
[12] Jeunet C., Lotte F., Batail J.-M., Philip P., Micoulaud-Franchi J.-A. Using Recent BCI Literature to Deepen our Understanding of Clinical Neurofeedback: A Short Review. Neuroscience. 2018;378:225–233.
[13] Jeunet C., N'Kaoua B., Subramanian S., Hachet M., Lotte F. Predicting Mental Imagery-Based BCI Performance from Personality, Cognitive Profile and Neurophysiological Patterns. PLOS ONE. 2015;10(12).
[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, 2015, vol. 9296, 488–505.
[15] Jeunet Camille, N'Kaoua Bernard, Lotte Fabien. Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates. In: Progress in brain research, 2016, vol. 228, 3–35.
[16] 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.
[17] Micoulaud-Franchi J.-A., Fovet T. A framework for disentangling the hyperbolic truth of neurofeedback: Comment on Thibault and Raz (2017). The American psychologist. 2018;73(7):933–935.
[18] Micoulaud-Franchi J.-A., Geoffroy P. A., Fond G., Lopez R., Bioulac S., Philip P. EEG neurofeedback treatments in children with ADHD: an updated meta-analysis of randomized controlled trials. Frontiers in Human Neuroscience. 2014;8.
[19] Micoulaud-Franchi J.-A., McGonigal A., Lopez R., Daudet C., Kotwas I., Bartolomei F. Electroencephalographic neurofeedback: Level of evidence in mental and brain disorders and suggestions for good clinical practice. Neurophysiologie Clinique/Clinical Neurophysiology. 2015;45(6):423–433.
[20] Pfurtscheller G., Neuper C. Motor imagery and direct brain-computer communication. Proceedings of the IEEE. 2001;89(7):1123–1134.
[21] 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.
[22] Ros T. et al. Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). 2019.
[23] Sitaram R. et al. Closed-loop brain training: the science of neurofeedback. Vol. 18. Nature Reviews Neuroscience. Nature Publishing Group, 2016.
[24] Tan L.-F., Dienes Z., Jansari A., Goh S.-Y. Effect of mindfulness meditation on brain–computer interface performance. Consciousness and Cognition. 2014;23:12–21.
[25] Thibault R. T., Lifshitz M., Raz A. Neurofeedback or neuroplacebo? Brain. 2017;140(4):862–864.
[26] Thibault R. T., Raz A. The psychology of neurofeedback: Clinical intervention even if applied placebo. The American psychologist. 2017;72(7):679–688.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-38

THE EFFECT OF 10 VIBRO-TACTILE P300 BCI SESSIONS IN PATIENTS WITH MINIMALLY CONSCIOUS STATE

Alexander Heilinger[1], Nensi Murovec[1,2], Ren Xu[2], Yangyang Miao[3], Jing Jin[3], Rupert Ortner[4], Rossella Spataro[5], Vincenzo La Bella[6], Christoph Guger[1,2]

1 g.tec medical engineering GmbH, Sierningstraße 14, A-4521 Schiedlberg
2 Guger Technologies OG, Herbertsteinstraße 60, A-8020 Graz
3 University of Science and Technology Shanghai, Department of Automation, East China
4 g.tec medical engineering Spain S.L., C/Plom 5-7, 4o-4a, 08038 Barcelona, Spain
5 IRCCS Centro Neurolesi Bonino Pulejo, Palermo, Italy
6 ALS Clinical Research Center, Bi.N.D., University of Palermo, Palermo, Italy

E-mail: heilinger@gtec.at

ABSTRACT: It has been shown that brain-computer interface (BCI) technology can help in assessing the cognitive abilities of patients with minimally conscious state (MCS) and can provide a platform for communication. In this study the effects of a vibro-tactile BCI oddball paradigm with three different stimulation location (VT3) were investigated. Seven MCS patients performed 10 VT3 P300 BCI sessions over 10 days with 8-12 runs each day. Changes in the classification accuracies over the 10 days were investigated. Coma Recovery Scale – Revised (CRS-R) was assessed before and after the 10 sessions. The average classification accuracy in the best sessions was 80±15%. Four out of seven patients showed an improvement in the CRS-R score of 2 to 7 points. This study shows the importance of repeated BCI measures in assessing MCS patients. The improvement in the CRS-R score is an important fact which should be considered in future studies.

INTRODUCTION

Some patients diagnosed with disorders of consciousness (DOC) recover to a chronic state of poor responsiveness to stimuli. Nevertheless these patients show evidence of awareness of themselves and their environment, depending on their motor control and cognitive abilities [1]. The fluctuations in responsiveness in these patients is a challenge for the diagnosis[2]. Behavioral tests such as the Glasgow Coma Scale or the Coma Recovery Scale – Revised (CRS-R) are commonly used tools to determine the diagnosis of these patients [2]–[5]. The limitation of using these scores is their dependence on voluntary motor control.

Alternatively, brain-computer interfaces (BCIs) have shown promising results in providing a platform for assessment and communication for DOC patients [6]–[8]. BCI systems rely on different neurophysiological phenomena and can thereby utilize different approaches like transient evoked potentials, for example the P300, a positive amplitude rising 300ms after the stimulus in the EEG signal [9]. Other techniques include slow cortical potentials and mu or beta rhythms [10], [11].

Visual P300 paradigms yielded high accuracies in patients with motor paralysis [12], [13]. However, DOC patients do not always have control of eye movements, which may cause difficulties in orienting attention to a specific location in the visual field. Therefore, we here chose to focus on a vibro-tactile paradigm.

The current study investigated the changes in the CRS-R score before and after ten sessions of a vibro-tactile P300 BCI paradigm with three stimulators used in seven MCS patients. Also, the changes in the classification accuracies during the training were explored. Our principal aim was to seek the possibility that training with a vibro-tactile BCI paradigm could facilitate recovery. Our secondary aim was to show that one or only a few sessions would not be sufficient for evaluating a patient's cognitive capabilities and that these results cannot be used potentially for BCI-based communication.

MATERIALS AND METHODS

Sample

This study included data acquired from 7 different MCS patients in a stable chronic stage. Patient Nr. 1 and 2 were located at the University of Palermo, Italy (PA) and patient Nr. 3-7 at the Shanghai Rehabilitation Hospital 3, China (SH). The following inclusion criteria were used: patients had to be over 18 years old, and diagnosed with MCS state according to the CRS-R scale administered by experienced neurologists. The CRS-R Score was measured the day before and after the 10 VT3 BCI sessions.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-38

Table 1 presents the patients' demographic data. We are presenting 7 MCS patients (Median age: 60; Min: 39; Max: 69).

Table 1: *Patients' demographic and CRS-R data. (A) Age (years); (B) Sex (M: Male, F: Female); (C) Diagnosis (CH: Cerebral Hemorrage, ENC: Mitochondrial Encephalopathy, TBI: Traumatic Brain Injury, HT: Hematoma, SH: Subdural hematoma); (D) Clinical State Before; (E) CRS-R before; (F) CRS-R after; (G) Δ CRS-R*

#	A	B	C	D	E	F	G
1	39	F	CH	MCS	11	11	0
2	30	F	ENC	MCS	12	14	2
3	66	M	TBI	MCS	7	9	2
4	60	M	CH	MCS	7	6	-1
5	56	M	HT	MCS	7	9	2
6	69	M	Anoxia	MCS	13	12	-1
7	61	F	SH	MCS	7	14	7

Brain-computer interface system

For data acquisition the mindBEAGLE system (Guger Technologies OG, Austria) was used. The gel-based EEG electrodes g.LADYbirds (Guger Technologies OG, Austria) were used. The electrodes were connected to a biosignal amplifier (g.USBamp, g.tec medical engineering GmbH, Austria). The amplifier has a 24-bit resolution combined with a high oversampling rate to increase the signal-to-noise ratio. The amplifier was connected to a computer using a USB cabel, and the EEG data were recorded with a sampling rate of 256 Hz. The EEG signal is presented for visual inspection on a monitor during the measurement. The data are stored in floating point format for later data analysis.

A bandpass filter between 0.1-30 Hz was used to filter the EEG signal. This was done to remove baseline shifts and eliminate most EEG artifacts. The electrode positions for recording were FCz, C3, Cz, C4, CP1, CPz, CP2 and Pz according to the extended international 10-20 electrode system. The reference electrode was mounted on the right earlobe, while the ground electrode was placed on the forehead. This system relies on P300 and motor imagery BCI approaches, and step-by-step explanations of system operation can be found in [7], [9]. In this study only the vibro-tactile P300 approach was used.

Paradigm

Three vibro-stimulators were placed on the left and right wrist of the patient. A third stimulator was placed on the foot to act as an additional distractor. The paradigm consists of 480 stimuli per run, with 60 groups of 8 stimuli. Via earbuds the patients were instructed to silently count vibro-tactile pulses on either their left or right wrist. All vibrotactile stimuli lasted 100 ms with a 100 ms pause between stimuli. The whole paradigm required around 2.5 minutes per run. Patient 1 and Patient 2 participated in 10 sessions over 10 consecutive days. For Patient 1 and Patient 2 each session consisted of 12 VT3 runs. Patient 3-7 participated in 10 sessions over 10 consecutive days. For Patient 3-7 one session consisted of 8 VT3 runs.

Data Analysis

Data segments of -100 ms to 600 ms around each stimulus were extracted and the EEG signal was averaged and baseline corrected. Trials with a signal amplitude +/- 100 µV were rejected from further processing. An automatic artifact detection was used during the run so the trial number varied.

To calculate the classification accuracy, the following steps were done. The target and nontarget trials are randomly assigned into two equal sized pools. One pool is used to train a classifier, and the other pool is used to test the classifier. The classifier is tested on an increasing number of averaged stimuli out of the test pool. At first, it is tested on only one target and seven nontarget stimuli. If the classifier detected the target stimulus correctly, the resulting accuracy is 100 %; otherwise, it is 0 %. This process is repeated for two averaged target stimuli and 14 averaged nontarget stimuli, for three nontarget stimuli and 21 target stimuli, and so on until the full test pool is used. This produces a plot of 30 single values (for 30 target stimuli in the test pool), each one either 100 % or 0 %. The averaging of 10 single plots results in values ranging from 0 % to 100 %. From these plots we selected the median value, the maximum value and first value on the x-axis the plot reached the maximum percentage. The accuracy value represents how well the data could be discriminated by the classifier, with a high value indicating a good separability of the EEG data.

The EPs from target and nontarget trials are averaged for all channels separately. Examples of EPs are shown in Figure 1. For each sample point, a Kruskal Wallis test ($p<0.05$) is done to find statistical differences between target and nontarget trials. A Wilcoxon rank sum test was performed to test the significance of the classification accuracies.

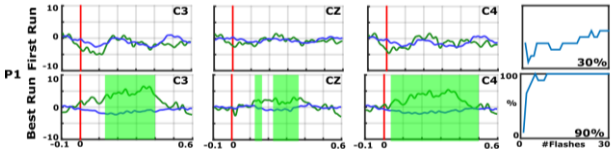


Figure 1: *EP's and classification accuracies of patient P1 in the first session compared to the best session on the electrodes C3, CZ and C4.*

A Wilcoxon rank sum test was performed to test both the significance of the classification accuracies and the significance of the CRS-R score. The factor was the timing of measurement (first run vs best run) for the classification accuracy and the timing of measurement (before vs after) for the CRS-R score.

RESULTS

Classification Accuracy

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-38

The results are shown in Figure 2. The classification accuracy in the first run ranged between 30% and 100% (Mean: 49%) cross all seven patients. The best classification accuracy of each patient yielded from 60% to 100% (Mean: 80%). Notably, P6 and P7 reached 100% in the best run. The median accuracy over 10 days ranged from 10% to 40% (Mean: 21%). A Wilcoxon rank sum test was performed on each single classification accuracy and revealed a significant difference between the accuracies in the first run compared to the accuracies in the best run ($p < 0.01$).

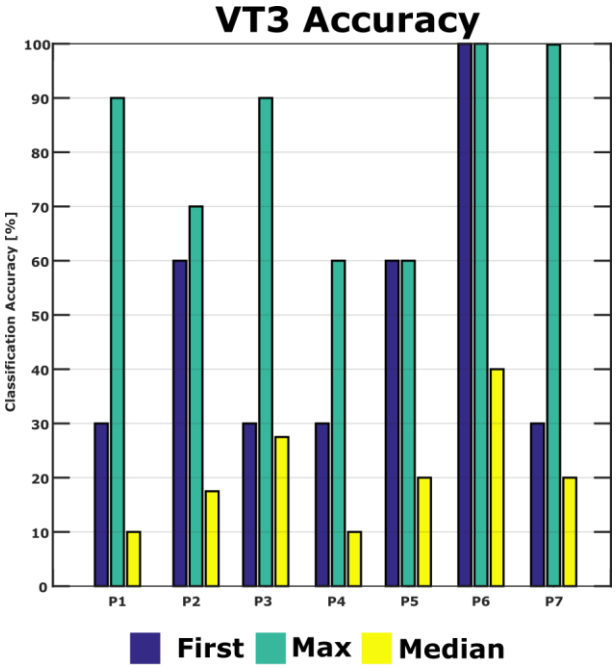


Figure 2: *Classification accuracies of all 7 MCS patients. Blue bars indicate the classification accuracies of the first run. Cyan bars indicate the best classification accuracy and the yellow bars show the median classification accuracy over all sessions.*

CRS-R score

Table 1 shows the total CRS-R scores of each patient before and after the 10 sessions. It shows also the delta of the CRS-R scores, i.e. the difference between the pre- and post- scores. The change in CRS-R score is on average 1.6 points (Max: 7; Min: 0). Specifically, 4 out of 7 patients (P1, P2, P5, P7) showed an improvement of two point or more after the VT3 BCI sessions. P1, P2 and P5 changed with a total improvement of 2 points. P7 showed the biggest improvement of 7 points. 1 patient did not show a change of the CRS-R score (P1) and 2 patients showed a decline in the score by 1 point (P4, P6). Of these 4 patients with improvement in CRS-R, 3 improved in auditory function, 3 in visual function, 2 in motor function, 1 in oromotor/verbal function, 2 in communication and 2 in arousal.

The Wilcoxon rank sum test revealed a significant difference in the CRS-R score before and after the 10 VT3 P300 BCI sessions ($p < .01$).

DISCUSSION

In this study the changes in the CRS-R score before and after ten VT3 P300 BCI sessions and the changes in the classification accuracies were investigated.

During the 10 VT3 P300 BCI sessions, the classifier resulted in accuracies of 60% or higher in the best runs of the patients. This is an indicator that the patients are able to follow commands as tested with the active P300 VT3 paradigm, according to the study by Guger et al[14].. This study also showed that a single session is not sufficient to assess command following in MCS since these patients have large fluctuations. By comparing the grand average accuracy of 49% in the first run with the grand average accuracy of 80% in the best run, the importance of repeated measurements are highlighted. This difference could be attributed to many factors in MCS patients, like motor or language impairments and vigilance fluctuations[15]. The best run is even in the range of healthy controls (N=6) who achieved 83 % classification accuracy after one VT3 run [16]. This highlights the importance of repeated testing of DOC patients to show command following [14]. It also shows that MCS patients can reach results similar to healthy controls.

An improvement in the CRS-R score was observed in 4 out of 7 MCS patients. In 3 patients the score improved for 2 points, in 1 patient it improved 7 points. We do not currently claim that the improvement was caused by the VT3 paradigm and that there is a causal link between the CRS-R scores and vibrotactile results. Although this should be investigated in future studies. In the past, studies indicated large-scale cerebral networks exist in MCS patients [17], [18]. These findings indicate that there might be residual functional capacity in some patients that could be supported by therapeutic interventions. Schiff et al. also showed activity in the postcentral gyrus during bilateral tactile stimulation of the hands[18]. A prior work reported that the repetition of behavioural assessments in DOC can influence the clinical diagnosis [19].

Following limitations of this study are listed. Firstly, the sample size of 7 MCS patients is too small. For the future, the effect on the CRS-R score will be investigated on more patients and compared to a control group. Secondly, CRS-R score that was used to measure the changes, was assessed only once on the day before and after the VT3 paradigm. Wannez et al. showed in a study that ingle CRS-R score evaluations are not reliable in about 35% of the patients [19], which may reflect fluctuations in conscious awareness and/or arousal that would also increase BCI performance variability. In the future repeated CRS-R score measurements will be done before and after the experiment to minimize the chance of fluctuations in the level of consciousness.

CONCLUSION

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-38

A tactile P300 BCI was performed on 7 MCS patients over ten consecutive days, showing significant changes in the CRS-R. 4 patients showed high classification accuracy and therefore are potential to use a BCI for communication. In the future the possible recovery effects of vibro-tactile P300 BCI paradigms will be further explored.

CONFLICT OF INTEREST

The authors AH, NM, RX, RO and CG belong to g.tec medical engineering GmbH. The other authors declare no conflict of interest.

ACKNOWLEDGMENT

NM, AH and YM recorded the data. RX, JJ, RO, FC, KM and CG contributed to study design, scientific protocols, and review/analysis of results. AH reviewed results and was primarily responsible for writing, and all authors discussed the results and implications. This research was supported by the EC SME Phase 2 project ComaWare, the Eurostars project ComAlert (Grant number E19361) and and the Marie Skłodowska-Curie grant DoCMA (agreement No 778234).

REFERENCES

[1] J. L. Bernat, "Chronic disorders of consciousness," *The Lancet*, vol. 367, no. 9517, pp. 1181–1192, 2006.

[2] J. T. Giacino *et al.*, "The minimally conscious state definition and diagnostic criteria," *Neurology*, vol. 58, no. 3, pp. 349–353, 2002.

[3] R. M. Gibson *et al.*, "Multiple tasks and neuroimaging modalities increase the likelihood of detecting covert awareness in patients with disorders of consciousness," *Frontiers in Human Neuroscience*, vol. 8, Nov. 2014.

[4] M. M. Monti *et al.*, "Willful modulation of brain activity in disorders of consciousness," *New England Journal of Medicine*, vol. 362, no. 7, pp. 579–589, 2010.

[5] M. Risetti *et al.*, "On ERPs detection in disorders of consciousness rehabilitation," *Frontiers in human neuroscience*, vol. 7, 2013.

[6] F. Aloise *et al.*, "P300-based brain–computer interface for environmental control: an asynchronous approach," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025025, Apr. 2011.

[7] J. Annen *et al.*, "MindBEAGLE: An EEG-based BCI developed for patients with disorders of consciousness."

[8] A. Heilinger *et al.*, "Performance Differences Using a Vibro-Tactile P300 BCI in LIS-Patients Diagnosed With Stroke and ALS," *Frontiers in Neuroscience*, vol. 12, Jul. 2018.

[9] Z. R. Lugo *et al.*, "A vibrotactile P300-based BCI for consciousness detection and communication," *Clin EEG and Neurosci*, 2014.

[10] A. Kübler and N. Birbaumer, "Brain–computer interfaces and communication in paralysis: Extinction of goal directed thinking in completely paralysed patients?," *Clinical Neurophysiology*, vol. 119, no. 11, pp. 2658–2666, Nov. 2008.

[11] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, Jun. 2002.

[12] Nijboer, "The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis - a longitudinal study," *Frontiers in Neuroscience*, no. 4, pp. 55, 2010.

[13] S. C. Kleih, F. Nijboer, S. Halder, and A. Kübler, "Motivation modulates the P300 amplitude during brain–computer interface use," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1023–1031, Jul. 2010.

[14] C. Guger *et al.*, "Assessing Command-Following and Communication With Vibro-Tactile P300 Brain-Computer Interface Tools in Patients With Unresponsive Wakefulness Syndrome," *Frontiers in Neuroscience*, vol. 12, Jun. 2018.

[15] R. T. Seel *et al.*, "Assessment Scales for Disorders of Consciousness: Evidence-Based Recommendations for Clinical Practice and Research," *Archives of Physical Medicine and Rehabilitation*, vol. 91, no. 12, pp. 1795–1813, Dec. 2010.

[16] R. Spataro *et al.*, "Preserved somatosensory discrimination predicts consciousness recovery in unresponsive wakefulness syndrome," *Clinical Neurophysiology*, vol. 129, no. 6, pp. 1130–1136, Jun. 2018.

[17] S. Laureys *et al.*, "Cerebral processing in the minimally conscious state," *Neurology*, vol. 63, no. 5, pp. 916–918, 2004.

[18] N. D. Schiff *et al.*, "fMRI reveals large-scale network activation in minimally conscious patients," p. 10.

[19] S. Wannez, L. Heine, M. Thonnard, O. Gosseries, S. Laureys, and Coma Science Group collaborators, "The repetition of behavioral assessments in diagnosis of disorders of consciousness: Repeated CRS-R Assessments for Diagnosis in DOC," *Annals of Neurology*, vol. 81, no. 6, pp. 883–889, Jun. 2017.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-39

ARE USERS' TRAITS INFORMATIVE ENOUGH TO PREDICT/EXPLAIN THEIR MENTAL-IMAGERY BASED BCI PERFORMANCES?

C. Benaroch[1], C. Jeunet[2], F. Lotte[1]

[1]Inria, LaBRI (Univ. Bordeaux, CNRS, Bordeaux INP), France
[2]CLLE (Univ. Toulouse Jean Jaurès, CNRS), France

E-mail: camille.benaroch@inria.fr

ABSTRACT: Mental-Imagery based Brain-Computer Interfaces (MI-BCIs) make use of brain signals produced during mental imagery tasks to control a computerised system. The current unreliability of MI-BCIs could be due, at least in part, to the use of inappropriate user-training procedures. In order to improve these procedures, it is necessary first to understand the mechanisms underlying MI-BCI user-training, notably through the identification of the factors influencing it. Thus, this paper aims at creating a statistical model that could explain/predict the performances of MI-BCI users using their traits (e.g., personality). We used the data of 42 participants (i.e., 180 MI-BCI sessions in total) collected from three different studies that were based on the same MI-BCI paradigm. We used machine learning regressions with a leave-one-subject-out cross validation to build different models. Our first results showed that using the users' traits only may enable the prediction of performances within one multiple-session experiment, but might not be sufficient to reliably predict MI-BCI performances across experiments.

INTRODUCTION

Brain computer interfaces (BCIs) enable users to interact with the environment using their brain activity alone (which is measured, most of the time, using electroencephalography - EEG) [1]. In this work we particularly focus on Mental-Imagery based BCIs (MI-BCIs), that require users to perform specific mental-imagery tasks, e.g., imagining movements of a hand or performing mental calculations, to control systems such as assistive technologies [2] or video games [3]. While promising, those new technologies remain barely used outside laboratories notably because of their low reliability [4]: the average performance of MI-BCI users is most of the time rather low, i.e., around 75% of classification accuracy for 2 class MI-BCIs [5]. In addition, a large proportion of MI-BCI users, between 15% to 30% [6], seems to be unable, while they are performing MI tasks, to produce brain activity patterns that can be discriminated by the system. To make MI-BCIs more reliable, researchers have mainly focused on hardware (e.g., electrodes) and software (e.g., signal processing algorithms) improvements, but less on the improvement of user-training procedures. Yet, this aspect is also essential. Indeed, if MI-BCI users cannot generate "understandable" signals (i.e. stable and distinct brain signals for each task), they will not be able to control the system, even if provided with the best hardware and software solutions. Producing such brain signals is a skill to be acquired by the MI-BCI user [7]. Each user having different skills, states and traits, the training procedure should be specifically adapted to each of them, which is not currently the case [8]. In order to better understand the mechanisms underlying MI-BCI control, and consequently design training strategies adapted to each user, several studies have investigated MI-BCI performance predictors [9]. These predictors could explain between-subject differences and thus variability in terms of MI-BCI control abilities. They can be related to demographic characteristics. For instance, in [10] a positive interaction was found between the participants' age and amount of daily hand-and-arm movements (e.g., practice of video games, musical instruments or sports) and their mu-power at rest, which itself has been shown to correlate with MI-BCI performances [11]. Moreover, Randolph et al. [12] have suggested that playing at least one instrument, not being on effective drugs, being a woman, and being over the age of 25 increased the likeliness of obtaining high MI-BCI performances. Beyond demographic variables, psychological traits like self-reliance and apprehension have been shown to linearly correlate with MI-BCI performances [13], just as mental rotation scores do, which suggests that spatial abilities influence MI-BCI performances. This last correlation was replicated in two further studies [14, 15]. Finally, [16] revealed a positive significant correlation between BCI performances and visuo-motor coordination abilities, which was replicated in [17], strengthening the fact that spatial abilities might strongly influence MI-BCI users' performances. Once the factors influencing MI-BCI performance have been identified, this influence can be quantified using modeling. For instance, [13] experimentally revealed a model including 4 factors (mental rotation scores, self-reliance, apprehension and the visual/verbal sub-scale of the Learning Style), using a step-wise linear regression. The average prediction error of this model was below 3%. Hammer et al. [16] proposed a model including the visuo-motor coordination factor and tested it across studies [18]. The average prediction error of this model

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-39

was below 10% for more than 50% of the participants. While they are potentially insightful, these results have been revealed in individual experiments, each including a small number of subjects and sessions or univariate models. Moreover, most of the highlighted factors have not been replicated since. Yet, to be useful, these correlations/models should be stable, accurate, should consider multiple variables and should generalize across experiments and datasets. Thus, in this paper, by combining data from three different experiments based on the same BCI paradigm, we explored the feasibility of determining stable, accurate and generalizable multivariate models that would explain/predict MI-BCI performance variability. The participants of the included datasets took part in 3 (for two of the datasets) or six (for one dataset) MI-BCI sessions, each session being structured into 5 runs. We gathered data from 42 subjects, for 180 BCI sessions in total. In these 3 experiments, the participants had to complete psychometric tests and were asked to learn to perform three MI tasks, namely, left-hand movement imagination, mental rotation and mental subtraction. We created six groups from the 3 datasets in order to pair the participants of the different experiments according to their specific experimental paradigms. We used a LASSO (Least Absolute Shrinkage and Selection Operator) regression to determine explanatory and predictive models of MI-BCI performance for each group.

MATERIALS AND METHODS

In order to build predictive and explanatory models of performance, we used a LASSO regression that only selected relevant features. We ensured the stability of the selected features by using a leave-one-subject cross validation. Then, to evaluate the reliability of the models and guarantee that the prediction was not due to chance, we empirically estimated the chance level in mean absolute error, based our data, using permutation tests. This approach is detailed in the following paragraphs.

Data sets: To maximize the number of subjects, we used data from three different experiments [13, 19, 20]. They were all based on the same BCI paradigm, as indicated before. The participants' personality and cognitive profiles were computed using different questionnaires (detailed in the *Variables and factors* section). Nonetheless, they were designed with some differences (see Fig. 1). The purpose of the first experiment (XP1 [13]), was to determine how users' cognitive and personality profiles influenced their MI-BCI performances. For this experiment, 18 BCI-naïve participants (9 women, 9 men; aged 21.5±1.2 year-old) took part in 6 MI-BCI sessions, on 6 different days. The second experiment (XP2 [19]) was designed to assess the influence of a Spatial Ability (SA) training procedure on MI-BCI performances. Fourteen participants (8 women, 6 men; aged 22.6±4.6 year-old) took part in this XP2. Each of them did 3 MI-BCI training sessions and 3 other cognitive training sessions (without BCIs), which consisted either

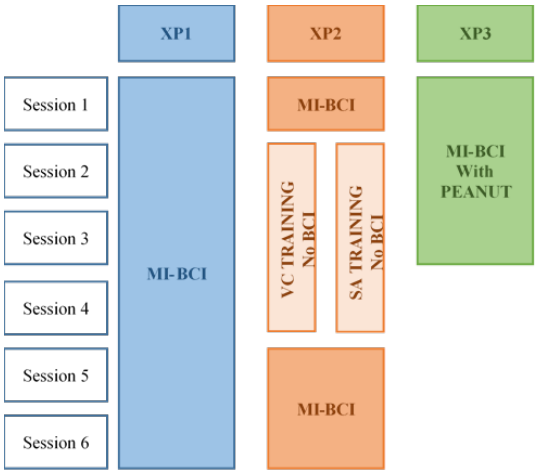


Figure 1: Details of the three different studies:XP1, XP2, XP3

in a SA training (7 participants) or in a verbal comprehension training (7 participants) procedure. In the third experiment (XP3 [20]), 10 subjects (5 women, 5 men; aged 20.7±2.1 year-old) were accompanied by a Learning companion called PEANUT (Personalized Emotional Agent for Neurotechnology User-Training) providing social presence and emotional support during 3 MI-BCI training sessions. The goal of this XP3 was to evaluate the influence that PEANUT had on MI-BCI performances. The 3 studies were conducted in accordance with the relevant guidelines for ethical research according to the Declaration of Helsinki and approved by the ethical committee of Inria. In total, we included the data of 42 participants (22 women, 20 men; aged 21.6 ±2.9 year-old) among which 18 took part in 6 MI-BCI sessions (with 5 runs per session, i.e. 30 runs) and 24 took part in 3 MI-BCI sessions (with 5 runs per session, i.e. 15 runs).

Experimental paradigm: Each BCI session was divided into 5 runs of 45 trials each. The paradigm was the same for all studies (see FIG. 2). For each trial, a cross was first displayed with on its left, a left hand pictogram (representing a L-HAND MI task); on top, a subtraction to perform (mental SUBTRACTION task) and on its right a 3D shape (mental ROTATION task). The MI task to be performed was then announced by a "beep" and a red arrow pointing towards the corresponding pictogram. Then, a blue bar was displayed as continuous visual feedback. The direction of this bar indicated the MI task recognized by the classifier and its length the classifier confidence in this recognition. The bar was displayed only when there was a match between the instruction and the recognized task. The first run of the first session was used as the calibration run to train the BCI classifier and a sham feedback (i.e. a blue bar) was provided to the user. For more details about the experimental paradigm, please refer to the related papers [13, 19, 20].

EEG recordings and pre-processing: For all studies, EEG signals were recorded using 30 active scalp electrodes. The EEG signal-processing pipeline used to classify the three mental imagery tasks online was the same one for all studies. EEG signals were spatially filtered

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-39

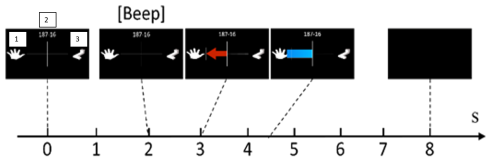


Figure 2: Timing of a trial. The first black screen shows each task, i.e., (1) L-HAND, (2) SUBTRACTION & (3) ROTATION.

using 3 sets of Common Spatial Pattern (CSP) filters [21] and classifier using a shrinkage Linear Discriminant Analysis (sLDA) classifier [22]. For more details about the preprocessing, please refer to [13, 19, 20].

Variables and factors: In the 3 studies, the participants were asked to complete psychometric and personality questionnaires, which aimed to assess different aspects of their personality and cognitive profile. The learning style inventory [23] was used to identify the participants' preferred learning styles according to four dimensions: visual/verbal, active/reflective, sensitive/intuitive and sequential/global. The 16 Personality Factors 5 (16 PF5 – 5) [24] provided a score for sixteen primary factors (warmth, reasoning, emotional stability, dominance, liveliness, rule consciousness, social boldness, sensitivity, vigilance, abstractness, privateness, apprehension, openness to change, self-reliance, perfectionism and tension) as well as for five global factors (extraversion, anxiety, tough mindedness, independence and self-control) of personality that are computed as linear combinations of the primary factors. The Mental Rotation test [25] assessed the participants' spatial abilities. In total, we used 21 parameters (only the 16 primary factors of 16PF5 were used in our study to avoid colinearities between primary and global factors) to represent the personality and cognitive profile of each participant. We thus had 21 features available to find a predictive/explanatory model.

Grouping the experiments: As not all the three studies had the same number of sessions, we made 6 groups out of the 3 datasets. The first three ones correspond to the BCI sessions of the three datasets taken separately (i.e. XP1, XP2 and XP3). The fourth group corresponds to the sessions 1, 5 and 6 of both XP1 and XP2 (the sessions 2, 3 and 4 of XP2 not being BCI sessions but SA or VC sessions, see Fig. 3). The fifth group gathers the sessions 1, 2 and 3 of XP1 and the three sessions of XP3. Finally, in order to have all the 42 subjects together, the last group includes the first session of XP1, XP2 and XP3 (Fig. 3) as it is the only session where all participants were trained to BCI control at the same time in the protocol (see Fig. 3).

Performances: MI-BCI performance was assessed in terms of mean classification accuracy $\overline{X}$ (mean performance measured over all the windows of the feedback periods of the runs 2 to 5 i.e., all 1s long sliding windows -separated by 0.0625s- between t=4.250s and t=8.250s of each trial). We used the mean performance $Perf_{real}$ over the different sessions as the target variable to be explained/predicted by our models. All participants managed to control the BCI interface. The outcome models provided us with a predicted performance $Perf_{pred}$. Our objective consisted in minimizing the mean absolute error ($|Perf_{real} - Perf_{pred}|$).

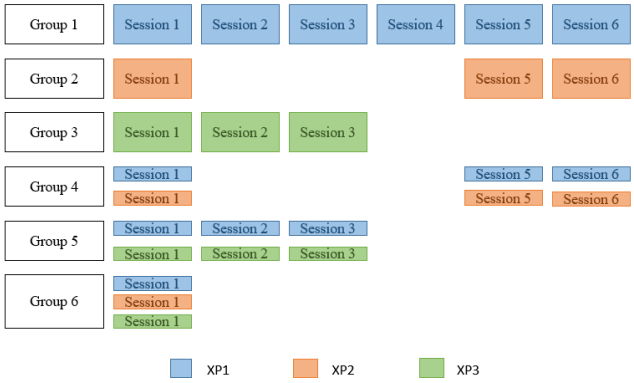


Figure 3: Details of BCI sessions for each group. In Blue sessions concerning XP1, in orange XP2 and in green XP3

LASSO regression: In order to have a stable and reliable model, we used a LASSO [26] regression to obtain models that could predict the performances of MI-BCI users from their profile. The LASSO regression uses a L1 regularization (see Eq. 1) that promotes sparse solutions, i.e., that selects only a small number of variables (many coefficients will be zero using this regularization). It is particularly adapted to reduce the number of relevant features (21 in this study) when the number of those features tend to be higher than the number of subjects [27]. In addition, the LASSO regression is more robust than a simple linear regression [28] and is easily interpretable as only few features are selected for the model. For a usual linear regression set up, we have a continuous output vector $Y \in \mathbb{R}^n$ (MI-BCI performance to be explained/predicted), a matrix $X \in \mathbb{R}^{n \times p}$ of p features (users' traits) for n examples (the users) and a coefficient vector $\beta \in \mathbb{R}^n$. The LASSO estimator is defined as:

$$\beta_{lasso} = \underset{\beta \in R^P}{\operatorname{argmin}} ||Y - X\beta||_2^2 + \lambda ||\beta||_1 \qquad (1)$$

where, $||u||_2^2 = \sum_{i=1}^{n} u^i$ for $u \in \mathbb{R}^n$ and $||\beta||_1 = \sum_{j=1}^{p} |\beta_j|$. For some values of the penalty parameter λ, some components of β_{lasso} will be set exactly to 0. Once β_{lasso} obtained, the MI-BCI performances of the i^{th} user $Perf^i_{pred}$ can be predicted from this user's traits x_i as $Perf_{pred} = \beta_{lasso} \times x_i$. In order to evaluate the stability and reliability of the different models, we use a leave-one-subject-out cross validation process. We also use an inner cross-validation (total number of subjects N - 2) to find the optimal λ (200 $\lambda \in [0.1; 0.5]$), i.e. the one that minimizes the mean absolute error and provides us with a small number of features. We use this optimal λ to build a model and then, the outer cross-validation (N-1 subjects for training, 1 for testing) is used to evaluate this model.

Random model: In order to determine the reliability of the models we estimated the empirical chance level in terms of mean absolute error, given our data. First, we

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-39

Table 1: Group details. Mean performance ($\overline{X_{Group}}$) and Standard Deviation (SD) for the 3-class MI-BCI over the sessions, for each group (chance level is estimated to be 37.7% [29])

Group	$\overline{X_{Group}}$	SD	Outlier (Mean Performance)
1	52.50%	5.62	Subject 1 (67.21%)
2	50.64%	9.47	-
3	50.74%	7.77	-
4	51.48%	7.87	-
5	52.04%	6.40	Subject 1 (67.21%) Subject 37 (38.97%) Subject 38 (38.49%)
6	53.27%	9.54	Subject 7 (32.80%)

randomly permuted the mean MI-BCI performances of the training sets, while keeping the profile variables identical, thus breaking the relationship between profile and performance. Then, we used the LASSO as explained above to predict the MI-BCI performance of the left-out subject. We repeated this process 10000 times and stored each mean absolute error to obtain the distribution of the prediction performance. Then, we sorted those values in descending order and the 99^{th}, 95^{th} and 90^{th} percentiles were used to identify the chance level for the mean absolute error for p=.01, p=.05 and p=.1, respectively.

RESULTS

Outliers' detection: We excluded from the analyses all the participants whose mean classification accuracy was above or below two standard deviations (SD) of the group's performance (see Table. 1).

Predictive models of MI-BCI Performances for each group: A LASSO regression and a leave-one-subject-out cross-validation (CV) were used to reduce the number of features and determine a reliable predictive model of each user's average MI-BCI performance obtained across the different training sessions. For each cross-validation of each group, different features were selected (see Fig. 4). We only showed three groups (1, 5 and 6) on Fig. 4: Group 1 because it was better than chance level ($p \leq .05$), *Group 5* as it failed to reach significance but still had a tendency towards significance and *Group 6* as the selected features are quite stable but the average prediction error is high. This first step allowed us to assess the stability of the results. For the *Group 1*, 16 models among the 17 generated included the same three factors: Mental Rotation scores, Self-reliance and Tension. Regarding the *Group 5*, 24 models among 25 included Warmth, Reasoning and Mental Rotation scores. Finally, for the *Group 6*, 36 models among 42 included Reasoning, Rule Consciousness, Social Boldness and Self-Reliance. The results or these 3 groups are depicted in Fig. 4. For the *Groups 2, 3* and *4*, the results were not conclusive (we decided not to show them due to space restrictions). Indeed, for the *Group 2*, an average of 9 features were selected for each generated model and 14 features (among 21) were chosen in total. For the *Group 3*, a different model was generated for each CV and 17 different features were selected in total and finally for the *Group 4*, rule-consciousness, Apprehension, Self-Reliance, the "Active/Reflective" subscale of the Learning Style were all selected in half of the models. In a second step, we determined the reliability of the models by testing each of them on the participant not included in the training set during the cross-validation process. We then computed the mean absolute error of all the models, i.e, $\sum_{i=1}^{n} \frac{|Perf_{pred(i)} - Perf_{real(i)}|}{n}$, n being the total number of models generated for the group. In order to ensure that the prediction of MI-BCI performances was not due to chance, we performed a permutation test (see Section *Random model* and Table. 2). The results indicated that only the *Group 1* was better than chance ($p<.05$), with a mean absolute error of 3.03%. The chance levels for each group are displayed on Table. 2. We also computed the correlation between the real and predicted MI-BCI performances for each subject. We only obtained a significant correlation for *Group 1* [r = 0.6, $p<.01$].

Table 2: Comparison of the Mean absolute error with the mean absolute error of the random model (after 10000 permutations)

Group	Mean absolute error (%) (pValue)	Mean absolute error (%) of random model ($p < .01$)	Mean absolute error (%) of random model ($p < .05$)	Mean absolute error (%) of random model ($p < .10$)
1	3.03 ($p = 0.047$)	2.60	3.05	3.21
2	7.98 ($p = 0.19$)	6.26	7.02	7.16
3	11.09 ($p = 0.86$)	3.28	4.51	5.24
4	6.69 ($p = 0.37$)	5.62	6.03	6.22
5	3.87 ($p = 0.11$)	3.35	3.68	3.84
6	7.85 ($p = 0.20$)	6.98	7.43	7.62

DISCUSSION

In this study, we gathered the data of 3 experiments in order to maximize the number of subjects, and investigated the feasibility of predicting/explaining MI-BCI performances, independently of the experiment, using a statistical model based on the participants' traits only. We were able to find a model reaching significance for the *Group 1* ($p<.05$) with an average prediction error of 3.03%. This model included three main factors: Self-reliance, Tension and the Mental Rotation scores. Those factors were only slightly different from the ones revealed, on the same dataset, by Jeunet et al. [13] using a step-wise-linear-regression. Indeed, the Self-Reliance, Apprehension, visual/verbal subscale of the Learning Style and the Mental Rotation scores were included in their model. It should be noted that both the Apprehension and Tension factors are related to the same global factor, Anxiety. Besides, the Apprehension and Self-Reliance factors were also selected in 80% of the CV models of the *Group 4* (even though the reliability for this group was not better than chance –p=0.37–). However, these factors were not automatically included in all the models for the other groups. For instance, in the *Group 5* (XP1 and XP3), no factor representing the anxiety of a subject was selected in the CV models (Fig. 4). Regarding the Mental Rotation scores, they were selected in both our models and in [13]. This result stresses that this parameter has a strong influence on MI-BCI performances. Further-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-39

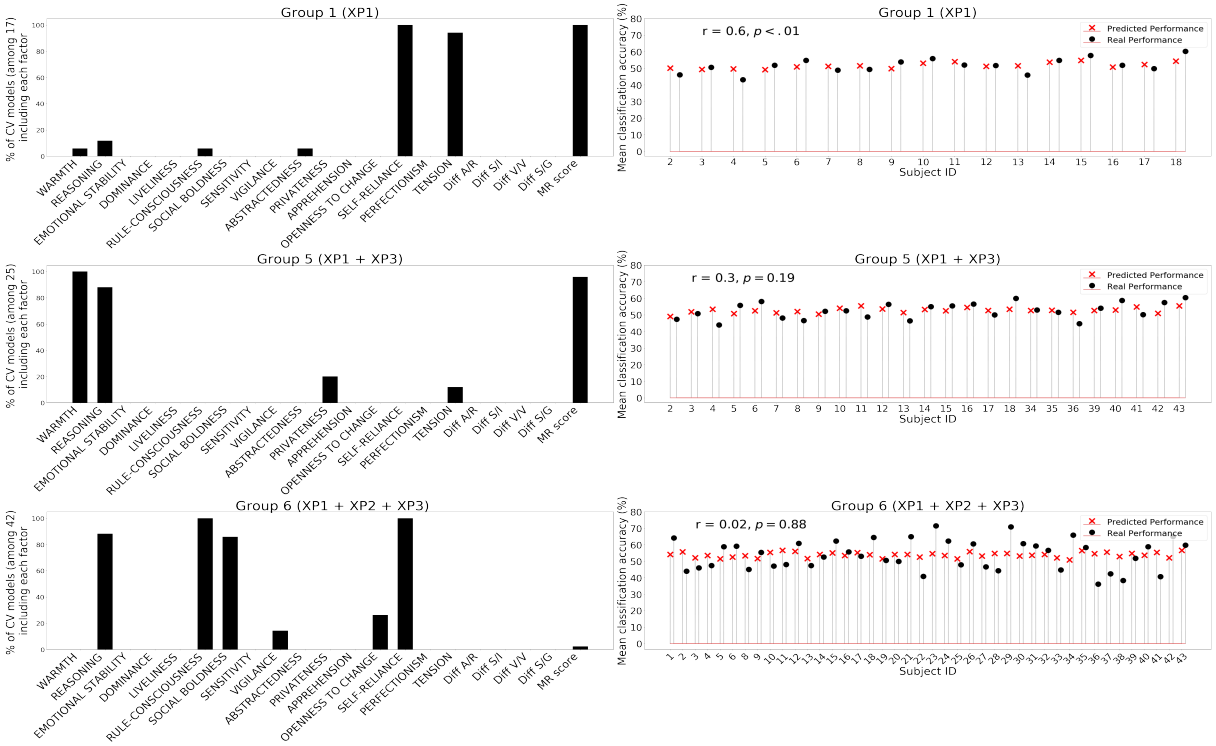


Figure 4: Results of the different models generated for Groups 1, 5 and 6. On the left, the percentage of Cross-Validation models including each factor. On the right, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). Finally, in the right plots, the correlation between the real and predicted performances. Only the models for group 1 had better than chance predictions.

more, even though the other models failed to reach significance, in five out of the six groups, the Mental Rotation scores were included in a large majority of the CV models. Interestingly enough, both XP2 and XP3 had been designed to influence the factors that had been identified in XP1 [13]: with a spatial ability training in XP2 dedicated to the improvement of the participants' Mental Rotation scores, and a learning companion in XP3 aiming to help the anxious and non self-reliant participants. Therefore, it is consistent to observe a reduced influence of the Mental Rotation scores and Apprehension/Self-Reliance factors in the groups including the data of XP2 and XP3, respectively. Still it is interesting to notice the tendency towards a stable and reliable model when grouping XP1 and XP3. In fact, the Mental Rotation scores were selected for all the CV models except one. It strengthens the fact that having good spatial abilities might have a positive influence on MI-BCI performances [14, 15].

Furthermore, only the average error of *Group 1* models reached significance, which could be due to the many differences existing between the 3 studies. Indeed, they did not include the same number of participants, nor the same number of MI-BCI sessions. The number of sessions over which performances are averaged is likely to influence the average performance variability (because of between-session variability due to, e.g., fatigue or motivation changes). Those variations can be significant and become an issue when computing the mean performance over all the MI-BCI sessions. Averaging performances over 6 sessions enabled us to reduce the between-session variability, and make the mean MI-BCI performance estimation more accurate. This might be one possible explanation of the fact that we only found stable and reliable models for the *Group 1* (XP1). As Traits are supposed to remain stable in time, having a more stable measure of performance (here with more sessions) could help us to find a more reliable model. Alternative metrics of performance, reflecting users performances rather than classifier performances, such as the ones proposed in [30], could also be used in the future to predict or explain BCI performances better.

Regarding the LASSO regression, it appears to be more stable than a regular linear regression as only a few factors were chosen for most of the groups. However, by using the LASSO we hypothesized that there was a linear correlation between those factors and MI-BCI performances, while it could be non-linear. In the future, combining users' traits with their states (e.g., inferred from neurophysiological data or questionnaires) could help us explain the between- and within-subject variability and better explain/predict MI-BCI performances over several sessions, but also better predict performances per session, per run and per trial.

CONCLUSION

In this study, we used a LASSO regression to determine experiment-independant predictive and explanatory

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-39

model of MI-BCI users' performances using their traits alone. Our results suggest that using only traits might not be sufficient to build such a model. Indeed, the between-session variability is high and seems to be multi-factorial. Further studies considering, for instance, an estimation of the users' states, the timing of the experiment and new metrics to assess performances are necessary to reveal more reliable models.

Acknowledgments This work was supported by the European Research Council with project BrainConquest (grant ERC-2016-STG-714567).

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] Pfurtscheller G, Neuper C, Flotzinger D, Pregenzer M. EEG-based discrimination between imagination of right and left hand movement. Electroencephalogr Clin Neurophysiol. 1997;103(6):642–651.
[3] Lécuyer A, Lotte F, Reilly RB, Leeb R, Hirose M, Slater M. Brain-computer interfaces, virtual reality, and videogames. Computer. 2008;41(10).
[4] Clerc M., Bougrain L., Lotte F. Brain-Computer Interfaces 1: Foundations and Methods. ISTE-Wiley, 2016.
[5] Guger C, Edlinger G, Harkam W, Niedermayer I, Pfurtscheller G. How many people are able to operate an EEG-based brain-computer interface (BCI)? IEEE Trans Neural Syst Rehabil Eng. 2003;11(2):145–147.
[6] Allison BZ, Neuper C. Could anyone use a BCI? In: Brain-computer interfaces, 2010, 35–54.
[7] Neuper C, Pfurtscheller G. Neurofeedback training for BCI control. In: Brain-computer interfaces.
[8] Lotte F, Larrue F, Mühl C. Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design. Front Hum Neurosci. 2013;7:568.
[9] 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.
[10] Randolph AB, Jackson MM, Karmakar S. Individual characteristics and their effect on predicting mu rhythm modulation. Int J Hum Comp Int. 27(1).
[11] Blankertz Benjamin et al. Neurophysiological predictor of SMR-based BCI performance. Neuroimage. 2010;51(4):1303–1309.
[12] Randolph AB. Not all created equal: individual-technology fit of brain-computer interfaces. In: Proc. HICSS. 2012, 572–578.
[13] Jeunet C, N'Kaoua B, Subramanian S, Hachet M, Lotte F. Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns. PloS one. 2015;10(12):e0143962.
[14] Jeunet C, Jahanpour E, Lotte F. Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study. J Neural Eng. 2016;13(3):036024.
[15] Pacheco K, Acuña K, Carranza E, Achanccaray D, Andreu-Perez J. Performance predictors of motor imagery brain-computer interface based on spatial abilities for upper limb rehabilitation. In: Proc. IEEE EMBC. IEEE. 2017, 1014–1017.
[16] Hammer EM et al. Psychological predictors of SMR-BCI performance. Biol. Psychol. 89(1).
[17] Botrel L, Kübler A. Reliable predictors of SMR BCI performance—Do they exist? In: Brain-Computer Interface (BCI), 2018 6th Int. Conf. on. IEEE. 2018, 1–3.
[18] Hammer EM, Kaufmann T, Kleih SC, Blankertz B, Kübler A. Visuo-motor coordination ability predicts performance with brain-computer interfaces controlled by modulation of sensorimotor rhythms (SMR). Front Hum Neurosci. 2014;8:574.
[19] Teillet S, Lotte F, N'Kaoua B, Jeunet C. Towards a spatial ability training to improve Mental Imagery based Brain-Computer Interface (MI-BCI) performance: A Pilot study. In: Proc. IEEE SMC. 2016, 003664–003669.
[20] Pillette L, Jeunet C, Mansencal B, N'Kambou R, N'Kaoua B, Lotte F. Peanut: Personalised emotional agent for neurotechnology user-training. In: 7th International BCI Conference. 2017.
[21] Ramoser H., Muller-Gerking J., Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Neural Syst Rehabil Eng. 2000;8(4):441–446.
[22] Lotte F. Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain–Computer Interfaces. Proc. IEEE. 2015 (;103):871–890.
[23] Kolb DA. Learning style inventory: Version 3. Hay/McBer Training Resources Group, 1999.
[24] Cattell RB, P. Cattell HE. Personality structure and the new fifth edition of the 16PF. Educ Psychol Meas. 1995;55(6):926–937.
[25] Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills. 1978;47(2):599–604.
[26] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. 1996;267–288.
[27] Fonti V, Belitser E. Feature selection using lasso. VU Amsterdam Research Paper in Business Analytics. 2017.
[28] Xu H, Caramanis C, Mannor S. Robust regression and lasso. In: Proc NIPS.
[29] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: a closer look on BCI results. Int J Bioelectromagn. 2008;10:52–55.
[30] Lotte F., Jeunet C. Defining and quantifying users' mental imagery-based BCI skills: a first step. J Neural Eng. 2018;15(4):046030.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-40

CLASSIFICATION OF VARIOUS GRASPING TASKS BASED ON TEMPORAL SEGMENTATION METHOD USING EEG AND EMG SIGNALS

J.-H. Cho, J.-H. Jeong, K.-H. Shim, and S.-W. Lee

Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

E-mail: sw.lee@korea.ac.kr

ABSTRACT: Electroencephalography (EEG)-based brain-computer interface (BCI) system is useful for rehabilitation and external device control. In this study, we analyzed the decoding of five different grasping tasks in actual movement and motor imagery (MI) paradigms. Eight healthy subjects participated in this experiment. They executed and imagined five sustained grasping tasks. In this actual movement and MI experiment, we proposed a muscle-related temporal segmentation method by detecting the electromyograph (EMG) signals. At the same time, we used common spatial patterns (CSP) and the regularized linear discriminant analysis (RLDA) for the EEG analysis. As a result, we classified the five different grasping tasks in the offline experiment and obtained average classification accuracies of 60.85± 7.71% for actual movement and 37.95±7.86% for MI, respectively. This result is encouraging, and the proposed method could potentially be used in future applications, such as a BCI-driven robot hand control for handling various daily use objects such as cup, ball, pencil, bolt, and card.

INTRODUCTION

Brain-computer interfaces (BCIs) are promising tools for detecting user intention and controlling robotic devices such as upper limb prosthesis [1]. Many research groups use electroencephalography (EEG) based BCI because of its cost-effectiveness and convenience. The EEG signals which are generated by brain activities can be separated from different paradigms or physiological phenomena such as visually evoked potentials, slow cortical potentials, and sensorimotor rhythms. In particular, sensorimotor rhythms include two kinds of amplitude modulations known as event-related desynchronization (ERD) and event-related synchronization (ERS) that are generated by sensory stimulation, motor behavior, and mental imagery [2, 3]. ERD and ERS are often used in such movement-related BCI studies and detected in the mu band [8-13] Hz. In this study, we used common spatial patterns (CSP) to avoid using the mu band only and extracted features from the extended frequency bands. The CSP method shows clear patterns and spatial features of brain activities so that it is effective in classifying motor imagery (MI) and actual movement [3].

Three related kinds of research inspired our study. Schwarz et al. [4] tried to decode natural reach and grasp actions from human EEG. They attempted to identify three different executed reach and grasp actions, namely lateral, pincer, and palmar grasp, utilizing EEG neural correlates. Throughout the experiment, they achieved binary classification accuracies of 74.2% between grasp types and 65.9% for the multi-class grasp types classification at the offline actual movement experiment. In the other case, Ofner et al. [7] had encoded single upper limb movements in the time-domain of low-frequency EEG signals. The primary goal of the experiment was to classify six different movements, and those are elbow flexion, extension, hand grasp, spread, wrist twist left, and twist right. They achieved the final classification accuracies of 55% in actual movement and 27% in MI.

Agashe et al. [8] had decoded hand motions with a different approach. They demonstrated that global cortical activity predicts the shape of the hand during grasping. It was an offline study, and they inferred from EEG hand joint angular velocities as well as synergistic trajectories as subjects perform natural reach-to-grasp movements. Classification accuracy between the predicted and the actual movement kinematics was 49%. They showed real-time closed-loop neuro-prosthetic control of grasping by an amputee and the feasibility of decoding brain signals of a variety of hand motions. However, these three related studies could not achieve robust decoding performance in MI paradigm due to its complex characteristic of the brain signal data. We tried to solve the problem with a new approach and perspective.

In this paper, we present the classification of grasping tasks within EEG signals for non-invasive BCI. The objective of this study is to confirm whether the method of muscle-related temporal segment selection by detecting the change of signals from each electromyograph (EMG) channel improves the BCI performance of each subject or not. At the same time, we proved the feasibility of classifying various grasping tasks in the right hand from EEG signals with the proposed method in the MI paradigm. The EEG signals from different participants were acquired, and we only selected muscle-related temporal segments from a whole-time domain. With the running of sufficient experiment trials and data analyses, we were able to construct a robust decoding model using our proposed

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-40

Figure 1: Description of the five grasping tasks and the experimental setup. (**left**) Five different reach-and-grasp tasks and the starting position. Five different grasping actions are matching to specific objects. (**right**) Experimental environment for EEG and EMG signals acquisition.

method to classify five different grasping tasks, which can be contributed to developing a robot hand control system driven by EEG signals and based on MI in the future.

MATERIALS AND METHODS

Participants: Eight healthy subjects with no history of neurological disease were recruited for the experiment (right-handed males aged 24-33 years). At the same time, we purposely chose every subject who already experienced the BCI experiment in order to maintain the stability of EEG and EMG signals as musch as possible. For the best results and to verify the feasibility of decoding complex grasping tasks, we only acquired data from trained subjects.
This study was reviewed and approved by the International Review Board, at Korea University [1040548-KU-IRB-17-172-A-2], and written informed consent was obtained from all participants before the experiements.

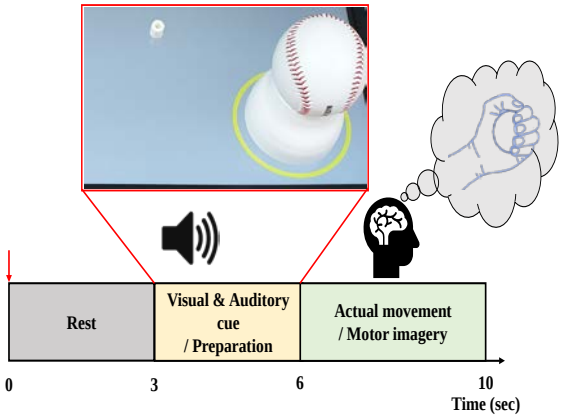


Figure 2: Illustration of experimental protocol to decode five class reach-and-grasp actions in actual movements and motor imagery paradigm.

Experimental Setup: During a session of the experimental protocol, the subjects sat in front of a 32-inch LCD monitor screen, in a comfortable chair. The screen was installed horizontally on the table to make subjects can see the objects and visual cue (flashing yellow circle around the targeted object) without moving their head. Figure 1 indicates the experimental setup and the environment during the entire session. The subjects were asked to perform or imagine a specific grasping task following each auditory and visual cue. The subject imitated the designated grasping tasks, including a half-opened hand shape that indicates muscle relaxation during the rest period (in the starting position). During the experiment, the subjects were asked to perform five different grasping tasks, which are illustrated in Figure 1.

Data Acquisition: EEG and EMG signals were collected using maximum possible channels for data collection for future analysis. During the experiment, the EEG data were collected at 1000 Hz using 64 Ag/AgCl electrodes in 10/20 international system via BrainAmp (BrainProduct GmbH). At the same time, a 50 Hz notch filter was used to remove power frequency interference. The FCz and FPz were used as reference and ground electrodes, respectively. All impedances were maintained below 10 kΩ. From the entire 64 EEG channels, only 20 channels (FC5, FC3, FC1, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, and CP5) were used for data processing, because we expected to suppress the artefacts such as signals from eye movements by removing peripheral channels except on the motor cortex. The 20 channels were located only on the motor cortex to make sure that the recorded EEG signals are highly related to the motor-related potentials, which are from the actual movement and MI, as shown in Figure 3.
In the case of EMG data acquisition, we used 5 Ag/AgCl electrodes (CH1: extensor pollicis brevis, CH2: extensor digitorum, CH3: flexor digitorum profundus, CH4: flexor digitorum superficialis, CH5: reference) from the same equipment during EEG data recording. The EMG channels were attached on the right arm to detect signal

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-40

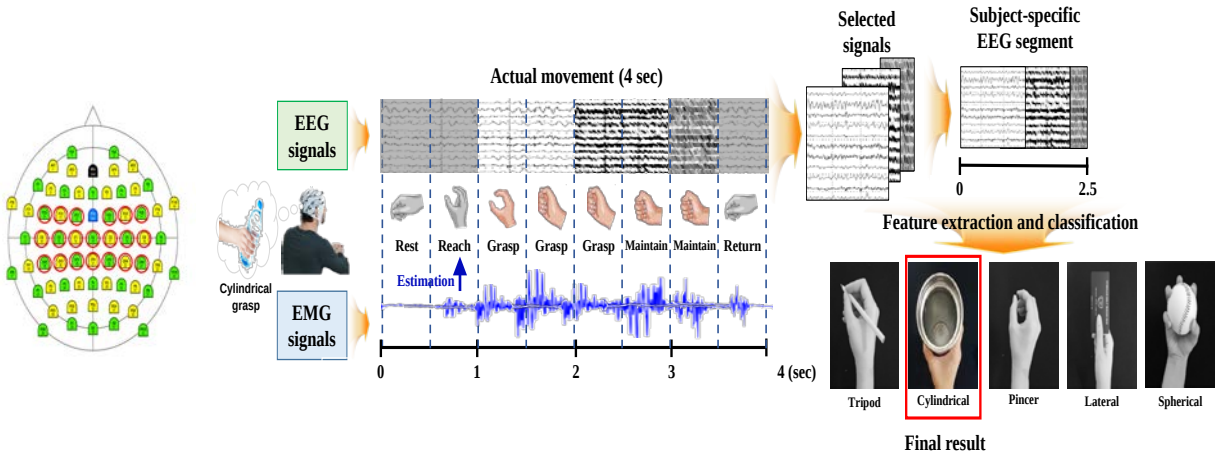


Figure 3: (**left**) EEG location of selected 20 channels that minimize the artefacts. (**right**) Proposed method to select subject-specific time segments during the classification. Measuring EMG signals to find the optimized time segments for feature selection on a trial. Select a specific 2.5-second length of time segment from 0 to 4s total length of a single trial.

change while the subject performed grasping tasks [7]. The one channel on extensor digitorum (CH2) was selected, and the EMG signals were first processed by 50 Hz notch and filtered by a 5-500 Hz band-pass filter [8, 9].

Every subject performed 75 trials per each grasping task (375 trials: 5 classes x 75 trials). They were also asked to perform or imagine a specific grasping task once but not repetitively in the actual movement/motor imagery period of 4 sec, shown in Figure 2.

Data Analysis: Filtering the EEG signals was a significant process in the experiment. We grouped the three most used motor-related EEG bands mu [8-13], beta [13-30], and delta [0.1-3] in Hz using a 2nd order of Butterworth filter [5, 10] and tested for the best performance. In particular, we chose [0.1-3] Hz to decode our EEG data because it showed the best classification result compared to using the other bands in this experiment. EMG signals were analyzed to detect muscle-related temporal segments by measuring the activation of muscles when the subject performed the actual movement. We calculated the mean integrated EMG (IEMG) as the indicator of the muscle activity during each grasping task. IEMG is capable of reflecting the contraction property of muslces [8-10].

It is defined as:

$$IEMG = \Sigma_{i=1}^{n} |x_i| \tag{1}$$

Here, x_i represents the ith sample of EMG signal and n represents the length of time segment.

We separated one trial of 4 seconds in eight (0.5 seconds for each segment) and selected only muscle activation point when performing grasp. We excluded the segmented EEG signals when the proposed model

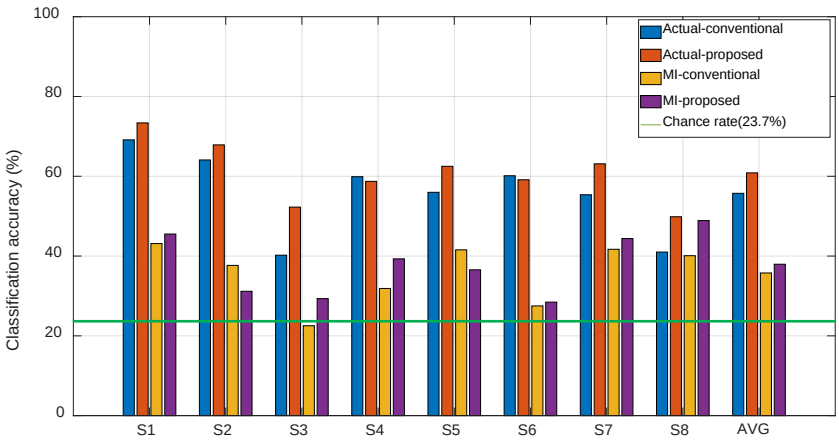


Figure 4: Classification accuracies for decoding 5-class grasping tasks in [0.1-3] Hz. Average MI decoding accuracy was 37.95% and average actual movement accuracy was 60.85%, respectively. 10 by 10 cross-validation was applied.
***Conventional: using fixed 4-second temporal segmentation to extract CSP features**
***Proposed : using subject-specific temporal segmentation of 2.5-second based on muscle-related signals**

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-40

Table 1: Classification accuracies of 5-class grasping task according to motor-related frequency bands

	Actual movement (Cylindrical / Spherical / Pincer / Tripod / Lateral)			**Movement imagery** (Cylindrical / Spherical / Pincer / Tripod / Lateral)		
	[0.1-3] Hz	**[8-13] Hz**	**[13-30] Hz**	**[0.1-3] Hz**	**[8-13] Hz**	**[13-30] Hz**
Sub 1	**73.38 %**	65.80 %	57.70 %	**45.52 %**	41.25 %	43.55 %
Sub 2	**67.87 %**	33.60 %	41.40 %	31.17 %	**36.01 %**	27.53 %
Sub 3	**52.27 %**	37.93 %	43.80 %	**29.33 %**	24.75 %	28.65 %
Sub 4	**58.73 %**	45.60 %	38.75 %	**39.30 %**	25.25 %	33.82 %
Sub 5	**62.50 %**	53.73 %	45.58 %	36.55 %	**37.09 %**	55.48 %
Sub 6	59.12%	**69.06%**	57.40%	**28.45%**	22.54%	21.99%
Sub 7	**63.12%**	51.43%	58.32%	**44.39%**	30.03%	29.32%
Sub 8	**49.85%**	43.44%	35.93%	**48.90%**	34.94%	44.33%
Avg.	**60.85 %**	50.07 %	47.36 %	**37.95 %**	31.48 %	35.58 %

estimates the stages are in progress of rest, reach, and return. This temporal segment indicates when the subjects actually move their hand to grasp the designated object. These temporal segments are subject specific so that if a subject has the temporal segment in between 1 second to 3.5 seconds (2.5-second length), we can apply this segment to analyze training and test data. However, we need to find differernt temporal segments for the other subjects.

After selecting these temporal segments, we applied the segment to cut the original EEG data in limited length so that eventually we could pick the data at the movement-related time point. Common spatial patterns (CSP) and regularized linear discriminant analysis **(**RLDA) were applied for further feature extraction and classification.

Of course, we could not measure any distinguishing characteristic in EMG data during the MI paradigm [10, 11]. However, applying the trained temporal segment using the data from the actual movement experiment was effective to most of our subjects when classifying MI.

RESULTS

Through the experiment, we confirmed the feasibility of classifying five different grasping tasks in the right hand. Figure 4 presents the classification accuracy results in actual movement and MI with bar graphs. In the actual movement session, in the red bars of Figure 4, every subject excluding S4 ans S6 showed increasing trends of accuracies approximately 3 to 12% compared to the conventional method. Classification accuracies of S3 and

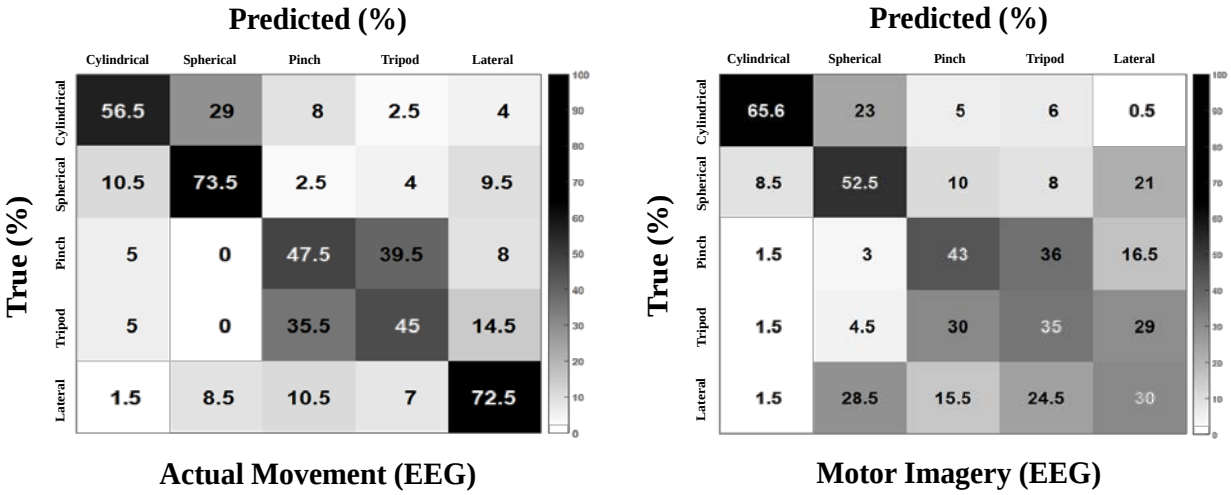


Figure 5: Confusion matrices of (**left**) actual movement classification and (**right**) MI classification results. Showing significant differences among each class especially on the power grasping tasks (cylindrical and spherical) and precision grasping tasks (pinch, tripod and lateral). The classification result of MI shows a similar tendency to the result from the actual movement.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-40

S7 when using 2.5-second length of time segmentation were dramatically improved and finally peaked at 63.12% while the other subjects showed little increase compared to the conventional method. On the other hand, during the MI session, the classification accuracies continuously increased, while some of the subjects showed a decrease. Most of them reached the higher classification accuracies when we applied our proposed decoding method. In the case of the S2 and S5, the classification accuracy started to decrease, but the subjects showed little decrease, while the other subjects showed a more significant increase in accuracy. For example, S4 and S8 achieved dramatic improvements of 7.42% and 8.81%, respectively.

According to the results, each subject showed the highest actual movement or MI performance at a different point in time, but most of them showed increase in performance with applying muscle-related temporal segment selection method.

The different results in five class multi-classification accuracies between actual movement and MI are strongly affected by the muscle-related time segment selections. For some subjects, such as S4 and S8, more improvement was shown in the classification accuracy when adopting the proposed method that was trained with actual movement data and applied in the MI paradigm than when using the same method with the actual movement paradigm. For instance, during the MI classification, the average accuracies of the proposed method increased by almost 2 to 9%. These were higher than the accuracies of the conventional model, using an ordinary fixed 4-second temporal segment.

Table 1 presents entire classification results of all subjects in three different motor-related frequency bands. First, we analyzed classification accuracies in delta band [0.1-3] Hz. Several related researches already showed the feasibility of decoding upper limb or hand movements in low frequency EEG [4, 5]. In the delta band most of subjects showed the best classification accuracy compared to the other bands. Traditionally the beta and mu bands are widely used to decode actual movement and MI from EEG signals. The classification results in the both two motor-related frequency bands are similar. Some of subjects showed the best results at the [8-13] or [13-30] Hz but most of them are performed the best in low frequency.

DISCUSSION

In this study, we confirmed the effectiveness of using our proposed muscle-related temporal segment selection to improve classification accuracy in both of actual movement and MI paradigms. As we can see in Figure 4, the results showed the feasibility of the proposed method as a temporal segment selection strategy feature to increase multi-class classification accuracy on set purpose, especially in actual movement paradigm. Although the visual cue-based synchronous system executed the experiment in an offline scenario, the subjects completed MI with different speed, trajectory, and force. We could see that the muslce-related temporal segment selection is working and adaptive in the MI paradigm even though it was trained with actual movement data. We suppose it is because the subjects imagine in a similar way and rhythm to actual movement. During each of the trials, the subjects only used intuitive motor imagery when they were asked to perform the specific grasping task. Hence, selecting the muscle-related temporal segment length to extract the apparent movement-related EEG features at the exact time interval for each particular individual during MI paradigm is useful when the situation in which the subjects perform is unobservable and there are variable motor-related brain activities.

CONCLUSION

Classifying five different grasping tasks from EEG signals within a hand is very challenging, but fortunately, we confirmed the feasibility. The results showed improvements to the overall classification accuracies with our proposed method. In the future work, we will focus on increases in reliability and efficiency to develop a robust MI decoding method based on muscle-related signal data with advanced machine learning and deep learning approaches. We will also modify our offline experimental protocol to be asynchronous and online-oriented for the development of practical BCI based robot hand control.

ACKNOWLEDGMENTS

This work was partly supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (No. 2017-0-00432, Development of Non-Invasive Integrated BCI SW Platform to Control Home Appliances and External Devices by User's Thought via AR/VR Interface) and partly funded by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (No. 2017-0-00451, Development of BCI based Brain and Cognitive Computing Technology for Recognizing User's Intentions using Deep Learning).

REFERENCES

[1] V. Gilja, C. Pandarinath, C. H. Blabe, P. Nuyujukian, J. D. Simeral, A. A. Sarma, B. L. Sorice, J. A. Perge, B. Jarosiewicz, L. R. Hochberg, K. V. Shenoy, and J. M. Henderson, "Clinical Translation of a High-Performance Neural Prosthesis," *Nature Medicine*, Vol. 21, No. 10, 2015, pp. 1142-1145.

[2] P. Gert, "EEG Event-Related Desynchronization (ERD) and Synchronization (ERS)," *Electroencephalography and Clinical Neurophysiology*, Vol. 103, No. 1, 1997, pp. 1-26.

[3] P. Gert, and A. Aranibar, "Evaluation of Event-Related Decynchronization (ERD) Preceding and Following Voluntary Self-Paced Movement," *Electroencephalography and Clinical Neurophysiology*, Vol. 2, No. 46, 1979, pp. 138-146.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-40

[4] A. Schwarz, P. Ofner, J. Pereira, A. L. Sburlea, and G. R. Müller-Putz, "Decoding Natural Reach and Grasp Actions from Human EEG," *Journal of Neural Engineering*, Vol. 15, No. 1, 2018, pp. 1-14.

[5] 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, 2017, pp. 1-24.

[6] H. Agashe, A. Y. Paek, Y. Zhang, and J. L. Contreras-Vidal, "Global Cortical Activity Predicts Shape of Hand During Grasping," *Frontiers in Neuroscience*, Vol. 9, No. 4, 2015, pp. 1-11.

[7] H.-I. Suk and S.-W. Lee, "A Novel Bayesian Framework for Discriminative Feature Extraction in Brain-Computer Interfaces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 2, 2014, pp. 286-299.

[8] K. Wang, Z. Wang, Y. Guo, F. He, H. Qi, M. Xu, and D. Ming, "A Brain-Computer Interface Driven by Imagining Different Force Loads on a Single Hand: an Online Feasibility Study," *Journal of NeuroEngineering and Rehabilitation*, Vol. 93, No. 14, 2017, pp. 1-10.

[9] A. Fu, C. Wang, H. Qi, F. Li, Z. Wang, F. He, P. Zhou, S. Chen, and D. Ming, "Electromyography-based Analysis of Human Upper Limbs During 45-Dat Head-Down Bed-Rest," *Acta Astronautica*, Vol. 120, 2016, pp. 260-269.

[10] X. Li, O. W. Samuel, X. Zhang, H. Wang, P. Feng, and G. Li, "A Motion Classification Strategy based on sEMG-EEG Signal Combination for Upper Limb Amputees," Journal of NeuroEngineering and Rehabilitation, Vol. 14, No. 2, 2017, pp. 1-13.

[11] K. Wang, Z. Wang, Y. Guo, F. He, H. Qi, M. Xu, and D. Ming, "A Brain Computer Interface Driven by Imaging Different Force Loads on a Single Hand: An Online Feasibility Study," Journal of NeuroEngineering and Rehabilitation, Vol. 14, No. 1, 2017, pp. 93-103.

[12] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Müller, "Single Trial Analysis and Classification of ERP Componenets – A Tutorial," *NeuroImage*, Vol. 56, 2011, pp. 814-825.

[13] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann, "One the Interpretation of Weight Vectors of Linear Models in Multivariate Neuroimaging," *Neuroimage*, Vol. 87, 2014, pp. 96-110.

[14] C. Umberto, "The Neuroscience of Grasping," *Nature Reviews Neuroscience*, Vol. 6, No. 9, 2005, pp. 726-736.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-41

PERFORMANCE, TRANSFER LEARNING AND UNDERLYING PHYSIOLOGY IN CHILDREN PLAYING P300 BCI GAMES

M. Fouillen[1,2], E. Maby[1,2], M. Partyka[1], V. Herbillon[1,2,3], J. Mattout[1,2]

[1] Lyon Neuroscience Research Center, CRNL; INSERM, U1028; CNRS, UMR5292; Brain Dynamics and Cognition Team, Lyon, F-69000, France

[2] University Lyon 1, Lyon, F-69000, France

[3] Hospices Civils de Lyon, 69000 Lyon, France

E-mail : melodie.fouillen@inserm.fr

ABSTRACT: P300-based BCI are widely explored for item selection but few studies have examined if this interface can be used by children... The aim of the current study was to evaluate the performance of healthy children playing three different calibration-free P300 BCI games. 19 children played all three games in a random order. EEG-based online selection relied on template signals derived from a previously acquired database. All children performed the task significantly well even though all children underwent an inevitable drop of performance when comparing offline (individual) with online (template based) accuracies. Offline analyses revealed no difference in performance between games. Transfer learning from one game to the others proved possible although one game appeared slightly less generalizable. Finally, offline ERP analyses revealed differences in the early (visual) components, which we relate to each game graphical specificity. In contrast, all games did involve a strong contribution of the P300 component, which is essential to support high attention-based control.

INTRODUCTION

The most well-known P300 BCI is the so-called P300-speller which allows the user to spell words without using the peripheral nervous and muscular pathways. This interface has been developed for people with very severe neuromuscular disorders, in the aim of enabling them to communicate [1]. A good control of this interface is partly based on the subject's ability to elicit well distinguishable brain signals after a target and after a non-target stimulus, respectively. Discriminating between those two classes highly depends upon the voluntary engagement of the subject in this selective attention task [2]. A way to enhance the motivation is to provide a more playful environment. With this in mind, P300-based BCI games have been developed [3]. Moreover, it seems also possible to increase the amplitude of the P300 with training, both in the auditory [4] and visual domains [5]. Although these trainings were short and have involved very few participants, an improvement over practice was reported which was concomitant with an increase in the P300 amplitude. With the aim of setting up a BCI-based training for children with ADHD [6] we decided to develop new P300-based games to increase the diversity of the games on offer. To propose various games could be essential to keep up the motivation over long training periods, and to avoid as much as possible the drop out of participants. Another particularity of our training is that children do not control the BCI based on their own calibration signal. Indeed, ADHD children having a diminished P300 compared to controls [7], the brain response of targets and non-targets might not be distinguishable and thus the subsequent classification might not be accurate. The whole purpose of the training is to restore a proper P300 signal in those children. Therefore, we decided to build a template of the expected electrophysiological responses in healthy children and use it as the target electrophysiological response for the training of children with ADHD [6]. We used covariance matrices as features and Riemannian geometry for subsequent online classification [8]. Such an approach has shown very good results for both classification and generalization [9]. The template was built from data of a previous study were children played with a P300-based connect 4 (Fig. 1) [6]. The aim of the present study was first to evaluate the classification accuracy with 3 new games and to assess whether the difference in game configuration would induce differences in electrophysiological responses. Finally, we wanted to evaluate the robustness of transfer learning between games, and between subjects.

MATERIALS AND METHODS

Template construction

The template was built on data from a previous experiment conducted in 34 healthy children. They had to control a Connect 4 game (Fig. 1a). After a short calibration, each child played for about 80 trials. The data from 5 children were finally discarded because of technical problems or too low performance. We thus ended up building template signals on data from 29

Published by Verlag der TU Graz
Graz University of Technology

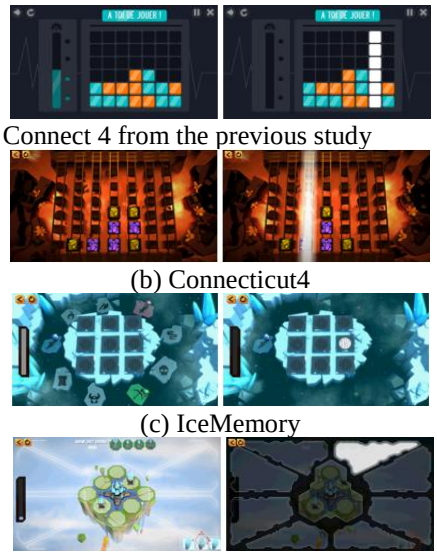
 DOI: 10.3217/978-3-85125-682-6-41

children (6-16 years old; 14 girls). It consists in two covariance matrices, one for the target and one for the non-target class. Precisely, the prototype ERP response is obtained by averaging the single trial responses from the target class. Then each single target trial is concatenated with this prototype ERP to create so-called super trials. These super trials are used to build covariance matrices thanks to the Sample Covariance Matrix estimator. These matrices are then averaged using the Riemannian mean to obtain the typical target covariance matrix. The same procedure applies to obtain the typical non-target covariance matrix [9].

Experimental setup: 19 healthy children (6-16 years old; 11 girls) took part in this new experiment. Children had never used a BCI before and reported normal or corrected-to-normal vision. The study was approved by the ethics committee n°2016-013B. EEG signals were recorded from 16 channels using an active EEG electrode system and a Vamp amplifier (Brain Products, Germany). EEG channel locations were Fz, Cz, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO9, O1, Oz, O2 and PO10 following the international 10-20 system.

Experimental procedure: The 3 games have been designed and implemented by Black Sheep Studio[1] using Unity 3D [10]. The first one, Connecticut4 (C4) (Fig. 1b) is a connect 4 game, whose aim is to align 4 pawns before the computer does. For the second game, IceMemory (IM) (Fig. 1c), the aim is to memorize and find cards to make one's own character move and grab the opponents. The third game named Armageddon (AR) (Fig. 1d), is a strategic game where the goal is to protect an island from asteroids. In each game, each possible target (7 in C4, 9 in IM and AR) was visually intensified 6 times in random order before a decision was made. Both the flash duration and the interstimulus interval (ISI) were set to 100ms. Children were instructed to focus their (overt) attention onto the target and count the number of times it was flashed. To ensure that the children did not suffer from attentional disorder, parents filled-up the ADHD rating scale. One child has been removed from the analyses because of a high inattention score. We used an eye-tracker (ET, SMI REDn scientific 60Hz) to know about the targets chosen by the children. After a short calibration of the ET, children could start playing directly, thanks to the above-described template. Each child played all three games (about 20 trials each). We counterbalanced the order of the games over children.

Online processing: EEG data were processed in real-time within a home designed pipeline coded in Python among which Pyacq (https://github.com/pyacq/pyacq). Data were sampled and bandpass filtered between 1 and 20Hz. After each flash, the epoched signal (0-600ms) was concatenated with the prototype ERP response (see



(a) Connect 4 from the previous study

(b) Connecticut4

(c) IceMemory

(d) Armageddon

Figure 1: Screenshots of the BCI games, in the absence of flashes (left panel) and during one flash (right panel).

template construction section) to build a covariance matrix. Using the Riemannian distance, this covariance matrix was then compared to the target and non-target covariance matrices, respectively. We then used as a feature for (probabilistic) classification, the log ratio between those two distances. The result of this online processing was sent to the BCI-game by TCP using a ZeroMQ socket [11]. The true target location was given by analyzing the eye-tracking measures. Therefore, seven or nine zones, depending on the game, were defined on the screen, corresponding to the seven or nine potential targets. At each trial we obtained a vector of seven or nine values, providing to amount of time the child had spent looking at each zone. We considered as the target, the area that was looked at the most.

Electrophysiological offline analyses: The EEG data were finely analyzed offline, with the MNE software package [12]. EEG data were filtered between 1 and 20Hz. Then the signal was segmented into epochs of 800ms (-200ms to 600ms peri-stimuli time) and a baseline correction was applied. Epochs with an EEG signal above 150 µV or below -150 µV were marked as artifactual and discarded. To identify spatio-temporal differences between target and non-target epochs, a cluster-based permutation test was performed at the group level. Then for each spatio-temporal significant cluster we computed the averaged amplitude of the signal, for each game (C4, IM, AR) and each class (Target and Non-Target). We constructed a linear model of the averaged amplitude as a function of Games and Classes. The between subject variability limits the comparison and means that data cannot simply be pooled for analysis. Using a linear mixed-effect model (lme4 package, Linear Mixed Effects version 4) [13] is the best

[1] Black Sheep Studio is an SME located in Paris, France specialized in video games (blacksheep-studio.com).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-41

way to deal with such datasets, as they allow for correction of systematic variability. We accounted for the heterogeneity of accuracy values across subjects by defining them as effects with a random intercept, thus instructing the model to correct for any systematic differences between subjects (interindividual variability). We used a binomial distribution to describe the model errors.
We then analyzed the influence of two possible fixed effects onto signal amplitude:
(i) the Game effect (three levels) (ii) the Class effect (two levels). We ran a type II analysis of variance. Wald chi-square tests were used for fixed effects in linear mixed-effect models. For post-hoc tests we used the Lsmean package (Lsmean version 2.20-23) [14] where effects were considered as significant when $p<0.05$ and adjusted for multiple comparisons (Tukey method). All statistical analyses were performed using the R Statistical Software.

Self versus template accuracy: To evaluate the efficacy of the template, we compared in each child, the self-accuracy with the template-based one. The latter corresponds to the actual online accuracy experienced by the children.
Self accuracy refers to a (theoretical) offline measured performance that is only based on the user data. We computed it following a cross-validation procedure. Therefore, we splitted each child's data into a training set (75% of the data) and a testing set (the remaining 25%). This random split was repeated 500 times to compute an estimate of what we refer to as self accuracy. This measure was used for comparison with the template accuracy.
It was also used for comparing performance over games using a linear model of self accuracy as a function of games. Finally, we also evaluated the self accuracy within games, following the same computational procedure as above, but considering each of the three games independently.
Computing the self accuracy this way does not allow us to take in account a potential effect of time. To evaluate such a possible additional effect, we used the data of the first game presented to the children as the training set and tested the performance on the subsequent two games. We compared the generalization performance to the second and third game using a two-sided t.test.

Transfer between games: In order to evaluate the transfer between games, we computed the target and non-target covariance matrices with the signal of one game and performed the classification on the two other games We then constructed a linear model of the accuracy as a function of the game used for training. For post hoc tests we used the Lsmean package where P-values were considered as significant at $P<0.05$ and adjusted for the number of comparisons (Tukey method).

Relating gaze fixation and BCI accuracy: To have a more precise measure of the gaze fixation we computed a gaze fixation index G as follows:

$$G = 1 - \frac{Sobs}{Smax}$$

where Sobs is the Shannon entropy of the actual gaze orientation over possible location and Smax is the theoretical maximum entropy used for normalization (it corresponds the case where the child would have looked at all targets for the same amount of time. A G close to 1 means a very good gaze fixation on the target. Conversely, a G close to 0 corresponds to a trial where children had a poor fixation performance. For each child, we averaged this index over all trials. We then computed the Pearson correlation coefficient between G and self-accuracy, over subjects.

Good and poor performers: Relating G and self-accuracy revealed two sub-groups, one of good performers and one of poor performers. We thus decided to test the difference between targets and non-targets signals for those two groups separately. We constructed a linear model for each cluster and each game of the average signal amplitude as a function of Classes and Performance (good vs. poor performers).
We then analyzed the influence of two possible fixed effects on signal amplitude as in the section: Electrophysiological offline analyses.

RESULTS

Electrophysiological analyses: Over all games and subjects, the cluster-based permutation test revealed 4 significant clusters showing differences between targets and non-targets. The first cluster corresponded to the P100 component in occipitals areas (Time: 67-98ms; sensors: O1 and O2). The second one corresponded to the left hemisphere N200 (Time: 160-229ms; sensors: CP5, P7 and P3) and the third one to the right hemisphere N200 (Time: 194-204; sensor: CP6). These two clusters have been combined into one, named further the N200 cluster (Time: 194-204ms; sensors: CP5, P7, P3 and CP6). The last cluster corresponded to the P300 component (Time: 244-481ms; sensors: CP5, CP6, P7, P3, P4, P8, O1 and O2) (Fig. 2).

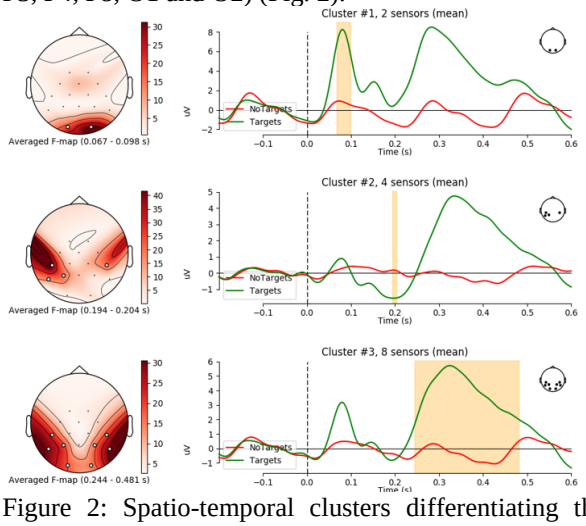


Figure 2: Spatio-temporal clusters differentiating the target and the non-target responses over all three games.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-41

For each of the 3 above clusters, each game and each class, we computed the mean amplitude of the ERP. For the P100, the 3*2 repeated ANOVA (game*class) revealed a main effect of class ($p < 0.001$) and a significant interaction. For both the N200 and P300, the ANOVAs revealed a main effect of the class only (Fig. 3).

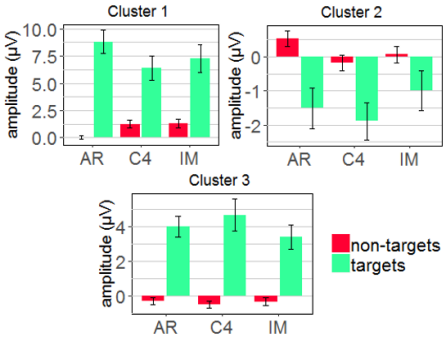


Figure 3: Averaged amplitude of the ERP for the targets and the non-targets on each cluster and for each game. There is a main effect of the class for each cluster and a significant interaction games*classes for the cluster 1. Error bars indicate S.E.M.

Evaluation of the games: Whether the learning was computed on all games or a single game only, the repeated measure ANOVAs showed no significant difference between the game classification accuracies. We found no difference between the accuracy computed on all games and the ones computed on each game independently (learning on all the games: mean = 84.79%; S.E.M = 1.82; learning on one game: mean = 86.45; S.E.M = 1.8). For subsequent analyses we thus considered the self-accuracy computed on all games only. Finally, we found no effect of time as we found no significant difference between the two classification accuracies when learning on the 1st game data (generalization to 2nd game: mean = 67.00%, S.E.M. = 5.34; generalization to 3rd game: mean = 70.33, S.E.M. = 4.90).

Transfer between games: To evaluate whether transfer is possible between games, we computed accuracies where the learning was based on trials from one game and the testing was performed on all trials from the other two games. The ANOVA showed a significant effect game used for learning. Post hoc analyses revealed higher accuracy when the learning was based on IM and AR trials compared to when it was based on C4 ones. Besides, we found a significant decrease of theses accuracy ($p < 0.001$) compared to self-accuracy.

Template evaluation: To test the possibility of using directly the template to control the games and thus get rid of the calibration, we compared the self-accuracy computed offline to the template-based accuracy. The t-test showed a significant decrease of accuracy when the children played with the template (mean = 0.55; S.E.M = 0.05). At the individual level, all children obtained a lower accuracy when playing with the template compared to self-accuracy. However, we observed a significant correlation between self-accuracy and template-based accuracy (cor = 0.62; p = 0.006). The more the children succeed in controlling the games (as measured by self-accuracy), the higher their performance calculated based on the template. When they played with the template all children but one performed above chance level. The comparison between the 3 games showed no significant difference between games.

Correlation between gaze fixation and BCI accuracy
There was a significant correlation between the gaze fixation index and the self-accuracy (cor = 0.77; p <0.001). Hence BCI accuracy increased with the gaze fixation. This correlation also revealed two groups, one of 11 children referred to as good performers, with accuracy above 85% and a gaze fixation index above 0.90; and a second of 7 children referred to as poor performers, with an accuracy below 85% and a gaze fixation index below 0.90 (Fig. 4).

Good vs. Poor performers: For the P100, the 2*2 repeated measure ANOVA (good/poor performers * classes) we obtained a main effect of class and no effect between good and poor performers for all games. For the N200, we obtained a main effect of class for C4 and a trend for IM (p = 0.061). For AR, the interaction proved significant and was driven by the target class: the N200 amplitude was

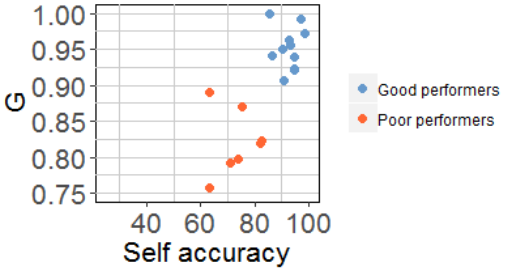


Figure 4: Correlation between the self accuracy and the gaze fixation index (cor = 0.77; p <0.001)

higher for good performers (p = 0.047). Finally, for the P300, we obtained a main effect of class for all three games, a main effect of group for IM and an interaction effect for C4 and IM, with a significantly higher amplitude in good performers compared to poor performers for class target (C4: p = 0.05; IM: p = 0.005) (Fig. 5).

DISCUSSION

This study had two main aims, first to evaluate the 3 new P300-based BCI games and then to test the transfer or generalizability between games. As shown in Fig. 1, stimulation configurations differ between games. In C4, a stimulus corresponds to the flash on one column, so it covers the full height of the screen. In IM the flashes are much smaller and grouped in the center of the screen. On the contrary, in AR the flashes are bigger and cover the

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-41

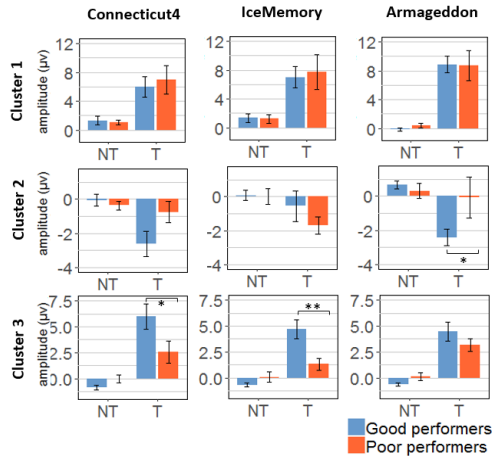


Figure 5: Averaged amplitude of the ERP for target and non-target, in each cluster and for good (blue) and poor (red) performers, respectively. We found a significant difference between good and poor performers for class target in C4 and IM, in the P300 cluster; and for AR in the N200 cluster. Error bars indicate S.E.M.

whole screen altogether. These differences between the flash configurations may induce differences in the ERPs. The ERPs corresponding to the average over target trials from all games, revealed the three components: early visual potential P100; the N200 and the P300 [15], [16]. To evaluate the implication of these three ERP components in differentiating target and non-target stimuli, we compared them for the three games, respectively. The significant difference between targets and non-targets for the three clusters suggests that the three components participate to the classification in all three games. The interaction between games and classes for the first cluster suggests that even if all the games elicited a P100, its relative contribution varies between games. In AR, the non-targets do not elicit a P100 at all, contrary to targets that elicit a high P100. In C4 and IM, the non-targets elicit a small P100 and the difference between targets and non-targets is less pronounced. Visual early potentials (P100 and N200) are larger when the stimulus is foveated [17], which may explain the larger influence of the P100 in AR, where stimuli are large and more far away from each other. Because of this configuration, it is easier for the subject to look at one flash only and to ignore the other (distracting) ones. This explains the absence of P100 for non-targets stimuli in AR. In other games, as the flashes are more grouped, it is more difficult to ignore the non-relevant flashes, that do yield a small P100.

In contrast, the fact that we observe a large P300 component in response to target stimuli, regardless of the game, indicates that the ability to selectively fixate the target is not affected by stimulus size and configuration. Indeed, we did not find any difference in self accuracy between games. A result in BCI performance that is in line with the ERP findings.

We then tested if transfer learning is possible between games. We found a significant difference in performance when learning from C4 trials compared to when learning for the other two game data. C4 appears less generalizable. Reasons for that are not obvious given the ERP findings. However, one should not forget that the ERP clusters do not exactly match the BCI features used online. Covariance matrices are built with the signal of all electrodes and a large time window, while the clusters are spatially and temporally more limited. Including all the electrodes, the design of C4 might induce covariance matrices more distant to the covariance matrices of the other games. When children played based on the template, which was the case online, we obtain a lower accuracy compared to the self-accuracy computed offline following a cross-validation procedure. However, despite this inevitable loss of accuracy when relying on template signals obtained in other children and BCI conditions [18], [19], all children but one performed above chance level online. Many studies have tried to reduce the calibration time without impacting the classification performance. One possibility is to use a subject dependent approach. It consists in using only a few epochs of each class from the subject, to tunes a template built on other subjects. Adding a small amount of data from the new subject seems to be sufficient to compensate for the inter-subjects variability [9]. An approach that could prove useful in this context too, in the near future.

Interestingly, the positive correlation between self-accuracy and template-based accuracy suggest that good performer behave alike, which fits perfectly with our objective to establish a template for the training of ADHD children in order to teach them how to produce a typical P300, which would be the hallmark of the ability to deploy sustained, spatial selective attention.

We also observed a positive correlation between stability in gaze fixation and BCI accuracy. It seems obvious that focusing gaze on the target concurs to focusing attention onto that target. Conversely, a lower gaze fixation index could indicate that children have been distracted by the non-target flashes and made saccade towards them, hence yielding BCI selection errors and a lower accuracy. This correlation also allowed to dissociate two groups: good performers (n = 11) showing a good classification accuracy and a good gaze fixation, and poor performers (n = 7) showing the reverse pattern. This allowed us to compare them in terms ERPs which revealed no difference in early visual potentials (P100) but mostly a difference in the P300 and N200 in a lesser extent. This suggest that although gaze fixation was high overall, including in poor performers (as testified by a large difference in P100), slight instability in gaze fixation is accompanied by some instability in attentional focus and has a dramatic effect on later, attention related, components, especially the P300. Although less reported in this context, the N200 is known to underline both processes of gaze fixation and attention to the target [20]. Altogether, these results emphasize that gaze orientation is mandatory for good BCI performance, but not sufficient as the most important ERP components are the ones related to (covert) sustained attention. A finding that supports the idea that P300 based BCI training could be efficient in ADHD children.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-41

CONCLUSION

We conclude that it is possible to use this kind of template to play BCI P300 games without calibration. There is an obvious drop of performance compared to an individual calibration, but the use of this kind of template may prove sufficient in the context of training. We found no difference in classification accuracies between games but differences in game design yielded differences at the physiological level. This should be further investigated in the future in order to optimize the training, for instance by presenting the games in a specific order. Theses results could also allow to guide the design of new games to target specific components, toward a more individualized approach.

ACKNOWLEDGEMENTS

This work was part of the FUI Mind Your Brain Project (FUI17 inter-ministry fund), funded by BPI-France and *Region Ile-de-France.*
Paul Bonnef, Martin Montrieul, François Kermorvant and Jean-Louis Verlaine from Black Sheep Studio (Paris) developed all the games used in this study under Unity 3D and enabled their interfacing with our BCI system.
The authors are also thankful to Judith Vergne and Alexandre Akel for their help in data acquisition.
Mélodie Fouillen held a doctoral fellowship from *Région Rhônes-Alpes*, from the Fédération pour la Recherche sur le Cerveau and from the FONDATION GROUPE EDF.

REFERENCES

[1] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 115–125, Jan. 2008.

[2] J. Mattout, M. Perrin, O. Bertrand, and E. Maby, "Improving BCI performance through co-adaptation: Applications to the P300-speller," *Ann. Phys. Rehabil. Med.*, vol. 58, no. 1, pp. 23–28, Feb. 2015.

[3] E. Maby, M. Perrin, O. Bertrand, G. Sanchez, and J. Mattout, "BCI Could Make Old Two-Player Games Even More Fun: A Proof of Concept with 'Connect Four,'" *Adv. Hum.-Comput. Interact.*, vol. 2012, p. 8, 2012.

[4] E. Baykara *et al.*, "Effects of training and motivation on auditory P300 brain-computer interface performance," *Clin. Neurophysiol.*, vol. 127, no. 1, pp. 379–387, Jan. 2016.

[5] J. D. Jacoby, M. Tory, and J. Tanaka, "Evoked response potential training on a consumer EEG headset," in *2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, 2015, pp. 485–490.

[6] M. Fouillen, E. Maby, L. Le Carrer, V. Herbillon, and J. Mattout, "ERP-based BCI training for children with ADHD: motivation and trial design," *7th Graz Brain-Comput. Interface Conf. 2017*, 2017.

[7] S. J. Johnstone, R. J. Barry, and A. R. Clarke, "Ten years on: A follow-up review of ERP research in attention-deficit/hyperactivity disorder," *Clin. Neurophysiol.*, vol. 124, no. 4, pp. 644–657, Apr. 2013.

[8] A. Barachant and M. Congedo, "A Plug&Play P300 BCI Using Information Geometry," *ArXiv14090107 Cs Stat*, Aug. 2014.

[9] F. Lotte and C. Guan, "Learning from other subjects helps reducing Brain-Computer Interface calibration time," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 614–617.

[10] U. Technologies, "Unity 3D," *Unity*. [Online]. Available: https://unity.com/frontpage. [Accessed: 16-Apr-2019].

[11] F. Akgul, *ZeroMQ*. Packt Publishing Ltd, 2013.

[12] A. Gramfort *et al.*, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, Feb. 2014.

[13] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, Oct. 2015.

[14] S. R. Searle, F. M. Speed, and G. A. Milliken, "Population Marginal Means in the Linear Model: An Alternative to Least Squares Means," *Am. Stat.*, vol. 34, no. 4, pp. 216–221, Nov. 1980.

[15] B. Z. Allison and J. A. Pineda, "Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: implications for a BCI system," *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.*, vol. 59, no. 2, pp. 127–140, Feb. 2006.

[16] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based brain-computer interface," *Behav. Brain Funct.*, vol. 6, no. 1, p. 28, May 2010.

[17] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk, "Does the 'P300' speller depend on eye gaze?," *J. Neural Eng.*, vol. 7, no. 5, p. 056013, Oct. 2010.

[18] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural Netw.*, vol. 22, no. 9, pp. 1305–1312, Nov. 2009.

[19] B. Reuderink, J. Farquhar, M. Poel, and A. Nijholt, "A subject-independent brain-computer interface based on smoothed, second-order baselining," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2011, pp. 4600–4604, 2011.

[20] I. A. Basyul and A. Ya. Kaplan, "Changes in the N200 and P300 Components of Event-Related Potentials on Variations in the Conditions of Attention in a Brain–Computer Interface System," *Neurosci. Behav. Physiol.*, vol. 45, no. 9, pp. 1038–1042, Nov. 2015.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-42

HETEROGENEITY OF EVENT-RELATED POTENTIALS IN A SCREEN-FREE BRAIN-COMPUTER INTERFACE

Henrich Kolkhorst[1,2], Joseline Veit[1], Wolfram Burgard[2], Michael Tangermann[1,2]

[1]Brain State Decoding Lab, Cluster of Excellence BrainLinks-BrainTools
[2]Autonomous Intelligent Systems Lab, Department of Computer Science
University of Freiburg, Freiburg, Germany

E-mail: kolkhorst@informatik.uni-freiburg.de
michael.tangermann@blbt.uni-freiburg.de

ABSTRACT: Interacting with the environment using a brain-computer interface involves mapping a decoded user command to a desired real-world action, which is typically achieved using screen-based user interfaces. This indirection can increase cognitive workload of the user. Recently, we have proposed a screen-free interaction approach utilizing visual *in-the-scene* stimuli. The sequential highlighting of object surfaces in the user's environment using a laser allows to, e.g., select these object to be fetched by an assistive robot.
In this paper, we investigate the influence of stimulus subclasses—differing surfaces between objects as well as stimulus position within a sequence—on the electrophysiological response and the decodability of visual event-related responses. We find that evoked responses differ depending on the subclasses. Additionally, we show that in the presence of ample data subclass-specific classifiers can be a feasible approach to address the heterogeneity of responses.

INTRODUCTION

In the context of assistive robotics, brain-computer interfaces (BCIs) can offer impaired users means to supervise or control the robot. Brain signals can be used to directly move the robot [1] or for high-level commands in a shared-control setting [2]. While user goals in robotic applications mostly correspond to objects or positions in the environment (e.g., which object to fetch and deliver to the user), most existing BCI setups require an auxiliary interface for stimulus presentation.

This interface is typically implemented using a graphical user interface on a screen, but auditory or haptic interfaces are also possible [3–5]. Actions in the world cannot be executed directly using the BCI but need to have a representation in this interface. In BCI paradigms based on visual event-related potentials (ERPs) [6], on visual steady-state evoked potentials [7], motion-onset visual evoked potentials [8] or visual noise codes [9], the screen has been considered an enabling, indispensable building block of the experimental paradigms. But even in BCI paradigms based on mental imagery tasks, the translation of these tasks into actions requires a supporting interface. The indirection step introduced by the interface, however, comes at a price. First, the user is required to switch attention between the screen representation and the real world in order to deliver commands. Second, the interface creates additional mental workload, as the user has to carefully plan a mapping from a desired real-world action into a sequence of potentially unrelated tasks.

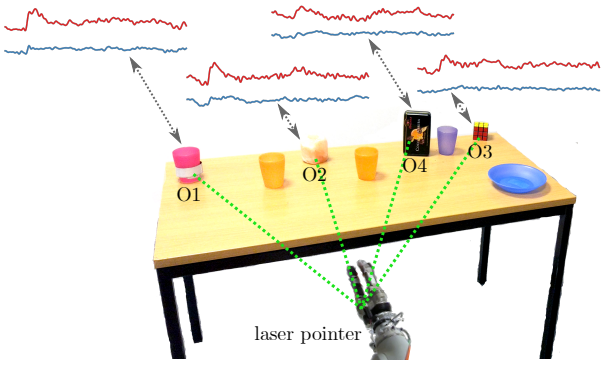


Figure 1: Overview of the experimental setup. In each condition, four objects were highlighted repeatedly with a laser pointer. Annotated objects and average responses correspond to condition C_{250}, the remaining objects correspond to condition C_{500}.

To address both issues, we recently proposed to select objects based on visual ERPs in a novel screen-free interaction [10]. We use a robotic arm to highlight objects *in the scene*. We attach a laser pointer to the arm to generate the visual stimuli. This novel screen-free approach circumvents the aforementioned indirection step, avoids frequent attention switches and may therefore reduce cognitive workload. In short, it allows for a more "intuitive" BCI usage.

While we could show the feasibility of the novel approach [10], the screen-free paradigm introduces other challenges. First, we found that decoding performance requires improvement specifically in the presence of heterogeneous candidate objects and surfaces (see Figure 1).

Second, as the novel paradigm interacts with the real world, it entails different experimental constraints than

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-42

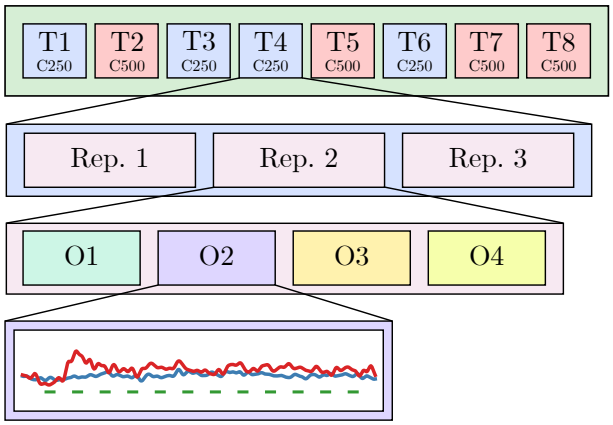


Figure 2: Structure of a single experiment block. Trials were alternated between two conditions (C_{250} and C_{500}) (row 1) in a pseudo-randomized manner. Each trial consisted of three repetitions (row 2) in which each object (row 3) was highlighted for a single sequence of 6 or 12 stimuli (row 4).

traditional screen-based ERP paradigms. Most importantly, the repositioning of the robotic arm between highlighting different objects as well as differing optical properties of stimulus objects necessitates a careful consideration of non-i.i.d. effects in the obtained ERP data.

Previously, the class-wise distributions of auditory ERPs with regard to individual stimuli within a sequence have been found to differ based on the position in the sequence, duration as well as target-to-target intervals [11]. Investigating the influence of subclasses (e.g., stimulus identity), Höhne and colleagues have proposed to incorporate subclass information into the decoding approach by selecting different shrinkage targets for Linear Discriminant Analysis (LDA) [12].

In the following sections, we characterize the influence of *within-sequence stimulus position* and *object surface properties* on the resulting evoked responses using data from six subjects who participated in an in-the-scene object selection paradigm. We present corresponding results in the form of decoding performances using a classification pipeline utilizing Riemannian geometry in differing data size regimes and discuss implications for usage in a BCI.

MATERIALS AND METHODS

Experimental Paradigm and Data Collection:

We recorded the electroencephalogram (EEG) data from six healthy subjects (aged 26 ± 3 years), who each participated in a single session of an experiment. None of the participants had previously used a BCI. Following the declaration of Helsinki, we received approval by the local ethics committee for this study and obtained written informed consent from participants prior to the session.

Subjects were seated in front of a table on which eight objects, denoted O1 to O8, were placed. A Kuka iiwa robotic arm outfitted with a laser pointer next to the end effector was positioned in front of the table and used to highlight objects (c.f., Figure 1). Each object was highlighted either six or twelve times before the robotic arm was repositioned to highlight the next object. Switching the arm position after each single highlighting event was prohibitive considering the delay introduced by the movement time of the robotic arm.

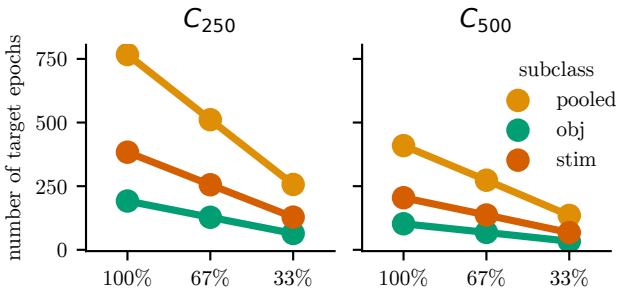


Figure 3: Number of target epochs per subclass based on condition (C_{250} and C_{500}) and data sizes (100 % to 33 %).

The surface types—and therefore, optical properties—of objects O1 to O4 differ from each other, while objects O5 to O8 are made of the same material (see Figure 1). Highlighting O1 resulted in a diffuse and partly specular reflection (away from the subject), whereas highlighting the other objects resulted in diffuse reflections only. Subjectively, O2 showed the strongest reflection among these three objects, the reflection intensity of O3 was reduced and the reflection intensity of O4 was weakest.

We acquired the brain signals using a cap holding $n_s = 31$ Ag/AgCl gel-based passive EEG electrodes positioned according to the extended 10-20 system with a nose reference. We kept channel impedances below 20 kΩ. The amplifier sampled the EEG signals at 1 kHz.

The experiment consisted of six blocks, each containing eight trials. Within a trial, we defined one out of four candidate objects as the target object. As depicted in Figure 2, in each of three repetitions per trial all four candidate objects were highlighted with a stimulus sequence that lasted 3 s per object. For each highlighting, the laser pointer was activated for 100 ms.

Trials were divided into two *conditions*: In condition C_{250}, the candidate objects consisted of O1 to O4 (i.e., they are heterogeneous) and highlighting stimuli were presented with a stimulus-onset asynchrony (SOA) of 250 ms (i.e., 12 stimuli per sequence). In the condition C_{500}, O5 to O8 were the (homogeneous) candidate objects and stimuli were presented with an SOA of 500 ms (i.e., 6 stimuli per sequence). The latter condition directly corresponds to the experiments performed in [10].

Target objects were balanced, i.e., each object was a target once per block. To assure this, a cue denoting the target object was presented to the user before each trial. We instructed users to attend the target object for the duration of the trial without mandating visual focus. After each trial, feedback on the decoded object was given (c.f., [10]). In a post-experiment questionnaire, subjects gave feedback on the experiment task using visual analogue scales.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-42

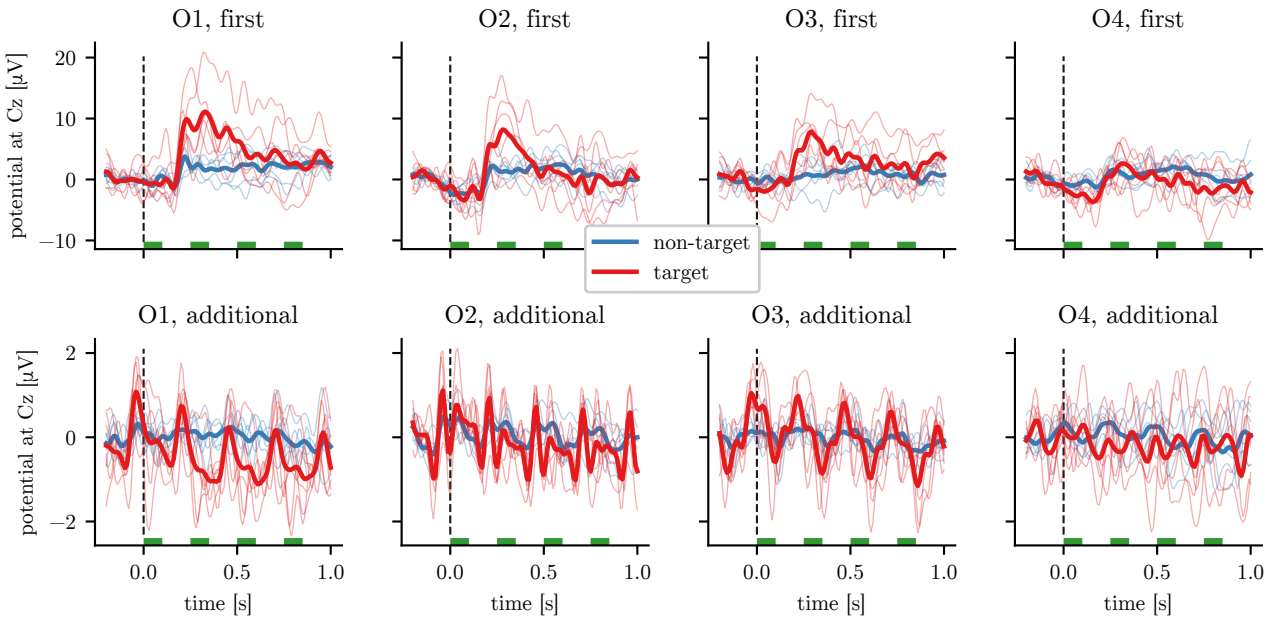


Figure 4: Evoked responses in condition C_{250} by object type (O1–O4) and stimulus position, i.e., first in sequence or additional (stimuli 2–12). The thin lines represent the class-wise average for each subject and the thick lines the grand average over all six subjects. Note the differing y-axis limits between rows. In the top row, each response is averaged based on 108 target and 324 non-target epochs, respectively. In the bottom row, averages correspond to 1,188 and 3,564 epochs. Stimulation intervals are marked with green bars.

Data Analysis:

We analyzed the data separately for each subject and each condition (SOA) in an offline manner. We filtered data to a band of 0.50 Hz to 16 Hz using a FIR filter before downsampling to 100 Hz.

We extracted epochs from -0.20 s to 1.00 s relative to each stimulus onset (leading to $n_s = 121$ samples per epoch) and subsequently corrected them separately for signal drifts using the first 0.20 s as a baseline. Hence, each epoch X_i can be associated with a class $y(i) \in \{target, non\text{-}target\}$, SOA $s(i) \in \{250, 500\}$, object $o(i) \in \{1, \dots 8\}$ as well as the position within the stimulus sequence $p(i) \in 1, \dots 12$. Since we expect most variation between the first and subsequent stimuli [11], stimulus positions can be aggregated to $\tilde{p}$ with $\tilde{p}(i) = first$ if $p(i) = 1$ and $\tilde{p}(i) = additional$ for $p(i) > 1$. We rejected epochs in which the peak-to-peak amplitude exceeded 100 µV in any channel.

For the epoch-wise decoding of the target/non-target response, we used a covariance-based feature representation [13, 14]. On training data, we extracted n_x xDAWN-filters [15] for both target and non-target epochs. Epoch data was projected using these filters and subsequently augmented with prototype responses based on the xDAWN-projected class means in the training data. We calculated the Ledoit-Wolf-regularized covariance of each $4n_x \times n_s$-dimensional augmented epoch separately. These covariances are projected into the Riemannian tangent space at the mean of the training data and subsequently classified using Logistic Regression.

We investigate three different subclass-specific regimes per condition: *pooled*, *obj* and *stim*. In all regimes we performed a chronological 5-fold cross-validation. For *pooled*, we train a single classifier per cross-validation fold. For *obj*, we trained four separate classifiers in each fold based on a partition of the data into objects $\{X_i | o(i) = k\}, k \in \{1, \dots, 4\}$ and classified each test epoch with its object-specific classifier. Similarly, for *stim*, we partitioned the data into $\{X_i | \tilde{p}(i) = l\}, l \in \{first, additional\}$ and trained two different classifiers per fold.

We used $n_x = 3$ components per class for the *pooled* classification and $n_x = 2$ components per class and subclass for *stim* and *obj* classification, resulting in 6 total components for *pooled*, 8 total components for *stim* and 16 total components for *obj*.

Note that this implies different amounts of training epochs per classifier in the three settings. To investigate the general performance degradation based on reduced training data, we performed the above analysis also with 67 % and 33 % of the data by only keeping the first two or only the first repetition of each trial, respectively. The size of the resulting data sets is depicted in Figure 3. We evaluate classification performance on an epoch level (i.e., keeping proportions between data partitions) using the area under the receiver-operating characteristic (AUC).

RESULTS

User Feedback:

All of the subjects were able to attend the stimuli and none reported difficulties perceiving the laser highlighting. On visual analogue scales (“easy” to “demanding”),

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-42

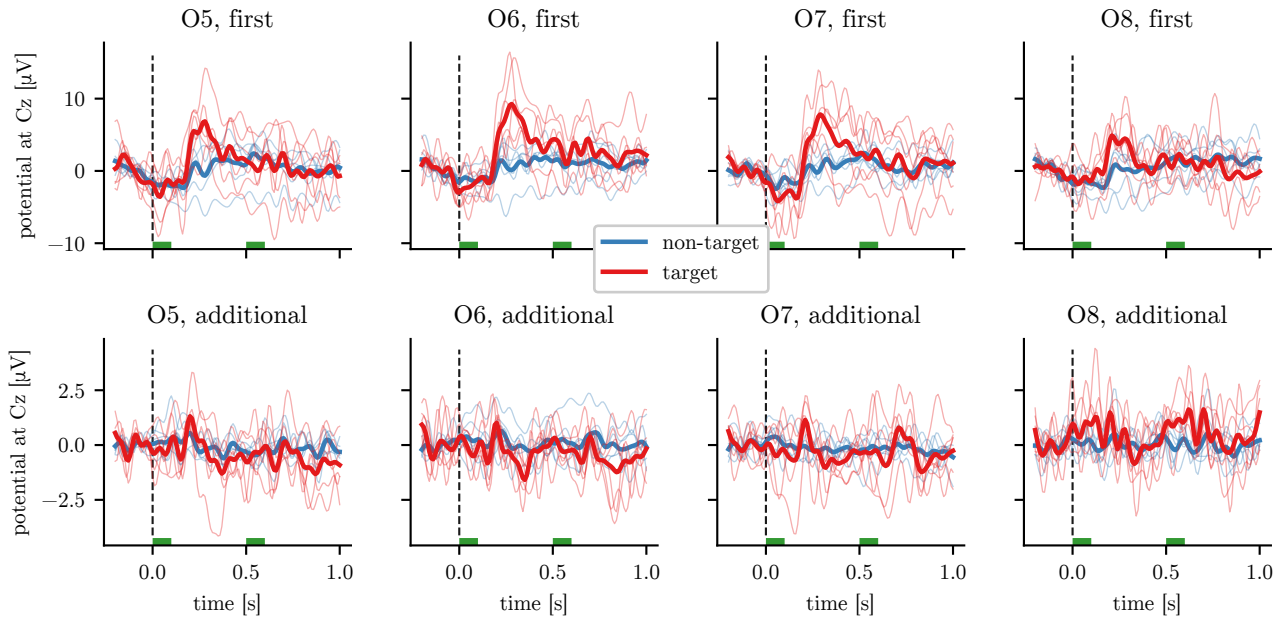


Figure 5: Evoked responses in condition C_{500} by object (O5–O8) and stimulus position, i.e., first in sequence or additional (stimuli 2–6). The thin lines represent the class-wise average for each of subject and the thick lines the grand average over all six subjects. Note the differing y-axis limits between rows. In the top row, each response is averaged based on 108 target and 324 non-target epochs, respectively. In the bottom row, averages correspond to 540 and 1,620 epochs. Stimulation intervals are marked with green bars.

users rated the task's mental demand with 26 ± 26 %, the physical demand with 52 ± 28 % and the temporal demand with 21 ± 21 %. The required effort was rated with 56 ± 31 % on a scale from "low" to "high".

Electrophysiology of Responses:

In both conditions, we observed strong visual evoked potentials which differ between target and non-target stimuli, most prominently approximately 200 ms after stimulus onset (see Figure 4 and Figure 5). The first stimulus in each sequence on average evoked higher amplitudes than subsequent ones, with initial amplitudes surpassing 10 µV compared to the preceding baseline for some objects.

The evoked responses for condition C_{250} with heterogeneous objects are shown as thick lines in Figure 4. Different objects resulted in differing grand average ERP waveforms. Specifically object O4—having a black surface—resulted in a much smaller amplitude, with target responses for later stimuli having similar amplitudes as non-target responses for the other objects. Looking at individual subjects (shown as thin lines in Figure 4), ERP latencies are similar for objects O1 to O3 yet appear to vary substantially between subjects for O4.

In condition C_{500} with homogeneous object surfaces, initial stimuli similarly evoked higher potentials than subsequent ones, yet the difference is smaller than in condition C_{250} (see Figure 5). The grand average responses to the four different but homogeneous objects O5–O8 have a higher similarity than the ones in C_{250}.

Decoding results:

The classification of single epochs into targets and non-targets is possible with a mean AUC of 0.81 for C_{250} and 0.86 for C_{500}, as determined by cross-validation on the *pooled* data sets of the respective conditions. Analyzing the classification performance in this setting on all epochs corresponding to a single object, we found that it is similar for all (homogeneous) objects in C_{500}, while it differed substantially between (heterogeneous) objects in C_{250} (in decreasing order of performance: O1, O2, O3, O4).

As the ERP responses differed for object surface types (O1–O4) and sequence positions, we investigated partitioning the data into different subclasses. As this involved a reduction of training data per sub-classifier (c.f., Figure 3), we also examined the influence of data set sizes.

Partitioning based on sequence position: We found that treating stimulus position ("first" vs "additional") with specialized classifiers did not yield a clear performance improvement or decrease for any of the three data set sizes (100 %, 67 % and 33 %, see Figure 6).

Partitioning based on object: Training separate classifiers for each object in condition C_{250}, as shown in the first column of Figure 6, increased the AUC from 0.81 to 0.84, whereas we observed a performance decrease in the C_{500} setting with homogeneous objects (from 0.86 to 0.81). We found that the improvements obtained with object-specific classifiers for heterogeneous objects (C_{250}) depend on the availability of sufficient training data. As shown in the top row of Figure 6, a reduction of data size to 67 % reduced performance of the *obj* classifier to 0.83, and on 33 % the *obj* classifier performed worse than the *pooled* one (0.78 compared to 0.79). Looking at classification performance for individual epochs on full data, we observed gains for all objects, with greatest gains for O1. In condition C_{500} with homogeneous objects, the performance differences between classifiers did not change with the different data sizes evaluated.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-42

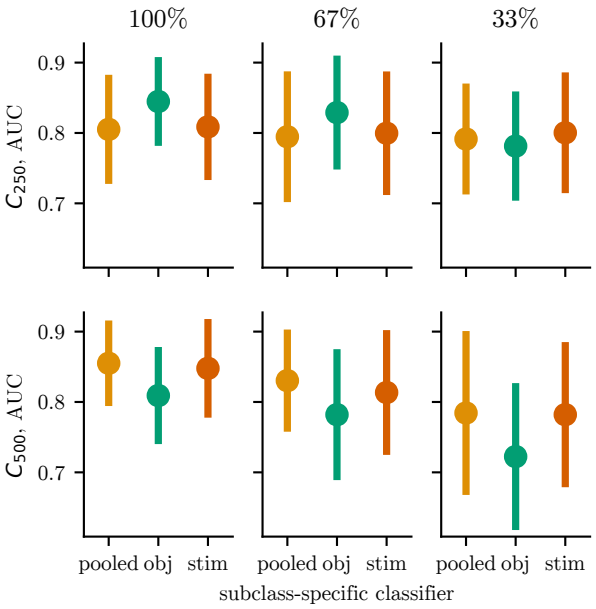


Figure 6: Classification performances of subclass-specific classifiers for different training data sizes (column) as well as SOA and object heterogeneity (rows). In the *pooled* case, data is not partitioned and a single classifier is trained. The means and standard deviations over the six subjects are reported.

DISCUSSION

Highlighting candidate objects directly reduces the levels of indirection of BCIs and can therefore entail greater usability. Yet it comes at the price of reduced control over stimulus presentation compared to traditional screen-based paradigms. This manifests itself in heterogeneous ERPs when object surfaces differ and similarly in distinct responses to initial stimuli (compared to subsequent ones). Whereas stimulus presentation is easily modifiable in screen-based solutions, it is neither desirable nor feasible to adapt properties of objects in the scene—leaving the question of how to handle the heterogeneity.

If maximal performance is not required, the non-i.i.d. characteristics of the data can be ignored, since using a single classifier based on the pooled data (i.e., ignoring the subclass information) still allows an effective decoding on average (however, performance on difficult subclasses suffers). The higher performance for C_{500} compared to C_{250} in the pooled setting can be attributed partly to the homogeneous object surfaces and partly to the larger SOA, which implies a larger fraction of (stronger) responses to initial stimuli in this condition.

Addressing the differing surfaces in the condition C_{250}, object-specific classifiers yield an improvement in AUC. This improvement by specialization is particularly interesting considering the reduced amount of training samples available for each classifier (c.f., Figure 3). The lack of heterogeneity in surfaces is the likely reason for the subpar performance of object-specific classifiers in condition C_{500}, since in condition C_{250} even reduced data allows for improvement compared to the pooled one.

Contrasting the improvement using object-wise classifiers to the lack thereof for sequence-specific ones, two observations can be made: First, the aforementioned reduced data size and therefore higher variance of subclassifiers can offer an explanation for the lack of improvement since the minority class (first stimulus) makes up only 17 % (C_{500}) or 8 % (C_{250}) of the data. Second, the (pooled) covariance-based features (as opposed to utilizing mean potentials) might be more robust to variations in amplitude (as observed between sequence positions) compared to variations in waveforms (for heterogeneous objects), avoiding the need for specialization in the former case.

Hence, subclass-specific classification of heterogeneous responses is feasible and can offer improved performance in the presence of highly differing ERPs and sufficient data. They are, however, not a “silver bullet” solution to this problem, and—without automatic relevance selection—inspecting the subclass data is necessary to identify suitable subclass partitions.

CONCLUSION

Utilizing in-the-scene stimuli for a brain-computer interface implies an influence of scene properties on observed brain responses. In this paper, we find that heterogeneous object surfaces as well as stimulus position within a sequence result in differing ERPs. Addressing this with subclass-specific classifiers can offer improved decoding performance, with the caveat of increased data requirements as well as unfavorable performance in the absence of subclass differences. For future work it would be interesting to automatically infer relevant subclasses and to investigate approaches capable of handling low-data scenarios [12].

ACKNOWLEDGMENTS

This work was (partly) supported by BrainLinks- BrainTools, Cluster of Excellence funded by the German Research Foundation (DFG, grant number EXC 1086). Additional support was received from the German Research Foundation through grant INST 39/963-1 FUGG and from the Ministry of Science, Research and the Arts of Baden-Württemberg for bwHPC.

REFERENCES

[1] Millan J., Galan F., Vanhooydonck D., Lew E., Philips J., Nuttin M. Asynchronous non-invasive brain-actuated control of an intelligent wheelchair. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009. EMBC 2009. Sep. 2009, 3361–3364.

[2] Burget F., Fiederer L. D. J., Kuhner D., *et al.* Acting thoughts: Towards a mobile robotic service assistant for users with limited communication skills. In: 2017 Euro-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-42

pean Conference on Mobile Robots (ECMR). Sep. 2017, 1–6.
[3] Höhne J., Schreuder M., Blankertz B., Tangermann M. A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. Front. Neurosci.. 2011;5.
[4] 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.
[5] Waal M. van der, Severens M., Geuze J., Desain P. Introducing the tactile speller: An ERP-based brain–computer interface for communication. J. Neural Eng.. 2012;9(4):045002.
[6] Bin G., Gao X., Wang Y., Hong B., Gao S. VEP-based brain-computer interfaces: Time, frequency, and code modulations [Research Frontier]. IEEE Computational Intelligence Magazine. 2009;4(4):22–26.
[7] Chen X., Wang Y., Nakanishi M., Gao X., Jung T.-P., Gao S. High-speed spelling with a noninvasive brain–computer interface. PNAS. 2015;112(44):E6058–E6067.
[8] Guo F., Hong B., Gao X., Gao S. A brain–computer interface using motion-onset visual evoked potential. J. Neural Eng.. 2008;5(4):477–485.
[9] 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.
[10] Kolkhorst H., Tangermann M., Burgard W. Guess What I Attend: Interface-Free Object Selection Using Brain Signals. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Oct. 2018, 7111–7116.
[11] Hübner D., Tangermann M. Challenging the assumption that auditory event-related potentials are independent and identically distributed. In: Proceedings of the 7th Graz Brain-Computer Interface Conference 2017. Verlag der Technischen Universität Graz, 2017, 192–197.
[12] Höhne J., Bartz D., Hebart M. N., Müller K.-R., Blankertz B. Analyzing neuroimaging data with subclasses: A shrinkage approach. NeuroImage. 2016;124:740–751.
[13] Barachant A., Bonnet S., Congedo M., Jutten C. Multiclass Brain-Computer Interface Classification by Riemannian Geometry. IEEE Transactions on Biomedical Engineering. 2012;59(4):920–928.
[14] Barachant A., Congedo M. A Plug&Play P300 BCI Using Information Geometry. ArXiv:1409.0107 [cs, stat]. 2014.
[15] Rivet B., Souloumiac A., Attina V., Gibert G. xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface. IEEE Transactions on Biomedical Engineering. 2009;56(8):2035–2043.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-43

THE EFFECT OF PERFORMANCE EXPECTANCY AND ACHIEVEMENT MOTIVE IN A P300 BASED BRAIN-COMPUTER INTERFACE

S. Kleih[1], A. Eder[1], & A. Kübler[1]

[1] Institute of Psychology, Julius-Maximilians University Würzburg, Würzburg, Germany

E-mail: sonja.kleih@uni-wuerzburg.de

ABSTRACT: We investigated the effect of performance expectancy and achievement motive on P300 BCI performance. Thirty-eight participants were separated in two groups according to their achievement motive and classified as either avoiding failure (AF) or approaching success (AS). Participants were presented with three different matrices in the colors red, green and blue and were told that spelling would be difficult, moderately difficult or easy depending on the color. We hypothesized AS participants to perform best and to show highest P300 amplitudes in a spelling condition perceived to be moderately difficult. AF participants were hypothesized to perform worst and to show lowest P300 amplitudes in the spelling condition perceived to be moderately difficult. Participants spelled six five-letter words in each perceived difficulty condition. We found highest P300 amplitudes in the easy condition irrespective of achievement motive; however, no differences concerning performance in percent correct responses. Even though we could not find an effect of performance expectancy on the behavioral level, we did show that performance expectancy does affect BCI performance on the physiological level.

INTRODUCTION

Motivation. Heckhausen and Heckhausen [1] describe motivation as a collective term for psychological processes that are necessary to choose a certain behavior and manage the required effort for pursuing that behavior after evaluation of expected results. Motivation, more precisely, summarizes processes such as stringency, beginning and finalizing a behavior, returning to a behavior after an interruption, possible conflict between several behavioral goals und the solution of this conflict. Heckhausen subsumed the cognitive processes related to motivation in his Cognitive Model of Motivation [2]. Based on this model, a person's motivation could be estimated (fig. 1). Four different components contributing to motivational force were distinguished: 1.) perceived situation, 2.) action, 3.) intended goal or outcome and 4.) consequences (see fig. 1). Transferred to a BCI context, the perceived situation would be the possibility to use a BCI system. The action to be taken would be the willingness to use the system and therefore, some kind of mental effort, such as focusing attention in a BCI system based on evoked potentials. The intended goal would be the successful selection of letters in a P300 BCI for spelling or, more general, BCI control. If successful, that would lead to the anticipated consequence of successful interaction with the environment. A person's subjective expectancy of a behavior leading to a certain outcome or consequences contributes to motivational force.

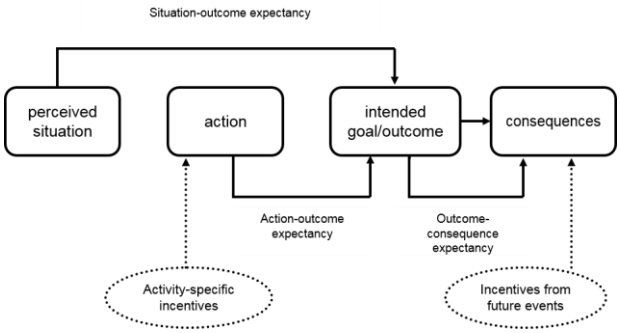


Figure 1: The Cognitive Model of Motivation [2].

In case, a BCI system would be usable but the user does not expect to be able to control the BCI system, motivational force would be low. Expectancies that influence the interaction between the mentioned components were classified by Heckhausen as follows: 1.) situation-outcome expectancy, 2.) action-outcome expectancy and 3.) outcome-consequence expectancy. Situation-outcome expectancy (SOE) describes the expectancy that an outcome will occur, even without taking any kind of action. In case such expectancy is high without changing behavior, motivation for action taking would be low. In a BCI context of course, without the user taking action, no BCI control will be possible, therefore the user can only develop action-outcome expectancy (AOE) by engaging in the BCI task. The user should expect a certain action, in this case focusing attention on the stimulation to influence the outcome, which would be P300 based spelling. Outcome-consequence expectancy (OCE) describes the expectancy that a desired consequence will follow the outcome. In a BCI context, the user might anticipate the desirable consequence of being able to interact with the environment based on his or her brain activity. Heckhausen also postulated two kinds of incentives that influence motivation: 1.) action specific incentives and 2.) incentives from future events. In a BCI context, action specific incentives might be interest in the

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-43

functionality of a BCI system and having the possibility to control such a system. In addition, of course, there are incentives from BCI performance as in a P300 speller, feedback directly informs the user about his or her performance level. Incentives from future events might be improvements in BCI use over time or achievement of goals based on BCI use, such as painting pictures [3]. Naturally, incentives from future events are more relevant in case BCI end-users with disease are involved as compared to healthy students who might never use a BCI system again after participation in one experiment. We cannot fully transfer Heckhausen's assumptions [2] postulated for non-BCI situations to a BCI context. When participating in a BCI study, some students volunteered without knowing anything about a BCI system. There is no deliberate choice to engage in BCI use. There even might be a deliberate choice to collect credit points by participation, such that there is a motivational goal underlying BCI use that is completely independent from BCI spelling. Thus, theoretical components such as situation-outcome expectancy might change during participation ("I volunteer for credit points" versus "I am really getting interested and believe that I can control this device by focusing attention"). Nonetheless, BCI use will most likely touch a persons' performance motivation specifically during P300 spelling, as it becomes immediately clear whether a person would be able to communicate via the BCI. Therefore, an inherent, cross-situational achievement motive might be triggered as a trait variable of motivation and should be addressed in this context.

Achievement motivation. In his theory of achievement motivation Atkinson postulated a person inherent achievement motive to influence motivation in achievement situations [4]. This motive consists of two specifications that might differ in every person and together add up to a person's individual tendency to undertake an achievement-oriented activity. One specification is the motive to approach to success, which describes a person's need to succeed or tendency to expect success. The other specification is the motive to avoid failure, which represents a person's fear of failure. Atkinson postulated persons with high tendency to approach success to prefer performance situations of medium difficulty. In such situations, the likelihood of success seems highest and chances to succeed are most promising. People with a high tendency to avoid failure were hypothesized to prefer situations of low or very high difficulty. Very easy tasks might be easily manageable; therefore, fear of failure would be low. Very difficult tasks might be judged as so difficult that failure would not result in a threat towards a person's abilities. Tasks of moderate difficulty are potentially threatening as in these tasks ability (and inability) might become obvious.

As postulated in the triarchic model of P300 amplitude [5], P300 amplitude varies depending on the value the stimulation represents for the BCI end-user. Based on the individual achievement motive as expressed by the motive to approach success or to avoid failure, the P300 speller stimulation might be perceived of different value when a BCI task is perceived to be easy, moderate or difficult which in turn will influence situation-outcome expectancy, action-outcome expectancy and outcome-consequence expectancy.

Studies on motivation in a BCI context. To our knowledge, there are no BCI studies yet, in which the influence of trait variables of motivation were investigated in the context of BCI use. Motivation was assessed as a state variable and it was shown that healthy participants usually highly motivated in a P300 spelling task [6]. In several studies [7, 8, 9] motivation was assessed using the Questionnaire for the assessment of Current Motivation during BCI use (QCM-BCI, [10]). The QCM-BCI is a BCI adapted version of the original QCM [11] to account for using a BCI instead of making predictions about performance and success in cognitive learning tasks. The QCM-BCI comprises 18 items divided into four subscales *incompetence fear*, *mastery confidence*, *interest* and *challenge*. Incompetence fear correlated negatively with BCI performance in healthy participants [12]. In stroke patients, mastery confidence correlated positively with BCI performance and negatively with spelling speed [12].

Therefore, state variables of motivation, such as mastery confidence and incompetence fear seem to influence BCI performance. It remains unclear whether trait variables of motivation such as the motive to approach success or the motive to avoid failure [13] also influence BCI performance. As P300 based spelling can be perceived as a performance task, trait variables might be activated. Dependent on the individual expression of the mentioned motives, P300 amplitudes and BCI performances might vary with perceived task difficulty. We hypothesized participants with a high motive to approach success to perform best and to show highest P300 amplitudes in a P300 spelling paradigm perceived to be moderately difficult. Participants with a high motive to avoid failure should perform worst in a task perceived to be moderately difficult and to show lowest P300 amplitude in this condition.

MATERIALS AND METHODS

Sample. Thirty-eight healthy students participated in this study, n=27 were female. Mean age was 22 (SD=3.00). All participants were naïve to BCI use and gave written informed consent to the study. The study was approved by the Ethical Review Board of the Medical Faculty, University of Tübingen, Germany.

Procedure. To measure the motive to approach success and the motive to avoid failure, the Multi-Motiv-Gitter [MMG, 14] was used. MMG test criteria

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-43

are satisfactory with a re-test reliability between .88 and .92. The MMG is a semi-projective questionnaire, in which a participant judges a situation shown as a drawing. Several statements to that drawing are offered and the subjectively fitting ones are to be selected. Exemplary statements are: "you can lose your reputation here" or "you can impress with performance". Based on the answers, a score is calculated for both motivation tendencies. The difference between the sums of both tendencies is the performance motivation score. As established and tested by the authors [14] a score above 0 indicates higher motive to approach success (AS) while a score below 0 indicates a higher motive to avoid failure (AF). The MMG is an instrument with which a motive is judged as a stable variable. As other trait variables, such as personality traits, it can be assumed that motive manifestation is permanent.

To manipulate participants' action-outcome expectancy, we developed a cover story. Participants were told that an effect of matrix color on BCI performance was found when using the P300 speller. The goal of the study would be to further investigate this finding. Therefore, the participants should spell the same words using three different matrix colors (red, green, blue). Participants were randomly assigned to one of six groups. In each group, the difficulties assigned to a color changed such that each color (red, green, blue) was assigned to each difficulty (easy, moderately difficult, and difficult). Physically, all matrices were of the same luminance to avoid real color effects on BCI performance. Only the cover story suggested three different difficulties where spelling with the matrices was of equal difficulty in all conditions. For calibration, participants spelled two words (BRAIN and POWER). For the copy-spelling runs, participants spelled six five-letter words in each spelling difficulty condition without receiving feedback. Data of all difficulty conditions were aggregated based on the respective task difficulty irrespective of the colors that were assigned to the difficulties.

P300 speller and stimuli. A 6x6 P300 speller matrix was used which included the letters of the alphabet, the numerals 1 to 9 and an underscore. For calibration, we used ten flash sequences. During copy-spelling, we used two sequences to avoid ceiling effects found in an earlier BCI study [15]. Flash duration was 62.5 ms and the inter-stimulus-interval was 250ms. The inter-trial-interval was 2500ms. Flash color was white as a positive effect of a colored matrix together with colored flashes on performance was suggested [16] and we did not intend to really manipulate task difficulty. The word to spell appeared in a line above the matrix and the letter to spell was displayed in parentheses next to the word to spell. For data acquisition and experimental control, we used the BCI2000 Software [17].

EEG data acquisition. For EEG measurement we used a g.USB amplifier (Guger technologies, Austria) with a low pass filter of 60Hz, a highpass filter of 0.1 Hz and a notch filter of 50Hz. We placed 12 Ag/AgCl ring electrodes on the positions Fz, FCz, C3, Cz, C4, CPz, P3, Pz, P4, PO7, PO8 and Oz. The right mastoid was used as reference, the left mastoid as ground. Impedances were kept below 5kΩ. To control for eye movement artefacts, four EOG electrodes were added.

Data analysis. EEG data were analyzed using BrainVision Analyzer Software. Data were artifact corrected (above 50 µV amplitudes), and baseline corrected (-100 ms). The P300 detection was performed semi-automatically with peak detection in a time window of between 250 and 600 ms after stimulus onset. For the research question addressed here, we used electrode position Pz. Statistical analysis was conducted with IBM SPSS © version 24. Concerning our statistical analysis we would have needed 60 participants when assuming a medium effect with a power of .8 and a significance level of α=.05. Our data met the criteria for parametric testing, however, the required sample size was not fulfilled.

RESULTS

To investigate our hypotheses we applied a mixed model with achievement motive as between subjects' factor and perceived difficulty as within subjects' factor. The dependent variables were the P300 amplitude and the performance in percent correct responses.

Mean performance ranged between 57% and 64% correct selections. Performance was not affected by the suggested difficulties ($F(2)=.06$, $p=.93$), and not by the achievement motive ($F(2)=1.72$, $p=.19$); the interaction was also not significant ($F(4)=1.84$, $p=.14$). For the P300 amplitude, we found a main effect of perceived difficulty ($F(2)=7.77$, $p<.01$) but no main effect of achievement motive ($F(2)=1.76$, $p=.19$).

The P300 amplitude found in the moderately difficult condition was lowest (M_{AS} = 6.10 µV, SD=3.13, M_{AF}=6.43 µV, SD=1.87, see fig. 2 and 3). It was found to be significantly lower as compared to the condition perceived easiest ($F(2)=13.01$, $p<.01$) in which highest P300 amplitudes were shown in both groups (M_{AS} = 6.85 µV, SD=2.81, M_{AF}=7.75 µV, SD=1.58, see fig. 2 and 3). In the difficult condition, P300 amplitudes were moderately high (M_{AS} = 6.54 µV, SD=2.39, M_{AF}=7.40 µV, SD=2.02, see fig. 2 and 3).

Irrespective of suggested difficulty, on average, the AS tendency group had P300 amplitudes of M=6.50 µV (SD=2.63) and the AF group showed P300 amplitudes of M=7.19 µV (SD=1.60).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-43

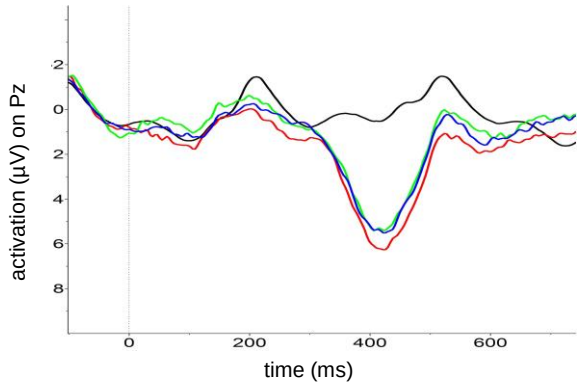


Figure 2: P300 amplitude for the approach success (AS) group. Black=non-targets averaged, red=condition perceived easy, green=condition perceived moderately difficult, blue=condition perceived difficult.

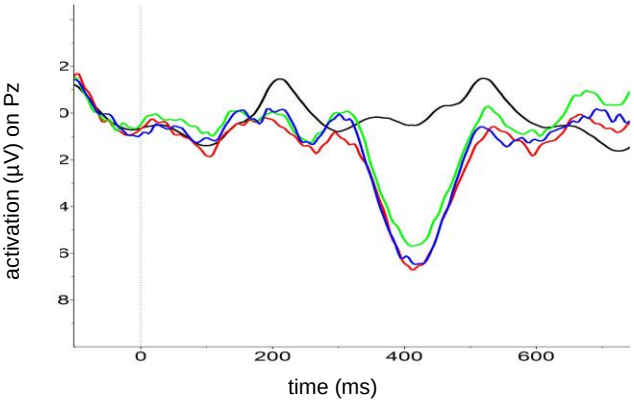


Figure 3: P300 amplitude for the avoid failure (AF) group. Black=non-targets averaged, red=condition perceived easy, green=condition perceived moderately difficult, blue=condition perceived difficult.

DISCUSSION

In this study, our goal was to investigate the effect of action-outcome expectancy and achievement motive in a BCI setting. We hypothesized action-outcome expectancy to affect P300 spelling performance and P300 amplitude and that there would be an interaction of this effect with achievement motive. While people who avoid failure should perform worst and show lowest P300 amplitudes in the condition suggested to be moderately difficult, we hypothesized that people who approach success to perform best and to show highest P300 amplitudes in this same condition.

We found no interaction of achievement motive with perceived difficulty. However, participants with the motive to avoid failure showed on average higher P300 amplitudes as compared to participants with the motive to approach success. We reject our hypothesis that participants who approach success to perform best in the condition perceived as moderately difficult and to show highest P300 amplitudes in this condition. Both groups showed highest P300 amplitudes in the condition perceived as easy. Irrespective of achievement motive, we see an effect of suggested difficulty on P300 amplitude even though this effect is not strong enough to affect performance on the behavioral level. It might be that our sample size was too small and potentially existing effects were not revealed in this data set.

Both groups of participants showed highest P300 amplitudes in the condition perceived as easy. Concerning performance, participants with a high motive to avoid failure performed best in the condition perceived as moderately difficult, while participants with the motive to approach success performed best in the condition perceived to be easy. These results are not in line with Atkinson's assumptions postulated in his theory of achievement motivation [4]. However, performance in this study was rather low and ranged between 57% and 64% correct. The number of event-related potentials to average might have been too small overall and therefore we possibly could not detect potential effects on performance.

It must be mentioned that Atkinson's theory was already challenged in the past. To name just two examples, self-efficacy beliefs [18], and interest [19] were found to influence performance motivation and therefore, to play a role in performance situations. In this study, we did not investigate other psychological variables that might affect achievement motivation. Especially, the role of incentives should be elucidated further as in a BCI spelling situation the incentive of a task might change according to personal performance expectancy [2]. Additionally, in a BCI situation not only performance motivation might be activated but also motivation components such as the need for affiliation [20]. Most participants ask about the goal of BCI research and the clinical applications and might experience compassion for the patients who are BCI end-users. Such influencing variables should be considered in future work.

Overall, there seems to be an effect of action-outcome expectancy on the P300 amplitude and expectancy value theories of motivation seem to be applicable in a BCI context [21]. Future research not only should address individual motives possibly influencing the perception of BCI situations, but also address explicit as implicit components of achievement motivation in a BCI situation. Only by identifying and investigating factors influencing BCI performance [22, 23], variance in BCI performance that can be explained by psychological factors, can be integrated into a theoretical framework [24] on the effect of motivation in BCI performance [21].

CONCLUSION

Performance expectation does influence BCI performance on a psychological level. More studies with higher numbers of participants are required to finally judge the influence of motivation on BCI performance. Creation of a theoretical framework on the

Published by Verlag der TU Graz
Graz University of Technology

effect of motivation in BCI performance seems useful and indicated.

REFERENCES

[1] Heckhausen J, Heckhausen H. Motivation und Handeln. Springer-Verlag; Germany (2006).

[2] Heckhausen H. Achievement motivation and its constructs: A cognitive model. Motivation and emotion. 1977; 1;1(4):283-329.

[3] Zickler C, Halder S, Kleih SC, Herbert C, Kübler A. Brain painting: usability testing according to the user-centered design in end users with severe motor paralysis. Artificial intelligence in medicine. 2013; 59(2):99-110.

[4] Atkinson JW. Motivational determinants of risk-taking behavior. Psychological review. 1957;64(6p1):359.

[5] Johnson Jr R. For distinguished early career contribution to psychophysiology: award Address, 1985: a triarchic model of P300 amplitude. Psychophysiology. 1986; 23(4):367-84.

[6] Sprague SA, Ryan DB, Sellers EW. The Effects of Motivation on Task Performance Using a Brain-Computer Interface. InProc. 6th Int. Brain Computer Interface Conf 2014.

[7] Kleih SC, Nijboer F, Halder S, Kübler A. Motivation modulates the P300 amplitude during brain–computer interface use. Clinical Neurophysiology. 2010; 121(7):1023-31.

[8] Hammer EM, Halder S, Blankertz B, Sannelli C, Dickhaus T, Kleih S, Müller KR, Kübler A. Psychological predictors of SMR-BCI performance. Biological psychology. 2012; 89(1):80-6.

[9] Kleih SC, Kübler A. Empathy, motivation, and P300 BCI performance. Frontiers in human neuroscience. 2013; 17;7:642.

[10] Nijboer F, Furdea A, Gunst I, Mellinger J, McFarland DJ, Birbaumer N, Kübler A. An auditory brain–computer interface (BCI). Journal of neuroscience methods. 2008; 167(1):43-50.

[11] Rheinberg F, Vollmeyer R, Burns BD. FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern-und Leistungssituationen (Langversion, 2001). Diagnostica. 2001;2:57-66.

[12] Kleih, SC, Kaufmann, T, Hammer, E., Pisotta, I, Picchiori, F, Riccio, A, Mattia, D, Kübler, A. Motivation and SMR-BCI: Fear of Failure Affects BCI Performance. In: J. d. R. Millan, S. Gao, G. Müller-Putz, JR Wolpaw & JE Huggins (Eds.). Proceedings of the 5th International Brain-Computer Interface Meeting: Defining the Furture, Asilomar Conference Center, Pacific Grove, California, USA, June, 3-7, 2013; pp.160-161.

[13] McClelland DC, Atkinson, JW, Clark, RA, Lowell, EL. The achievement motive. East Norwalk, CT, US: Appleton-Century-Crofts (1953).

[14] Langens TA, Sokolowski K, Schmalt HD. Das Multi-Motiv-Gitter (MMG). J. Erpenbeck & L. von Rosenstiel (Herausgeber), Handbuch Kompetenzmessung. 2003:71-9.

[15] Kleih SC, Kübler A. Empathy, motivation, and P300 BCI performance. Frontiers in human neuroscience. 2013; 17;7:642.

[16] Takano, K., Komatsu, T., Hata, N., Nakajima, Y., & Kansaku, K. (2009). Visual stimuli for the P300 brain–computer interface: a comparison of white/gray and green/blue flicker matrices. *Clinical neurophysiology*, *120*(8), 1562-1566.

[17] 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):1034-43.

[18] Zimmerman BJ, Bandura A, Martinez-Pons M. Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. American educational research journal. 1992;29(3):663-76.

[19] Ainley M. Connecting with learning: Motivation, affect and cognition in interest processes. Educational Psychology Review. 2006 Dec 1;18(4):391-405.

[20] Murray, H. A. Explorations in personality. New York: Oxford University Press (1938).

[21] Kleih, SC, Kübler, A. Motivation matters: Psychological models in brain-computer interfacing. In: G. Müller-Putz, J. E. Huggins & D. Steryl (Eds.). *Proceedings of the Sixth International Brain-Computer Interface Meeting: BCI Past, Present, and Future*, Asilomar Conference Grounds, Monterey, CA, USA, May 30- June 03., (2016); p. 141. DOI: 10.3217/978-3-85125-467-9-141.

[22] Jeunet C, N'Kaoua B, Subramanian S, Hachet M, Lotte F. Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns. PloS one. 2015 Dec 1;10(12):e0143962.

[23] 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. InProgress in brain research 2016 Jan 1 (Vol. 228, pp. 3-35). Elsevier.

[24] Jeunet C, N'Kaoua B, Lotte F. Towards a cognitive model of mi-bci user training. International Graz BCI Conference , Sep 2017, Graz, Austria. 2017

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-44

EVALUATION OF AUDITORY BCI SYSTEM BASED ON STREAM SEGREGATION

S. Kanoh [1,2], S. Kojima [2]

[1] College of Engineering, Shibaura Institute of Technology, Tokyo, Japan
[2] Graduate School of Engineering, Shibaura Institute of Technology, Tokyo, Japan

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

ABSTRACT: The auditory brain-computer interface (BCI) system based on auditory stream segregation was tested. In this system, three oddball sequences consisted of musical tones at different frequency bands were presented alternately to subject's right ear. Stimulus parameters were set that subjects could perceive these three oddball sequences as streams (stream segregation). Subjects were requested to pay attention to one of the streams during experiments. It was shown that the P300 was elicited only by the deviant stimuli in the attended target stream in two out of three subjects, and the target stream could be detected by pattern classification with high accuracies on all subjects.

INTRODUCTION

The brain-computer interface (BCI) system based on selective attention to specific sensory input has attracted interests. The selective attention to the sensory stimulus of user's interests is reflected by the P300, one of the event-related potential (ERP) components of human EEG. The P300 and an earlier ERP component, mismatch negativity (MMN), can be elicited when user perceives an oddball sequence, on which frequently presented stimuli (standard stimuli) are presented while some of which were randomly replaced by infrequently presented stimuli (deviant stimuli). Such a P300-based BCI has been practically applied as a virtual keyboard [1] or other types of menu selection systems [2].

On almost all of the P300-based BCIs, visual domain has been used. Visual stimulation contains rich information comparing to other sensory domains and it can be easily controlled and delivered to users. Another reason is that it is easy to pay attention to the specific visual targets. However, on P300-based BCI with visual stimuli, the visual functions are "occupied" by the BCI system and not available during operation.

The BCI system which uses auditory stimulation (auditory BCI) can be used without restricting user's visual function. And it can be applied to blind patients. However, only a few types of auditory BCI have been realized. Hill et al. proposed a P300-based auditory BCI [3]. In this system, two kinds of auditory oddball sequences are presented to user's left and right ear, respectively, and user was asked to pay attention to one of the sequences presented to left or right. Schreuder et al. proposed multi-class auditory BCI with multiple oddball tone sequences presented from different locations by speakers placed around subjects [4].

The authors proposed the novel auditory BCI based on stream segregation [5]. Stream segregation is the phenomenon that the series of tones at similar frequencies tend to be bound together and are perceived as a stream of tones, even if tones at different frequencies were inserted between segregated tones. This phenomenon is one of the auditory illusions which have been discussed in the studies of auditory scene analysis [6] of the Gestalt Psychology. The mechanism of the phenomenon can be understood that the two constraints for perceptual grouping of multiple tones, similarity of tone frequency and the factor of time continuity, compete with each other [7]. In this auditory BCI system, two individual oddball sequences at different frequency bands were presented alternately to user's single ear to realize the stream segregation. Subjects were requested to pay attention to one of the tone streams. By classifying P300 and MMN components, high detection accuracy was achieved. However, only a few EEG channel data was recorded in the previous study, and the detailed investigations were left for further study.

In this study, the auditory BCI system proposed by the authors [5] was evaluated by multichannel EEG measurements. The three streams which consisted of musical tones were used as auditory stimuli. It was shown that the P300 was elicited only by the deviant stimuli in the attended target stream (target stream) in two out of three subjects, and the target stream could be detected by pattern classification with high accuracies on all subjects.

MATERIALS AND METHODS

Stimulus presentation: MIDI musical tones were used for presenting stimuli. All the tones were generated by using digital auditory workstation (Cakewalk by BandLab, BandLab Technologies, Singapore) and software sound source (SampleTank3, IK Multimedia Production, Italy). Piano tones (Grand Piano 1 SE) were used to enhance the stream segregation. All the tones were presented to subjects by DSP (System3, Tucker-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-44

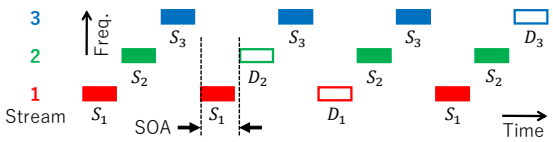


Figure 1: Schematic diagram of tone sequences

Table 1: Frequencies of the tones used in this study

Stream	Standard tone		Deviant tone	
1	S_1: C3	(131 Hz)	D_1: G3	(196 Hz)
2	S_2: D5	(587 Hz)	D_2: A5	(880 Hz)
3	S_3: E7	(2637 Hz)	D_3: B7	(3951 Hz)

Davis Technologies, USA) and headphones (HDA200, Sennheiser, Germany). Timing and order of presented tones were controlled by Arduino UNO.

Stimulus setting: Stimulation tone sequences consisted of three streams with different frequency bands. Fig.1 shows a schematic diagram of stimulation tone sequence. Stream n $(n = 1, 2, 3)$ consisted of standard tone S_n $(p = 0.9)$ and deviant tone D_n $(p = 0.1)$. The frequencies of standard and deviant tones in the three streams were shown in Tab. 1. The duration and rise/fall time of each tone was 150 ms and 10 ms, respectively. SOA was set to 180 ms.

Experiments: Three volunteers (one female, 21 ~ 22 years old) with normal hearing functions took part in the experiments as subjects. Subject sat on a comfortable armchair in a sound-proof electromagnetically-shielded room (Music Cabin, Takahashi Kensetsu Co. Ltd., Japan). During experiment, subject was requested to pay attention to one of the three streams (indicated by experimenter before starting experiments) and to count the number of deviant tones in the target stream. The experiment to each subject consisted by two sessions, which were performed in a day. In each session, three sets of experiment were executed. In each set in a session, the object of selective attention was stream 1, 2 and 3, respectively.

EEG measurement: 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) was measured during experiments by Ag-AgCl electrodes (EasyCap, EasyCap GmbH, Germany). Reference and ground electrodes were placed at right and left earlobe, respectively. Measured signal was amplified at 0.1 Hz to 100 Hz, and acquired to PC with sampling frequency of 1000 Hz (BrainAmp MRplus/DC/ExG, Brain Products GmbH, Germany).

Data analysis: Acquired data was analyzed by MATLAB. The data was band-pass filtered (2nd order Butterworth filter, 1~40 Hz), and the responses to the tone were extracted (-100~500 ms from the stimulus onset). The extracted data was referred to the mean amplitude in the pre-stimulus baseline (-50~0 ms). EEG epochs with amplitudes exceeded ± 100 μV were excluded from further analysis. The pre-processed epoch data was averaged over trials, and the difference between the responses to standard and deviant tones was statistically tested.

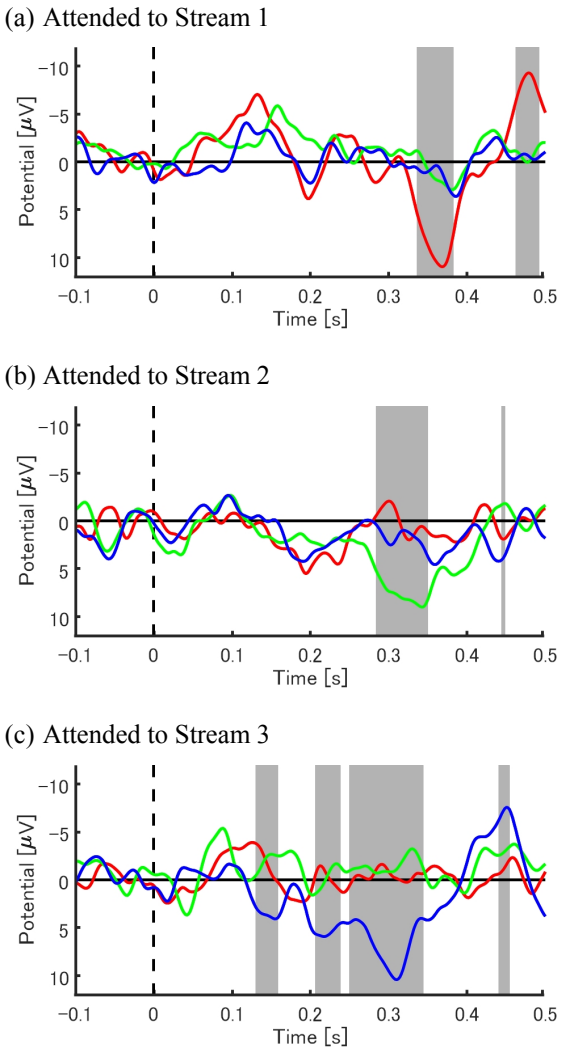


Figure 2: Averaged responses to deviant tones in Stream 1 (D_1: red), Stream 2 (D_2: green) and Stream 3 (D_3: blue) on Subject A (electrode Pz). Figures (a), (b) and (c) show responses when the subject paid attention to Stream 1, 2 and 3, respectively. Grey box denotes the significant difference between responses to target (attended) and non-target deviant tones.

Pattern classification: The ability to detect the selective attention to the target stream was evaluated by pattern classification. Three linear discriminant analysis (LDA) classifiers [8] to detect event-related potentials (P300) elicited by the deviance on stream 1, 2 and 3 were designed.

The feature vectors were calculated in the following way. The pre-processed data (shown above) at electrodes F3, Fz, F4, Cz, P5, Pz and P6 was used for pattern

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-44

Table 2: Classification accuracy to detect the target stream for selective attention. Number of the data set for sample and test data, and the accuracy to detect Streams 1, 2 and 3 are shown.

Subject	Data Set		Accuracy [%]		
	Sample	Test	Stream 1	Stream 2	Stream 3
A	1	2	93.3	93.3	87.1
	2	1	93.5	96.7	96.7
B	1	2	100.0	90.0	93.5
	2	1	100.0	80.0	96.8
C	1	2	93.5	73.3	60.0
	2	1	77.4	96.7	90.0

classification. Data taken from each channel was down-sampled to 500 Hz. The resulting data was concatenated ($300 \text{ samples} \times 7 = 2100$ dimensions). The principal component analysis (PCA) was applied to sample data which has dimensions of $2100 \times \mathrm{N}$ (N: number of samples) to reduce dimensions and to determine a linear transformation matrix by limiting the variance to 99%. Dimensions of test data was reduced by the linear transformation matrix determined above.

A set of the data taken in a session was used as a sample data set, and that in another session was used as a test data set. For evaluating detection rate, all the responses to deviant tones in 10 s time window were classified and a classification accuracy was calculated in each stream [5].

RESULTS AND DISCUSSION

Fig. 2 shows the examples of averaged responses to deviant tones in three streams. Grey box denotes the significant difference between responses to target (attended) and non-target deviant tones (Student's t-test, $p < 0.01$). On Figs. 2(a), 2(b) and 2(c), the responses to the deviant tones when this subject paid attention to Stream 1, 2 and 3 are shown, respectively. In each of the graph, the responses to the deviant tones in Stream 1 (D_1: red), Stream 2 (D_2: green) and Stream 3 (D_3: blue) are displayed. From these figures, it was found that only the corresponding deviants to the target stream for selective attention elicited P300 at the latencies around 300 to 400 ms (D_1 in Fig. 2(a), D_2 in Fig. 2(b) and D_3 in Fig. 2(c)). This result was observed clearly on Subjects A and B.

Fig. 3 shows the topographies of the averaged responses to deviant tones in Stream 1, 2 and 3 (D_1, D_2 and D_3 respectively) on the same subject if Fig. 2 (Subject A). On Figs 3(a), 3(b) and 3(c), time courses of the topographies on three conditions when the subject paid attention to Stream 1, 2 and 3 are shown. From this figure, it can be found that the P300 component was only observed in the response to the deviant tones in the target (attended) stream. Like the result shown in Fig. 2, such a phenomenon could be observed clearly on Subject A and B. On Subject C, although no such a clear phenomenon was observed, the unique and corresponding responses were shown on the ERP responses to the deviant tones according to the target stream to pay attention.

Tab. 2 shows the classification accuracy to detect the object stream for subject's selective attention. There are two sessions that were executed for each subject, and the data sets taken from each of the two sessions were used as sample and test data for pattern classification. The classification accuracies on Subjects A and B were high, and the result on Subject C was much higher than chance level.

On the results shown above, it was shown that if the subject paid attention to one of the three tone streams, the deviant tone in the attended target tone stream elicited P300 activities and the target tone stream could be detected by pattern classification.

CONCLUSION

In this study, the auditory brain-computer interface (BCI) system based on auditory stream segregation was tested. This BCI system was the extension of the BCI proposed by the authors, on which two kinds of streams were used [5]. In this study, three oddball sequences consisted of musical tones at different frequency bands were presented alternately to subject's right ear, and stimulus parameters were set so the subjects could perceive these tone sequences as three segregated tone streams. Subjects were requested to pay attention to one of the three streams during experiments.

It was shown that the P300 was elicited only by target stream in two out of three subjects, and the target stream could be detected by pattern classification with high accuracies on all subjects.

The results in this study indicate that such an auditory stimuli based on auditory scene analysis [6] or auditory illusion can extend the auditory sound space for stimulus presentation and improve the current auditory BCI system. The evaluation of the present system by many subjects, and the detailed investigation on elicited ERP components and their pattern classification are left for further study.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-44

REFERENCES

[1] Donchin E, Spencer KM, Wijesinghe R. The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. IEEE Trans. Rehab. Eng., 2000; 8(2): 174-179.

[2] Acqualagna L, Blankertz B. Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). Clin. Neurophysiol., 2013; 124(5); 901-908.

[3] Hill NJ, Lal N, Bierig K, Birbaumer N, Scholkof B. An auditory paradigm for brain-computer interfaces. Advances in Neural Information Processing Systems, 2005; 17: 569-576.

[4] Schreuder M., Blankertz B., and Tangermann M. A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue. PLoS ONE, 2010; 5(4): e9813.

[5] Kanoh S, Miyamoto K, 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] Bregman AS, Auditory scene analysis: The perceptual organization of sound. The MIT Press, 1990.

[7] Kanoh S, Futami R, Hoshimiya N. Sequential grouping of tone sequence as reflected by the mismatch negativity. Biological Cybernetics, 2004; 91(6): 388-395.

[8] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd Edition, Wiley Interscience Publication, 2000.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-44

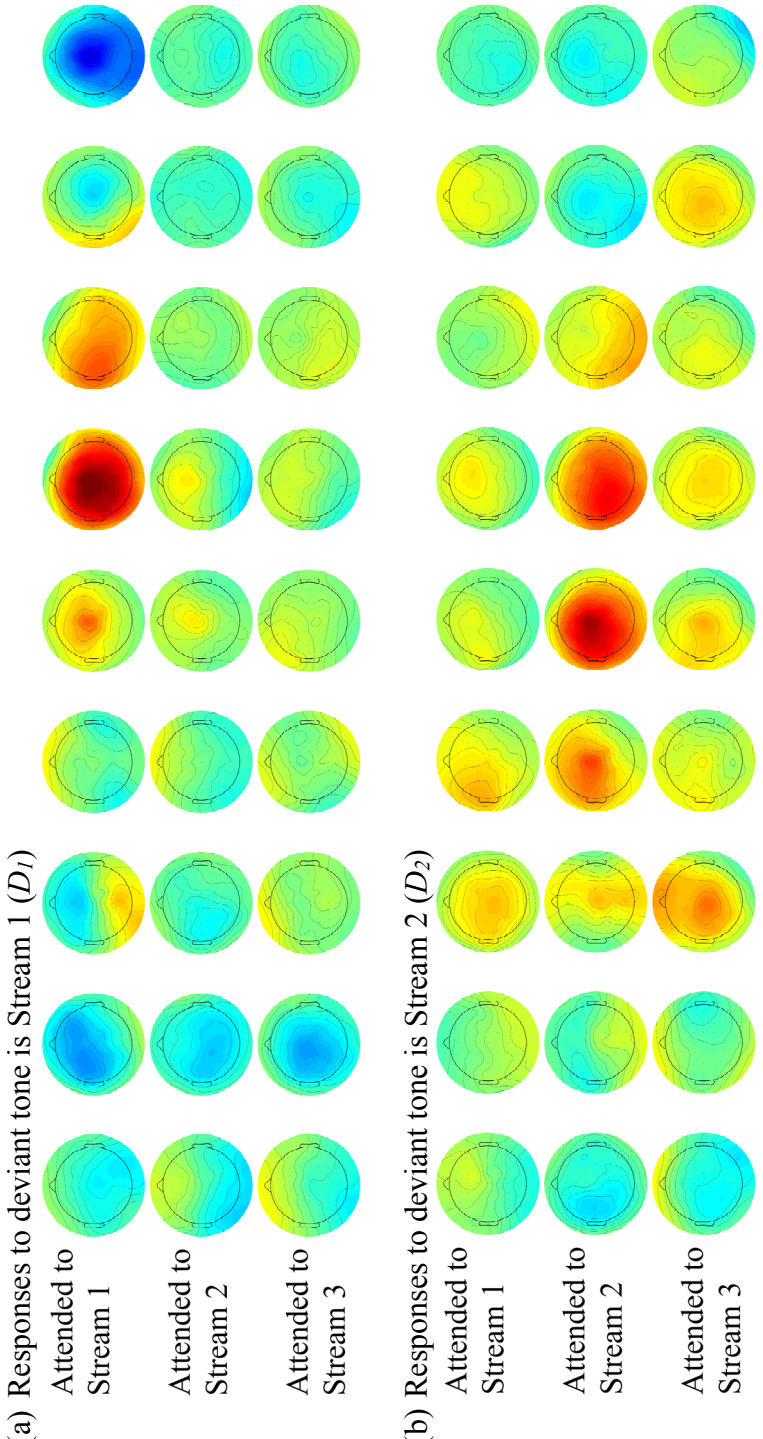


Figure 3: Topography of the averaged responses to deviant tones in Stream 1 (D_1: a), Stream 2 (D_2: b) and Stream 3 (D_3: c) of Subject A. In Figures (a), (b) and (c), time courses of the responses to corresponding deviant tones when the subject paid attention to Stream 1, 2 and 3 are shown.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-45

TUNING OF PARAMETERS FOR A VIBROTACTILE KINAESTHETIC FEEDBACK SYSTEM UTILIZING TACTILE ILLUSIONS

L. Hehenberger[1], A. I. Sburlea[1], G. R. Müller-Putz[1]

[1]Institute of Neural Engineering, Graz University of Technology, Graz, Austria

E-mail: gernot.mueller@tugraz.at

ABSTRACT: Kinaesthetic and haptic somatosensory feedback is an integral part of the natural movement feedback loop. Afflictions like spinal cord injury potentially disrupt both efferent and afferent pathways, and thus neuroprosthesis research must address both control and feedback. Artificial somatosensory feedback has great potential to inform the user in an intuitive way and facilitate the integration of a prosthesis into their body image.
Our aim is to provide kinaesthetic feedback of arm movements via a sparse grid of vibrating actuators, manipulating actuator intensities in such a way that tactile illusions of temporally and spatially continuous movement are evoked. To this end, we examine parameter spaces of apparent tactile motion and phantom sensations, in order to design a comprehensive feedback system.

INTRODUCTION

The ability to move is a vital prerequisite to leading a self-determined life. The goal of neuroprosthesis research is to provide a person afflicted with e.g. spinal cord injury with a tool to reclaim part of their autonomy in their daily life. An ideal neuroprosthesis would closely mimic the feedback loop of natural movement.
In recent years, progress has been made concerning intuitive control parameters from non-invasive electroencephalography, with respect to detecting goal-directed movement intention [1], [2], decoding grasps and upper-limb movements [3], [4] and decoding of kinematics [5], [6], as well as concerning the potential use of Error-Related Potentials to trigger corrective actions [7].
Apart from the challenges posed by the aspect of control, it is imperative that we also address feedback in order to facilitate intuitive use of neuroprosthesis. In many scenarios, neuroprostheses provide no natural means of somatosensory feedback to the user. Most commonly, (natural or artificial) visual feedback is used to compensate. While the visual sense has a high capacity to process a variety of information, the absence of other feedback modalities restricts the use of the visual sense for other tasks by mandating that the actions of the prosthesis be monitored at all times. Furthermore, matching the feedback modality and properties as closely as possible to the natural scenario would go a long way towards making prosthesis use more intuitive.
The absence of kinesthetic feedback has been shown to negatively impact the ability to produce cortical motor control commands [8]. Somatosensory feedback also plays an integral part in perceptually incorporating a prosthesis into one's own body image. In fact, a lack of tactile feedback is a prominent factor in prosthesis abondonment [9].
In non-invasive tactile stimulation, the most common modalities are mechanotactile [10], [11], electrotactile [12], [13], and vibrotactile. The vibrotactile modality has been used to provide force feedback [14] or discrete coded feedback for BCI applications [15], [16], but has also been demonstrated to be suitable to produce moving sensations by exploiting the inaccuracy of tactile perception.

As [17] and [18] demonstrated, when two stimuli are active with a certain temporal offset, termed Stimulus Onset Asynchrony (SOA) or Inter-Stimulus Onset Interval (ISOI), a moving sensation between stimulus locations ("apparent tactile motion") is perceived. If the SOA is too large, the two stimuli are perceived separately. If on the other hand it is too small, the two stimuli feel like one merged stimulus. In [19], a linear equation of optimal SOA control is proposed in dependence of the stimulus duration to reliably produce an apparent tactile motion.
Furthermore, two simultaneously active stimuli in relative proximity are perceived as a "phantom sensation" [20] in between the two stimulation sites. The location of the phantom sensation depends on the relative intensities of the physical stimuli. If they are equally strong, the perceived location of the phantom sensation is at the midpoint between the physical stimuli, while if they are uneven, the perceived location moves closer to the stronger stimulus. Concerning the mapping between stimulus intensities and the location of a phantom sensation, [20] examined linear and logarithmic relations, and claimed that a logarithmic mapping maintains constant intensity. Similarly, [21] compared one linear and three logarithmic models via a rating system with respect to consistency of perceived strength, location of the phantom sensation, and direction of movement, and found that the linear model fell short in all categories. In [19], a third mapping model was introduced, based on the energy summation model in the Pacinian channels. This publication concluded that this model can accurately

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-45

predict the intensity of the phantom sensation, and works better to map the phantom sensation location than both the linear and logarithmic models. Recently, [22] compared the three models with different actuator layouts and stimulation locations, in conjunction with subject-specific sensitivity adjustment, and concluded that the power model performs best for circular layouts while the logarithmic model is preferred for straight layouts. Sensitivity-adjusted models outperformed generic models across the board.

We are developing a vibrotactile stimulation system providing spatiotemporally continuous kinaesthetic feedback via a sparse tactor grid. To this end, we wish to exploit the aforementioned tactile illusions (i.e. apparent tactile motion and phantom sensations).

This work presents the results of two experiments that were performed to inform design choices for a vibrotactile stimulation system providing kinaesthetic feedback of arm movements.
In experiment 1, we employed a similar paradigm as proposed in [19] to determine the limits of SOA to reliably produce apparent tactile motion. The main objective was to verify that we can obtain comparable results with considerably larger stimulus durations.
Experiment 2 served the purpose of finding out which phantom location mapping model was best suited for our feedback system.
Experiment 1 was performed with a preliminary setup, while experiment 2 was performed using a prototype of a stimulation device for comprehensive feedback.

MATERIALS AND METHODS

Experiment 1: Seven participants (7 female, average age 27 years) took part in the first experiment. Two C-2 tactors (Engineering Acoustics Inc., Casselberry, USA) were placed on the subjects' right shoulder blade, with a horizontal inter-tactor spacing of 5 cm. The control signal was produced using an Arduino UNO (Arduino, Turin, Italy).
The stimulation frequency was fixed to 250 Hz, the recommended driving frequency of the C-2 tactors which is near the sensitivity peak of rapidly adapting Pacinian corpuscles [23]. Three stimulus durations were tested: 1200 ms, 2000 ms and 2800 ms. The order of the tactors and thus the direction of the perceived movement was varied pseudorandomly between subjects. In the beginning of the experiment, subjects completed a practice run to get used to the procedure, followed by one run per stimulus duration.

Participants were seated in front of a computer screen. They were subjected to a 1I-2AFC (one-interval two-alternative forced choice) paradigm with one-up one-down adaptive procedures. One run consisted of a test for the upper SOA threshold (continuous moving sensation vs. two discrete stimuli) and the lower SOA threshold (moving sensation vs. one perceptually merged stimulus) for a given stimulus duration.
In the test for the upper threshold, the initial SOA was considerably larger than the expected threshold (which was estimated in pre-tests). The stimulation sequence was presented to the subject, along with the instruction to focus on whether the stimuli felt apart or whether they could feel a transition. Then, they were prompted to answer the question of whether they perceived the stimuli as separate on a keyboard with "y" (two separate stimuli) or "n" (transition between the two stimuli). Subjects were allowed to take their time and replay the stimulation sequence as often as they wanted in order to produce a decisive answer. When the answer was "y", the SOA was decreased, in the opposite case it was increased again. After the SOA was decreased and then increased twice, the step size was decreased. After three more changes of direction at the smaller step size, the iteration was terminated.
The test for the lower threshold worked in a similar fashion. It started at a low SOA, and subjects were asked whether they perceived a movement sensation ("y"), as opposed to the stimuli feeling simultaneous ("n"). At every "y" answer, the SOA was decreased, while at every "n" answer it was increased.
The subject-specific thresholds are determined by averaging over the decisions after the first change of direction at the smaller step size.

The responses were recorded to a log file and analyzed using Matlab (Mathworks, Inc., Nattick, USA).

Experiment 2: The second experiment was conducted with six participants (4 female and 2 male, average age 27). Tactors were arranged as illustrated in Fig. 1, and held in place on the right shoulder blade using a custom shirt. To assure good contact of the tactors as well as comfort for the participants, shirts in several sizes were prepared. The inter-tactor spacing was between 5 cm and 6 cm, depending on the shirt size.
The stimulation frequency was 250 Hz. Each stimulation sequence lasted for 2 s. The direction of the simulated movement and the pairing of models (as well as their order) was varied pseudorandomly.

The tactors were controlled via an ARM Cortex M4 micro-controller (STMicroelectronics, Geneva, Switzerland) and custom amplifiers.

Participants were seated facing a computer screen. Tactor intensities were calibrated such that they were well perceived but not uncomfortable, and perceived equally strongly.
In each trial, the participant was presented with two stimulation sequences evoking the sensation of a unidirectional movement, starting from the middle tactor to one of the outer ones. Each sequence was computed to represent a movement with constant velocity, according to one of the three intensity models (linear, logarithmic,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-45

Figure 1: The stimulation device and tactor layout used in experiment 2. Inter-tactor spacing: 5-6 cm.

power). After the presentation of the sequence pair, the participant was prompted to choose which sequence they perceived to have a more constant velocity by clicking on the corresponding button on the screen using a computer mouse. Participants were allowed to replay sequence pairs as often as they needed.

The responses were recorded to a log file and analyzed using Matlab.

RESULTS

Experiment 1: Fig. 2 shows the distributions of the thresholds across subjects, where the upper box plots correspond to the upper thresholds, and the lower box plots to the lower thresholds, respectively. Median values are indicated in red. Fig. 3 depicts the mean thresholds over subjects with black circles. The shaded area between represents the parameter space where a movement sensation exists, with the center of this space indicated by the black line. This line represents the optimal offsets to evoke apparent tactile motion. The dashed blue line is an extrapolation of the analogous line determined in [19] for stimulus durations of 40 ms and 160 ms.

Experiment 2: The distribution of subject choices is illustrated in Fig. 4. The colours in the left-hand plot identify the individual subjects. The preferred model of each subject is marked with a star.
Three subjects preferred the power model (on average by 15% compared to the logarithmic model), and two subjects narrowly preferred the log model over the power model (by 3%), while one subject equally preferred those two models. The linear model was the least preferred by every subject.
Subjects 1-5 most frequently replayed sequences for the pairing of the logarithmic vs. the power model, while subject 6 replayed more often for the linear vs. logarithmic pairing. Furthermore, it took subject 1-4 the longest to reach a decision between the logarithmic and the power model; subjects 5 and 6 took longer to decide between the linear and the logarithmic model. The individual average

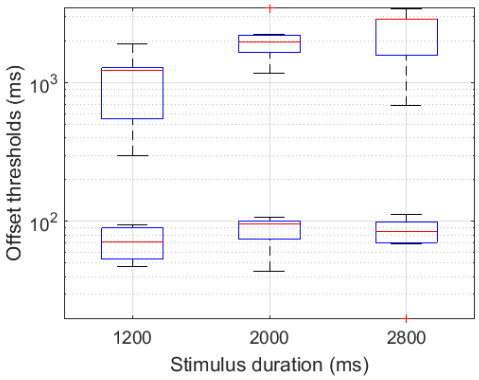


Figure 2: Upper and lower SOA thresholds. The top box plots represent the distribution of the upper thresholds across subjects, and the bottom box plots the distribution of the lower thresholds, respectively, each for the three durations that were tested.

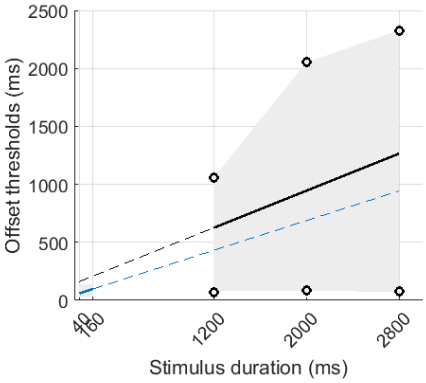


Figure 3: Mean thresholds and optimal SOA for movement sensation. The black circles mark the average thresholds, with the parameter space in between (where a movement sensation exists) shaded in grey. The black line indicates the optimal SOA as the center of this space, while the dashed blue line identifies an extrapolation of the analogous result obtained for stimulus durations of 40 ms and 160 ms in [19].

response times and number of repeats are listed in Fig. 5, where the response time is defined as the total amount of time needed to reach a decision, including the presentation of the stimulation sequences.

DISCUSSION

We conducted two experiments to tune parameters for a vibrotactile feedback system utilizing tactile illusions to provide smooth kinaesthetic feedback.

In experiment 1, we identified the range of SOA to produce apparent tactile movement, employing a similar paradigm as in [19], but with longer stimulus durations (by a factor of $\approx$ 20-30). On average, a line of optimal SOA control can be determined that is similar to the extension of the one proposed in [19], albeit with a rather large inter-subject variance.
Some subjects initially struggled to identify the transition

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-45

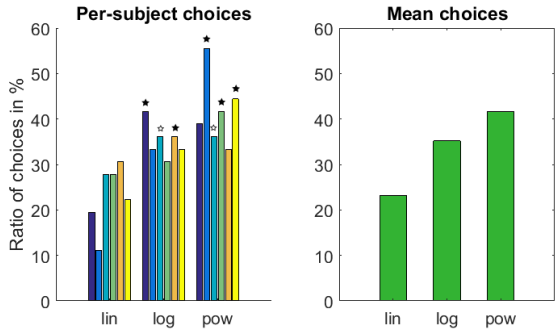
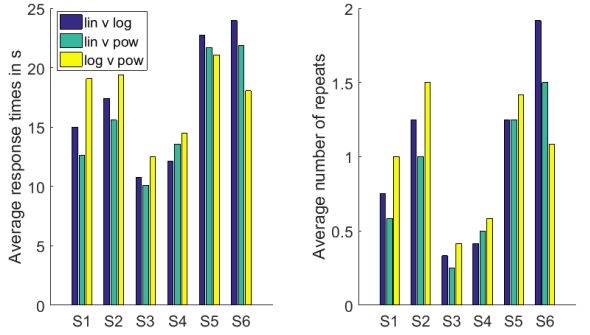


Figure 4: Distribution of model choices. Black stars mark the preferred model for each subject (white stars indicate equal preference of two models).



(a) Average response times (b) Average number of repeats

Figure 5: Average response times and number of repeats of individual subjects. The response time is the time needed to reach a decision, including the presentation of stimuli.

around either the lower or upper threshold, but proceeded to oscillate around their perceived threshold after some time.

The objective of experiment 2 was to determine the optimal mapping of tactor intensities to produce a smooth movement sensation. To this end, we chose a paradigm using dynamic stimulation sequences rather than temporally discrete stimuli. Our results are in agreement with previous ones obtained in different testing scenarios with discrete stimuli ([19], [20], [22], [21]), in that logarithmic and power models are found to be superior to a linear model.
Most subjects found it harder to decide between logarithmic and power model sequences, as reflected in the longer response times and the higher number of replays for these pairings.

The first experiment was conducted with a makeshift setup that permitted fewer options to adapt to subjects' sensitivity profiles. However, this did not constitute a problem, since a calibration of the output intensities of both tactors to the same value proved to be sufficient, and all intensities were kept constant over the duration of the experiment.
Experiment 2, on the other hand, demanded a more flexible setup with fine control of individual tactor intensities. In both instances, subjects reported the stimulation to be pleasant.

CONCLUSION

The results presented above are used to tune parameters for a vibrotactile feedback system intended to provide artificial kinaesthetic somatosensory feedback of arm movements.
During experiment 2, a prototype of a custom stimulation device was utilized. We conclude that the device, as well as the tactor shirts and calibration procedure are well suited to our purposes.

ACKNOWLEDGEMENTS

This work was supported by the ERC Consolidator Grant 681231 "Feel Your Reach".

REFERENCES

[1] Pereira J, Ofner P, Schwarz A, Sburlea AI, Müller-Putz GR. Eeg neural correlates of goal-directed movement intention. NeuroImage. 2017;2017(149):129–140.
[2] Pereira J, Sburlea AI, Müller-Putz GR. Eeg patterns of self-paced movement imaginations towards externally-cued and internally-selected targets. Sci. Rep.. 2018;8(13394).
[3] Ofner P, Schwarz A, Pereira J, Müller-Putz GR. Upper limb movements can be decoded from the time-domain of low-frequency eeg. Plos One. 2017;12(8).
[4] Schwarz A, Ofner P, Pereira J, Sburlea AI, Müller-Putz GR. Decoding natural reach-and-grasp actions from human eeg. J. Neural Eng.. 2017.
[5] Ofner P, Müller-Putz GR. Decoding of velocities and positions of 3d arm movement from eeg. Conf. Proc. IEEE Eng. Med. Biol. Soc.. 2012;2012:6406–6409.
[6] Kobler RJ, Sburlea AI, Müller-Putz GR. Tuning characteristics of low-frequency eeg to positions and velocities in visuomotor and oculomotor tracking tasks. Sci. Rep.. 2018;2018(17713).
[7] Lopes Dias C, Sburlea AI, Müller-Putz GR. Masked and unmasked error-related potentials during continuous control and feedback. J. Neural Eng.. 2018;2018(15).
[8] Galán F, Baker MR, Alter K, Baker SN. Degraded eeg decoding of wrist movements in absence of kinaesthetic feedback. Hum. Brain Mapp.. 2015;2015(36):643–654.
[9] Biddiss E, Beaton D, Chau T. Consumer design priorities for upper limb prosthetics. Disabil. Rehabil. Assist. Technol.. 2007;6(2):346–57.
[10] Patterson PE, Katz JA. Design and evaluation of a sensory feedback system that provides grasping pressure in a myoelectric hand. J. Rehabil. Res.. 1992;29(1):1–8.
[11] Antfolk C, D'Alonzo M, Controzzi M, Lundborg G, Rosén B, Sebelius Fea. Artificial redirection of sensation from prosthetic fingers to the phantom hand map on tran-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-45

sradial amputees: Vibrotactile versus mechanotactile sensory feedback. IEEE T. Neur. Sys. Reh.. 2013;21(1):112–120.
[12] Rita P Bach-y, Kercel SW. Sensory substitution and the human–machine interface. Trends Cogn. Sci.. 2003;7(12):541–546.
[13] Franceschi M, Seminara L, Dosen S, Strbac M, Valle M, Farina D. A system for electrotactile feedback using electronic skin and flexible matrix electrodes: Experimental evaluation. IEEE Trans. Haptics. 2016;10(2):162–172.
[14] Antfolk C, Balkenius C, Lundborg G, Rosén B, Sebelius F. Design and technical construction of a tactile display for sensory feedback in a hand prosthesis system. Biomed. Eng. OnLine. 2010;9(50).
[15] Cincotti F, Kauhanen L, Aloise F, Palomäki T, Caporusso N, Jylänki Pea. Vibrotactile feedback for brain-computer interface operation. Comput. Intel. Neurosc. 2007;2007(48937).
[16] Chatterjee A, Aggarwal V, Ramos A, Acharya S, Thakor NV. A brain-computer interface with vibrotactile biofeedback for haptic information. J. Neuroeng. Rehabil. 2007;4(40).
[17] Sherrick CE, Rogers R. Apparent haptic movement. Percept. Psychophys. 1966;1(3):175–180.
[18] Kirman JH. Tactile apparent movement: The effects of interstimulus onset interval and stimulus duration. Percept. Psychophys. 1974;15(1):1–6.
[19] Israr A, Poupyrev I. Tactile brush: Drawing on skin with a tactile grid display. In: Proc. SIGCHI CHI'11. New York, NY, USA, 2011, 2019–2028.
[20] Alles DS. Information transmission by phantom sensations. IEEE T. Hum.-Mach. Syst. 1970;11(1):85–91.
[21] Leeb R, Gwak K, Kim DS, Millán JdR. Freeing the visual channel by exploiting vibrotactile bci feedback. In: Conf. Proc. IEEE EMBS 2013. Osaka, Japan, 2013.
[22] Luzhnica G, Stein S, Veas E, Pammer V, Williamson J, Murray Smith R. Personalising vibrotactile displays through perceptual sensitivity adjustment. In: Proc. ISWC '17. Maui, Hawaii, USA, 2017.
[23] Keidel WD. Vibrationsreception, der erschütterungssinn des menschen. Universitätsbund Erlangen, Erlangen, Germany (1956).

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-46

INTRODUCING A MOTIVATING TRAINING STUDY DESIGN TO COMPARE AUDITORY AND TACTILE STREAMING-BASED P300 BCIS

P. Ziebell[1], J. Stümpfig[1], S.C. Kleih[1], M.E. Latoschik[2], A. Kübler[1], S. Halder[3]

[1] Institute of Psychology, University of Würzburg, Würzburg, Germany
[2] Institute of Computer Science, University of Würzburg, Würzburg, Germany
[3] School of Computer Science and Electronic Engineering (CSEE), University of Essex, Colchester, United Kingdom

E-mail: philipp.ziebell@uni-wuerzburg.de

ABSTRACT: Since neurodegenerative diseases or brain injuries causing locked-in-syndrome (LIS) might lead to loss of vision, different approaches of vision-independent BCIs were developed, for example P300 BCIs using auditory or tactile stimuli. However, BCI-inefficiency has been reported in these approaches as well, with high workload proposed as an underlying problem.
In this contribution, a motivating training study design is introduced to compare auditory with tactile P300 BCIs using a streaming-based approach of stimulus presentation that was proposed as a relatively low-workload alternative to classic approaches of sequential stimulus presentation.
First performance results of $N = 6$ healthy participants were examined case-wise – all participants were able to use at least one BCI version successfully. The preliminary results indicate high motivation and show that there is no superior modality per se but individual preferences in stimulus modalities. This should be considered for future research with healthy as well as LIS users.

INTRODUCTION

The P300 BCI enables users to control various applications, for example text spelling devices, via the non-muscular pathway of electroencephalography (EEG), primarily relying on event-related potentials (ERP), in particular the P300 component [1].
To elicit the P300, an external stimulation is necessary – in P300 BCIs, visual stimulation has until recently been the major focus of research. However, since the eye-gaze dependence of these visual approaches has been pointed out – which is especially problematic for a main target population of potential BCI users, namely severely paralyzed patients suffering from locked-in-syndrome (LIS), a condition often accompanied by a drastic loss of gaze control – eye-gaze independent BCIs exploring alternative P300 stimulation methods have become a new focus of research [2].
Promising results with LIS patients have been reported in a training study using a multi-class auditory P300 BCI; however, cases of BCI-inefficiency occurred as well, with high workload of the auditory multi-class stimulation approach being identified as a problematic factor [3]. A multi-class tactile P300 BCI also showed encouraging results in a training study with elderly healthy participants aged between 50 and 73, but BCI-inefficiency occurred in this approach, too [4]. Eventually, it also yet has to be tested with LIS users and might be problematic with regard to a potentially high workload for this user group in a similar way as the auditory sequential multi-class approach.
To decrease workload problems, an alternative approach, the streaming-based auditory P300 BCI, has been developed. This approach alters the common sequential multi-class presentation of stimuli in one stimulus stream by arranging them in a two-class stimulus presentation with two stimulus streams. First positive results with healthy and LIS users could be shown as well, indicating its potential as an intuitively usable communication device [5].
The current study aimed to replicate the positive results of this intuitive two-class auditory streaming-based P300 BCI with alternative stimuli and furthermore to create a two-class streaming-based tactile P300 BCI in a detailed training case-study approach. Since potential pitfalls of BCI training studies have been pointed out [6], attention was focused on developing a user-friendly study design. Therefore, outcome measures were selected based on the user-centered design approach to evaluate exhaustively the usability of BCI-controlled applications [7]. To ensure high motivation during the whole study, the guidelines by [8] were followed and ideas from BCI gaming literature were taken into account (e.g. [9]), resulting in a “Star Wars”-themed study design.
We hypothesized that the streaming-based P300 BCI version with auditory stimulation as well as the version with tactile stimulation are both intuitively usable and that user performance would further increase via training. Subsequently, preferences for one of the BCI versions and a potential transfer of learning were explored in an additional transfer session after three training sessions, where users switched from the auditory version to the tactile version or vice versa.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-46

MATERIALS AND METHODS

Participants: Herein reported are data from *N* = 6 healthy participants between 20 and 41 years (2 female). Participants were recruited via internet advertisement and compensated financially for participation. Exclusion criteria were: no auditory or neurological impairments, no use of psychotropic substances, no left-right disorientation and no previous BCI-use. All participants spoke German at native-speaker level. The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki [10].

EEG data collection: EEG was recorded using a g.USBamp-amplifier and 16 Ag/AgCl active electrodes (g.Ladybird electrodes on a g.Gamma cap; products of g.tec Medical Engineering GmbH, Austria, http://www.gtec.at/). Electrodes were placed on the following 10-20-system positions (modified international standard [11]): AF7, Fpz, AF8, F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4 and Pz. The reference electrode was attached to the right earlobe, the ground electrode to position AFz. The signal was band-pass filtered between 0.1 to 30 Hz, notch-filtered at 50 Hz and recorded with a sampling rate of 256 Hz. A Hewlett–Packard ProBook 6460b with Dual-Core-CPU, 4 GB RAM and 64-Bit Windows 7 was used for data collection, BCI2000 was used for P300 stimulus presentation as well as recording and processing of the signal [12]. These EEG data formed the basis for calculating objective measures for successful BCI-use, covering the BCI-usability criteria effectiveness and efficiency.

Design of the streaming-based P300 BCIs: As a first step, the intuitive streaming-based P300 BCI version with auditory stimulation was designed as a slight modification of the one introduced by [5]. The streaming-based tactile P300 BCI was inspired by the auditory version and furthermore incorporated ideas from earlier tactile P300 BCI research [4]. A major focus while designing the two BCI versions was to maximize their comparability. In both versions, an object on a computer screen had to be moved toward a target, via a predetermined pathway of single steps leading left, right, up or down. The participants' task was to choose the correct direction using the BCI. In the current study only left and right selections had to be actively chosen. Figure 1 shows a schematic of the task accompanied by a more detailed description.

For the auditory streaming-based P300 BCI version, one stimulus-channel was presented to the right ear and the other stimulus-channel to the left ear, via stereo-headphones (Sennheiser HD 280 PRO, http://de-de.sennheiser.com/hd-280-pro). On the right ear, the German word "rechts" (meaning "right") served as a target-stimulus and the English word "right" served as a non-target-stimulus, while on the left ear, the German word "links" (meaning "left") served as target-stimulus and the English word "left" served as non-target stimulus. The words "links" and "left" were spoken by a female voice, the words "rechts" and "right" by a male voice, duration of each word was standardized to 500 ms, volume and tone of voice were kept neutral. At the end of the setup, volume was individually adjusted for each participant, so that all stimuli were pleasantly audible, easy to discriminate and none more salient in comparison to the other stimuli.

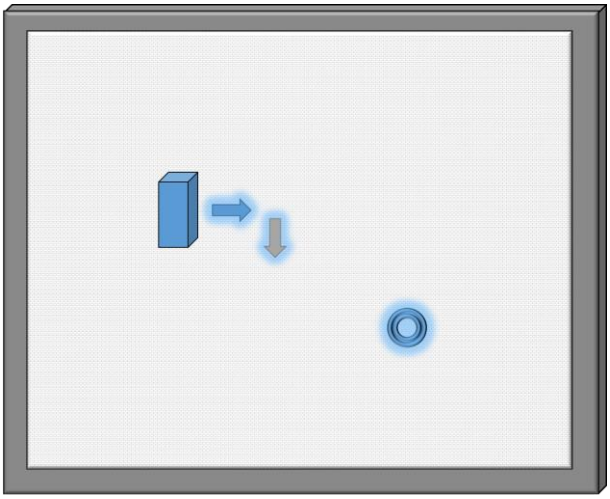


Figure 1: Schematic of the users' task. Since the current study involved only healthy participants without severe visual impairments, the direction selections that had to be chosen via BCI were announced visually on a computer screen. An object (cuboid) had to be moved along a predetermined pathway of glowing left/right arrows to a goal (glowing ring). In addition, up/down arrows were integrated, to create more diverse pathways and to potentially expand the task and the P300 BCI by including two more direction selection possibilities. In the current study the up/down arrows were automatically shown and selected with the previous left or right arrow. Only the currently relevant direction selection arrow was shown on the computer screen, to facilitate concentration on the currently relevant direction selection and to not forestall the whole pathway of arrows, which might have induced boredom. After every correct direction selection, users received auditory feedback in form of a positive sound. If users selected the wrong direction, the sound remained absent. Then the hitherto existing arrow vanished, the object was moved to the next position, the next and now currently relevant arrow appeared and the next direction selection began. The glowing goal ring was presented permanently on the screen. For the object, the screen wallpaper, the arrows and the goal, different pictures were inserted to increase task variability.

The tactile streaming-based P300 BCI version used the right forearm as one stimulus-channel and the left forearm as the other stimulus-channel. Therefore, four tactile stimulators which were able to induce tactile sensations via vibration were attached using elastic bands (C-2 Tactors, Engineering Acoustics, Inc., https://www.eaiinfo.com/). One stimulus-channel consisted of two tactile stimulators on the inner side of

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-46

the right forearm and the other of two tactile stimulators on the inner side of the left forearm. On each of those two stimulus channels one tactile stimulator was placed near the right/left wrist. These two tactile stimulators were defined as the emitters of the tactile target-stimuli to select a right direction selection (right wrist) or a left direction selection (left wrist). The remaining tactile stimulators were placed near the inner side of the elbow on the right/left forearm and defined as the emitters of the tactile non-target-stimuli. All four tactile stimulators vibrated with a stimulus duration of 125 ms and an inter-stimulus-interval of 250 ms from the end of one stimulus to the beginning of the next stimulus. A paper tissue was placed between the skin and each tactile stimulator to reduce EEG artifacts that could potentially originate from the tactile stimulators. As in the auditory version, participants were eventually asked, if the tactile stimuli were easy to discriminate from each other and if the caused vibrations were perceived as equally strong and pleasant.

In both versions, auditory as well as tactile, a stimulus-selection-sequence consisted of 10 stimuli in total with one target stimulus and four non-target-stimuli per channel. After the start of the first stimulus on one channel, the stimuli of the other channel were started in constant anti-phase to the stimulus-stream of the first channel – in both streaming paradigm versions, the stimuli of the second channel started 375 ms after the corresponding stimuli of the first channel, which lead to the total length of each stimulus-selection-sequence being 3.75 seconds. Within each stimulus-selection-sequence, the target-stimuli and non-target-stimuli were presented in a pseudo-randomized order to enlarge the P300 effect via heightened unpredictability. To give participants a chance to refocus, a break of four seconds was included after the last stimulus of each stimulus-selection-sequence, before the next stimulus-selection-sequence could get started.

Training protocol: Since recent training studies mainly showed increases until the third training session, but mostly plateau effects in the sessions thereafter (e.g. [3, 4]) and the streaming-based P300 BCI approach has been shown to be intuitively quick to learn [5], the current study included three training sessions, t1, t2 and t3, before the additional transfer session, t4, where users switched from the auditory version to the tactile version or vice versa, so all in all four sessions. The sessions were individually assigned on four different days with at least one day and at most six days without BCI-use between each session. Before the first session, the participants were pseudo-randomly assigned to one of two groups: Group AT, using the auditory BCI version at t1, t2 and t3, transferring to the tactile version at t4, and Group TA, using the tactile BCI version at t1, t2, and t3, transferring to the auditory version at t4.

Each training session started with an individual calibration to estimate the optimal parameters for EEG signal classification (based for example on [3]). First, users sat down in front of the computer screen in a comfortable position, then EEG, headphones, and in case of the tactile version, tactile stimulators were set up as described above. To prevent EEG artifacts, participants were instructed to keep their eyes open and to avoid unnecessary movements during the stimulus-selection-sequences. Thereafter the experimenters explained the BCI, recommending two strategies, first, focusing on the relevant stimulus-channel and blending out the irrelevant stimulus-channel, and second, counting the target-stimuli of the relevant stimulus channel as a way to focus attention on them. Then the participants could ask questions. The calibration consisted of 24 direction-selections, each consisting of 10 stimulus-selection-sequences, divided in two runs with 12 direction-selection-sequences and a one-minute break in between. Based on that, the optimal classification weights and the number of optimal stimulus-selection-sequences for the subsequent online-classification-runs were estimated using the heuristic described by [3]. The online-classification-runs had 12 direction selections and per training session, four online-classification-runs had to be completed, resulting in 48 direction selections. For both the calibration-runs as well as the online-classification-runs, the total number of left direction selections was exactly equal to the total number of right direction selections. Sequence-effects were balanced and each direction selection did not occur more than three times in a row, to avoid monotony which might lead to a weaker P300 response.

To embed the calibration-runs into the motivational context of "Star Wars", they were named "Introduction", like the introduction to the magical "Star Wars" power called "the Force", which some "Star Wars" characters experience in their adventure stories. The object that needed to be moved was a "Star Wars" light saber and it was moved before a "Star Wars" background, a "Jedi Knight Temple". The online-classification-runs were called "Missions", where different objects out of the "Star Wars" movies had to be moved through different "Star Wars" backgrounds, with each object and background forming a scene from a "Star Wars" movie. To give immediate feedback on successful direction selection, a one-second long feedback sound was given via headphones after each successful selection, in the auditory as well as the tactile version. A positive sound of the popular "Star Wars" droid "R2D2" was chosen ("R2D2a.wav", http://www.galaxyfaraway.com/gfa/1998/12/star-wars-sounds-archive/), to fit into the "Star Wars" context. If the selection was incorrect, no feedback sound was given. To further support the atmosphere, short pretexts inspired by the introduction text from the "Star Wars" movies were created and presented, to announce the upcoming task (example: "Move the Millennium Falcon through the clouds of Bespin…"). All object and background pictures were collected via internet search engines, each object-background combination was checked for perceptibility and the backgrounds were checked to be free of potential distractions (as recommended by [9]).

Data analysis: To evaluate the effects of our

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-46

experimental design consisting of two independent variables, group (AT, TA) and session (t1, t2, t3, t4), the following dependent variables were analyzed: EEG ERPs following targets and non-targets, online-selection-accuracy (P), information transfer rate (ITR) and the number of possible selections per minute (SPM), which was selected based on the users' performance during the offline-calibration.

EEG data were analyzed offline with the MATLAB toolbox EEGLAB and additional MATLAB scripts ([13], https://sccn.ucsd.edu/eeglab/). Artifacts were removed using artifact subspace reconstruction on the continuous data [14]. The data were segmented in epochs ranging from -200 ms to 800 ms surrounding stimulus-presentation, the epoch of -200 ms to 0 ms before stimulus presentation served for baseline correction. Epochs with extreme amplitudes or distributions were rejected. Epochs related to target-stimuli were grouped and averaged; the same procedure was applied to the non-target-stimuli. Since the P300 was expected to be most pronounced at electrode position Cz (e.g. [3, 4]), analysis was focused on Cz.

To determine the weights and the number of necessary stimulus-selection-sequences for online-run signal classification, a stepwise linear discriminant analysis (SWLDA) was applied to the EEG data from the calibration-runs, for a detailed description of the used algorithm see [15]. SWLDA has been used successfully in earlier auditory and tactile paradigms, with the best results obtained if calculated anew for each individual session, allowing the weights and number of stimulus-selection-sequences to change from session to session via training (e.g. [3]).

To explore successful BCI-use on an objective level, P and ITR were calculated. P served as a measure for the BCI-usability criterion effectiveness and was defined as the percent value of correct selections out of all given selections of a session. As a measure of BCI-inefficiency it was analyzed how many participants could exceed the threshold of 70% selection-accuracy introduced by [16]. ITR served as a measure for the BCI-usability criterion efficiency, and was defined as the amount of correctly transferred information during the time interval of one minute, using the following formula (1) as recommended by [17]:

$$B = \log_2 N + P * \log_2 P + (1-P) * \log_2 \left(\frac{1-P}{N-1}\right) \qquad (1)$$

With B standing for bits per selection, N standing for number of possible selection-targets and P standing for the estimated probability of a correct classification, based on the online-selection-accuracy. To calculate the ITR, B was multiplied with the SPM, using the following formula (2), with S as the number of stimulus-selection-sequences that was chosen for each participant at each session and taking into account the duration of each stimulus-selection-sequence (3.75 s) as well as the post-stimulus-selection-sequence break (4 s):

$$SPM = \frac{60\ \text{s}}{S * 3.75\ \text{s} + 4\ \text{s}} \qquad (2)$$

RESULTS

EEG ERP data averages for each participant are depicted in Figure 2. Visual inspection suggests that the participants AT-1, AT-2, AT-3 and TA-2 were able to produce a solid P300 ERP pattern, with AT-2 slightly increasing and TA-2 slightly decreasing between t1, t2 and t3, while AT-1 and AT-3 remained stable between t1, t2 and t3. AT-1 is the only participant that increased to t4, while AT-2, AT-3 and TA-2 show a lesser to equally pronounced P300 ERP at t4. TA-1 and TA-3 do not show a solid P300 pattern over t1, t2, t3 and t4.

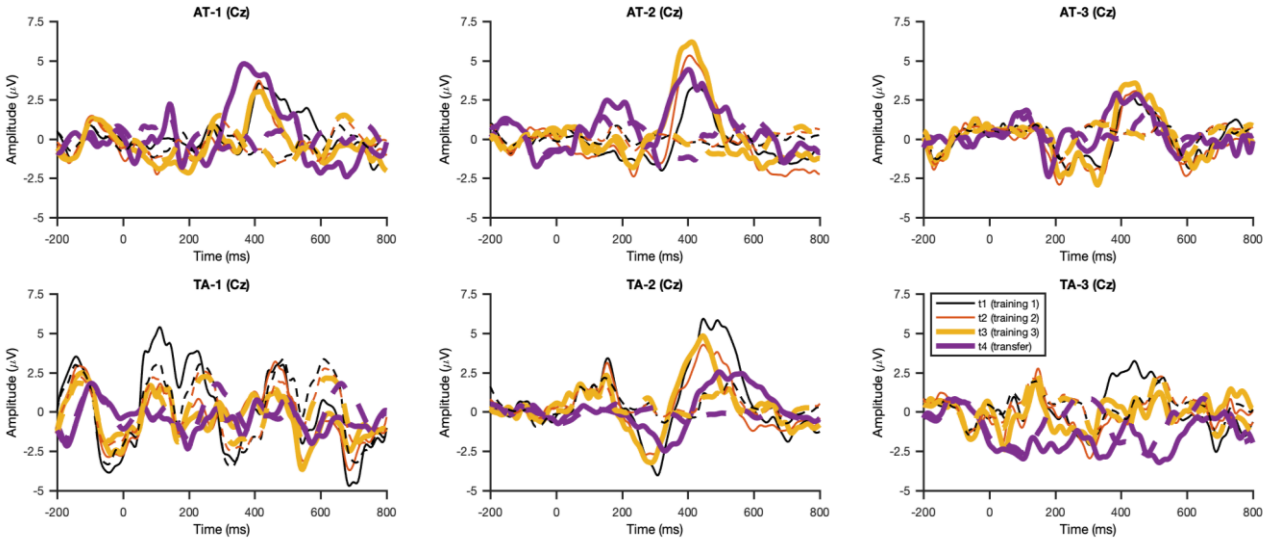


Figure 2: Average EEG ERP patterns for each participant over sessions t1, t2, t3 and t4, indicated by color, see legend. ERPs following target stimuli occurring at 0 ms are indicated by continuous lines, ERPs following non-target stimuli are indicated by dotted lines. Since earlier studies found the most pronounced P300 ERP pattern at electrode position Cz (e.g. [3, 4]), current analysis was focused on Cz.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-46

Table 1 shows the development of P over t1, t2, t3 and t4 for each participant. While AT-1 shows a steadily increasing value of P that is above the BCI-inefficiency-threshold of 70% at t1, t2, t3 and t4, AT-2 shows a value of P above the BCI-inefficiency-threshold at t1, t2 and t3, but below the BCI-inefficiency-threshold of 70% selection-accuracy at t4. AT-3 shows values slightly above 70% at t1 and t3, the lowest value of P of all participants at t2 (27.08%), but the highest value of P of all participants at t4 (97.92%). TA-1 and TA-3 show values of P above the 70%-criterion at session t2 and t4 (TA-1) and t1, t2 and t3 (TA-3). TA-2 shows the highest P of all participants at t1 (89.58%), but values below the 70%-threshold at t2, t3 and t4.

Table 1: Development of P (% correct selections) over training sessions t1, t2 and t3 to transfer session t4 for the participants of group AT and group TA, values below the BCI-inefficiency-threshold of 70% selection-accuracy [16] are highlighted by grey background

Participant	*t1*	*t2*	*t3*	*t4*
AT-1	77.08	79.17	87.50	95.83
AT-2	75.00	93.75	81.25	64.58
AT-3	70.83	27.08	72.92	97.92
TA-1	56.25	87.50	64.58	87.50
TA-2	89.58	68.75	68.75	66.67
TA-3	75.00	81.25	83.33	58.33

ITR values over t1, t2, t3 and t4 for each participant are listed in Table 2. Since P contributes a substantial part to the calculation of the ITR, it is to be expected, that if participants had low values of P at a certain session, the ITR is relatively low as well. This pattern is evident in Table 2, especially if participants were not able to exceed the 70%-criterion.

Table 2: Development of ITR (bits/min) over training sessions t1, t2 and t3 to transfer session t4 for the participants of group AT and group TA, values affected by values of P below the BCI-inefficiency-threshold of 70% selection-accuracy [16] are highlighted by grey background

Participant	*t1*	*t2*	*t3*	*t4*
AT-1	0.51	0.69	1.20	1.98
AT-2	0.43	2.09	1.20	0.16
AT-3	0.29	0.00	0.31	1.93
TA-1	0.02	1.03	0.14	1.20
TA-2	1.17	0.21	0.27	0.16
TA-3	0.50	0.96	1.10	0.05

SPM is reported in Table 3 over t1, t2, t3 and t4 for each participant. Since the ITR is determined by P and the SPM, ITR scores are relatively high, if a relatively high value of P is achieved in combination with a relatively high SPM value, which lead to ITR values that are more than twice as high as the ITR at t1, for example at t2 (AT-2, TA-1, TA-3), t3 (AT-1, AT-2, TA-3) or t4 (AT-1, AT-3, TA-1).

Table 3: Development of SPM (1/min) over training sessions t1, t2 and t3 to transfer session t4 for the participants of group AT and group TA, numbers in brackets indicate the corresponding number of chosen stimulus-selection-sequences based on the offline-calibration-heuristic [3]

Participant	*t1*	*t2*	*t3*	*t4*
AT-1	2.26 (6)	2.64 (5)	2.64 (5)	2.64 (5)
AT-2	2.26 (6)	3.16 (4)	3.93 (3)	2.64 (5)
AT-3	2.26 (6)	1.98 (7)	1.98 (7)	2.26 (6)
TA-1	1.45 (10)	2.26 (6)	2.26 (6)	2.64 (5)
TA-2	2.26 (6)	1.98 (7)	2.64 (5)	1.98 (7)
TA-3	2.64 (5)	3.16 (4)	3.16 (4)	2.26 (6)

DISCUSSION

The descriptive analysis of the six participants revealed a high variety of inter-individual differences of BCI-usability, measured by EEG ERPs at electrode position Cz, P, ITR and SPM. The analysis of P, ITR and SPM revealed that five participants were able to use the BCI version they were first assigned to intuitively with above 70% accuracy in the first session (all except TA-1) and five participants profited from training at the second and/or third session (all except TA-2). Three participants profited from transfer to the BCI version they did not use before in the last session (AT-1, AT-3, TA-1), while three were not able to transfer successful BCI-use in the last session (AT-2, TA-2, TA-3). Even though BCI-inefficiency occurred in some sessions, all participants showed successful BCI-use in at least one session, using either the developed streaming-based P300 BCI version with auditory stimulation or the tactile version or in three cases with both versions, where also successful transfer of learning was possible (AT-1, AT-3, TA-1). Despite some participants partly showing BCI-inefficiency, there were no drop-outs and, defying potentially discouraging results, every participant completed all sessions, leading to success for example in the case of AT-3, who showed the lowest value of P of all participants at the second training session (27.08%), but eventually the highest value of P of all participants at the transfer session (97.92%), potentially facilitated by transfer of learning.

Yet, several limiting aspects have to be noted, indicating the need for further research. After showing the best results in the first training session, participant TA-2 failed to replicate these results in the following three sessions, even though TA-2 produced a relatively solid ERP pattern. TA-2 reported emerging monotony throughout the paradigm post-hoc after the last session, indicating that the used tactile BCI version might not be optimal. Furthermore, the P300 BCI is thought to be driven by P300 elicitation, but high values of P, ITR or SPM did not always correspond with a clear P300 pattern at Cz (e.g. TA-1 at t2 and t4, TA-3 at t2 and t3) or vice versa (e.g. AT-2 at t4, AT-3 at t2, TA-2 at t2, t3 and t4). This should be further examined, especially

Published by Verlag der TU Graz
Graz University of Technology

since only one electrode position was considered in the current analysis.

Training lead to relatively high ITR increases, but ITR in general was relatively low in comparison to earlier studies (e.g. [3, 4]). Both used BCI versions should be optimized with regard to ITR maximization, for example by lowering the amount of time needed for target-selection (e.g. using less non-targets) and by updating the signal processing pipeline (e.g. using shrinkage LDA). Eventually, both BCI versions have to be tested by LIS users, for example in multiple training sessions or long-term independent home use [1, 3] and therefore include a completely non-visual task design.

For future research, a more exhaustive analysis on a higher number of participants is planned, taking into account data not yet reported in this paper, like questionnaire measures of motivation and of the user-centered design criteria efficiency and satisfaction, further ERP analysis (e.g. different electrode positions) as well as including inferential statistics.

CONCLUSION

The presented results indicate that the streaming-based P300 BCI is a promising approach with auditory stimuli as well as with tactile stimuli and that the study design motivated the participants; all could achieve successful BCI-use in at least one session of the current study. It seems that there is no superior modality, but individual preferences. These findings should be considered for future research with healthy and LIS users.

REFERENCES

[1] Kübler, A., Müller, K.-R., & Guan, C. (2017). The P300 BCI: on its way to end-users? In Müller-Putz, G.R., Steyrl D., Wriessnegger S.C., & Scherer, R. (Eds.), *Proceedings of the 7th Graz Brain-Computer Interface Conference 2017.* Graz: Verlag der Technischen Universität Graz. doi: 10.3217/978-3-85125-533-1-00.

[2] Riccio, A., Mattia, D., Simione, L., Olivetti, M., & Cincotti, F. (2012). Eye-gaze independent EEG-based brain–computer interfaces for communication. *Journal of Neural Engineering, 9(4),* doi: 10.1088/1741-2560/9/4/045001.

[3] Halder, S., Käthner, I., & Kübler, A. (2016). Training leads to increased auditory brain–computer interface performance of end-users with motor impairments. *Clinical Neurophysiology, 127,* 1288-1296.

[4] Herweg, A., Gutzeit, J., Kleih, S., & Kübler, A. (2016). Wheelchair control by elderly participants in a virtual environment with a brain-computer interface (BCI) and tactile stimulation. *Biological Psychology, 121,* 117-124.

[5] Hill, N.J., Ricci, E., Haider, S., McCane, L.M., Heckman, S., …, & Vaughan, T.M. (2014). A Practical, Intuitive Brain-Computer Interface for Communicating "Yes" or "No" by Listening. *Journal of Neural Engineering, 11,* doi: 10.1088/1741-2560/11/3/035003.

[6] Chavarriaga, R., Fried-Oken, M., Kleih, S., Lotte, F., & Scherer, R. (2017). Heading for new shores! Overcoming pitfalls in BCI design. *Brain-Computer Interfaces, 4,* doi: 10.1080/2326263X.2016. 1263916.

[7] Kübler, A., Holz, E.M., Riccio, A., Zickler, C., Kaufmann, T., …, & Mattia, D. (2014). The user-centered design as novel perspective for evaluating the usability of BCI-controlled applications. *PLoS One, 9(12),* doi: 10.1371/journal.pone.0112392.

[8] Lotte, F., Larrue, F., & Mühl, C. (2013). Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design. *Frontiers in Human Neuroscience, 7,* doi: 10.3389/fnhum.2013.00568.

[9] Marshall, D., Coyle, D., Wilson, S., & Callaghan, M. (2013). Games, gameplay, and BCI: the state of the art. IEEE Transactions on *Computational Intelligence and AI in Games, 5(2),* 82-99.

[10] World Medical Association. (2013). World Medical Association Declaration of Helsinki ethical principles for medical research involving human subjects. *JAMA: Journal of the American Medical Association, 310(20),* 2191-2194.

[11] Sharbrough, F., Chatrian, G. E., Lesser, R.P., Lüders, H., Nuwer, M., & Picton, T.W. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology, 8,* 200-202.

[12] Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., & Wolpaw, J.R. (2004). BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering, 51,* 1034-1043.

[13] Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134(1),* 9-21.

[14] Kothe, C.A.E. & Jung, T.P. (2016). *U.S. Patent Application No. 14/895,440.* https://patentimages.storage.googleapis.com/c3/6b/dc/7dadfae33c0062/US20160113587A1.pdf

[15] Halder, S., Rea, M., Andreoni, R., Nijboer, F., Hammer, E.M., ..., & Kübler, A. (2010). An auditory oddball brain–computer interface for binary choices. *Clinical Neurophysiology, 121(4),* 516-523.

[16] Kübler, A., Neumann, N., Kaiser, J., Kotchoubey, B., Hinterberger, T., & Birbaumer, N.P. (2001b). Brain-computer communication: self-regulation of slow cortical potentials for verbal communication. *Archives of Physical Medicine and Rehabilitation, 82(11),* 1533-1539.

[17] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., & Vaughan, T.M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology, 113,* 767-791.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-47

FLIP-THAT-BUCKET
A FUN EEG-BCI GAME ON GOOEY MOVEMENT INTENTIONS

C.S. Verbaarschot[1], A.B.W. Gerrits[1], W.F.G. Haselager[1], J.D.R. Farquhar[1]

[1] Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

E-mail: c.verbaarschot@donders.ru.nl

ABSTRACT: "Flip-that-Bucket" is an open source, portable and enjoyable BCI game suitable for investigating or demonstrating movement intentions using scientific experiments or educational demonstrations in noisy environments. In the game, a sneaky virtual robot aims to predict a player's intentions to act based on their action history, muscle activity, or brain activity. The game can be used to assess (1) the accuracy of brain-based movement predictions, (2) the timing of these predictions relative to movement onset, (3) the potential benefit of brain-based over behavior-based predictions, and (4) the correlation between certain brain signals (e.g. the readiness potential and event-related desynchronization between 8-30Hz over the pre- and primary motor areas) and the experience of an intention to move. Answering each of these questions may greatly benefit future applications in prosthetics and motor rehabilitation. Flip-that-Bucket is made as an extension to the open-source buffer_bci toolbox, encouraging further development. Here, we demonstrate the idea, its practical implementation, the 'fun factor' and a first analysis of experimental results.

INTRODUCTION

When we move, imagine movement or observe movement, specific parts of our brain activate: the premotor, supplementary and primary motor cortices (termed 'motor cortex' in the remainder of this paper). When we perform self-paced voluntary movements, we typically see a readiness potential (RP) and event-related desynchronization (ERD) at 8-30Hz across the motor cortex [3, 5, 6, 8, 11]. Interestingly, these brain signals are visible on average around 1.5s before a person reports a conscious feeling of wanting to move. This suggests that the brain starts preparing a movement before a person reports a conscious intention to move [7]. Other studies show that using a real-time probing method, awareness of an intention to move can be reported up to 2s prior to movement onset [12]. How exactly a conscious intention to move relates to these neural signals remains unclear.

In this paper, we present "Flip-that-Bucket": a BCI game that serves as a tool to investigate or demonstrate the relation between the neural preparation and reported awareness of movement intentions (see Figure 1). In contrast to previous research [1, 2, 3, 4, 5, 6, 7, 8], the game provides an engaging real-time set-up to measure spontaneous self-paced right hand movements in an intuitive way. In the game, players try to beat a virtual robot opponent in a slime-bucket challenge. Across multiple rounds, both the participant and robot try to gather as much green slime as possible to throw it over their opponent's head. Both the player and robot have

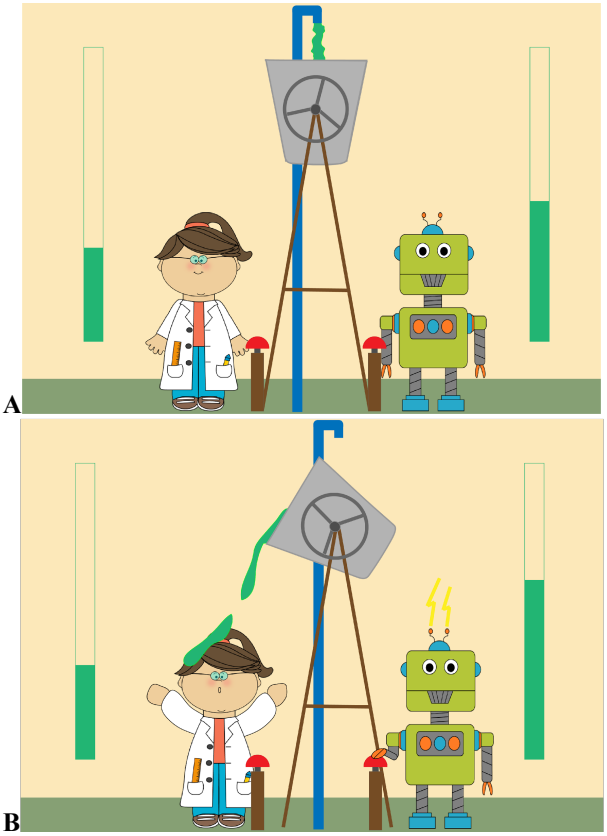


Figure 1. Screenshots of Flip-that-Bucket. The game consists of several rounds during which both the player (the scientist) and the robot can gather as much gooey green slime as possible. At the start of each round, an empty bucket is displayed for 2s. (A) This bucket fills with green slime over time, although the exact content of the bucket remains hidden. (B) The player and robot can flip over the bucket and empty its contents onto their opponent any time they want. Whoever flips the bucket first, will get points for the amount of slime they threw over their opponent. Whoever has been slimed the least wins the game.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-47

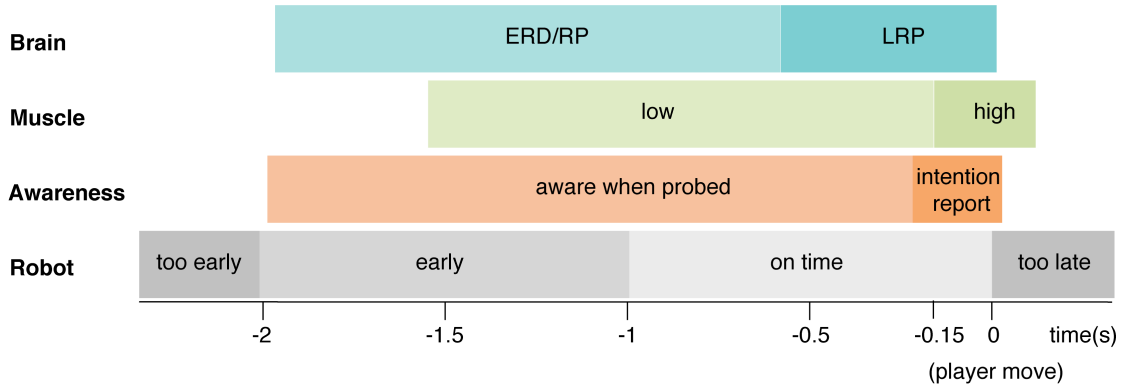


Figure 2. Schematic visualization of the possible moments in time at which a robot could predict a player's intention to move (movement onset is at time 0s). We distinguish four possible categories of robot predictions. (1) Too early: no visible ERD/RP, no visible muscle activity, only negative intention reports. (2) Early: no or weak ERD/RP, no visible muscle activity, more negative/unknown intention reports than positive ones. (3) On time: medium to strong ERD/RP (possibly including an LRP), low to high muscle activity, more positive intention reports than negative/unknown ones. (4) Too late: strong (lateralized) ERD/RP, high muscle activity, no intention reports collected.

access to the same bucket, which can flip over only once each round to spill out its gooey content. The robot is sneaky: he tries to predict a player's intention to move using their action history, muscle activity, or brain activity. As soon as the robot detects a player's intention to move, he will flip over the bucket of slime. Because the robot uses different types of predictions (based on action history, muscle or brain activity), the game serves as a thought-provoking means to explain to a general public how movement intent travels from the brain to the muscles in order to perform a voluntary movement.

In a first experiment, we use robot predictions based on action history. We collect EEG data to assess whether there is a strong correlation between the RP and/or ERD and the awareness of an intention. To do so, participants are asked to report whether they experienced an intention to move at the moment when a prediction is made (i.e. when the robot moves). This ante-hoc probing strategy measures awareness of movement intent prior to movement performance, which avoids the potential confound of movement execution on post-hoc awareness reports [4]. We expect to see a clear RP and ERD prior to a movement intent prediction that is reported as "intended" and no (or a weak) RP and ERD prior to a movement intent prediction that was reported as "unintended" (see Figure 2).

Flip-that-Bucket is implemented as an extension to an existing open-source BCI development toolbox called "buffer_bci",[1] encouraging further development of our project ideas by both academics and the general developer community. Anyone with access to an EEG system can try the game. It can be used at home, during public events or in scientific experiments.

MATERIALS AND METHODS

Participants: 41 healthy volunteers were tested at the InScience festival in Nijmegen, the Netherlands.[2] The experiment was conducted in accordance with the ethical standards provided by the 1964 Declaration of Helsinki. The study protocol was approved by the local Ethics Committee of Faculty of Social Sciences of the Radboud University Nijmegen. All participants gave their written informed consent. Ten participants were excluded from analysis because they did not follow instructions correctly or would not stop talking or moving during the experiment.

Task: Participants play 4 blocks of Flip-that-Bucket[3] (see Figure 1): a practice block of 3 trials, a training block of 60 trials, a hidden validation block of 15 trials and a test block of 60 trials. This experimental structure is used to accommodate future experimentation including brain-based robot predictions. In between the training and validation block is a self-paced break and a short questionnaire. Participants are informed that we are testing a new BCI that attempts to predict the moment at which they intend to move and asked to report whether or not they wanted to move when a prediction (i.e. robot move) was made. During the training block, predictions of movement intent are made based on the participant's action history. Based on [12], we expected awareness of intending no earlier than 2s prior to action onset. Each trial, a minimum cost function takes the current mean and standard deviation of the player's action times (relative to trial start) and calculates a distribution of possible robot action times such that: 1/5 of robot acts are performed earlier than 2s prior to the average scientist move, 3/5 of robot acts between 2s and 0s prior to the average scientist move, and 1/5 of robot acts are performed after the scientist moves (i.e. the robot loses). The robot action time for a

[1] www.github.com/jadref/buffer_bci

[2] www.insciencefestival.nl

[3] Flip-that-Bucket can be found at: www.github.com/jadref/buffer_bci/tree/master/matlab/movementBCIgame

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-47

given trial was drawn randomly from this distribution. After the training block, a combined features classifier (incorporating both RP and 8-30Hz ERD features of pre-move and non-move data) was trained on the collected labeled EEG data and used to provide brain-based predictions during the test block. Unfortunately, due to technical errors (i.e. accidentally including post-move data in the training set and switching class labels during real-time prediction) the brain-based predictions were not executed properly and the predictions made in the second block were roughly random.
At the end of the train and test block, participants fill in a short questionnaire asking them (1) what they thought about the game on a scale of 1 (boring) to 5 (fun), (2) whether they felt free to do what they want (Yes/No), (3) how difficult it was to win on a scale of 1 (easy) to 5 (difficult), and (4) how accurate the robot was in predicting their actions on a scale of 1 (inaccurate) to 5 (accurate). At the end of the test block, they were asked how good they thought the robot predictions were in the second block compared to the first on a scale of 1 (worse) to 5 (better). For additional motivation, a high-score list across all players and robots is maintained. In total, the experiment took 24 minutes (excluding cap fitting).

Data acquisition: The experiment was run in Matlab.[4] Instructions and visual stimuli were displayed using a 17 inch TFT screen with a resolution of 800 by 600 pixels and a refresh rate of 60Hz that was placed roughly at 70cm directly in front of the participant. To flip the bucket, participants press SPACE with their right hand on a regular keyboard. EEG data was collected using the TMSi Porti system,[5] with water-based electrodes sampled at 512 Hz placed at Fp1, Fp2 F3, Fz, F4, C4, Cz, C4, P3, Pz, P4, POz, TP9 and TP10 (according to the International 10/20 system). In addition, muscle activity was recorded using two EEG electrodes in a bipolar pair on the wrist and the center of the right forearm (flexor pollicis longus).

Analysis: All analyses were performed on the 31 participants who followed instructions correctly. To assess the correlation between the RP and/or ERD and a reported intention to move, epochs of -15 to 15s around a player and robot act are analyzed. Epochs around a robot act are further subdivided in epochs in which the robot acted when the participant did or did not want to move. To ensure a decent baseline period, only epochs in which the player or robot acted slower than 4s after trial start are kept for analysis. Linear trends are removed from the data. Subsequently, the data is re-referenced by subtracting the average signal from all outer channels (Fp1, Fp2, F3, P3, POz, F4, P4, TP9, TP10) from each individual channel. This was done to subtract as much noise as possible without subtracting the signals of interest. Since most recorded channels cover the motor cortex, a full common average reference would subtract much of the signal of interest along with the noise (leading to a decrease in RP amplitude). Eye-artifacts are removed using linear de-correlation of channels Fp1 and Fp2 with respect to the other EEG channels. Only the central channels (F3, C3, P3, Pz, Fz, F4, Cz, C4 and P4) are kept for further analysis. Channels that differ more than 2 times the standard deviation in power from the median are removed. If necessary, spherical spline interpolation is used to reconstruct missing central channels. The data is band-pass filtered between 0.2 and 35Hz. Bad trials that differ more than two times the standard deviation in power from the median are removed. For the ERD, frequencies of interest are defined from 4 to 30Hz using 2Hz bins. A flexible Hanning window is used such that it includes at least 7 cycles of each frequency of interest. The baseline activity is defined per electrode, frequency and trial as the median power between 3.5 and 2.5s prior to action. A relative baseline (where a value of 1 means no signal change compared to baseline) is subtracted from the data. The ERD is calculated per participant by taking the median power across trials for each electrode, frequency and trial.

Offline classification: For each participant, a linear classifier is trained using 10-fold cross-validation to distinguish baseline data from pre-movement data. Baseline data is extracted from the last 500ms of the baseline period. Pre-movement data is extracted from the last 500ms prior to a player move. All training data is extracted from the training block. The data is pre-processed using the same steps as described in the *Analysis* section. The RP features consist of 257 time points for each channel and epoch. The ERD features consist of the average power of each of the 14 frequencies (4,6,…,12,14Hz) for each channel and epoch. The classifier is trained on both RP and ERD features, giving a total of 2439 features per epoch. Data from the validation block is used to set an optimal threshold for classification. This threshold is set such that the number of *on time* (between 1 and 0s prior to a player move) predictions is maximized whereas the number of *too early* (more than 2s prior to a player move) predictions is minimized (see Figure 2). The accuracy of the classifier is assessed using data from the

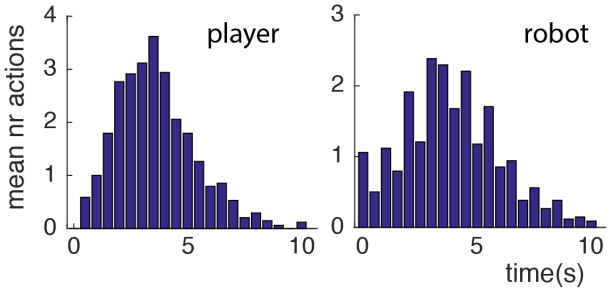


Figure 3. Distribution of player and robot moves relative to trial start (0s) across all participants during the training block.

[4] www.mathworks.com
[5] www.tmsi.com/products/porti/

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-47

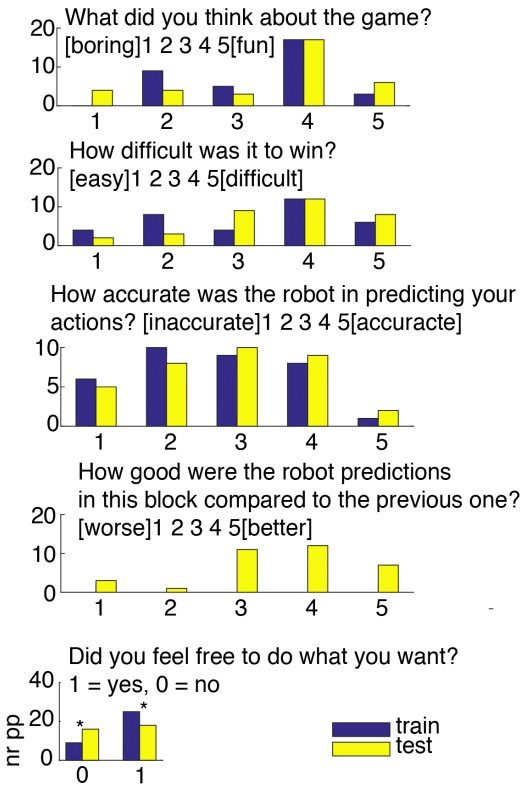


Figure 4. Questionnaire results. Participants felt significantly less free to do what they wanted during the test compared to the train block. In question 4, "this block" refers to the test block and "the previous one" refers to the training block.

test block. The trained classifier is applied to non-overlapping 500ms epochs starting from the start of a trial until a player or robot move. For each trial that includes an intention to move (the player moved or reported an intention to move at the time of a robot move), the first classifier prediction that exceeds the optimal threshold for motor intention detection is selected. The timing of each first motor intention prediction is determined relative to the corresponding robot or player move and categorized as *too early*, *early*, *on time* or *too late* (see Figure 2).

RESULTS

An average of 69 (min:42, max:104) player moves, 20 (min:5, max:54) robot moves with player intent and 33 (min:9, max:67) robot moves without player intent are collected across the experiment. An overview of player and robot move times during training is provided in Figure 3.

The majority of participants reported the game as fun and slightly difficult to win (see Figure 4). Although opinions on the accuracy of the robot predictions varied greatly, participants judged the robot predictions of the second experimental block as more accurate than the first (even though these predictions were roughly random). Furthermore, a within-subject t-test on the questionnaire data of the first en second experimental blocks revealed that participants felt significantly ($p < .05$) less free to act during the second block of the experiment compared to the first. No significant differences were found between the number of robot actions that happened at a time when the player did or did not intend to move during the first and second experimental blocks.

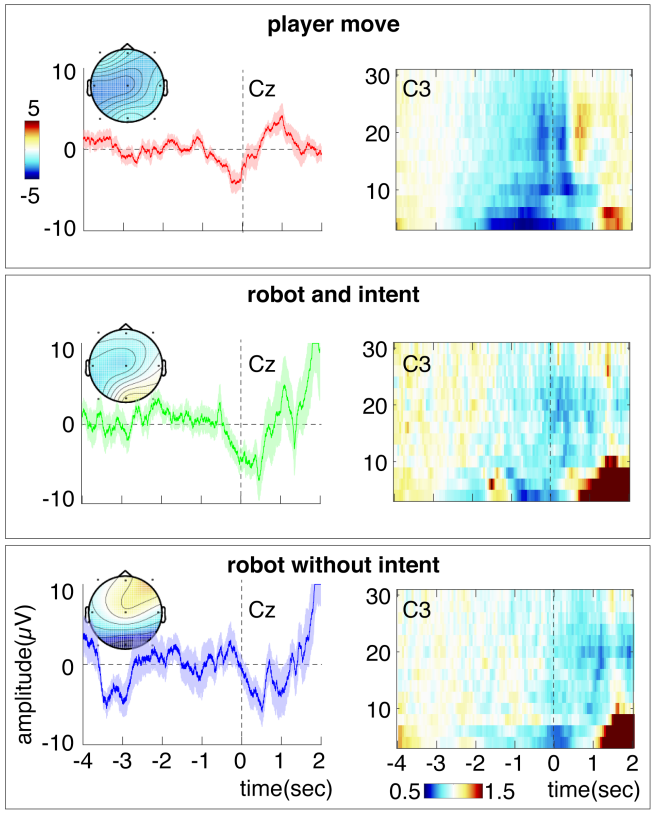


Figure 5. Grand average EEG data of Flip-that-Bucket. Left column: grand average ERPs including the standard error, and topoplots across the last 500ms prior to a player/robot move until action onset (0s). Right column: grand average spectrogram across 4 to 30Hz before player/robot action onset (at 0s). Baseline activity between 3.5 and 2.5s prior to action onset is subtracted from the ERD data. Although a clear RP and ERD is visible prior to a player move (top row), no RP and or ERD is visible prior to a robot act that happened when the player did not intend to move (bottom row). A weak RP and ERD is visible when the player did intend to move when the robot acted (middle row).

A clear RP and 8-30Hz ERD are visible in the grand average prior to a player move, starting 1 to 2s prior to movement onset respectively (see Figure 5). A weak RP and ERD is visible prior to a robot act that happened at a time when the player intended to move (i.e. a correct prediction), and no RP or ERD is visible prior to a robot act that happened when the player did not intend to move (i.e. an incorrect prediction). After a robot acts, a big positive response is visible in the recorded EEG. This response is due to the additional button press that is required to report whether the robot moved at a time when the player wanted to move. Any differences

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-47

between the post robot move responses may be due to the presence or absence of an intention to move, leading to an superimposed error potential in one case rather than the other. Moreover, a player may show different levels of surprise or frustration after the robot moved, depending on the presence or absence of an intention to move.

Classifiers are trained on an average of 48 (min:44, max:50) pre-move and 47 (min:42, max:50) baseline epochs from the training block. A mean classifier performance of 74% across all participants is found on the training data. Although these results seem promising, the performance of the classifier on the sequential test data are rather poor. On average a mere 12% of all motor intentions is detected *on time* (within 1s prior to a player or robot move). The majority of motor intentions is detected *too early* (24%) or *too late* (54%).

PP	Perf.	Too early	Early	On time	Too late
1	76	7	0	0	93
2	92	0	8	13	79
3	64	6	8	8	78
4	90	17	0	12	71
5	78	15	6	17	62
6	90	18	0	20	63
7	79	45	15	10	30
8	61	64	31	6	0
9	72	21	38	29	13
10	60	41	14	5	41
11	69	0	0	0	100
12	83	56	12	9	23
13	88	4	0	2	93
14	66	29	3	12	56
15	60	28	26	16	30
16	74	70	17	4	9
17	75	85	5	5	5
18	61	21	3	18	59
19	65	2	2	2	93
20	84	0	0	0	100
21	76	21	50	25	4
22	59	35	13	15	37
23	98	19	2	7	71
24	71	17	22	26	35
25	68	0	0	0	100
26	80	24	10	19	48
27	68	6	2	10	82
28	60	32	6	3	58
29	82	17	9	20	54
30	70	23	12	37	28
31	65	35	7	11	48
mean	**74**	**24**	**10**	**12**	**54**

Table 1. Offline classification results. For each participant the cross-validated test performance of the classifier on the data of the training block is provided. Furthermore, for each trial in the test block that includes a motor intention, the first classifier prediction that exceeds the chosen threshold for motor intent is determined. These predictions are categorized as too early, early, on time or too late relative to the robot or player move. The percentage of predictions within each category compared to the total number of motor intentions present is provided in the remaining columns.

DISCUSSION

Moving your body at will seems trivially easy. You probably do it all the time. You can knock on a table, stomp with your feet and flap your arms whenever you want to. But what mechanism enables you to initiate these movements? Moreover, how does this mechanism relate to your conscious experience of wanting to move? Flip-that-Bucket is a fun and thought-provoking EEG-BCI game on movement intentions. In the game, a virtual robot tries to predict a player's intention to move using the player's history of action times, the onset of muscle activity or the neural preparation for movement. Flip-that-Bucket can serve as a means to educate a general public about the neural mechanisms that underlie our ability to perform voluntary movements: the connection between a participant's brain signals and their voluntary movements are made explicit to them by means of the robot opponent.

Here, we demonstrate that predictions based on action history are effective to create a competitive game. Furthermore, we demonstrate that both the RP and 8-30Hz ERD are clearly visible across the motor cortex prior to a player move, whereas it is not visible prior to an incorrect robot prediction (i.e. the robot acted when the participant did not experience an intention to move). In case of a correct robot prediction (i.e. the robot acts when the participant experienced an intention to move), a weak RP and 8-30Hz ERD are visible across the motor cortex. Along with previous research [1, 2, 9, 10], these results suggest that brain-based predictions of movement intent may be reasonably successful. If this is the case, brain-based predictions may be used to assess the relation between the RP and ERD in real-time in further experiments. It would be interesting to see whether brain-based predictions are more accurate compared to behavior-based predictions (i.e. the robot predictions are more often reported to be correct when using brain-based rather than behavior-based data).

Flip-that-Bucket can also facilitate scientific investigations on the neural preparation for a voluntary movement. Although some studies aimed to predict movement onset in real-time prior to movement performance based on the RP and 8-30Hz ERD across the motor cortex, e.g. [1, 2, 9, 10], the accuracy and timing of single-trial predictions remains difficult to assess since often only averages across participants are reported. Flip-that-Bucket can be used to assess the accuracy and timing of brain-based predictions on single-trial continuous data. A correct prediction of movement intent would happen at a time when a participant reports that they want to move, their muscles are active in preparation of the upcoming movement or when it coincides with movement performance. Furthermore, predictions that happen very early (more than 2s) prior to movement onset would be considered incorrect (false positive), whereas those happening shortly (within 1s) prior to movement onset would be correct (true positive). A first offline analysis of these prediction results suggests that our combined features

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-47

classifier would perform rather poorly, detecting only about 12% of all motor intentions. The current discrepancy between the performance on the training and test data, may be due to a potential expectation effect that builds up during a trial: the longer a trial develops, the more likely a player or the robot is to act. This expectation effect may induce additional brain responses that are maximally different between baseline and pre-move training data, but less so between subsequent epochs of test data. Based on previous research [1, 2, 9, 10], we expect that better results should be possible. Possible improvements could be made by (1) including more instances of "non-move" data during classifier training, (2) implementing a more sophisticated method to set an optimal threshold for detecting a motor intention or (3) using different brain signals that are indicative of movement preparation.

Assessing the accuracy and timing of real-time predictions of movement intent on continuous data may greatly benefit future applications in prosthetics and motor rehabilitation. Delays in activating an assistive device could be minimized by detecting movement intent early on, which potentially increases the therapeutic benefit by minimizing the time between motor planning in the cortex and the execution of that plan with the assistive device.

Flip-that-Bucket is made as an extension to the existing open-source buffer_bci development toolbox, encouraging further development of the game by both academics and the general developer community. This extended toolbox is available to anyone with access to an EEG system. It can be used at home, during public events or in scientific experiments.

REFERENCES

[1] Bai, O., Rathi, V., Lin, P., Huang, D., Battapady, H., Fei, D. Y., ... & Hallett, M. (2011). Prediction of human voluntary movement before it occurs. *Clinical Neurophysiology*, *122*(2), 364-372.

[2] Blankertz, B., Dornhege, G., Lemm, S., Krauledat, M., Curio, G., & Müller, K. R. (2006). The Berlin brain-computer interface: Machine learning based detection of user specific brain states. *J. UCS*, *12*(6), 581-607.

[3] Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron*, *69*(3), 548-562.

[4] Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature neuroscience*, *5*(4), 382.

[5] Kornhuber, H. H., & Deecke, L. (2016). Brain potential changes in voluntary and passive movements in humans: readiness potential and reafferent potentials. *Pflügers Archiv-European Journal of Physiology*, *468*(7), 1115-1124.

[6] Lew, E., Chavarriaga, R., Silvoni, S., & Millán, J. D. R. (2012). Detection of self-paced reaching movement intention from EEG signals. *Frontiers in neuroengineering*, *5*, 13.

[7] Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act. *Brain*, *106*(3), 623-642.

[8] Pfurtscheller, G., & Aranibar, A. (1979). Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalography and clinical neurophysiology*, *46*(2), 138-146.

[9] Schneider, L., Houdayer, E., Bai, O., & Hallett, M. (2013). What we think before a voluntary movement. *Journal of cognitive neuroscience*, *25*(6), 822-829.

[10] Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., ... & Haynes, J. D. (2016). The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences*, *113*(4), 1080-1085.

[11] Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential?. *Clinical neurophysiology*, *117*(11), 2341-2356.

[12] Verbaarschot, C., Haselager, P., & Farquhar, J. (2016). Detecting traces of consciousness in the process of intending to act. *Experimental brain research*, *234*(7), 1945-1956.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-48

WHY BCIS WORK POORLY WITH THE PATIENTS WHO NEED THEM THE MOST?

P. Séguin[1,2,3], E. Maby[1,2], J. Mattout[1,2]

[1]Lyon Neuroscience Research Center, CRNL; INSERM, U1028; CNRS, UMR5292;
Brain Dynamics and Cognition Team, Lyon, F-69000, France
[2]University Lyon 1, Lyon, F-69000, France
[3]Department of Physical Medicine and Rehabilitation, Faculty of Medicine, University Jean Monnet of Saint Etienne, F-42023 Saint-Etienne, France

E-mail: jeremie.mattout@inserm.fr

ABSTRACT: A major objective of Brain-Computer interfaces (BCI) is to restore communication and control in patients with severe motor impairments, like people with Locked-in syndrome. These patients are left only with limited eye and eyelid movements. However, they do not benefit from efficient BCI solutions, yet. Different signals can be used as commands for non-invasive BCI: mu and beta rhythm desynchronization, evoked potentials and slow cortical potentials. Whatever the signal, clinical studies show a dramatic loss of performance in severely impaired patients compared to healthy subjects. Interestingly, the control principle is always the same, namely the replacement of an impossible (overt) movement by a (covert) attentional command. Drawing from the premotor theory of attention, from neuroimaging findings about the functional anatomy of spatial attention, from clinical observations and from recent computational accounts of attention for both action and perception, we explore the hypothesis that these patients undergo negative plasticity that extends their impairment from overt to covert attentional processes.

INTRODUCTION

Since its early days, brain computer interfaces (BCI) aim at improving autonomy for people with paralysis. However and despite numerous efforts and improvements, BCI are still not used in clinical routine [1], [2]. In this article, after a brief overview of non-invasive BCI for communication with patients with severe motor impairment, we show how empirical and theoretical results in cognitive neuroscience could help explain the failure of technological transfer that our field is facing.

THE TARGETED CLINICAL POPULATION

The « classical » locked-in syndrome (LIS) is caused by a lesion of the brainstem, most of the time due to a stroke [3], [4]. The patient is totally paralyzed. Only the vertical eye movements and blinks are spared and allow him to communicate. This condition can also be encountered in other pathologies, like in amyotrophic lateral sclerosis (ALS), a neurodegenerative disease of the motor neurons, at the late stage. Different forms of this disease exist, some affect more the lower motor neurons (progressive muscular atrophy), cortico-spinal motor neurons (primary lateral sclerosis), brainstem motor neurons (bulbar ALS) or cortico-frontobulbar motor neurons (pseudobulbar palsy) [5]. In all these forms, oculomotor muscles are usually preserved, except in the extremely late stages.

EEG-BASED BCI

To control non-invasive BCI, users have to learn unusual means of interactions. As we will highlight in this article, the signals used as commands for BCI all rely on some form of (covert) selective attentional process.

Synchronous BCI:

P300: The most exploited neurophysiological marker in this context is arguably the P300, a positive wave observed in response to external stimuli (about 300ms later), if rare and relevant to the user, surrounded by more frequent but irrelevant stimulations (distractors). If the "P300-speller" mainly uses the visual modality, the P300 wave can also be observed after auditory or tactile stimuli. For selection, the patient has to actively attend the expected target stimulus.

Steady-state visual evoked potential (SSVEP): SSVEP are characterized by a rhythmic neural responses to a flickering stimuli at the frequency of the flicker and its harmonics [6]. BCI protocols use several targets with different flickering frequencies. The amplitude of the SSVEP is increased when the subject pay attention to the target, which allows to detect which stimulus the subject is attending.

Asynchronous BCI:

Motor imagery: The imagination of a movement produces a brain activity globally similar to that observed during the actual production of that same movement [7]. A desynchronization (i.e. a drop in EEG signal strength) in the mu and beta frequency bands occurs during the preparation and during the execution or the imagination of a movement. It is measured next to the primary motor cortex, mainly on the contralateral side of the movement. I If the number of commands remains limited, many publications have shown, in healthy volunteers, the

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-48

ability to control a cursor on a screen. Studies in patients are fewer but the feasibility has been demonstrated [8]. However, these interfaces require a long training, offer limited performance and above all a large number of healthy volunteers are unable to produce the desired signals, to the point of "illiteracy in BCI" for nearly 30% of users [9].

Slow cortical potentials: These potentials result in a slow drift of the signal that reflects a change in cortical excitability. A depolarization of a large neuron assembly corresponds to greater excitability, whereas a hyperpolarization corresponds to a greater inhibition. These signals were among the first used in patients. Several healthy subjects and patients managed to communicate using a BCI exploiting these potentials, but at the cost of several months of training and with a writing speed of the order of one character per minute, well below the means for alternative communication when these are accessible [10], [11].

BCI PERFORMANCE IN PATIENTS WITH SEVERE MOTOR IMPAIRMENT

The communication with LIS patients through non-invasive BCI is often put forward by the press when the publication of encouraging but mitigated results is released. Truth is that BCI has been used routinely by patients only in very rare cases [12]. Researchers and clinicians point to the lack of usability of current systems that are still difficult to use on a daily basis without the help of experts [13], [14].

However, looking more closely at the results of the few studies already conducted in patients, and despite their heterogeneous conclusions, another observation is necessary. These interfaces work well, on average, in severely disabled patients who, due to the residual muscle activity they still enjoy, could use another type of interface (lighter and more efficient like a gaze tracking system). Conversely, in patients who most need these BCI because they have more or almost no means of communication, i.e. with abnormal oculomotor movements or in a complete LIS (CLIS, no residual movements, including oculomotor ones), possibilities of restoring communication remain controversial.

Most studies tend to show that brain computer interfaces work poorly with CLIS patients [2], [15]–[17]. Even in LIS, some patients do not achieve a sufficient accuracy. For example a recent study revealed that amongst 37 persons at a late stage of ALS, 9 (24%) could not control a P300-BCI [12].

Two studies published in 2017 claimed that a communication was restored with CLIS patients [18], [19], but some of the methodological aspects of these studies remain unclear, and their results are debated in the BCI community[20].

A longitudinal study, where a patient could learn to use a BCI at the LIS stage and then maintain some control when becoming a CLIS, was more convincing. This study included three subjects with ALS at the LIS state, and one became a CLIS over the 27 months follow up [21].

This patient, a 37-year-old woman who had ALS for 6 years, was asked to control a binary SSVEP-BCI by either focusing on one LED, or ignoring it. From month three onwards, the electro-oculogram could not show any difference between the attended and ignored conditions (confirming the CLIS). However, she performed BCI control with 79% accuracy online, with accuracies above the confidence limits in 18 out of the 27 months and in 27 out of 40 sessions.

This case of communication in CLIS remains an exception, and the reliability of communication happened to be fluctuant.

Some researchers raised the question of a possible "extinction of goal directed thinking" that would accompany the occurrence of a total handicap [16]. This concept was refined later by the same team in terms of “ideomotor silence”, which would lead to a loss of voluntary responses and operant learning in long-term paralysis of human patients [23], [24].

WORKING HYPOTHESIS

Severe motor impairment come with and/or yield cognitive deficits such that handling a BCI is made difficult if not impossible. In particular, we postulate that severe motor impairments lead to altered spatial and/or temporal selective attention abilities.

SUPPORTING EVIDENCE

Persons with severe motor disabilities often present with cognitive impairments

The neurological examination of these patients is limited by the lack of communication. Some adaptations of the usual technics were necessary to overcome this impairment [25] and to have a better overview of their neurological status.

ALS

Sensory functions: In ALS patients, the oculomotor system was classically considered as spared [26], [27], but the development of life support systems, exceeding the natural course of the disease, led to observation that oculomotor impairments can be present at the late stage [28]. These impairments can be both peripheral (nuclear) and supranuclear. It is noticeable that the oculomotor system is crucial to maintain interaction with the world. An earlier onset of oculomotor troubles in the ALS has a pejorative effect: it is correlated with a higher probability to progress from the locked-in state to CLIS [29], [30]. ALS subjects present other sensory abnormalities [31], [32].

Neuropsychological assessment: 30% of ALS patients (without fulfilling the criteria for dementia) show deficits in fluency, language, social cognition, executive functions and verbal memory. Amongst all the function tested, only visuoperceptive functions were

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-48

preserved [33].

Electrophysiological assessment:
A longitudinal observation in electrocorticography was conducted in a patient with ALS, around the time when he lost all muscle control [22]. Whereas the P300 wave was still observable at the time of this dramatic evolution; three months later, it was no longer detectable.
A recent MEG study showed that the late stage of ALS is associated with higher connectivity in all frequency bands, more scale-free and disassortative brain networks [34]. The auditory event-related potentials are different than in control subjects, namely the location and amplitude of the late positivity, the amplitude of the early negativity (N200), the latency of the late negativity [35] and the mismatch negativity (related to automatic change detection) [36].

LIS

Sensory functions: neuro-ophtalmological evaluations suggest visual impairment in all locked-in syndrome subjects [37]. This is a multifactorial impairment including binocular diplopia or oscillopsia, refractive errors, dry eye syndrome, keratitis or visual field defect.
Tactile perception is usually spared in LIS patients, as the typical cases are due to bilateral ventral pontine lesions, which usually preserve the posterior part of the brainstem that mediates sensory afferences [3]. However, studies assessing the ability of the patients to discriminate the precise location of a stimulus are missing.

Neuropsychological assessment: First studies reported intact cognitive abilities in LIS with classical neuropsychological assessments [38]–[41]. More recent studies [42],[43] revealed disparity between patients. Study [42] revealed a decreased short-term memory in 4 out of 10 persons with LIS, and an alteration of executive functions in 2 out of 10. The authors assumed that these impairments could be related to additional cortical or thalamic structural brain lesions [42]Another study highlighted impairments in auditory recognition (associative level), oral comprehension of complex sentences, delayed visuospatial memory, mental calculation and problem solving, compared to matched healthy subjects [43].
Studies testing specifically embodied cognition hypothesis found some abnormalities. LIS subjects have severe troubles to mentally manipulate hand images (a task that imply a mental manipulation of the subject's own hand), whereas they are still able to mentally manipulate images [44]. Contrary to healthy subjects, having the presentation of the hand in the correct physiological side (e.g. left hand on the left) doesn't help them to solve the task. This failure of action simulation was interpreted as a defect of embodied cognition. This points to aspects that are usually not tested in these patients, namely modifications of their peripersonal space.

Neurophysiological and anatomical data obtained in LIS reveal some abnormalities that are not yet well explained. Cortical neuronal synchronization mechanisms in the resting state condition are altered [45]. Lugo et al, 2016, report an absence of auditory P300 in four LIS subjects amongst seven tested in a passive condition where subjects were asked to listen only. In an active condition, where the subjects were asked to count the deviants, there were still two LIS subjects out of seven who did not present a P300 [46].
Finally, an anatomical MRI study showed a selective cortical volume loss in patients, in regions that could be linked to an alteration of the mirror neuron system [47], further supporting the idea that embodied simulation process is altered in these patients.

Chronic complete SCI
There are contradictory results between behavioral and electrophysiological assessments in SCI subjects.

Behavioral assessment: Studies report that patients experienced movement duration and sensation of effort as normal subjects do [48].

Electrophysiological assessment: There are abnormal event-related potentials [49] and patterns of cortical activation during motor imagery tasks [50]. In 2013, Lazzaro et al showed some modifications of the auditory P200 latency (earlier than control) and a diminished bilateral posterior P300 amplitude. This was correlated behaviorally with increased false-positive errors and greater variability of response time in the SCI group [51]. These observations were interpreted as a defect of inhibitory functions in both early perceptual encoding processes and later executive functioning that engages contextual and memory-updating operations.

Impaired motor execution alters motor planning and action selection

Studies with induced limb immobilization on healthy subjects [52] show that limb nonuse may affect both motor execution and motor imagery performances. These changes happen within days and even hours of immobilization.

Impaired motor preparation leads to impaired (covert) attention

Rizzolatti et al. postulated in 1987 that spatial attention is the consequence of activation of the premotor system. They proposed this framework based on the observation of the oculomotor system, especially covert attention. Covert attention is an attentional effort, without any concomitant movement (except involuntary micro-saccades). They assumed a strict link between the covert orienting of attention and the programming of explicit ocular movements. The fact that oculomotor mechanisms and visuospatial attention share similar cortical networks was confirmed by other studies [53]–[55].
Interestingly, motor preparation is so much related to covert attention, that when eyes cannot move to a

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-48

location, also the attention cannot move [56], [57].
Furthermore, in healthy subjects, when movements are preceded by an invalid motor cue (Posner like paradigm in the motor domain), an increase in reaction times is observed [58], suggesting that motor preparation is a top-down attentional process and that this biasing is mediated at low level in the motor hierarchy.
Despite criticisms of some aspects of the premotor theory of attention, Smith and Schenk proposed in 2012 that a limited version of this theory is still valuable, namely the tight link between exogenous attention and motor preparation [59]. This could be sufficient to impact the P300 marker, which is simultaneously associated with attentional, exogenous and endogenous processes and action selection processes [60], [61].
This is confirmed by some BCI studies addressing the problem of covert attention. Be it in healthy subjects [62] or patients with oculomotor impairments [63], it was observed that BCI performance are much higher or involve much less mental workload when the gaze can be directed toward the target.
To overcome these visual impairments, some auditory BCI were proposed. Surprisingly, in the few clinical studies with auditory BCI, including ours, most of the patients continue to show poor performance compared to healthy subjects [64]. This multi-modal impairment could be explained by the fact that attention would be mediated by a supra-modal network. For instance, orienting attention toward a tactile target triggers an automatic displacement of spatial attention in the visual modality [65]–[67]. Furthermore, a subpopulation of neurons in the frontal eye field (FEF, orienting the gaze) is directly activated by auditory cues. As the FEF is involved in the planning of both covert and overt attentional orienting, an impairment could prevent accurate orientation of attention in the auditory domain when the oculomotor modality is impaired.

Connectivity changes induced by focal lesions in motor pathways

The motor networks and the attentional ones are so interdependent that the question remains about the attentional consequences of chronic severe motor impairments. The general concept of connectional diaschisis (opposed to focal diaschisis) was recently proposed to address this question [68]. Connectional diaschisis is defined as the changes of structural and functional connectivity between brain areas distant to a lesion. The authors consider that those changes should be maximal immediately after the insult and then progressively improve, or sometimes even normalize in parallel with clinical function. Diaschisis is only one phenomenon of connectomal changes impact after cerebral injury, amongst other confounding factors like neuroplasticity and vicariation. All of these factors combined could induce important changes in all systems tightly linked with the motor system, including attention networks.

An action oriented view of cognition

Our hypothesis is in line with an action-oriented view of cognition, which is currently experiencing a strong revival of interest [69]–[71]. This theory postulates that cognition should not be understood as providing models of the world, but as subsuming action and being grounded in sensorimotor coupling. Thus, a bundle of theoretical as well as empirical and clinical arguments contribute to reinforce the hypothesis that a negative plasticity, generated by the occurrence of an extreme sensorimotor disability including oculomotor ones, could be at the origin of the degradation of some cognitive capacities at the heart of which lie attentional processes. We especially suspect diminished spatial selection abilities with detrimental consequences for the control of non-invasive brain-machine interfaces.

ADDITIONAL FACTORS AND OBSERVATIONS THAT NEED TO BE ACCOUNTED FOR

Cognitive impairments due to the physiopathological causes of the disease

ALS: ALS is not only a motoneuron disease. There is a continuum between ALS and other degenerative disorders, the most frequent being fronto-temporal dementia. About 20% of ALS patients meet the criteria for fronto-temporal dementia This implies that even in patients that do not fulfil the whole criteria for fronto-temporal dementia, the observed cognitive impairments could be due to some incomplete form of the disease.

LIS: BCI studies approach LIS patients as a homogenous group. However, from a clinical point of view, the variety of etiologies should be taken into account when estimating their cognitive status. Indeed, the most common cause of LIS is stroke. Stroke patients have a bad cardio-vascular condition that can lead to multiple small cerebral stroke that impair cognitive status [72]. The problem is similar with traumatic lesions that can induce acute diffuse axonal lesions, frontal and/or occipital cortical lesions. Detecting of these lesions is difficult, as the routine (anatomical) imaging technics lack sensibility in that respect [73]. The finest assessment is done by neuropsychological and ecological assessments in communicating patients with traumatic brain injury, but with LIS patients, this is often not feasible, except in expert teams. Other etiologies like infectious and/or toxic ones share the same risk of multiple direct injuries of the central nervous system.

Chronic complete SCI: when the etiology is traumatic, a traumatic brain injury is frequently associated, in at least one third of the cases [74]. Some studies highlighted the cognitive impact of breathing disorders like sleep apnea induced by the tetraplegia in chronic [75]. Sleep apnea concerns up to 91% of these chronic complete SCI patients [76]. More severe chronic apnea was associated with decrease performance in verbal attention and concentration, immediate and short-term memory, cognitive flexibility, internal scanning and working

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-48

memory.

Differences in performance due to a bias in patient selection

Only a few BCI studies addressed the direct comparison between performance of healthy subjects and patients, and even less did control for age, gender and socio-professional status. Other authors noted that the healthy subjects included in the BCI studies are on average younger and more educated than patients, whereas these factors impact cognitive abilities and electrophysiological data. This bias could explain in part the decrease of performance reported in patients.

Besides, the publication bias favors the reporting and diffusion of successful BCI studies. This yields a probable overestimation of patients' performance.

CONCLUSION

Years of BCI attempts to help patients with severe motor impairment have revealed a paradox. These patients have long been regarded as cognitively intact and motivated enough to be the ideal candidates to benefit from neurotechnologies made for communication. However, the performance of BCI in these patients is lower than in healthy controls and no BCI solution is routinely used by these patients who have no other solution yet. This is endorsed by the BCI studies we briefly reviewed. Most importantly, it urges our community to consider the possible factors that beyond imperfect BCI systems and algorithms, may explain this striking observation.

In this short paper, we first question the simplistic view of those patients as being purely motor impaired, with intact cognitive function. It turns out that whatever the etiology, cognitive impairments do come with motor ones.

Precisely, we hypothesize that the attentional processes could be altered in these patients, concomitantly with their motor disorders or because of a negative plasticity resulting from motor disorders. In particular deficits in oculomotricity as selective spatial attention has been shown to be closely related to overt eye movements. This is endorsed by prevalent theories in cognitive neuroscience, namely the premotor theory of attention which is supported by both behavioral and neurophysiological evidence. And this could explain poor BCI performance as most protocols heavily rely on spatial attention.

Interestingly, this view is also endorsed by recent computational theories of the brain that tightly couple perception and action, and namely attention as a gain control process to select visual features and eye movements towards salient spatial locations [77].

We conclude that improving the way we approach those patients requires addressing those questions in all their complexity, by using neuropsychological assessments, by considering the heterogeneity of etiologies and patient's trajectories, by testing patients in a longitudinal and multi-dimensional manner and also by reporting negative BCI results. This might be the only way to design new and efficient BCI solutions.

This work was supported by the following grant from the French government: ANR-17-CE40-0005, Mind Made Clear.

REFERENCES

[1] M. Marchetti et K. Priftis, « Brain–computer interfaces in amyotrophic lateral sclerosis: A metanalysis », *Clin. Neurophysiol.*

[2] S. Silvoni, M. Cavinato, C. Volpato, C. A. Ruf, N. Birbaumer, et F. Piccione, « Amyotrophic lateral sclerosis progression and stability of brain-computer interface communication », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 14, nº 5-6, p. 390-396, sept. 2013.

[3] F. Plum et J. B. Posner, « The diagnosis of stupor and coma », *Contemp. Neurol. Ser.*, vol. 10, p. 1-286, 1972.

[4] 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, nº 7, p. 571-582, juill. 2002.

[5] J. P. Taylor, R. H. Brown, et D. W. Cleveland, « Decoding ALS: from genes to mechanism », *Nature*, vol. 539, nº 7628, p. 197-206, nov. 2016.

[6] M. P. Regan et D. Regan, « Objective investigation of visual function using a nondestructive zoom-FFT technique for evoked potential analysis », *Can. J. Neurol. Sci. J. Can. Sci. Neurol.*, vol. 16, nº 2, p. 168-179, mai 1989.

[7] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, et C. Neuper, « Brain oscillations control hand orthosis in a tetraplegic », *Neurosci. Lett.*, vol. 292, nº 3, p. 211-214, oct. 2000.

[8] A. Kübler *et al.*, « Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface », *Neurology*, vol. 64, nº 10, p. 1775-1777, mai 2005.

[9] C. Vidaurre et B. Blankertz, « Towards a cure for BCI illiteracy », *Brain Topogr.*, vol. 23, nº 2, p. 194-198, juin 2010.

[10] N. Neumann, A. Kübler, J. Kaiser, T. Hinterberger, et N. Birbaumer, « Conscious perception of brain states: mental strategies for brain-computer communication », *Neuropsychologia*, vol. 41, nº 8, p. 1028-1036, 2003.

[11] A. Kuebler *et al.*, « Self-regulation of slow cortical potentials in completely paralyzed human patients », *Neurosci. Lett.*, vol. 252, nº 3, p. 171-174, août 1998.

[12] J. R. Wolpaw *et al.*, « Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis », *Neurology*, vol. 91, nº 3, p. e258-e267, juill. 2018.

[13] F. Nijboer, « Technology transfer of brain-computer interfaces as assistive technology: barriers and opportunities », *Ann. Phys. Rehabil. Med.*, vol. 58, nº 1, p. 35-38, févr. 2015.

[14] A. Kübler *et al.*, « The User-Centered Design as Novel Perspective for Evaluating the Usability of BCI-Controlled Applications », *PLoS ONE*, vol. 9, nº 12, p. e112392, déc. 2014.

[15] N. Birbaumer, « Breaking the silence: brain-computer interfaces (BCI) for communication and motor control », *Psychophysiology*, vol. 43, nº 6, p. 517-532, nov. 2006.

[16] A. Kübler et N. Birbaumer, « Brain–computer interfaces and communication in paralysis: Extinction of goal directed thinking in completely paralysed patients? », *Clin. Neurophysiol.*, vol. 119, nº 11, p. 2658-2666, nov. 2008.

[17] I. Lazarou, S. Nikolopoulos, P. C. Petrantonakis, I. Kompatsiaris, et M. Tsolaki, « EEG-Based Brain–Computer Interfaces for Communication and Rehabilitation of People with Motor Impairment: A Novel Approach of the 21st Century », *Front. Hum. Neurosci.*, vol. 12, janv. 2018.

[18] C. Guger *et al.*, « Complete Locked-in and Locked-in Patients: Command Following Assessment and Communication with Vibro-Tactile P300 and Motor Imagery Brain-Computer Interface Tools », *Front. Neurosci.*, vol. 11, p. 251, 2017.

[19] U. Chaudhary, B. Xia, S. Silvoni, L. G. Cohen, et N. Birbaumer, « Brain-Computer Interface-Based Communication in the Completely Locked-In State », *PLoS Biol.*, vol. 15, nº 1, p. e1002593, 2017.

[20] M. Spüler, « Questioning the evidence for BCI-based communication in the complete locked-in state », *PLOS Biol.*, vol. 17, nº 4, p. e2004750, avr. 2019.

[21] Y. Okahara *et al.*, « Long-term use of a neural prosthesis in progressive paralysis », *Sci. Rep.*, vol. 8, nº 1, p. 16787, nov. 2018.

[22] 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, nº 2, p. 026006, avr. 2014.

[23] 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, nº 2, p. 183-191, mars 2012.

[24] N. Birbaumer et F. C. Hummel, « Habit learning and brain–machine interfaces (BMI): a tribute to Valentino Braitenberg's "Vehicles" », *Biol. Cybern.*, vol. 108, nº 5, p. 595-601, oct. 2014.

[25] N. Neumann et B. Kotchoubey, « Assessment of cognitive functions in

Published by Verlag der TU Graz
Graz University of Technology

severely paralysed and severely brain-damaged patients: neuropsychological and electrophysiological methods », *Brain Res. Protoc.*, vol. 14, nº 1, p. 25-36, nov. 2004.

[26] B. Brownell, D. R. Oppenheimer, et J. T. Hughes, « The central nervous system in motor neurone disease », *J. Neurol. Neurosurg. Psychiatry*, vol. 33, nº 3, p. 338-357, juin 1970.

[27] A. J. Hudson, « Amyotrophic lateral sclerosis and its association with dementia, parkinsonism and other neurological disorders: a review », *Brain J. Neurol.*, vol. 104, nº 2, p. 217-247, juin 1981.

[28] T. Mizutani *et al.*, « Amyotrophic lateral sclerosis with ophthalmoplegia and multisystem degeneration in patients on long-term use of respirators », *Acta Neuropathol. (Berl.)*, vol. 84, nº 4, p. 372-377, sept. 1992.

[29] Y. Nakayama, T. Shimizu, K. Hayashi, Y. Mochizuki, M. Nagao, et K. Oyanagi, « [Predictors the progression of communication impairment in ALS tracheostomy ventilator users] », *Rinshō Shinkeigaku Clin. Neurol.*, vol. 53, nº 11, p. 1396-1398, 2013.

[30] K. Hayashi *et al.*, « [Communication disorder in amyotrophic lateral sclerosis after ventilation--a proposal of staging and a study of predictive factor] », *Rinshō Shinkeigaku Clin. Neurol.*, vol. 53, nº 2, p. 98-103, 2013.

[31] K. Pugdahl *et al.*, « Generalised sensory system abnormalities in amyotrophic lateral sclerosis: a European multicentre study », *J. Neurol. Neurosurg. Psychiatry*, vol. 78, nº 7, p. 746-749, juill. 2007.

[32] J. D. Isaacs, A. F. Dean, C. E. Shaw, A. Al-Chalabi, K. R. Mills, et P. N. Leigh, « Amyotrophic lateral sclerosis with sensory neuropathy: part of a multisystem disorder? », *J. Neurol. Neurosurg. Psychiatry*, vol. 78, nº 7, p. 750-753, juill. 2007.

[33] E. Beeldman, J. Raaphorst, M. Klein Twennaar, M. de Visser, B. A. Schmand, et R. J. de Haan, « The cognitive profile of ALS: a systematic review and meta-analysis update », *J. Neurol. Neurosurg. Psychiatry*, vol. 87, nº 6, p. 611-619, juin 2016.

[34] P. Sorrentino *et al.*, « Brain functional networks become more connected as amyotrophic lateral sclerosis progresses: a source level magnetoencephalographic study », *NeuroImage Clin.*, vol. 20, p. 564-571, 2018.

[35] L. M. McCane *et al.*, « P300-based brain-computer interface (BCI) event-related potentials (ERPs): People with amyotrophic lateral sclerosis (ALS) vs. age-matched controls », *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 126, nº 11, p. 2124-2131, nov. 2015.

[36] R. McMackin *et al.*, « Dysfunction of attention switching networks in amyotrophic lateral sclerosis », *NeuroImage Clin.*, vol. 22, févr. 2019.

[37] 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, nº 5, p. 437-440, mai 2016.

[38] P. Allain, P. A. Joseph, J. L. Isambert, D. Le Gall, et J. Emile, « Cognitive functions in chronic locked-in syndrome: a report of two cases », *Cortex J. Devoted Study Nerv. Syst. Behav.*, vol. 34, nº 4, p. 629-634, sept. 1998.

[39] S. F. Cappa et L. A. Vignolo, « Locked-in syndrome for 12 years with preserved intelligence », *Ann. Neurol.*, vol. 11, nº 5, p. 545, mai 1982.

[40] S. F. Cappa, C. Pirovano, et L. A. Vignolo, « Chronic "locked-in" syndrome: psychological study of a case », *Eur. Neurol.*, vol. 24, nº 2, p. 107-111, 1985.

[41] F. Gayraud *et al.*, « Written production in a case of locked-in syndrome with bilateral corticopontic degeneration », *Neuropsychol. Rehabil.*, vol. 25, nº 5, p. 780-797, 2015.

[42] C. Schnakers *et al.*, « Cognitive function in the locked-in syndrome », *J. Neurol.*, vol. 255, nº 3, p. 323-330, mars 2008.

[43] M. Rousseaux, E. Castelnot, P. Rigaux, O. Kozlowski, et F. Danzé, « Evidence of persisting cognitive impairment in a case series of patients with locked-in syndrome », *J. Neurol. Neurosurg. Psychiatry*, vol. 80, nº 2, p. 166-170, févr. 2009.

[44] M. Conson, S. Sacco, M. Sarà, F. Pistoia, D. Grossi, et L. Trojano, « Selective motor imagery defect in patients with locked-in syndrome », *Neuropsychologia*, vol. 46, nº 11, p. 2622-2628, sept. 2008.

[45] C. Babiloni *et al.*, « Resting state eyes-closed cortical rhythms in patients with locked-in-syndrome: an EEG study », *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 121, nº 11, p. 1816-1824, nov. 2010.

[46] Z. R. Lugo *et al.*, « Cognitive Processing in Non-Communicative Patients: What Can Event-Related Potentials Tell Us? », *Front. Hum. Neurosci.*, vol. 10, p. 569, 2016.

[47] F. Pistoia *et al.*, « Disembodied Mind: Cortical Changes Following Brainstem Injury in Patients with Locked-in Syndrome », *Open Neuroimaging J.*, vol. 10, p. 32-40, mai 2016.

[48] J. Decety et D. Boisson, « Effect of brain and spinal cord injuries on motor imagery », *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 240, nº 1, p. 39-43, 1990.

[49] M. G. Lacourse, M. J. Cohen, K. E. Lawrence, et D. H. Romero, « Cortical potentials during imagined movements in individuals with chronic spinal cord injuries », *Behav. Brain Res.*, vol. 104, nº 1-2, p. 73-88, oct. 1999.

[50] S. C. Cramer, E. L. R. Orr, M. J. Cohen, et M. G. Lacourse, « Effects of motor imagery training after chronic, complete spinal cord injury », *Exp. Brain Res.*, vol. 177, nº 2, p. 233-242, févr. 2007.

[51] I. Lazzaro, Y. Tran, N. Wijesuriya, et A. Craig, « Central correlates of impaired information processing in people with spinal cord injury », *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.*, vol. 30, nº 1, p. 59-65, févr. 2013.

[52] H. Burianová *et al.*, « Adaptive Motor Imagery: A Multimodal Study of Immobilization-Induced Brain Plasticity », *Cereb. Cortex N. Y. N 1991*, vol. 26, nº 3, p. 1072-1080, mars 2016.

[53] M. Corbetta *et al.*, « A common network of functional areas for attention and eye movements », *Neuron*, vol. 21, nº 4, p. 761-773, oct. 1998.

[54] A. C. Nobre, D. R. Gitelman, E. C. Dias, et M. M. Mesulam, « Covert visual spatial orienting and saccades: overlapping neural systems », *NeuroImage*, vol. 11, nº 3, p. 210-216, mars 2000.

[55] M. S. Beauchamp, L. Petit, T. M. Ellmore, J. Ingeholm, et J. V. Haxby, « A parametric fMRI study of overt and covert shifts of visuospatial attention », *NeuroImage*, vol. 14, nº 2, p. 310-321, août 2001.

[56] L. Craighero, A. Carta, et L. Fadiga, « Peripheral oculomotor palsy affects orienting of visuospatial attention », *Neuroreport*, vol. 12, nº 15, p. 3283-3286, oct. 2001.

[57] L. Craighero, M. Nascimben, et L. Fadiga, « Eye position affects orienting of visuospatial attention », *Curr. Biol. CB*, vol. 14, nº 4, p. 331-333, févr. 2004.

[58] H. Brown, K. Friston, et S. Bestmann, « Active inference, attention, and motor preparation », *Front. Psychol.*, vol. 2, p. 218, 2011.

[59] D. T. Smith et T. Schenk, « The Premotor theory of attention: time to move on? », *Neuropsychologia*, vol. 50, nº 6, p. 1104-1114, mai 2012.

[60] J. Polich, « Updating P300: an integrative theory of P3a and P3b », *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 118, nº 10, p. 2128-2148, oct. 2007.

[61] R. Verleger, L. M. Hamann, D. Asanowicz, et K. Śmigasiewicz, « Testing the S-R link hypothesis of P3b: The oddball effect on S1-evoked P3 gets reduced by increased task relevance of S2 », *Biol. Psychol.*, vol. 108, p. 25-35, mai 2015.

[62] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, et G. Schalk, « Does the "P300" speller depend on eye gaze? », *J. Neural Eng.*, vol. 7, nº 5, p. 056013, oct. 2010.

[63] M. Marchetti, F. Piccione, S. Silvoni, L. Gamberini, et K. Priftis, « Covert Visuospatial Attention Orienting in a Brain-Computer Interface for Amyotrophic Lateral Sclerosis Patients », *Neurorehabil. Neural Repair*, p. 1545968312471903, janv. 2013.

[64] P. Séguin *et al.*, « Évaluation clinique d'une interface cerveau–machine auditive à destination des personnes en Locked-in syndrome complet », *Neurophysiol. Clin. Neurophysiol.*, vol. 46, nº 2, p. 92-93, avr. 2016.

[65] E. Larson et A. K. C. Lee, « The cortical dynamics underlying effective switching of auditory spatial attention », *NeuroImage*, vol. 64, p. 365-370, janv. 2013.

[66] I. Koch, V. Lawo, J. Fels, et M. Vorländer, « Switching in the cocktail party: exploring intentional control of auditory selective attention », *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 37, nº 4, p. 1140-1147, août 2011.

[67] C.-T. Wu, D. H. Weissman, K. C. Roberts, et M. G. Woldorff, « The neural circuitry underlying the executive control of auditory spatial attention », *Brain Res.*, vol. 1134, nº 1, p. 187-198, févr. 2007.

[68] E. Carrera et G. Tononi, « Diaschisis: past, present, future », *Brain J. Neurol.*, vol. 137, nº Pt 9, p. 2408-2422, sept. 2014.

[69] A. K. Engel, A. Maye, M. Kurthen, et P. König, « Where's the action? The pragmatic turn in cognitive science », *Trends Cogn. Sci.*, vol. 17, nº 5, p. 202-209, mai 2013.

[70] G. Buzsáki, A. Peyrache, et J. Kubie, « Emergence of Cognition from Action », *Cold Spring Harb. Symp. Quant. Biol.*, vol. 79, p. 41-50, 2014.

[71] T. M. Press, « The Pragmatic Turn », *The MIT Press*. [En ligne]. Disponible sur: https://mitpress.mit.edu/books/pragmatic-turn. [Consulté le: 16-févr-2019].

[72] P. B. Gorelick *et al.*, « Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the american heart association/american stroke association », *Stroke*, vol. 42, nº 9, p. 2672-2713, sept. 2011.

[73] M. Wintermark, P. C. Sanelli, Y. Anzai, A. J. Tsiouris, C. T. Whitlow, et American College of Radiology Head Injury Institute, « Imaging evidence and recommendations for traumatic brain injury: advanced neuro- and neurovascular imaging techniques », *AJNR Am. J. Neuroradiol.*, vol. 36, nº 2, p. E1-E11, févr. 2015.

[74] B. Budisin *et al.*, « Traumatic Brain Injury in Spinal Cord Injury: Frequency and Risk Factors », *J. Head Trauma Rehabil.*, vol. 31, nº 4, p. E33-42, août 2016.

[75] D. Sajkov *et al.*, « Sleep apnoea related hypoxia is associated with cognitive disturbances in patients with tetraplegia », *Spinal Cord*, vol. 36, nº 4, p. 231-239, avr. 1998.

[76] A. Sankari, J. L. Martin, et M. S. Badr, « Sleep Disordered Breathing and Spinal Cord Injury: Challenges and Opportunities », *Curr. Sleep Med. Rep.*, vol. 3, nº 4, p. 272-278, déc. 2017.

[77] T. Parr et K. J. Friston, « Attention or salience? », *Curr. Opin. Psychol.*, vol. 29, p. 1-5, oct. 2019.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-49

SIMPLE DEEP NEURAL NETWORKS SHOW STATE-OF-THE-ART PERFORMANCE IN ERP-BASED BCI

L. Delobel[1], E. Maby[1] & J. Mattout[1]

[1]Lyon Neuroscience Research Center, CRNL; INSERM, U1028; CNRS, UMR5292;
Brain Dynamics and Cognition Team, Lyon, F-69000, France
[2]University Lyon 1, Lyon, F-69000, France
E-mail: loic.delobel@gmail.com

ABSTRACT: Deep neural networks (DNN) have great success in solving difficult recognition tasks such as speech and image processing. However, performance depends on the amount of data available for training the network. In BCI, very large datasets are still missing and EEG data are particularly complex and noisy. Nevertheless, it is of interest to investigate whether DNN may prove efficient in this context. We tested 6 different deep learning models to accomplish binary classification of single-trial ERPs and compared them with Riemannian Geometry based classifiers. Each model implements a different architecture and uses two different input formats: an image (`2D`) or a video (`3D`). All models were tested on two different datasets, and under three different scenarios: within-subject, cross-subject and cross-experiment classification. Finally, to get insights about the decision process of the most successful DNN, we visualized the learned features using saliency maps. This revealed informative and interpretable differences between the two empirical datasets used for evaluation.

INTRODUCTION

This work focuses on the application of deep learning methods to EEG recordings for Brain-Computer Interfaces (BCIs) [1]. A BCI is a device that allows interaction with a machine through brain signals, bypassing normal neuromuscular outputs. For example, the P300 Speller [2] is a BCI which helps to restore communication with people who cannot control their muscles anymore (e.g. patients with Amyotrophic Lateral Sclerosis). The P300 Speller exploits the P300 event-related potential (ERP) and performs a binary classification task: detecting the presence or not of a P300 wave. To accomplish this task, Riemannian Geometry classifier (RGC) and shrinkage based linear discriminant analysis classifier (sLDA) seem to be, to date, the current state-of-the-art method [3].

Deep learning methods have emerged from the connectionism movement in cognitive science that hopes to explain intellectual abilities using connectionist architectures known as artificial neural networks (ANN [4]). Connectionist architectures have existed for more than 70 years, but new architectures and graphical processing units (GPUs) brought them to the forefront of artificial intelligence. Depending on their architecture, neural networks excel at speech and language recognition (RNN, LSTM [5]) or image and video recognition with convolutional neural networks (CNN [6]). CNN exploit the local connectivity concept that makes them suitable for EEG data. Although still few in number, recent studies have shown that deep learning methods can be effective in classifying ERP components [7–10].

In the current study, we compare the performance of 6 deep learning network architectures for application to a P300 Speller [2] (26 subjects) and a RSVP [2] (11 subjects) paradigm, both exploiting the P300 and N200 components. This makes it possible to observe the behavior of deep learning methods in two different paradigms. We introduce different deep learning network architectures and compare them with state-of-the-art methods that combine XDAWN algorithm [11] and Riemannian geometry [12]. In order to explore the intra-subject and inter-subject generalization abilities, we compare the performance of the classifiers for two different scenarios: Within-Subject testing and Cross-Subject testing. In addition, the best model resulting from these scenarios is tested in a cross-experiment scenario where it is trained with data coming from one experiment and tested on the other.

MATERIALS AND METHODS

Datasets: Two datasets were used for evaluation and model comparisons. The first one is denoted as `P300 Speller` and implements the P300 speller protocol [2]. This protocol consists in displaying a grid of 36 symbols on screen (the 26 letters of the alphabet, the underscore and the 9 digits) and asking the user to focus attention on the desired item. The flashing of the items was similar to that reported in [13], i.e., the splotch stimulus presentation. Twelve pre-defined groups of 6 non-adjacent items are flashed (one at a time) in a pseudo-random order. Each item belongs to two groups only and these two groups have only this particular item in common. A P300 evoked potential is expected when the desired symbol is flashed. A single pass on all groups (all symbols twice) is referred to as a *repetition*. The `P300 Speller` dataset is an electroencephalographic (EEG) collection of 56-channel recorded from

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-49

Table 1: Description of the datasets

Name	Rsvp	P300 Speller
Type	RSVP	P300 Speller
Subjects	11	23
Sensor	48	48
Repetitions	10	2 to 4
Sampling Frequency	200	600
Down-sampled to	100	100
Band pass filtered	1Hz-20Hz	1Hz-20Hz
Epoched between	0ms and 600ms	0ms and 600ms

23 healthy subjects during one of our previous studies. We made the spelling challenging by considering very short (2 repetition-long) and short (4 repetition-long) trials [14].

The second dataset denoted as `Rsvp` and available online [15], was acquired during a Rapid serial visual presentation (RSVP) task. The RSVP paradigm is a variant of the P300 speller where all the symbols are presented one-by-one in a serial manner and in the center of the screen (on fovea). This dataset is made of data from 12 healthy subjects but we discarded the data of subject `VPgce` because of the use of a different set of channels. In this experiment, item selection relied on 10 repetitions (see [15] for a thorough description of the paradigm).

Data formats and pre-processing: Table 1 summarizes the main characteristics and pre-processing steps for the two studied datasets. They were preprocessed in the same manner: for each subject, raw signals were band-pass filtered between 1Hz and 20Hz and downsampled to 100Hz. Epochs were extracted from 0ms to 600ms after stimulus onset, resulting in two big matrices (containing the Target/Non-Target samples) $X \in R^{N \times C \times T}$ with the number of EEG sensors denoted by C, time samples denoted by T and the number of epochs denoted by N. We selected 48 electrodes common to both datasets (F1-F8-Fz, FC1-FC6-FCz, T7-T8, C1-C6-Cz, TP7-TP8, CP1-CP6-CPz, P1-P8-Pz, PO7-PO8-POz, O1-O2).

We considered two different data formats for subsequent classification with deep learning models. The first one is the standard method where each sample is a matrix $X \in R^{C \times T}$ with the number of EEG sensors denoted by C and time samples denoted by T. This is more convenient for an online application where no additional transformations are needed. But the disadvantage is that the classifier is dependent on the number of electrodes on which it has been trained. This format is denoted by `2D` from now on. The second format aims at alleviating the disadvantages of the first format. Considering the X matrix presented above, it consists of interpolating the space vector (sensors) into a square matrix (16×16) for each time sample. It results in a matrix $X \in R^{H \times W \times T}$ with $H = 16$ and $W = 16$ which correspond to a topographic scalp map evolving over time (a video). Therefore we used a 2D clough tocher (cubic) interpolation [16]. Having a spatial representation (as a pixelized image) allows better processing of datasets with different EEG electrode montages. It also makes perfect sense to apply such a spatial smoothing, because of the blurring effect of head tissues. Because of their excellent performance on natural images, we expected convolutional neural networks to better handle this kind of format [17, 18]. However, we must bear in mind that such a transformation has a cost (computation time) and can complicate the real-time pipeline. It also increases the number of degrees of freedom, hence the complexity and training time. This format is denoted by `3D` from now on.

Models and training procedures: Riemannian geometry based classifiers – We compared our deep learning models with two classifiers based on Riemannian geometry which outperformed other approaches in BCI competitions. This method is considered as a strong reference in the field [12]. We used a Riemannian method based on tangent space mapping which shows overall better performance compared with the minimum distance to mean algorithm. A recommended pipeline is to estimate XDAWN covariances and project them into the tangent space, then classify with logistic regression [12]. We used 2 components to estimate the XDAWN covariances corresponding to the 2 classes (Target/Non-Target). XDAWN reduces the dimensionality of data, which facilitates and accelerates training. This pipeline is denoted as `XDAWN + Riemann`. We also used a variant with no XDAWN filters as XDAWN is not very effective in transfer learning situations. This pipeline is denoted as `Riemann`.

Deep learning classifiers – With the use of the two data formats, we developed 6 deep learning models that exploit the temporal and spatial dimensions of the data. Also, EEG signals are known to have inherent temporal and spatial smoothness properties. In other words, close features in space and time dimension are dependent. Thus, we wanted to test different architectures with different assumptions about local dependencies over space and time. CNN and GRU (Gated Recurrent Unit) layers are both strong candidates to accomplish this task. CNN layers are known to be very effective for the classification of image-like data with spatial pixels. As for the GRU layers, they enable the learning of temporal dependencies. GRU layers are a less resource-intensive alternative to LSTMs, while maintaining good performance. We first developed three models that take the `2D` format as input. `MLP` is a multilayer perceptron (MLP) composed of 1 hidden layer. This is the simplest architecture: it does not take into account the temporal and spatial dependencies in EEG data. The `CNN1D_T2` is composed of a convolutional layer. It is inspired by the proposed EEGNet one [9] but only one convolution operation was done to learn both spatial and temporal features in one go. We found empirically that splitting the operation into two steps does not improve performance. The learned filters are also more easily interpretable. This is a 1-dimension convolution because the convolutional filters are moved in only one direction. The 10 filters of size 48×2 represent a

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-49

time window of $\simeq 10ms$ (2 timesteps at 100Hz). The feature maps are passed directly to the final `LogSoftMax` layer with no additional dense layers in-between (non-linear activations). This has been proven to work effectively while reducing the number of parameters [19]. The `CNN1D_T2_GRU` is the `CNN1D_T2` with an additional output GRU layer. A GRU layer is a recurrent layer that allows the learning of temporal dependencies [20]. We then developed three other models that take the `3D` format as an input: The `CNN2D_S3` exploits the 3-dimensional input format by learning spatial features with a 2-dimensional convolutional layer applied to the spatial dimension. The weights of the filters are shared across the spatial dimension but not across the temporal dimension. The `CNN2D_S3_T2` is a variant of `CNN2D_S3` that shares the weights of the convolutional filters over the temporal dimension. A 3-dimensional convolutional layer is needed to perform this operation. The `CNN2D_S3_T2_GRU` is the `CNN2D_S3_T2` with an additional output GRU layer.

All models also include a batch normalization layer that influences the training by normalizing the data before they enter the hidden layer and the LogSoftmax layer at the end, before predicting the output probabilities pertaining to the two classes.

All deep learning models were trained with the `RMSProp` optimization algorithm that allows adapting the learning rate during training. We used an early stopping strategy during training, meaning that the training was interrupted as soon as the valid loss started to increase. Precisely, we train for N epochs (big enough not to fall into an under-fitting situation) and we go select the epoch with the smallest valid loss. The learning rate was set to 10^{-4} or 10^{-5} depending on the experiment. We set the batch size to 128 and the optimizer momentum to 0.9 as recommended in [21]. We performed L2-norm regularization with a weigh decay of 5^{-4}. All deep learning models terminate with a `LogSoftMax` layer. The negative log-likelihood loss was used for our two-class classification problem (Target/Non-Target). No manual rescaling weight or over/undersampling methods were used to balance the samples between the two classes. The unbalanced dataset is not problematic because of the final averaging (in a Bayesian fashion, see below).

Evaluation metrics and statistics: In such a BCI protocol, there are two levels of classification, hence two levels of accuracy. The first one is the accuracy at the level of the binary classifier (number of correctly classified Target/Non-Target trials). The second one is the accuracy corresponding to the number of correctly classified (or spelled) symbols in a given sentence. We only report the latter that corresponds to the one of interest for BCI use. Initially, all items are assumed to be equiprobable targets. At each new observation (after computing the probability for the current sample to be a target), this belief is updated following Bayes rule, by optimally combining the data likelihood and prior and by considering the posterior belief as the prior for the next observation. In this way, all items of the sentence are predicted. At the end of the sentence, we compute the final spelling accuracy (number of correct symbols) with different repetitions. Obviously, better accuracy is expected as repetitions increase. For statistical significance, we used a one-way ANOVA (III) and post-hoc Tukey Test on the spelling accuracy.

Evaluation procedures: We evaluated the spelling accuracy of the classifiers with three different scenarios: within-subject, cross-subjects and cross-experiments.

Within-subject testing – We first explored the intra-subject generalization. In both datasets, each subject completed a calibration phase and a test phase. In order to simulate online classification, we trained our classifiers using calibration data only and tested on the test phase data. To avoid overfitting, we perform cross-validation in a way that the data from the subject's calibration phase is split to form the training (80%) and validation (20%) sets. The dataset's test phase simply formed the test set. Thus, N tested subjects lead to N fits composed by a training, validation and test set.

Cross-subjects testing – Many studies attempt to reduce the time of the calibration phase. In fact, the calibration stage is exhausting and demotivating for the subject. The ideal situation would be to completely free oneself from it. To assess the ability of the classifiers to generalize from a pool of subject to a new subject, we built a scenario where each subject is tested with a classifier which has been trained with the calibration plus test data of all other subjects. More specifically, the data from each subject's calibration and test phase minus the tested one were merged to form the training and validation sets (80%/20%). The data from the test phase of the tested subject simply form the test set. Thus, N tested subjects lead to N fits composed by a training, validation and test set.

Cross-experiment testing – To further assess the ability of the models to generalize across different experiments involving different subjects but also different protocols, we chose the model that performed best in the two previous scenarios and tested it in a cross-experiment scenario. We trained the chosen model on the calibration data of the `P300 Speller` dataset plus the full `Rsvp` dataset except tested subject.

Exploring the learned features:

Deep learning models are able to automatically extract the relevant features for classification, in an unsupervised manner. Apart from the choice of design and hyperparameters, no a priori knowledge was injected into the model to accomplish the classification task. Although interesting and useful, this property makes it difficult to apprehend the rationale of the decision process that the model implemented. Thus, the purpose of this experiment was to reveal the features that the neural networks identified as most relevant in order to classify a sample as a target or non-target response. Also, understanding what neural networks have learned may point to interest-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-49

Table 2: Number of model parameters and training time (approximate ratio) on Rsvp dataset.

Classifiers	#Params	Within	Cross
Riemann XDWAN	-	1x	1x
Riemann	-	12x	150x
MLP	5862	6x	11x
CNN1D_T2	1582	11x	28x
CNN1D_T2_GRU	1662	52x	85x
CNN2D_S3	4242	44x	192x
CNN3D_S3_T2	3287	83x	450x
CNN3D_S3_T2_GRU	2317	152x	780x

ing and unknown neurophysiological phenomena. In image classification, a common approach to understand the decisions of a classifier is to find regions of an image that were particularly influential (saliency maps). We chose a simple technique which allows elucidating the relevant pixels in the input image [22]. We chose the CNN1D_T2 model to test this method. It takes the 2D format as an input.

RESULTS

We used PyTorch to build the models. The code to reproduce the results and a more detailed description of the models are available online.[1] Figure 1 shows the accuracy performance at 2 repetitions for all classifiers applied to the two datasets in Within-subject testing and Cross-subject testing. Accuracies can be interpreted as online accuracies because only calibration data were used for training. Table 2 shows the training time for each classifiers on Rsvp dataset. The number of model parameters is also reported. The training times are approximated and have been calculated by multiplying the time of an iteration by the number of iterations (for deep learning models). Note that the training time of a model has been chosen to ensure that the model has been fully trained. Therefore, the optimal training time of a model may be less than the one indicated.

Within-Subject testing:

Result 1 – Deep learning classifiers that use the 2D format as input do as well as Riemann's classifiers on both datasets.

Result 2 – Deep learning models that use the 3D input format are statistically less effective than others on both datasets (p-value < 0.05).

Result 3 – Training deep learning classifiers took substantially longer than Riemann ones, at least for the most complex ones.

Cross-Subject testing:

Result 4 – XDAWN + Riemann is significantly worse than Riemann (p-value < 0.05).

Result 5 – Overal Cross-Subject testing performance is significantly worse than the Within-Subject testing (p-value < 0.05).

[1] https://git.io/fj317

Result 6 – All classifiers (except CNN3D_S3_T2_GRU) do as well as Riemann on both datasets (p-value < 0.05).

Cross-Experiment testing: Figures 2 shows the accuracy performance for the Cross-Experiment testing.

Result 7 – Cross-Experiment performance is no better than Cross-Subject performance.

Exploring the learned features: – Figure 3 shows, for CNN1D_T2 model, the absolute value of the difference between the mean gradient of the Target samples and the mean gradient of the Non-Target samples. In other words, pixels with relatively high gradient values are useful for distinguishing target from non-target samples.

Result 8 – CNN1D_T2 has learned well known features on both datasets (N200 around 200ms on P300 Speller dataset and P300 around 300ms on Rsvp dataset).

Result 9 – CNN1D_T2 focuses on earlier features (around 200ms) when trained on P300 Speller dataset compared to when trained on Rsvp dataset (around 300ms).

DISCUSSION

This study explored the performance of deep neural networks for feature extraction and classification in the context of ERP-based BCIs, namely the P300-Speller and RSVP paradigms. Overall, we showed that deep learning methods with an appropriate architecture can perform as well as Riemannian geometry for intra-subject and cross-subject generalization, on both paradigms.

A surprising result is that the MLP model, a very simple one, proves able to perform as well as the best models, even in a transfer learning situation. Keeping in mind that the raw data are band-pass filtered in the low frequencies and downsampled, this certainly eases the learning in the time domain. Besides, we noticed that the Batch Normalization step plays a major role in the convergence of the models. Hence a MLP could be a strong candidate for BCIs, since its simplicity implies short training time.

In this study, We also compared two different input formats for EEG data: an image (2D) versus a topographic video of the scalp (3D). From a practical point of view, the first one is easy to use and does not require much computing power but requires that the model is compatible with the montage (the number and position of sensors). Conversely, the latter is convenient for mixing datasets from different experiments with potentially different setups. However, transforming EEG signals into a topographic video is time-consuming and can become challenging in real-time. In terms of performance, we found that the suggested models that use the 3D format are no better than those that use the 2D format. Although it might be that other model architectures might be able to extract more information for the video format, our results suggest that despite its flexibility with regard to the number of channels, it did not yield higher performance in the tasks we studied.

Note that obviously, these results do not reflect a fully

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-49

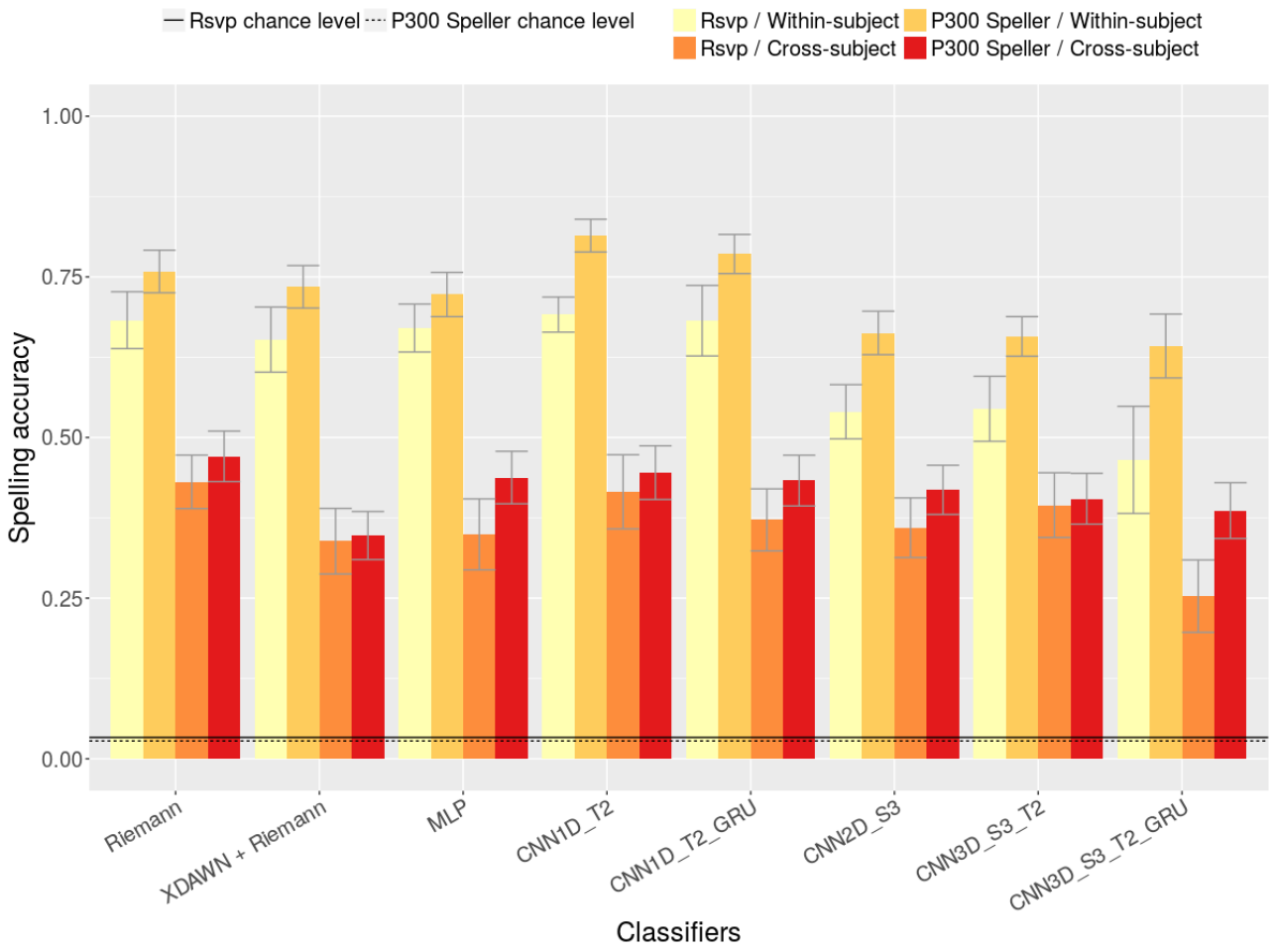


Figure 1: Spelling accuracy performance of all classifiers for Within-subject testing and Cross-subject testing on the two datasets, at 2 repetitions.

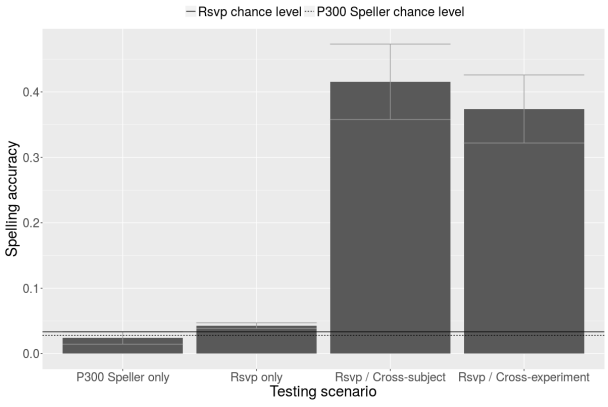


Figure 2: Accuracy performance of model CNN1D_T2 for cross-experiment testing. `RSVP only` (resp. `P300 Speller only`) is the performance of the model when trained on the `Rsvp` data (resp. `P300 Speller` data) and tested on the `P300 Speller` (resp. `Rsvp`) data. `Rsvp / Cross-experiment` refers to the performance of the model when trained on data samples from the two datasets and tested on the `Rsvp` dataset. For comparison `Rsvp / Cross-subject` is the one obtained when trained and tested on the `Rsvp` data only.

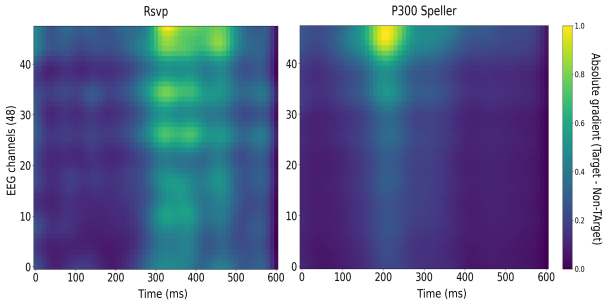


Figure 3: Absolute value of the difference between the mean gradient of the Target samples and the mean gradient of the Non-Target samples (normalized between 0 and 1).

realistic simulation of an online experiment. These deep learning models have a well-known limit that could affect their real-time use: the training time. Although the proposed 2D models (and the MLP) seem applicable using a laptop computer with few resources, this will probably not be the case for 3D models. But this will have to be carefully evaluated in practice, online.

Explaining the decisions of the neural networks is very interesting, notably to have more confidence in the model but also to discover new relevant features in the data. Vi-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-49

sualization techniques can help to tackle this difficulty. In this aim, we applied a well-known approach coming from image classification in deep learning to visualize the inner space of trained neural networks. Our results show that, when trained on P300-Speller and Rsvp data, deep learning models learn well-known discriminating features such as the N200 and P300 components. In addition, the models trained on the `P300 Speller` dataset seem to focus on the N200 ERP while the models trained on the `Rsvp` dataset rather exploit the P300 ERP. This actually fits quite well with the specificity of these two paradigms. Indeed, in Rsvp, the target and non-target stimuli are all displayed at the same location on the screen, and at the center of the fovea, with the same intensity. This means that only attention related components will contribute to the classification, hence mostly the P300. Conversely, in P300-Speller, only the target is flashed at the center of the fovea, which implies that early visual components will also contribute to the classification.

CONCLUSION

To conclude, we showed that fairly simple Deep learning models are serious candidates for ERP-based BCI, showing similar performance as state-of-the-art methods. In addition, their reasonable calibration times make them suitable for real-time application. On the other hand, we have shown that visualization techniques are promising for explaining the decisions of neural networks and for identifying possible new electrophysiological markers.

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] 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.

[3] 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.

[4] Haykin S. Neural Networks: A Comprehensive Foundation. 2nd. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

[5] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997;9(8):1735–1780.

[6] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444.

[7] Cecotti H, Graeser A. Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces. IEEE transactions on pattern analysis and machine intelligence. 2011;33:433–45.

[8] Manor R, Geva AB. Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI. Frontiers in Computational Neuroscience. 2015;9.

[9] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces. arXiv:1611.08024 [cs, q-bio, stat]. 2016.

[10] Tal O, Friedman D. Using recurrent neural networks for p300-based brain-computer interfaces. 6.

[11] Rivet* B, Souloumiac A, Attina V, Gibert G. xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain–Computer Interface. IEEE Transactions on Biomedical Engineering. 2009;56(8):2035–2043.

[12] Congedo M, Barachant A, Bhatia R. Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. Brain-Computer Interfaces. 2017;4(3):155–174.

[13] 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.

[14] Perrin M, Maby E, Daligault S, Bertrand O, Mattout J. Objective and Subjective Evaluation of Online Error Correction During P300-based Spelling. Adv. in Hum.-Comp. Int. 2012;2012:4:4–4:4.

[15] Acqualagna L, Blankertz B. Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology. 2013;124(5):901–908.

[16] Alfeld P. A trivariate clough—tocher scheme for tetrahedral data. Computer Aided Geometric Design. 1984;1(2):169–181.

[17] Bashivan P, Yeasin M, Rish I, Codella N. Learning representations from eeg with deep recurrent-convolutional neural networks. 2016;15.

[18] Maddula RK, Stivers JT, Mousavi M, Ravindran S, Sa VRd. Deep Recurrent Convolutional Neural Networks for Classifying P 300 Bci Signals. In: 2017.

[19] Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for Simplicity: The All Convolutional Net. arXiv:1412.6806 [cs]. 2014.

[20] Cho K et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. en. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, 2014, 1724–1734.

[21] Tieleman T, Hinton G. *Lecture 6.5 - RmsProp: Divide the gradient by a running average of its recent magnitude*. COURSERA: Neural Networks for Machine Learning. 2012.

[22] Simonyan K, Vedaldi A, Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv:1312.6034 [cs]. 2013.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-50

SPATIAL FILTERS FOR AUDITORY EVOKED POTENTIALS TRANSFER BETWEEN DIFFERENT EXPERIMENTAL CONDITIONS

Jan Sosulski[1], Michael Tangermann[1,2]

[1]Brain State Decoding Lab, Cluster of Excellence BrainLinks-BrainTools, Dept. Computer Science, University of Freiburg, Freiburg, Germany
[2]Autonomous Intelligent Systems Lab, Dept. Computer Science, University of Freiburg, Freiburg, Germany

E-mail: { jan.sosulski | michael.tangermann } @blbt.uni-freiburg.de

ABSTRACT: To interact with the brain in closed-loop applications despite low signal-to-noise ratios, brain-computer interfaces can make use of data-driven linear spatial filtering approaches to improve the single-trial classification performance. While the transfer of spatial filters between users and within multiple sessions of the same user under the same experimental paradigm is feasible to some extent, it is unclear, whether changing experimental conditions affects this transferability.

To investigate this question with regards to spatial filters, we evoke event-related potentials by an auditory oddball paradigm under various stimulus onset asynchronies (SOAs). Using four similarity and distance measures, we analyze the between and within subject transferability of xDAWN filters.

We found that the used measures reflect the differences of spatial filters between subjects. Within the same subject, the measures indicate similarity of the spatial filters for almost all SOA conditions. We conclude, that our proposed measures can be used to give indications under which circumstances spatial filters can be transferred.

INTRODUCTION

Brain-computer interfaces (BCIs) are systems which use machine learning to interpret the brain signals of its users in single trial in order to obtain information and possibly act upon this information. Popular applications for BCI are the control of wheelchairs [1], spelling applications for patients [2, 3], stroke rehabilitation [4] and non-clinical applications [5, 6]. One non-invasive way of obtaining the brain signals is to use electroencephalogram (EEG) recordings which are obtained via electrodes placed on the scalp of user.

In order for the brain signals to be recorded by the EEG, they have to travel through the brain, the skull and the skin of the subject. Having low signal amplitudes, they are easily masked by non-neural signal sources as the sensitive electrodes collect all sorts of electric activity, such as power line frequency, muscle activation or currents induced by electrode movements. These interfering electric activities are called artifacts. As a result, the obtained neural signals have generally a poor signal quality, i.e., a bad signal-to-noise ratio. In BCI, one is usually interested only in a specific type of brain signal while background EEG activity and artifacts are to be ignored. Frequently used techniques to deal with these problems are frequency filtering [7], spatial filtering [8, 9] and artifact removal [10]. Applying these methods is oftentimes necessary to extract useful information from the EEG signal.

The aforementioned spatial filtering methods try to find weights on electrodes that minimize the influence of noise while maximizing the brain signal of interest. General purpose approaches such as Laplacian filtering [11] can usually be applied without special considerations. However, subject-specific data-driven spatial filtering methods, e.g., common spatial patterns [8] or SPoC [12] for oscillatory brain signals or xDAWN [9] for event-related potentials (ERPs), usually outperform these general approaches. The drawback of subject-specific methods is that the filters have to be trained on data from the subject, therefore requiring additional training data.

One experiment to elicit auditory ERPs is the simple two-class auditory oddball paradigm, where the subject is asked to attend to rare high-pitched target tones and to ignore frequent low-pitched non-target tones. While each tone stimulus elicits an ERP response, they have different spatio-temporal characteristics depending on the class (i.e., target or non-target) of the played tone. Additionally, the waveform of these ERPs can look differently depending on other stimulation parameters. In this work, we investigate the influence of stimulation onset asynchrony (SOA), which denotes the time between the onset of two successive tone stimuli. The SOAs we examine are in the range of 60 ms to 600 ms. The so-called P300 response is a positive potential occurring approximately 300 ms after a target stimulus over central electrodes. This ERP component can be evoked by oddball tasks. Its delay relative to the stimulus onset has been reported to vary with task difficulty [13], among others. Thus the P300 rise may occur earlier in simple auditory oddball tasks. Besides the P300, other modality-dependent and task-dependent ERP components can be elicited, which have shorter or longer delays compared to

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-50

Figure 1: Averaged ERP waveforms elicited by auditory target stimuli delivered at time t=0 ms for two SOAs. Top: waveforms observed at channel *Cz* (no spatial filter was applied). Bottom: waveform observed after applying an xDAWN spatial filter.

the P300. For multiple stimuli presented at short SOAs, it thus is known, that elicited ERP responses do overlap. The used xDAWN method is specifically suited to reduce this overlap in the filtered ERP responses.

Figure 1 shows examples, how changing the SOA parameter can influence the averaged waveform of the ERP responses recorded at the *Cz* electrode. However, when looking at the same signals filtered using an xDAWN filter, the average ERP responses look more similar. Please note, that in this example one xDAWN filter was determined specifically for each SOA condition based on training data.

Finding solutions to reduce the amount of necessary training data is beneficial in BCI for many reasons. Transfer learning of spatial filters and full classifiers between repeated sessions of one user or even between users has been investigated in this context [14]. For changing experimental conditions, however, it is unclear, if such transfer is feasible. Thus we will focus on the question, how well spatial filters trained under a specific SOA condition can be transferred to a different SOA condition, but within the same user.

To investigate this, we will first propose four measures that quantify transferability. Finally, we will apply these measures to within subject transfers of xDAWN filters obtained under different SOA conditions in an auditory oddball experiment.

MATERIALS AND METHODS

Data-driven spatial filtering algorithms learn a weight matrix W with $N_w \times N_c$ dimensions that is applied to the $N_c \times N_t$-dimensional EEG signal X with N_c channels and N_t number of time samples. The filtered signal

$$S = WX \tag{1}$$

retains $N_w \times N_t$ dimensions, while the number of filters N_w is usually much smaller than the original number of channels N_c. In this work we focus on a single individual spatial filter $\mathbf{w}$ only, such that we retain a weight vector instead of a matrix, i.e., $N_w = 1$. The physiological interpretation of spatial filter weights is not straightforward. As described by Haufe et al. [15], one usually calculates the activation pattern of a filtered signal to support the interpretation. This activation pattern A can be calculated using the channel covariance matrices of the original EEG signal Σ_x and the filtered EEG signal Σ_s

$$A = \Sigma_x W \Sigma_s^{-1}. \tag{2}$$

The covariance matrices are calculated for a time interval of interest. As we only consider a single spatial filter $\mathbf{w}$, we also look at only a single spatial pattern $\mathbf{a}$.

We use the xDAWN algorithm to find spatial filters $\mathbf{w}$ that enhance target ERPs elicited in an auditory oddball paradigm. The xDAWN spatial filter is obtained by solving a generalized eigenvalue problem. Therefore, the resulting spatial filters do not always have the correct—in the sense that the P300 ERP is actually positive—sign and have no longer interpretable amplitudes. We usually chose the spatial filter on the first rank, unless visual inspection suggested that the filters on second or third rank represented the P300 ERP better.

After receiving an ethics vote from the local ethics committee in Freiburg, and after obtaining written informed consent, EEG data from 13 subjects (six female, seven male, mean age: 24.9 years, standard deviations: 5.1 years) was recorded. Within a single session 20 trials (split in four blocks with five trials each) of an auditory oddball paradigm were conducted for three to five different SOA conditions, where one trial consisted of 15 target and 75 non-target stimulus presentations. In total this yields 20 trials for each SOA condition. The order of the stimuli was pseudo-randomized such that at least two non-target stimuli were presented between any two target stimuli. The recorded EEG data was preprocessed for offline analysis by zero phase band-pass filtering in a frequency band from 1.5 Hz to 40 Hz. While this frequency range is atypical for traditional ERP analysis, the usage of very short SOAs down to 60 ms required this higher-frequent upper limit of the low-pass. After frequency filtering, the signal was downsampled from 1000 Hz to 100 Hz. The signal was then epoched from −200 ms to 1000 ms around a stimulus. Each epoch was baseline corrected relative to the interval of −200 ms to 0 ms. We chose not to perform any artifact removal to keep the amount of data comparable between all subjects and to estimate how robust the xDAWN filter is under the influence of potential artifacts. For each subject and SOA condition we trained an individual xDAWN filter in the interval of 0 ms to 600 ms as suggested in [9]. Additionally, we trained an xDAWN filter on a mixture data set

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-50

containing trials from each SOA condition such that its size was equal to the number of trials for the individual SOAs. As one xDAWN filter is obtained per SOA and one for the mixed data set, we finally obtain four to six spatial filters per subject. This yields a total of 66 different xDAWN spatial filters.

On this data we applied several similarity and distance measures that should allow to determine whether a spatial filter can be transferred to other conditions or subjects.

Comparing spatial filter weights:

In order to quantify commonalities or differences between two spatial filters, we can employ a similarity measure. The spatial filters determined by xDAWN are represented as weight vectors that unfortunately are determined only up to sign and amplitude. However, this representation still allows us to use the cosine similarity measure, i.e., the cosine of the angle θ between two weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$:

$$\text{cossim}(\mathbf{w}_1, \mathbf{w}_2) = \cos(\theta) = \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{||\mathbf{w}_1|| \cdot ||\mathbf{w}_2||} \qquad (3)$$

Angles between weight vectors has been proposed for comparing spatial filters before [16, 17]. A cosine similarity of 1 indicates collinearity of the vectors, while 0 indicates orthogonality. Negative values indicate that the vectors point to opposite half spaces. Please note, that the cosine similarity disregards spatial relations between electrodes.

Applied to our setup we expect smaller angles between xDAWN filters within one subject, i.e., $\cos(\theta) \to 1$. For analysis, we calculate pairwise cosine similarities of all possible pairs within the 66 xDAWN filters obtained over subjects and SOA conditions.

Comparing elicited patterns:

Aside from similarity of the actual spatial filters, the similarities of the elicited patterns may be informative. Patterns tend to be more homogeneous than spatial filters and the interesting information is the distribution and location of the pattern on the scalp. Therefore, instead of using the location invariant cosine similarity to compare patterns, we utilize the optimal transport (OT) distance [18] (also known as earth mover's distance or Wasserstein distance). This OT distance tries to find the minimal transport plan or flow F that transforms a distribution A into distribution B. In our case A and B are the elicited spatial patterns.

For our data set we expect smaller OT distances when comparing two patterns obtained for different SOAs within one subject than when comparing two patterns for different subjects.

Comparing ERP waveforms:

We also need to quantify differences between ERP waveforms, that result from spatially filtered EEG signals. As amplitude values generally are not comparable anymore after the application of xDAWN filters, we first rescale each filtered waveform to a standard deviation of one.

To lower the influence of noise prior to applying a waveform-specific distance metric, we reduce the temporal resolution of the waveforms to six predefined time intervals by averaging over multiple samples. Furthermore, we average the waveform features over trials of the same condition and subject. Thus, we obtain six amplitude features per SOA and subject.

Then the difference between the amplitude features of two filtered ERP waveforms $\mathbf{s}_1$ and $\mathbf{s}_2$ is quantified by using the Euclidean distance.

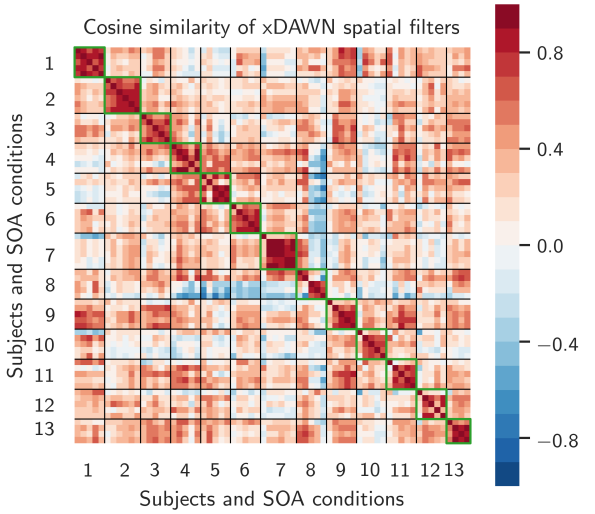


Figure 2: Matrix of all pairwise cosine similarities between xDAWN filters trained for each subject and SOA condition. The thick black lines separate subjects from each other. Note that the green-bordered blocks on the main diagonal correspond to within subject filter similarities under different SOA conditions.

Classification performance:

We have limited our analysis to a single spatial filter per SOA condition and subject. However, if a spatial filter $\mathbf{w}$ captures the relevant features of the ERP responses evoked by the experimental paradigm, then applying it to a data set X should nevertheless lead to a target/non-target classification performance well above chance level. We will make use of a linear discriminant analysis classifier with covariance shrinkage regularization (rLDA) to estimate the class discriminative information provided by a filter, given by the area under receiver-operator characteristic curve (AUC). As a single xDAWN component $\mathbf{w}$ may not capture the full ERP information, we generally must expect a slight performance decrease by applying a single spatially filter compared to using the high-dimensional unfiltered signal [9]. To quantify this performance drop, we trained an rLDA classifier on each data set, first without applying any spatial filtering and then (consecutively) using each of the derived 66 xDAWN spatial filters. Utilizing a 4-fold chronological cross-validation scheme we can report the estimated performance loss $\Delta\text{AUC} = \text{AUC}_{\text{filt}} - \text{AUC}_{\text{nofilt}}$ for each spatial filter when transferred to each data set.

Successful transfer of filters should be expressed by an only small drop in performance.

RESULTS

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-50

Table 1: List of SOA conditions (in ms) used for each subject.

Subject	SOAs in ms				
1	60	175	200	460	—
2	60	111	200	218	446
3	60	123	175	380	—
4	60	193	286	506	—
5	60	208	226	508	—
6	60	175	296	474	—
7	60	226	235	518	596
8	60	177	256	446	—
9	60	180	400	447	—
10	60	220	325	402	—
11	60	179	194	513	—
12	60	193	220	600	—
13	60	211	518	—	—

First, we utilize the cosine similarity to investigate the similarity between xDAWN filters. In Figure 2, the cosine similarities between all spatial filters calculated for each subject and SOA condition are visualized. Black lines indicate the boundary between subjects. Within one subject, the SOA conditions on which individual filters were trained are in ascending order from top to bottom / left to right. The concrete SOAs for each subject are shown in Table 1. The last filter of each subject was trained on a mixed data set of all SOAs with an equivalent number of trials.

Intense red entries in the diagonal blocks indicate, that the similarities between spatial filters obtained from within a single subject are generally high. Specifically, they are considerably larger than filter similarities between subjects as indicated by the off-diagonal blocks. Within subjects, the fastest SOA of 60 ms seems to take a special role: its spatial filters are substantially different from the ones calculated under slower SOA conditions for most subjects. Additionally, for subjects 4, 5, and 12, there exist some SOA conditions which do not produce a spatial filter comparable to the rest. Interestingly, subject 8 produces spatial filters that are specific for the two fast SOAs and different spatial filters specific for the two slower SOAs. Additionally, the last row/column of each subject generally shows consistent similarities to other filters for this subject (data not shown). This observation is in line with our expectations, as this filter had been trained on the mixed data set consisting of all SOAs of the subject.

The histograms in Figure 3 provide a closer look at the within subject vs. between subject cosine similarities of the xDAWN filters. As expected, we can observe higher similarities of filters within a subject (blue color, median similarity 0.695) than across subjects (orange color, median similarity of 0.218). Introspection into the data revealed that values close to zero cosine similarity were caused mostly by filters of the very fast 60 ms SOA.

The optimal transport distance between patterns provides a novel view upon the similarity of xDAWN components. The histogram in Figure 4 shows, that the differences of the distance of patterns between and within subjects are not as emphasized as for cosine similarities which acts on

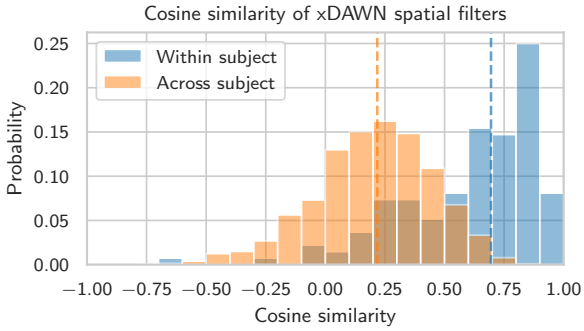


Figure 3: Distribution of the similarities between spatial filters within and across subjects. Dashed lines indicate median values of 0.218 (between subjects) and 0.695 (within subjects).

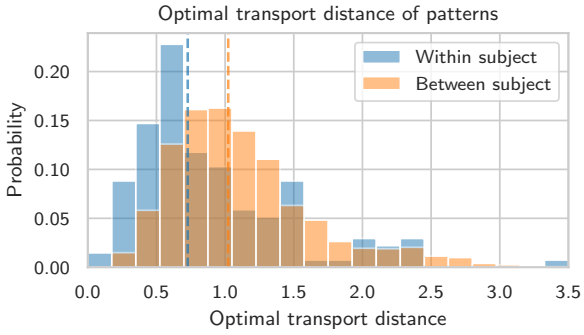


Figure 4: Distribution of the optimal transport distances between spatial patterns within and across subjects. Dashed lines describe median values of 0.728 (within subjects) and 1.023 (between subjects).

the filters. This is due to the fact, that the xDAWN algorithm has almost always found a component representing the P300 evoked potential for each data set (although we had to manually select this component in few cases). As the pattern topology does not differ greatly between subjects and between conditions, we obtain small optimal transport distances. The peak in the within subject histogram values around distances >1.5 are caused by patterns that do not show a typical P300 topology. Overall we conclude, that looking at the pattern alone is not informative enough to describe (dis)similarities in the ERP domain.

To determine the general variability of ERP waveforms and how it affects the Euclidean distance between two ERPs, we calculated the Euclidean distances for three different settings: (1) how ERPs differ between subjects (including different SOA conditions), (2) how ERPs differ within subject under different SOA conditions and (3) how changing the spatial filter but not the underlying data within subjects but across SOA conditions affects the ERP waveform. Hereinafter, we will call this last setting *within subject filter transfer*.

Figure 5 shows the histograms for the three aforementioned settings. Interestingly, the distances resulting from within subject filter transfer are the lowest on average. This indicates that the spatial filters trained on different

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-50

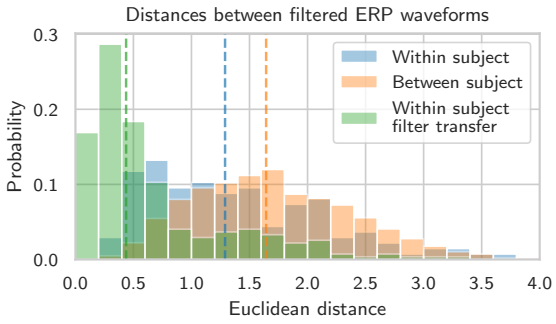


Figure 5: Distribution of the Euclidean distances between ERP waveforms. Blue shows the within subject distances, while orange shows between subject distances. The distances resulting from within subject filter transfer are shown in green. Median values (left to right): 0.437, 1.289, 1.643.

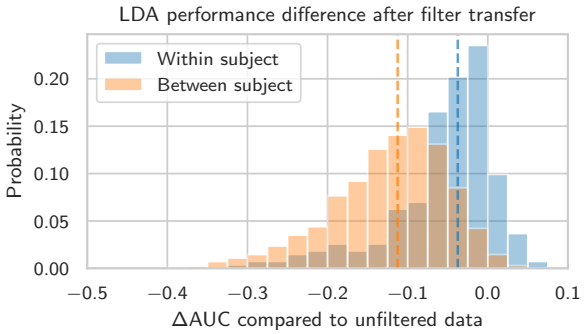


Figure 6: Distribution of ΔAUC for transferring spatial filters within a subject and across subjects. Dashed lines indicate the median values at -0.113, -0.037.

SOAs of one subject can be re-used for other SOA conditions of the same subject without affecting the evoked ERP waveform drastically.

Finally, we quantified how strong the rLDA classifier performance of the xDAWN-filtered data deteriorates when applying a spatial filter trained on a different data set. As shown in Figure 6, the performance losses (provided as ΔAUC values) are much larger when spatial filters from a *different* subject had been used, than when using spatial filters from the same subject. This fits with our previous observations, that spatial filters derived within a subject but across different SOAs still are very similar.

An example for within subject filter transfer is given in Figure 7. We can see that most xDAWN filtered ERP responses look highly similar regardless of which spatial filter we use. Only the filter for SOA 60 seems to extract a different ERP waveform—it generally prolongs the latencies of the ERPs. Applying the filter trained on SOA 60 to the data of SOA 513 even eliminates the P300 ERP completely. This indicates that the filter extracted for this extremely fast SOA value cannot be transferred to the other SOA conditions.

As a contrast, in Figure 8 it is shown how the data recorded for the first four SOA conditions of subject 11 changes when using spatial filters obtained from subject 13. The ERP curves look very different compared

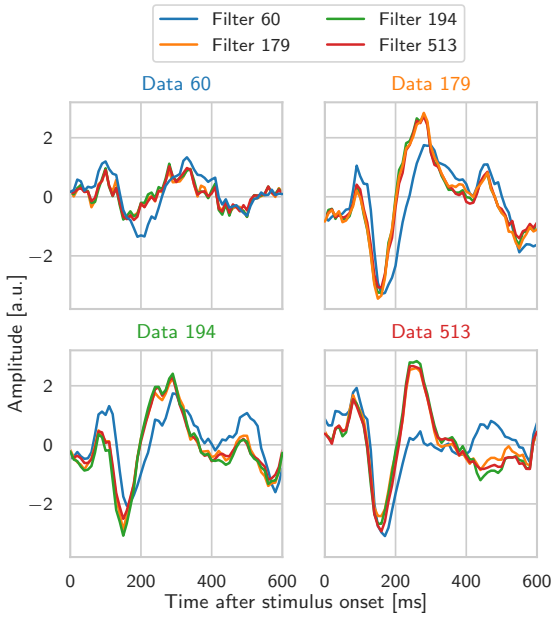


Figure 7: ERP waveforms as a result of filter transfer within subject 11. Each subplot shows a data set on the SOA given in the title. The different colors indicate which spatial filter was applied to the data set. If title and line have the same color, this means that the spatial filter trained specifically on this data set was used to create the filtered ERP signal.

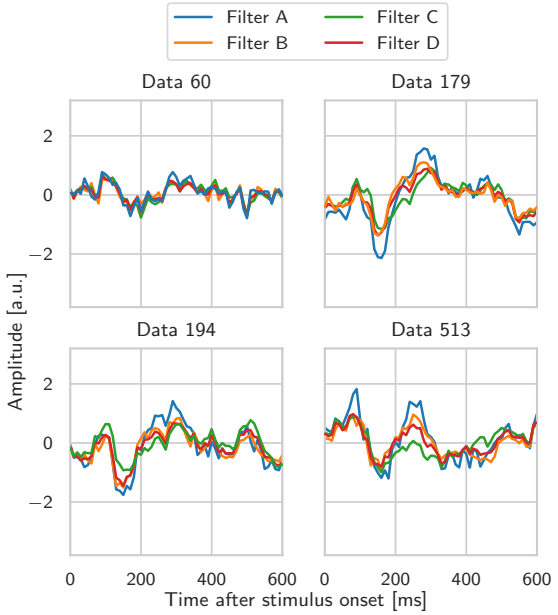


Figure 8: Applied filter transfer from subject 13 to subject 11. Each subplot shows a data set of subject 11 on the SOA given in the title.

to the within subject filter transfer case. Additionally, the difference of ERPs for one data set is much more pronounced by changing the spatial filter.

DISCUSSION AND CONCLUSION

Motivated by the widespread use of data-driven spatial

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-50

filter methods in the field of BCI, we have described methods to quantify similarities and differences of spatial filters and have applied it to xDAWN filters in the ERP domain. Making use of these methods, we could make a novel contribution to the field by showing that the transfer of spatial filters is feasible for spatial filters trained under different SOA conditions but within one subject—at least if the SOAs are not prohibitively fast. In this context we could report evidence that using solely spatial patterns is not sufficiently informative to characterize spatial components obtained under ERP paradigms.

Our finding may pave the way for a more flexible parameterization of ERP paradigms over multiple sessions of a subject, as it reduces the effort of recording training data for each SOA condition separately in order to get a specialized filter for each condition. An empirical question for future research is, how well this within subject transfer performs if experimental conditions other than SOA are changed, and if the observed successful transfer of filters will be possible also for session-to-session transfer within subjects and between conditions.

In future work, the measures for filter comparison will be examined whether they are applicable also with other spatial filter algorithms and under different experimental paradigms, e.g. CSP, SPoC or ICA in mental imagery tasks or whether there exist measures that can even better reflect transferability properties. Furthermore, an automatic framework that can decide for few data points obtained under varying conditions, whether the resulting spatial filters are transferable or not could help to reduce unnecessary calibration effort.

ACKNOWLEDGMENTS

This work was supported partly by the Cluster of Excellence BrainLinks-BrainTools funded by the German Research Foundation (DFG), grant number EXC 1086, by the DFG project SuitAble (TA 1258/1-1) and by the state of Baden-Württemberg, Germany, through bwHPC and the German Research Foundation (DFG) through grant no INST 39/963-1 FUGG.

REFERENCES

[1] Wolpaw J, Wolpaw E. Brain-computer interfaces: Principles and practice. Oxford University Press, USA (2012).

[2] 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.

[3] Schreuder M, Rost T, Tangermann M. Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. Frontiers in Neuroscience. 2011;5.

[4] Ang KK, Guan C. Brain-computer interface in stroke rehabilitation. 2013.

[5] Blankertz B, Tangermann M, Vidaurre C, *et al.* The Berlin brain–computer interface: Non-medical uses of BCI technology. Frontiers in neuroscience. 2010;4:198.

[6] Krauledat M, Grzeska K, Sagebaum M, *et al.* Playing pinball with non-invasive BCI. In: Advances in neural information processing systems. 2009, 1641–1648.

[7] Nitschke JB, Miller GA, Cook EW. Digital filtering in EEG/ERP analysis: Some technical and empirical comparisons. Behavior Research Methods, Instruments, & Computers. 1998;30(1):54–67.

[8] Blankertz B, Kawanabe M, Tomioka R, Hohlefeld F, Müller K-R, Nikulin VV. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In: Advances in neural information processing systems. 2008, 113–120.

[9] Rivet B, Souloumiac A, Attina V, Gibert G. xDAWN algorithm to enhance evoked potentials: Application to brain–computer interface. IEEE Transactions on Biomedical Engineering. 2009;56(8):2035–2043.

[10] Winkler I, Brandl S, Horn F, Waldburger E, Allefeld C, Tangermann M. Robust artifactual independent component classification for BCI practitioners. Journal of neural engineering. 2014;11(3):035013.

[11] McFarland DJ, McCane LM, David SV, Wolpaw JR. Spatial filter selection for EEG-based communication. Electroencephalography and Clinical Neurophysiology. 1997;103(3):386–394.

[12] Dähne S, Meinecke FC, Haufe S, *et al.* SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. NeuroImage. 2014;86:111–122.

[13] Polich J. Task difficulty, probability, and interstimulus interval as determinants of P300 from auditory stimuli. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section. 1987;68(4):311–320.

[14] Kindermans P-J, Tangermann M, Müller K-R, Schrauwen B. Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. Journal of neural engineering. 2014;11(3):035005.

[15] Haufe S, Meinecke F, Görgen K, *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage. 2014;87(0):96 –110.

[16] Krauledat M, Tangermann M, Blankertz B, Müller K-R. Towards zero training for brain-computer interfacing. PloS one. 2008;3(8):e2967.

[17] Xu Y, Wei Q, Zhang H, *et al.* Transfer learning based on regularized common spatial patterns using cosine similarities of spatial filters for motor-imagery BCI. Journal of Circuits, Systems and Computers. 2018:1950123.

[18] Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. International journal of computer vision. 2000;40(2):99–121.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-51

HOW SIMILAR ARE THE NEURAL PATTERNS WHEN OBSERVING GRASPING HAND POSTURES TO THE BEHAVIORAL PATTERNS WHEN EXECUTING THE GRASP?

Andreea I. Sburlea [1], Gernot R. Müller-Putz[1]

[1] Institute of Neural Engineering, Graz University of Technology, Graz, Austria

E-mail: {andreea.sburlea, gernot.mueller}@tugraz.at

ABSTRACT: It is well-known that human neural representations of grasping movement observation and execution share similarities. However, it remains unclear to what extent neural patterns of movement observation relate with behavioral covariates of the movement, such as muscle or kinematic activity, or with categorical models that describe hand posture or object properties. In this study, we explored the relation between EEG neural representation of observing a large variety of hand-object interaction images and two representations of the observed movement's execution: the behavioral covariates and categorical models. We found that the EEG representation of the observation phase was correlated with the muscle representation during the execution most strongly in the movement holding phase. Furthermore, we found similarities with the categorical model that reflects the shape and the size of the object. With these findings we gain a joint understanding of the relation between movement observation and execution and a mean to facilitate an intuitive control of neuroprostheses in motor impaired individuals.

INTRODUCTION

Movement is the main way we can interact with the world. The motor system is involved in all types of movements, including speech, walking, reaching, grasping and many others. In particular, the ability to perform grasping movements plays an essential role in independent everyday activities. Spinal cord injury (SCI) or stroke can cause motor impairments that hinder the execution of these movements. Therefore, great effort is devoted towards the control of robotic limbs or neuroprostheses in people with motor impairments [1], [2]. Since these individuals cannot execute the desired movements, they can instead attempt or imagine the movement. When performing these tasks similar brain activations have been found as in an actual movement execution task, leading to the discovery of the mirror-neuron system [3], [4]. The existence of the neural mirroring activity and the observation-execution matching has been first reported in macaques [5], [6], and later in humans [7], [8], [9]. In humans, the mirroring system is thought to play a role in understanding others' actions [10] and in a range of social cognitive processes, such as action planning or anticipation [11], [12], perceiving intention [13], [14], [15], imitation [16], or in other social interactions [10], [17]. Moreover, the observation of movements or hand-object interactions, even in static pictures has been shown to involve the motor and visuo-motor areas [4], [18].

In individuals with SCI, we lack information about muscle activations during movement execution and for the control of neuroprostheses we can rely on the EEG correlates of the imagination or attempt of hand movements [19]. We have recently shown that brain activity shares similar activation with the muscle activity during grasping and holding phases of the movement [20]. While observation and execution also share similar brain activations, the relation between the brain patterns associated with the observation of hand postures and the actual muscle components during movement execution remains unclear. We believe that a better understanding of the relation between these movement stages and their covariates could lead to an intuitive control of neuroprostheses or robotic limbs.

Explaining how behavior is implemented by neural mechanisms is a challenging task. One possibility is that tackling both of these challenges simultaneously may be more tractable than addressing each separately at its own level of analysis [21]. Hence, it is essential to establish a reliable relation between observation and execution in terms of both neural and behavioral representations.

In this study, we explored the similarity between the neural representation while observing static images of hand-object grasping interactions and the associated behavioral representation (in terms of muscle and kinematic activations) while executing the observed grasping interaction. We also categorized the grasps based on the shape and size of the object, position of the thumb relative to the palm, in addition to the more conventional grasp type categorization in power, pincer and intermediate, and then built three different categorical models. We applied representational similarity analysis (RSA) [22], [23] to investigate the similarities among the effects of the neural, behavioral (muscle and kinematic) representations and categorical models at different time stages of grasping movements.

MATERIALS AND METHODS

A previously recorded dataset has been used in this study [20]. The dataset contains simultaneously acquired neural (EEG) and behavioral (muscle and kinematic) data of thirty one participants, in a task that involves observation and

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-51

execution of different grasping movements. The experimental protocol was approved by the ethical committee of Medical University of Graz (approval number: 29–352 ex16/17), and all subjects gave written informed consent before participating in the experiment. Fig. 1A illustrates the structure of the experimental protocol.

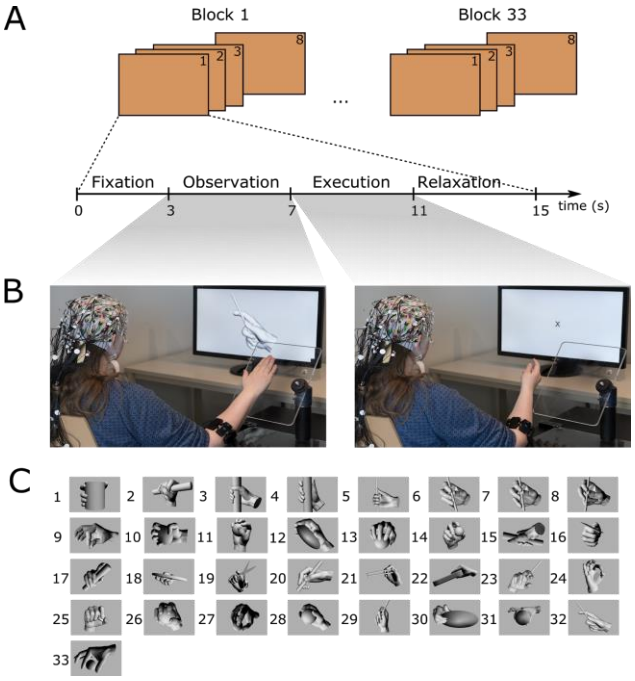


Figure 1: (**A**) Experimental protocol. Each of the 33 blocks contained eight consecutive repetitions (trials) of the same grasp. Each trial had four phases: fixation (three seconds long), observation (four seconds long), execution (four seconds long) and relaxation (four seconds long). (**B**) Experimental setup. Photos of one participant during the observation and execution phases, and the materials used during recording. (**C**) Pictograms of the grasping movements.

During the fixation period, participants were instructed to focus their gaze on a cross located in the middle of the screen and avoid eye movements for three seconds. Next, during the observation phase which lasted four seconds, participants were presented with a static image showing a hand in a final grasping position together with the grasped object. During this observation phase, participants could move their eyes and observe the shape and size of the object, as well as the position of the fingers and the overall shape of the hand. During the fixation and observation phases, as well as during the break, participants rested their hand on a transparent, custom-made plexiglass hand support, shown in Fig. 1B. We built this hand support to reduce hand fatigue and recorded the resting position using the Leap Motion optical tracking device. During the execution phase which was also four seconds long, participants were instructed to focus their gaze on the "x" symbol located in the middle of the screen and to lift their hand from the resting support, move it slightly to the left and perform the grasping movement that they had observed during the previous phase. The acquisition of the blocks was divided in 4 parts. The first three parts contained 8 blocks and lasted 16 minutes each, while the last part contained 9 blocks. After each part, we offered longer breaks to the participants. The total duration of one recording was, on average, an hour and a half. The participant shown in Fig. 1B gave her informed consent for the photo to be made available in the conference publication. Fig. 1C shows the pictograms of the 33 grasping movements with their ordinal numbers. The order of the grasping conditions (blocks) was randomized among participants.

In the dataset, EEG and EOG data was recorded using a 64-channel ActiCap System with two BrainAmp amplifiers (BrainProducts, Germany). The ground sensor was placed on AFz and the reference sensor on the right mastoid. To record the muscle activity, we used a Myo armband (Thalmic Labs Inc., Canada). The armband was located on the right arm close to the elbow, above the extrinsic hand muscles. To record the kinematics related to the grasping movements, we used the Leap Motion controller (Leap Motion Inc., CA, USA).

For all data preprocessing and analyses, we used Matlab R2016b (Mathworks, Inc. USA). EEG data was first filtered using a Butterworth fourth-order, zero-phase, band-pass filter between 0.1–40 Hz and then downsampled to 100 Hz. We rejected the trials in which the task was incorrectly performed (e.g., start of movement before the associated cue). From the EEG data we extracted the first seven seconds associated with the fixation and observation period and performed a similar cleaning processing pipeline as described in [20]. Next, we computed the time-frequency representation using Morlet wavelets with a resolution of 0.5 Hz in between 0.1 and 40 Hz. To compute the ERD/S patterns, we used the fixation period (1-3s) as baseline.

The eight EMG data channels were processed using Hilbert transform, standardized using *z*-score and, finally, the envelope of the data was computed. From the signals recorded using the Leap Motion controller, we used nineteen joint angles for the rest of the analysis. These joint angles correspond to the five proximal (one artificially created for the thumb and four for the other fingers), five intermediate and five distal joints of each finger, as well as the four in-between fingers joints. Given the high degree of correlation between some of the explored joints, we applied principal component analysis (PCA) to reduce the dimensionality of the kinematic data. We retained for each subject the first five components, which accounted, on average, for more than 95% of the data variance.

For each of the three acquisition modalities (EEG, EMG and joint angles), we averaged single repetitions of the same grasp condition both within subject and between subjects leading to one group-level representation per grasp condition. For the averaging of the joint angles, we used the circular mean, as implemented in the CircStat toolbox [24].

For the extraction of the behavioral features of interest we defined three temporal windows in the movement execution segment. The windows were consecutive and each window was 500 ms long, which is in line with the latencies of the grasping phases previously reported in the literature [25]. The first window was associated with the reaching and pre-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-51

shaping of the hand, and it started from the movement onset. The second window was associated with the finalization of the grasping movement. The third window contained information about the holding position of the grasping. From the three windows we extracted the features of interest from the two behavioral data acquisition modalities: EMG and joint angles. For the neural data, we extracted the 4-second window corresponding to the observation phase.

Our strategy consisted in comparing the neural information from the observation period with the behavioral information from the three windows of the execution period.

For the EEG pattern extraction, we implemented a searchlight technique to extract ERD/S patterns simultaneously at different spatial locations and frequency bands. We present the searchlight approach in more detail in [20]. For the analysis, we defined 31 centroids in the channel space and 26 in the frequency space. One neighborhood, in either of the spaces, had five members: one centroid and four equidistant neighbors, and had 2 members in common with any other neighbourhood. The two-dimensional window defined by the channel and frequency neighborhoods was slid across the two dimensions. Each frequency neighborhood corresponded to a 2 Hz-wide band and had an overlapping step of 0.5 Hz.

We applied the searchlight technique on the entire 4-second long window from the observation segment. All the information within a two-dimensional window was concatenated into a vector. This vector defined the EEG pattern of activity for the region with a given centroid in space and in frequency. For the EMG envelope and for the kinematic joint angles, the behavioral patterns resulted from the concatenation of all channels (components) of data over each of the three inspected temporal windows.

For each of the three temporal windows, we compared the behavioral patterns of different grasping conditions with the neural patterns of the observation period. We used 1-*r* (where *r* is the Pearson correlation) as the distance metric between the patterns of different conditions. Next, we ranked and scaled the distances between 0 and 1, and computed the representational dissimilarity matrix (RDM) [23] between pairs of grasping conditions. In addition to the neural and behavioral representations, we also implemented categorical models. These models were built from previous categorizations based on the type of grasp, position of the thumb relative to the palm, defined in the largest pre-existing taxonomy of static and stable grasps [26], and on the shape and size of the object [20]. The RDMs of these categorical models were binary representations, which contained a 0 for each pair of stimuli falling into the same category and a 1 for each pair of stimuli falling into different categories. We define as reference RDM the neural representation during the observation phase and as candidate RDMs both the behavioral representations during the execution phase and the categorical models.

To evaluate the consistency of our obtained RSA effects we conducted a bootstrap analysis with 500 iterations. We performed the bootstrap analysis on a pool of 31 different samples corresponding to our subject number. In each iteration, we replaced the 31 samples with other randomly pulled 31 samples. Through bootstrapping, we could estimate the confidence interval, as the variance of the mean distance between the reference and candidate RDMs in our population of 31 subjects. We report the 95% confidence interval (CI) of the procedure together with the obtained mean effect.

RESULTS

We computed the reference representations based on the searchlight analysis of the time-frequency representation of the EEG data during the observation period. We then computed the representational similarity between the reference (EEG) RDM and the behavioral (EMG and joint angles) representations in the three time windows of interest of the movement execution period: hand pre-shaping (0 – 0.5 s), reaching of the final grasping posture (0.5 – 1 s) and holding (1 – 1.5s). We also computed the RSA between the reference representation and the three categorical models.

Fig. 2 describes, at a population level, the results of the representational similarity analysis between the reference neural representation (during the observation period) and the candidate behavioral representations (described by muscle and joint angles information) and the categorical models (type of grasp, position of the thumb relative to the palm and shape of the object). We observed that the reference neural representation during movement observation is more similar (smaller 1-*r* distance) to the muscle than to the kinematic representation during movement execution. This similarity becomes even stronger during the holding phase of the movement, in which the hand maintains the final grasping posture (1-*r* = 0.69, 95% CI: [0.66, 0.72]) in the mu and low beta frequency bands in parietal and occipital areas. In the previous movement phases the values of similarity were: 1-*r* = 0.74 during hand pre-shaping and 1-*r* = 0.71 during the reaching period of the final grasping posture. Moreover, from the categorical models, the reference representation is more similar to the model that encodes the shape and the size of the object also in mu and beta frequency bands over parieto-occipital brain regions (1-*r* = 0.79, 95% CI: [0.77. 0.8]) compared to the other categorical models.

In Fig. 3, we focused on the previously observed effects and assessed their significance at the population level. Fig. 3A shows the topographical representation of the similarity effects relative to the behavioral representations (top and middle) and categorical model Object shape (bottom). Next, we conducted a bootstrapping analysis and we show in Fig. 3B the effects and their 95% confidence interval.

DISCUSSION

In this exploratory study, we investigated the similarities among neural, behavioral and categorical representations of observing and executing a large repertoire of unique grasping movements [26]. Using a searchlight approach, we extracted EEG patterns from the observation period at different frequencies and brain areas, and compared them with behavioral (muscle and kinematic) representations of the executed movements, as well as with three categorical

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-51

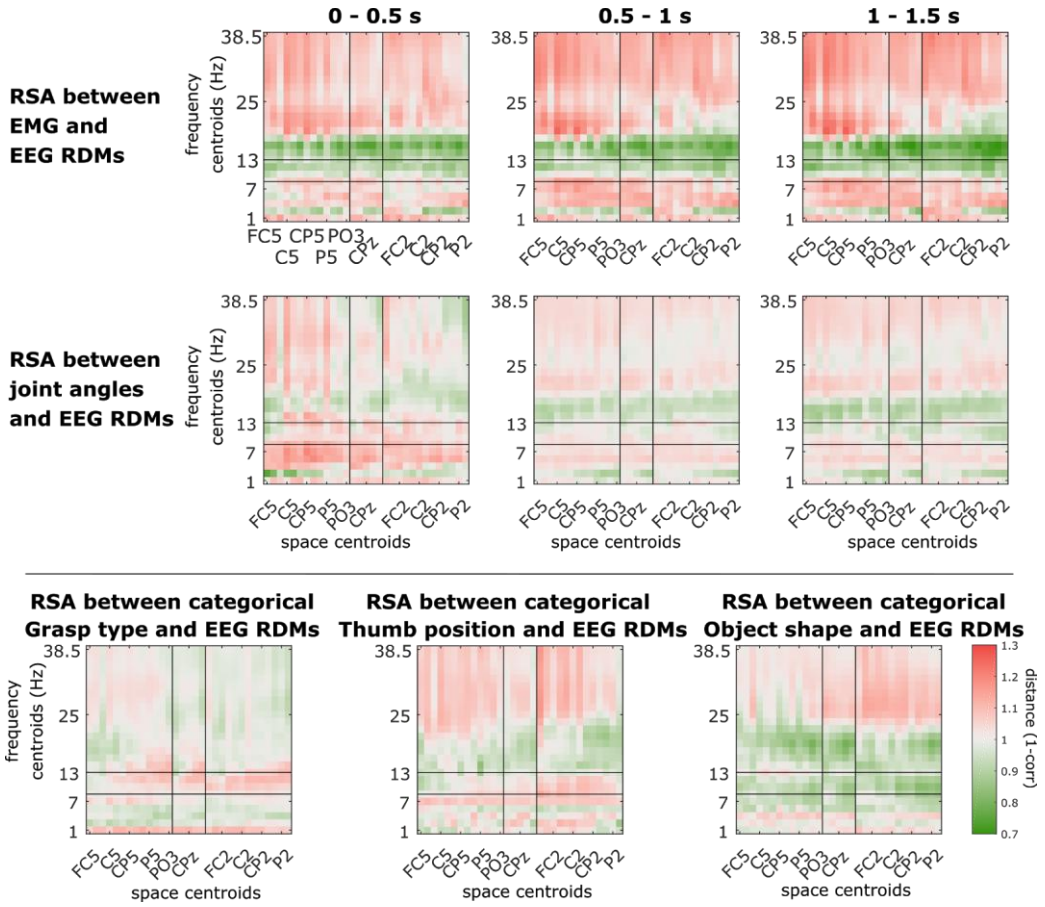


Figure 2. Representational similarity analysis between the candidate representations and the reference representations extracted through a searchlight implementation. The vertical black lines mark the area covered by the midline centroids. The horizontal lines mark the mu frequency band. The space centroids are sorted from periphery-to-midline and anterior-to-posterior.

models that describe the type of grasp, the position of the thumb during grasping and the intrinsic properties of the object (shape and size).

We found that the parietal and occipital areas encode information about grasping movements during the observation period. Moreover, we found that the muscle representation during the holding phase of the movement showed similarities with the EEG representation during the observation period in the mu and low beta frequencies. We also found similarities between the EEG representation during observation and the categorical model that describes the shape and the size of the objects. These similarities were more pronounced in the contralateral parietal and occipital regions in the mu and beta frequency bands. EEG activation in parietal and occipital brain regions in mu and beta frequency bands have been reported also by [27] when observing hand-object interactions. According to the ideomotor theory [28], [29], actions are represented by their perceived effects. Hence, external or internal events can trigger the corresponding actions.

During the observation of grasping hand postures, the neural processes related to the concept of movement result from peripheral sensations generated by previous knowledge about how the movement would be executed, plus proprioception [29]. In our study, all the observed grasping postures were familiar to the participants. It has been shown that attention to kinesthetic perception of the movement also triggers activations of motor areas in similar frequency bands [30]. Desynchronization of the mu and beta rhythms is a commonly used measure of motor cortex activation [31], [32] in movement observation tasks. Mu desynchronization has been identified as a possible measure of the mirroring system [33]. We also found activations in the frequency range of the mu band (8-13 Hz) both in the parietal and occipital brain areas. Indeed, the two activations originate from distinct processes, the first being related to the motor planning and the second to the visual processing. Furthermore, Pfurtscheller et al. [31] made a division into lower and upper alpha (mu) bands. The upper alpha band is shown to be associated to stimulus identification and response preparation in a movement task. Indeed, our findings show that in this frequency band the ERD/S representation during observation has similarities with the muscle representation during execution. In particular, we found stronger similarities during the holding phase of the movement, which indicates that when visually processing

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-51

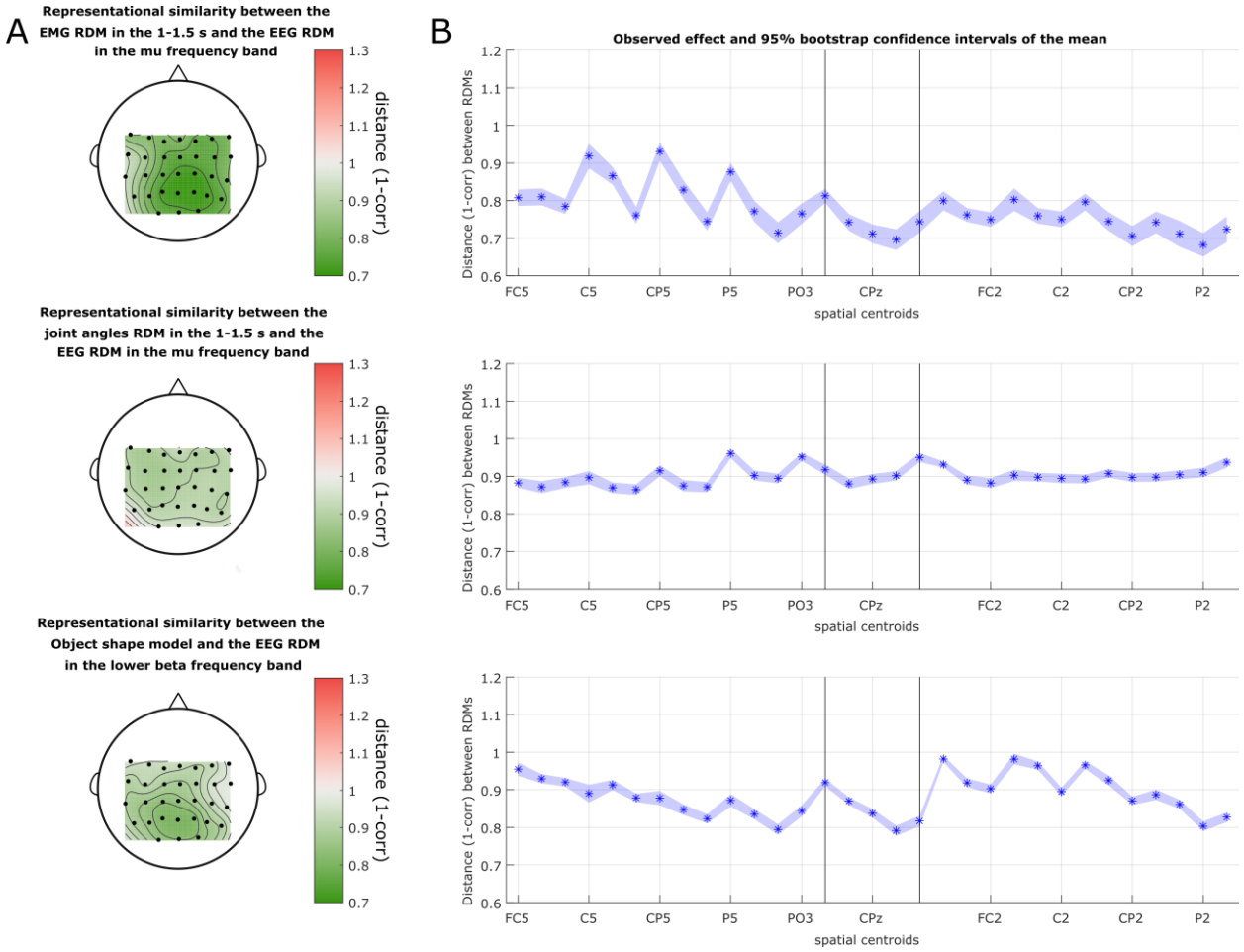


Figure 3. (A). Topographical RSA in the specified frequency bands. On the top, the dissimilarity w.r.t. the EMG representation in the time window between 1–1.5 s. The middle topographical plot shows the dissimilarity w.r.t. the joint angles representation in the time window 1–1.5 s. The bottom topographical plot shows the dissimilarity w.r.t. the categorical Object shape model. (B) Bootstrapping results of the RSA for the three representations shown on the left (A) panel. Centro-parietal and occipital brain regions show smaller distances (higher correlations) between the reference (EEG) representation and the candidate (behavioral EMG - top left and categorical Object shape – top right) representations. The shaded area indicates the 95% confidence interval computed after 500 bootstrapping iterations.

the hand-object grasping interaction, we focus on the final grasping posture.

As we have already observed before in [20], the kinematic representations based on joint angles were similar to the EMG representations; however, the similarity between the kinematic representation and the EEG representation during the observation period was lower, indicating that the kinematic representations may contain additional grasping information that is not encoded in the EEG brain patterns related with visual processing or motor planning.

Concerning the categorical models, we found that the model describing the type of grasp is the least correlated with the EEG representation during the observation of grasping hand postures, the model describing the thumb position during grasping movements showed slight similarities with the EEG representation during observation on central and ipsilateral regions in the low beta frequencies. Furtheremore, the model based on the object's shape showed the largest similarities (smallest distances) to the EEG representation during the observation period.

These similarities were found in occipital areas in the low beta frequency range. The large similarities in both mu and beta bands with the Object shape model indicate that during the observation phase there are neural processes related with visual processing (potentially due to a sequential search [34]) as well as motor planning or imagination of movement.

CONCLUSION

To sum up, our findings indicate that the neural representation of observing images of hand-object grasping interactions shares similarities with the muscle representation during the grasping posture holding phase as well as with the model that encodes the information about intrinsic properties of the object, such as shape and size. These findings are of particular relevance for the control of neuroprostheses based on EEG signals and shed some light on the joint relation between neural processes such as planning or imagination (inevitable when observing a hand-object interaction) and the actual behavioral representations

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-51

associated with grasping execution.

ACKNOWLEDGEMENTS

Authors want to thank Filip Melinscak and Catarina Lopes Dias for carefully reading the manuscript. This work was supported by the ERC Consolidator Grant ERC-681231, Feel Your Reach.

REFERENCES

[1] Taylor, D. M., Tillery, S. I. H. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science* 296, 1829–1832 (2002).

[2] Müller-Putz, G. R., Scherer, R., Pfurtscheller, G. & Rupp, R. EEG-based neuroprosthesis control: a step towards clinical practice. *Neurosci. Lett.* 382, 169–174 (2005).

[3] Rizzolatti, G. & Craighero, L. THE MIRROR-NEURON SYSTEM. *Annu. Rev. Neurosci.* 27, 169–192 (2004).

[4] Craighero, L., Bello, A., Fadiga, L. & Rizzolatti, G. Hand action preparation influences the responses to hand pictures. *Neuropsychologia* 40, 492–502 (2002).

[5] di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180 (1992).

[6] Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. Action recognition in the premotor cortex. *Brain* 119 (Pt 2), 593–609 (1996).

[7] Hari, R. *et al.* Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences* 95, 15061–15065 (1998).

[8] Iacoboni, M. *et al.* Cortical mechanisms of human imitation. *Science* 286, 2526–2528 (1999).

[9] Lepage, J.-F. & Théoret, H. EEG evidence for the presence of an action observation-execution matching system in children. *Eur. J. Neurosci.* 23, 2505–2510 (2006).

[10] Rizzolatti, G., Fogassi, L. & Gallese, V. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670 (2001).

[11] Rizzolatti, G. & Luppino, G. The cortical motor system. *Neuron* 31, 889–901 (2001).

[12] Kilner, J., Hamilton, A. F. de C. & Blakemore, S.-J. Interference effect of observed human movement on action is due to velocity profile of biological motion. *Soc. Neurosci.* 2, 158–166 (2007).

[13] Fabbri-Destro, M. & Rizzolatti, G. Mirror neurons and mirror systems in monkeys and humans. *Physiology* 23, 171–179 (2008).

[14] Ortigue, S., Sinigaglia, C., Rizzolatti, G. & Grafton, S. T. Understanding actions of others: the electrodynamics of the left and right hemispheres. A high-density EEG neuroimaging study. *PLoS One* 5, e12160 (2010).

[15] Rizzolatti, G. & Sinigaglia, C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat. Rev. Neurosci.* 11, 264–274 (2010).

[16] Iacoboni, M. Neurobiology of imitation. *Curr. Opin. Neurobiol.* 19, 661–665 (2009).

[17] Rizzolatti, G. & Arbib, M. A. Language within our grasp. *Trends Neurosci.* 21, 188–194 (1998).

[18] Johnson-Frey, S. H. *et al.* Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron* 39, 1053–1058 (2003).

[19] Müller-Putz, G. R., et al. "From classic motor imagery to complex movement intention decoding: The noninvasive Graz-BCI approach." *Progress in brain research.* Vol. 228. Elsevier, 2016. 39-70.

[20] Sburlea, A. I. & Müller-Putz, G. R. Exploring representations of human grasping in neural, muscle and kinematic signals. *Sci. Rep.* 8, 16669 (2018).

[21] Turner, B. M., Forstmann, B. U. & Steyvers, M. A Tutorial on Joint Modeling. in *Computational Approaches to Cognition and Perception* 13–37 (2019).

[22] Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4 (2008).

[23] Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* 10, e1003553 (2014).

[24] Berens, P. CircStat: A MATLAB Toolbox for Circular Statistics. *J. Stat. Softw.* 31, (2009).

[25] Pellegrino, J. W., Klatzky, R. L. & McCloskey, B. P. Timecourse of Preshaping for Functional Responses to Objects. *J. Mot. Behav.* 21, 307–316 (1989).

[26] Feix, T., Romero, J., Schmiedmayer, H.-B., Dollar, A. M. & Kragic, D. The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions on Human-Machine Systems* 46, 66–77 (2016).

[27] Wriessnegger, S. C., et al. "Watching object related movements modulates mirror-like activity in parietal brain regions." Clinical neurophysiology 124.8 (2013): 1596-1604.

[28] Greenwald, A. G. Sensory feedback mechanisms in performance control: with special reference to the ideo-motor mechanism. *Psychol. Rev.* 77, 73–99 (1970).

[29] James, W. *The principles of psychology, Vol I*. (1890).

[30] Melinscak, F., Montesano, L. & Minguez, J. Asynchronous detection of kinesthetic attention during mobilization of lower limbs using EEG measurements. *J. Neural Eng.* 13, 016018 (2016).

[31] Pfurtscheller, G., Neuper, C. & Mohl, W. Event-related desynchronization (ERD) during visual processing. *Int. J. Psychophysiol.* 16, 147–153 (1994).

[32] Koelewijn, T., van Schie, H. T., Bekkering, H., Oostenveld, R. & Jensen, O. Motor-cortical beta oscillations are modulated by correctness of observed action. *Neuroimage* 40, 767–775 (2008).

[33] Fox, N. A. *et al.* Assessing human mirror activity with EEG mu rhythm: A meta-analysis. *Psychol. Bull.* 142, 291–313 (2016).

[34] Hoppe, D. & Rothkopf, C. A. Multi-step planning of eye movements in visual search. *Sci. Rep.* 9, 144 (2019).

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-52

SWLDA OFFERS A VALUABLE TRADE-OFF BETWEEN INTERPRETABILITY AND ACCURACY FOR REHABILITATIVE BCIS

E. Colamarino[1,2], T. Colombo[1], F. Pichiorri[2] , D. Mattia[2] , L. Palagi[1], F. Cincotti[1,2]

[1] Department of Computer, control, and management engineering Antonio Ruberti, Sapienza, University of Rome, Italy

[2] Neuroelectrical Imaging and BCI Lab, Fondazione Santa Lucia IRCCS, Rome, Italy

emma.colamarino@uniroma1.it

ABSTRACT: Interpretability, accuracy and a solid neurophysiological basis can be considered as the main requirements for the classification model to monitor motor imagery tasks in post-stroke motor recovery paradigms supported by the brain-computer interface technology.
This study aimed at comparing the accuracy performance of different classification approaches applied on a dataset of 15 stroke patients. We also explored how the variation in the dimensionality of the feature domain would influence the different classifier performance.
To this purpose, stepwise linear discriminant analysis (SWLDA), shrinkage linear discriminant analysis, logistic regression, support vector machine, multilayer perceptron, decision tree and random forest classifiers entered in the performance analysis.
SWLDA statistically outperformed the classifiers commonly used in sensorimotor-BCI paradigms, achieving 80% in classification accuracy even in case of feature domain dimensionality reduction. The linearity, the interpretability and the accuracy of the SWLDA model, even just by means few EEG electrodes, yielded to consider SWLDA an optimal solution to fulfil the main requirements of the rehabilitation context.

INTRODUCTION

Electroencephalogram (EEG)-based brain-computer interfaces (BCIs) have been recently proposed to assist motor recovery training in stroke patients [1], [2]. In this context two main approaches have been identified: the first employs brain activity to control devices to assist movement [1], the second aims at modifying brain activity to improve motor behaviour [2].
In [2], BCIs monitor the modulation of brain activity induced by the movement imagination. Indeed, motor imagery (MI) practice as well as motor execution elicits event-related desynchronization that occurs within EEG frequency bands (alpha and beta) and primarily over the scalp in sensorimotor cortical regions contralateral to the part of the body involved in the task. Therefore, the rationale behind such BCI approach is based on the assumption that the practice of mental imagery with motor content could influence brain plasticity and thus, enhance post-stroke functional motor recovery.

Although many approaches have already been proposed to detect and classify EEG signals [3], i.e. sensorimotor rhythms (SMR), classification algorithms are still being investigated. Low signal-to-noise ratio of EEG signals, their non-stationarity over time and the limited amount of training data available to calibrate the classifiers are the main challenges faced by classification methods for BCI [4].
In the context of SMR classification, although a survey of classifiers' performances has been already approached [3], no conclusive results were achieved because of the different context, i.e. different set of subjects, set of features and parameters, in which the studies took place. To overcome that limitation, Bashashati et al. [5] built a unified comparison framework to evaluate the performance of different classifiers (two feature extraction methods and seven classification methods) on several sensorimotor BCI dataset collected from healthy subjects. Multilayer perceptron, logistic regression, support vector machine and linear discriminant analysis classifiers resulted in the best performance in synchronous BCI paradigms. Moreover, the authors concluded that, since pre-processing of the data, feature extraction and feature selection all change the distribution of the data in the feature space, a BCI system should be viewed as a unit consisting of different blocks in which all the block settings and parameters should be adjusted jointly for each individual subject.
In contrast to other fields of application where optimal control is pursued, in the rehabilitation context the aim is also to reinforce the appropriate sensorimotor activation in terms of both topographical and spectral characteristics. Therefore, physiological constraints should be considered in the classification process to take into account neurophysiological evidences and rehabilitation principles.
In [2], neurologists selected the proper EEG features, i.e. BCI control parameters, basing on the visualization of EEG pattern in form of statistical index matrix obtained by the comparison between two conditions (task and rest). Those features and their weights (conventionally fixed to -0.5) were merged in a linear classifier.
From this approach, we have recently moved toward a semiautomatic selection, based on physiological constraints, able to reproduce the choices of neurologists [6]. The possibility to combine a proper feature selection

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-52

and a linear classifier (fast and powerful in interpretative terms) led us to apply the stepwise linear discriminant analysis [7], [8].
Since the specific rehabilitative context and the neurophysiological constraints, this study aimed to compare our classification approach and those proposed for the classification of sensorimotor rhythms. The experimental group, 15 stroke patients, made unique this analysis. Seven classification methods were compared, even on varying of the feature space size (two, ten or all available features). The methods, proposed in [5] as the best for synchronous BCI paradigms, i.e. the linear discriminant analysis, the logistic regression, the support vector machine and the multilayer perceptron were considered. The random forest classifier and its elementary module, the decision tree, were included in the comparative framework because of good classification performance achieved by the former [9] in binary classification of imaginative tasks.

MATERIALS AND METHODS

Data collection: EEG data were previously collected from fifteen stroke subjects according the procedure and the protocol in [2]. Scalp EEG potentials were acquired from 61 electrodes, assembled on an electrode cap (according to an extension of the 10–20 International System, linked ears reference, mastoid ground) and bandpass filtered between 0.1 and 70Hz.
All signals were digitalized at 200 Hz and amplified by a commercial EEG system (BrainAmp; Brain Products, Gilching, Germany).

Experimental protocol: During the acquisition all subjects were comfortably seated in an armchair in a dimly lit room with their upper limbs resting on a desk. Visual cues were presented on a screen on the desk.
All subjects were trained to perform the kinesthetic motor imagery of the hand movements (grasping and finger extension) with their affected upper limbs. Each run comprised 30 trials (15±1 rest, 15±1 motor imagery). The total duration of each trial was 7 seconds with an inter-trial interval of 3.5 seconds. In each trial the experimental task took up just 4 seconds.

Pre-processing and Feature Extraction: Runs collected during the motor imagery of grasping and finger extension movements were concatenated.
EEG data were notch filtered at 50 Hz. EEG signal intervals containing artefacts (muscular, environmental) were identified, using a semi-automatic procedure based on the definition of a voltage threshold, and discarded. EEG data were spatial filtered by means the common average reference spatial filter.
EEG data were divided into epochs 1 second long and spectral features were extracted using the maximum entropy method (16th order model, 2 Hz resolution, no overlap) [10]. Two hundred and forty epochs were at most available for each dataset (one for each subject).
Given the specific motor rehabilitation context, spectral features belonging to the sensorimotor strip (i.e. FC, C, CP electrodes) in the contralateral scalp area to the hand involved in the task and in the range from 7 Hz to 25 Hz were used for the following steps.
After extracting features from each relevant channel (i.e. FC, C, CP in the affected hemisphere, 12 channels) and frequency bin (10 frequency bins), the features were assembled to build a feature vector, which was supplied to each classifier.

Feature selection: All classifiers were tested, even on varying of the feature number. Two features, ten features and all features (120 features) were considered.
To select the best two or ten features the recursive feature elimination cross-validation approach was used [11].
Recursive feature elimination (RFE) is a feature selection method that fits a model and removes the weakest features until the required number of features is reached. Features are ranked by the model's attributes, and by recursively eliminating a small number of features per loop, RFE attempts to eliminate dependencies and collinearity existing in the model. To find the optimal number of features, cross-validation strategy is applied. RFE cross-validation (RFECV) scores different feature subsets and selects the best scoring collection of features. In our approach the number of folds for the cross-validation was set to 3. Although in its first formulation RFE was applied jointly with the support vector machine (SVM) [11], we opted for the decision tree (DT) as estimator. The choice was based on the best classification results obtained in the comparison between SVM and DT as estimators. For conciseness, results of those analyses were not reported in this work.
RFECV was applied before all classifiers with the exception of the stepwise linear discriminant analysis, since the last intrinsically includes a feature selection process. This process can be controlled by the experimenter in term of number of features to be selected. As above reported, two or ten features was set also for the stepwise linear discriminant analysis.

Classification: Seven classifiers were compared in terms of classification performance.
Stepwise linear discriminant analysis (SWLDA) is an extension of the Fisher's Linear Discriminant that performs feature space reduction by selecting suitable features to be included in the discriminant function. In this classifier, a combination of forward and backward stepwise analysis is implemented. The input features are weighted using the Fisher's Linear Discriminant to predict the target labels. Starting with an empty model, the most statistically significant input feature for predicting the target label (having a p-value < 0.05) is added to the discriminant function. After each new entry to the discriminant function, a backward stepwise analysis is performed to remove the least significant features, having a p-value > 0.1. This process is repeated until the discriminant function includes a predetermined number of features, or until no additional features satisfy the entry/removal criteria [7].
Shrinkage linear discriminant analysis (sLDA) is a standard linear discriminant analysis (LDA) classifier in which the class-related covariance matrices used in its optimization have been regularized using shrinkage. The

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-52

sLDA classifier has been demonstrated to be more effective and more robust for small dataset than LDA [4]. In our implementation, the optimal shrinkage parameter was determined following the lemma introduced by Ledoit and Wolf.

Logistic regression (LR) is a discriminative learning classifier that directly estimates the parameters of the posterior distribution function. The maximum likelihood method is used to approximate such parameters [12]. In our implementation, the regularized logistic regression was optimized by means the liblinear [13] algorithm that supports both L1 and L2 regularization.

Support vector machine (SVM) is a classifier that uses a discriminant hyperplane to identify classes. The selected hyperplane is the ones that maximizes the distance (margin) from the nearest data points (support vectors) of each class [12]. In this study, we implemented the linear SVM and set the penalty parameter to c=1. Before selecting the value, three values (c=0.1, c=1 and c=10) were tested. No statistically significant differences were found among values.

Multilayer perceptron (MLP) is a neural network. For this analysis, we implemented a MLP that trains using a quasi-Newton algorithm which uses a backpropagation implementation of the gradient. We considered one hidden layer MLP with 20 neurons, ReLU (rectified linear unit) as activation function of the neurons and L-BFGS solver. This final setting has been defined after testing a combination of different values for the number of neurons of the hidden layer (10, 20, 40 and 80 neurons), the activation functions (ReLU, sigmoid) and the solvers (L-BFGS, Adam and RMSProp). The combination that gave the best results, in terms of classification accuracy average across subjects, was that used in this analysis.

Decision tree (DT) is a classifier which partitions the feature space until terminal nodes, each one assigned to a predicted value. Although decision trees are very easy to use for no-statisticians, they work for non-linear functions and the treatment of missing values is more satisfactory than most other model classes, we might not be able to find the best model at all. Moreover, results can be quite variable: small changes in the data can potentially lead to completely different splits (i.e. trees) [14].

Random Forest (RF) classifier is a set of decision trees merged by a probabilistic scheme. To classify an epoch, the corresponding feature vector is the input for each tree in the forest. Each tree makes a prediction and the forest chooses the prediction having the most votes over all the trees in the forest. RF can work on high-dimensional data and it can be applied to any model. Despite of its ability to returns the variable importance, it is very hard to interpret [14].

Many variant RF parameters impact on the algorithm accuracy. For each subject, we tested both the number of trees (10, 20, 50, 100, 200, 500, 1000, 2000) and the minimum number of samples required to split an internal node (from 2 to 24 in steps of two). The best set in terms of accuracy average (across subjects) was considered for the following analysis: 2000 trees and 4 samples required to split an internal node.

Validation: A 10-times cross-validation (Fig. 1) was implemented to compare classifiers and number of features. For each iteration the feature domain (epochs x number of features, at most 240 x 120) was shuffled along the first size (epochs). The first ninety and the last ten percent of the data have been the training and testing dataset, respectively.

Training dataset was the input for the feature selector (RFE-CV based on DT). It performed the feature selection ten times using the same dataset and returned the list of features sorted according to the more selected among the feature selection iterations.

The first two or ten features were considered to reduce the feature domain or all features if feature selection was not required.

Figure 1: Validation approach. For each iteration (10 in total) the steps in the hatch block were repeated. Specifically, feature domain was shuffled and divided in training and testing dataset. The former was used to train the classifier (and to select the best two or ten features, if that was the condition under investigation), the latter to test the classifier. The performance index was computed for each iteration.

The feature domain, properly reduced (only for 2 or 10 features analysis), was the input for each classifier. Each classifier was trained from the training dataset. The testing dataset, never seen before, was used to test the model and compute the performance index.

For each pair number of features-classifier (e.g. 2 features – MLP) the average of the performance index across all iterations (10 in total) has been considered the emblematic value for that pair.

Performance Measures: For each pair number of features (2, 10, 120 features) and classifier (SWLDA, sLDA, LR, SVM, MLP, DT, RF) classification accuracy was computed.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-52

Statistical Analysis: For each pair number of features-classifier the Shapiro-Wilk test was applied to assess the normality of each performance index distribution.
To investigate the effect of the number of features as well as of the classifier and their potential interaction, classification accuracy was analysed by the repeated measure two-way analysis of variance (ANOVA). The Tukey HSD post hoc analysis was applied to assess pairwise differences. The threshold for statistical significance was set to $p<0.05$. Results are presented as mean ± standard error (SE) across subjects.

RESULTS

The statistical analysis revealed the significant effect of the classifier factor ($F=77.22$, $p<0.001$) as well as the number of the features ($F= 19.11$, $p<0.001$) on the classification accuracy and the significant interaction among factors ($F= 13,20$, $p<0.001$).
Figure 2 shows for each pair (classifier-number of features) the results, presented as average and standard error across subjects.
The post-hoc analysis, applied to the *classifier* factor, pointed out the overall superiority of the SWLDA classifier over all classifiers as well as better performance obtained by the sLDA respect to those of the LR, SVM, MLP and DT classifiers. All classifiers outperformed DT classifiers.
Moreover, better performances were globally (*number of features* factor) achieved when ten or all available features were used than those obtained for two features.
Since the number of features directly impacts on the number of physical EEG electrodes required to collect EEG data and, therefore, extract the needed features, results from interaction between classifiers and number of features were deeply analysed and reported in Fig. 3.
With equal number of features (both for 2 and 10 features), SWLDA (accuracy average= 0.78 evaluated for 2 features, accuracy average= 0.79 evaluated for 10 features) statistically outperformed all classifiers. DT classifier, instead, (accuracy average= 0.62 evaluated for 2 features, accuracy average= 0.67 evaluated for 10 features) did not reached good performance, revealing to be the worst classifier (Fig. 3 upper panel, left side).
When all available information were used to train the classifiers, no statistically significant differences emerged among SWLDA, sLDA and RF (Fig. 3 upper panel, right side).
Increasing the size of the feature domain significantly improved performances in both sLDA (0.70, 0.75, 0.80 accuracy average for 2, 10 and 120 features) and RF classifiers (0.67, 0.73, 0.80 accuracy average for 2, 10 and 120 features). Conversely, SWLDA performances (0.78, 0.79, 0.80 accuracy average for 2, 10 and all features) did not differ among them varying on feature number.

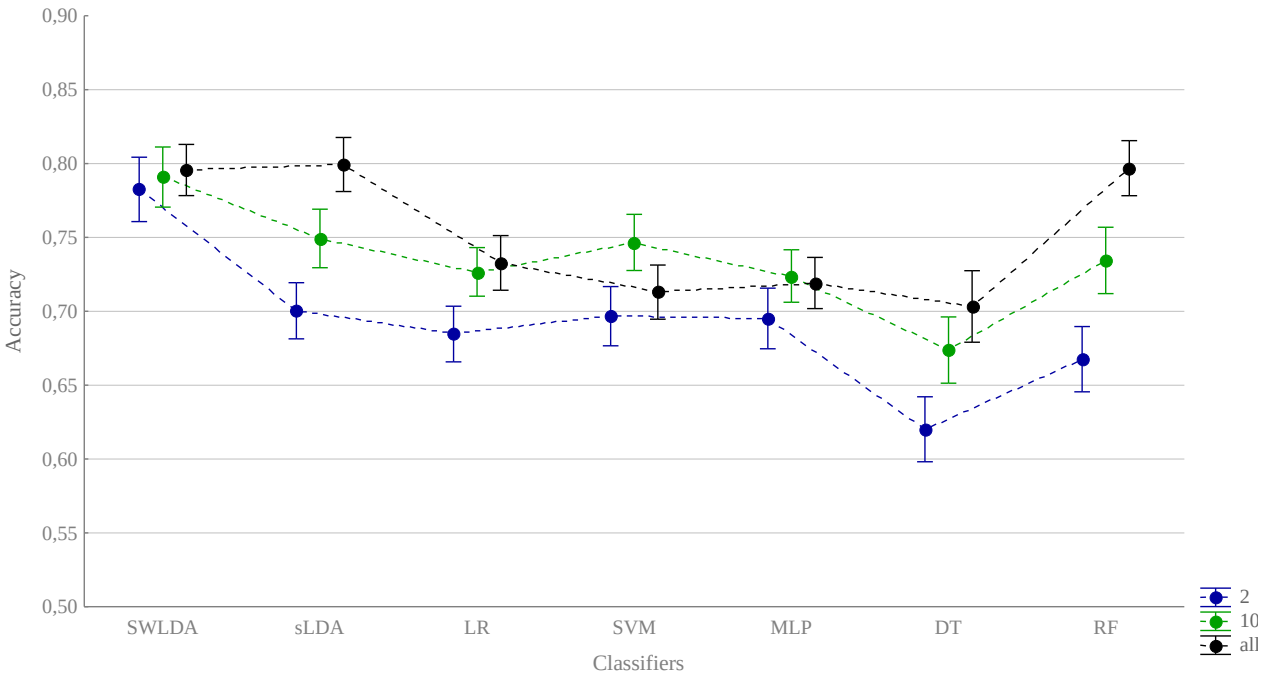


Figure 2: Classification accuracy, presented as mean ± standard error (15 stroke patients), computed for seven classifiers: stepwise linear discriminant analysis (SWLDA), shrinkage linear discriminant analysis (sLDA), logistic regression (LR), support vector machine (SVM), multilayer perceptron (ML), decision tree (DT), random forest (RF). For each classifier, accuracy was evaluated when the feature domain had been reduced to include two features (in blue), ten features (in green) and all available (black) features, i.e. no feature domain dimensionality reduction.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-52

Figure 3: Post-hoc test results of the ANOVA interaction factor. Upper panel, left side: comparison among classifiers with the same number of features. Pairwise differences among classifiers, for 2 and 10 features, were presented in same matrix since they had returned equal results. Upper panel, right side: comparison among classifiers when all features were considered. Lower panel, left side: comparison between classifiers trained from 2 features and those trained from 10 features. Lower panel, centre: comparison between classifiers trained from 2 features and those trained from all available features. Lower panel, right side: comparison between classifiers trained from 10 features and those trained from all available features.
Matrix reading: The classifier in each column header statistically differed/ did not differ from the classifier reported in the row header. Significant/no-significant differences correspond to coloured/white boxes. Green (orange) boxes means that the classifier in the column (row) header outperformed that in the row (column) header.

The increasing trend in accuracy was observed also for LR, DT and MLP classifiers: for the first the increase from 2 to 10 or 120 features statistically improved classification accuracy, for the neural network model the trend was not supported by the statistical results. SVM seemed, instead, to be prone to overfitting (0.70, 0.75, 0.71 accuracy average for 2, 10 and 120 features).
Lastly, the cross-check between classifier and number of features revealed that even the best model of the sLDA and RF (120 features) did not significantly differ from the SWLDA based just on two or ten features (Fig. 3, lower panel, centre and right side). Therefore, even if sLDA and SWLDA are both linear and, therefore, interpretable models, the last reached comparable performance by means few features (i.e. 10 features, less than 10 EEG channels).

DISCUSSION

Identifying the optimal classification method, based on relevant features, fast and able to provide an interpretable model of the EEG reinforced pattern, is a milestone in post-stroke rehabilitation protocols supported by BCI technology. In contrast to other fields of application where optimal cursor control is pursued, in a rehabilitation context the reinforcement of the proper sensorimotor activation in terms of both topographic and spectral characteristics is the main aim.
Spectral features belonging to the sensorimotor area of the affected hemisphere, in alpha and beta bands, were extracted from EEG data of 15 stroke subjects to compare seven classifiers in terms of classification accuracy. Performance was also analysed varying on the number of features considered in the feature domain.
Stepwise linear discriminant analysis (SWLDA) revealed being the best classifier even just considering two or ten features. Considering all available features (120 features) shrinkage linear discriminant analysis (sLDA) and random forest (RF) achieved good results and comparable to those of SWLDA. Nevertheless good results, linear models (SWLDA and sLDA), resulting from the linear combination of features properly weighted, are more interpretable than RF model.
In our approach, indeed, monitoring the cursor trajectory, feedback provided to the therapist and directly related to the combination of proper features, allows to explain single trial and rehabilitative session performances.
RF belongs to the bootstrap aggregating methods based

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-52

on decision tree classifiers and, although decision tree is the simplest model because the intuitive interpretation, the structure of the RF (in our case 2000 decision trees) decreases the interpretability of the model by the clinicians. Among linear models, instead, even if SWLDA and sLDA reached the same performance, SWLDA built the model by less than 120 features: in most cases, the embedded feature selection process, starting from the empty model, did not add all predictors to the model. Moreover, the interaction among ANOVA factors did not highlight differences between SWLDA, trained with two or ten features, and both sLDA and RF trained by all features.

The linearity, the interpretability and the possibility to achieve good classification results also thank to the embedded selection process (not covered by the nature of the other classifiers) yielded SWLDA to be considered the best classification approach in the rehabilitation context. Furthermore, the possibility to monitor EEG patterns to reinforce by means few electrodes (10 features i.e. less than 10 EEG channels,) matches the use of BCI technology in clinical context.

Focusing for each classifier on the number of features factor, if from one hand for SWLDA the increasing trend, justified by the increasing number of features (2, 10, less than 120 features), was not supported by the statistical analysis, from the other hand the trend revealed being significant for both sLDA and RF. Moreover, with equal number of features, no differences were observed between sLDA and RF, supporting, therefore, the use of RF model as a good approach to the binary classification of motor imagery tasks, as proposed in [9]. Although Steyrl et al. observed the superiority (3% in accuracy) of the RF approach to the sLDA, the characteristics of their approach, i.e. different tasks, number of channels, pipeline of EEG signal processing, should be considered in the comparison. For similar reasons, our results did no confirm results in [5]. Although we used similar spectral features, the application of the common spatial pattern filter and the lower number of the recorded EEG channels may be the reason why the multilayer perceptron and the logistic regression did not show good performances.

CONCLUSION

SWLDA classifier statistically outperformed those commonly used in SMR-BCI paradigms, achieving good performance even in case of feature domain dimensionality reduction. Monitor the brain activity by means few EEG electrodes, indeed, is the key to use BCIs in clinical realm. Linearity, interpretability and impact on the usability yielded to positively evaluating SWLDA approach in the upper limb post-stroke motor recovery protocols supported by BCI.

ACKNOWLEDGEMENTS

We thank Angelica Cottarelli for support in the preliminary analysis of EEG data. This work was partially supported by the project APOSTROPHES, Sapienza, University of Rome, Progetti di Ateneo and Promobilia Foundation (2018-H1 ref. 18076).

REFERENCES

[1] Ramos-Murguialday A et al., Brain-Machine-Interface in Chronic Stroke Rehabilitation: A Controlled Study, Ann. Neurol., vol. 74, no. 1, pp. 100–108, Jul. 2013.

[2] Pichiorri F et al., Brain-computer interface boosts motor imagery practice during stroke recovery, Ann. Neurol., vol. 77, no. 5, pp. 851–865, 2015.

[3] Lotte F, Congedo M, Lécuyer A, Lamarche F, and Arnaldi B, A review of classification algorithms for EEG-based brain-computer interfaces, J. Neural Eng., vol. 4, no. 2, pp. R1–R13, Jun. 2007.

[4] Lotte F 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.

[5] Bashashati H, Ward RK, Birch GE, and Bashashati A, Comparing Different Classifiers in Sensory Motor Brain Computer Interfaces, PLOS ONE, vol. 10, no. 6, p. e0129435, Jun. 2015.

[6] Colamarino E, Pichiorri F, Schettini F, Martinoia M, and Matti D and Cincotti F, Guider: A Gui for Semiautomatic, Physiologically Driven Eeg Feature Selection for a Rehabilitation Bci, Proc. 7th Graz Brain-Comput. Interface Conf. 2017 Vis. Real.

[7] Krusienski DJ et al., A comparison of classification techniques for the P300 Speller, J. Neural Eng., vol. 3, no. 4, pp. 299–305, Dec. 2006.

[8] Krusienski DJ, McFarland DJ, and Wolpaw JR, Value of amplitude, phase, and coherence features for a sensorimotor rhythm-based brain-computer interface, Brain Res. Bull., vol. 87, no. 1, pp. 130–134, 2012.

[9] Steyrl D, Scherer R, Faller J, and Müller-Putz GR, Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: A practical and convenient non-linear classifier, Biomed. Tech., vol. 61, no. 1, pp. 77–86, 2016.

[10] McFarland DJ, Lefkowicz AT, and Wolpaw JR, Design and operation of an EEG-based brain-computer interface with digital signal processing technology, Behav. Res. Methods Instrum. Comput., vol. 29, no. 3, pp. 337–345, Sep. 1997.

[11] Guyon I, Weston J, Barnhill S, and Vapnik V, Gene Selection for Cancer Classification using Support Vector Machines, Mach. Learn., vol. 46, no. 1, pp. 389–422, Jan. 2002.

[12] Bishop C, Pattern Recognition and Machine Learning. New York: Springer-Verlag, 2006.

[13] 'LIBLINEAR -- A Library for Large Linear Classification'. [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/liblinear/.

[14] Breiman L, Classification and Regression Trees. Routledge, 2017.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-53

AN EXPECTATION-BASED EEG MARKER FOR THE SELECTION OF MOVING OBJECTS WITH GAZE

D.G. Zhao[1], A.N. Vasilyev[1], B.L. Kozyrskiy[1], A.V. Isachenko[1,2], E.V. Melnichuk[1], B.M. Velichkovsky[1,2], S.L. Shishkin[1]

[1] Lab. for Neurocognitive Technologies, NRC "Kurchatov Institute", Moscow, Russia

[2] Institute of NBCIS S&T, Moscow Institute of Physics and Technology, Moscow, Russia

E-mail: sergshishkin@mail.ru

ABSTRACT: The use of an EEG expectation-related component, the expectancy wave (E-wave), in brain-machine interaction was proposed more than 50 years ago, but active exploration of this possibility has started only recently, in the context of developing passive brain-computer interfaces for the enhancement of gaze interaction. We report, for the first time, the results of a systematic experimental study that revealed an EEG marker for selecting intentionally an object among other moving objects using smooth pursuit eye movements. This marker appeared to have the same nature as the E-wave previously observed in the EEG accompanying the selection of static objects with gaze fixations. A convolutional neural network classified the intentional and spontaneous smooth pursuit eye movements with average ROC AUC 0.69±0.13 (M±SD). These results suggest that the E-wave might be robust enough to serve, after further improvement of the methodology, as the basis of hybrid eye-brain-computer interfaces applied for selection in dynamically changing visual environments.

INTRODUCTION

One of the oldest proposals for the design of a brain-computer interface (BCI), put forward by Grey Walter more than 50 years ago, was the use of the expectancy wave (E-wave), the electroencephalogram (EEG) slow negative wave that arises when a person is waiting for a stimulus. In this paradoxical BCI design expectation of the interface triggering was proposed to be used to trigger the interface [1]. Although in the subsequent decades the E-wave (later known also under the name of stimulus preceding negativity, SPN) and the more complex phenomenon, the contingent negative variation (CNV), were among the most studied EEG phenomena, they still not received much attention in the BCI world. One important exception was the proposal to use an expectation based passive BCI to support gaze interaction [2, 3] – more specifically, to solve the Midas touch problem, one of the most serious obstacles for using gaze as a tool to control computers and other devices. The Midas touch problem is the inability of gaze based human-machine interfaces to differentiate, in many cases, eye movements and fixations intentionally used for control from spontaneous ones, e.g. those used for visual scene exploration [4]. Later, an EEG component with the time course and topography similar to a typical SPN/E-wave and enabling classification of the gaze fixation intentionally used to make moves in a game vs. spontaneous gaze fixations was described [5] and first online tests of a hybrid eye-brain-computer interface (EBCI) based on this design were run [6]. In these studies, static visual layout was used, while the Midas touch problem may be more severe when the scene is changing dynamically.

Recently, the selection of moving objects with smooth pursuit eye movement was found to be very effective, in addition to the long-used gaze fixations and saccades [7]. In the presence of multiple moving objects, however, any of such objects can easily attract attention and become pursued with gaze automatically. The use of an EBCI instead of a merely gaze-based interface for moving object selection thus seems reasonable, however, only experimental studies can show if this is feasible, as multiple factors may change EEG components in such dynamical settings (see discussion on such factors for the case of selection of moving objects with the P300 BCI in [8] and for the case of using the fixation-related EEG P300 wave in the context of visual search in [9]). The EEG under smooth pursuit eye movements was studied very little, and we found no prior studies of EEG accompanying intentional object selection with the smooth pursuit eye movements.

In our preliminary studies [10, 11] we observed an EEG component resembling the SPN/E-wave prior to moving object selection by gaze, but a similar wave also accompanied smooth pursuit eye movements in control conditions, where the intention to select an object could not appear. We concluded that certain form expectation or at least estimation of time to execute certain operation (a related cognitive activity) easily appears under various conditions when a pursuit is intentionally used to follow instruction. In the current study, we designed a control condition that was free from this shortcoming and corrected some other details of the study design. With the new design, we collected data

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-53

from a larger group of participants, which allowed us to analyze the difference between the EEG accompanying spontaneous smooth pursuit and intentional selection of moving objects by gaze, and also to model their classification in the framework of an EBCI using a convolutional neural network.

METHODS

Participants: 14 healthy volunteers (8 male, 6 female), aged 22–52 (M±SD 28±8) years participated in our study (participant #1 participated in two sessions on different days). Except for four participants, others had prior experience with a gaze-based interface. Ten participants had normal vision, others had corrected to normal vision (3 with glasses and 1 with contact lenses).

Data acquisition: A consumer grade eye tracker *Tobii 4C* (*Tobii*, Sweden) was attached to the lower edge of the monitor. EEG was recorded from 19 locations (F3, Fz, F4, C3, Cz, C4, P1, P3, Pz, P2, P4, PO7, PO3, POz, PO4, PO8, O1, Oz, O2) at 500 Hz with *NVX52* amplifier (*MCS*, Russia). Digitally connected electrodes at earlobes served as reference. The vertical electrooculogram (EOG) was recorded using electrodes about 2 cm below the right eye, and the horizontal EOG with electrodes above the eyebrow and near the outer canthi, both with the same amplifier. Impedance was kept below 20 KOhm. In the last four experiments, a signal from a microphone was also recorded with the same amplifier to enable detection of time periods when the participants reported their answers.

Task: The participants were presented with 15 "balls" (circles), each 80 px in diameter (2.8°), displayed on an 18.5" monitor, at a distance of about 64 cm from their eyes (Fig. 1). The balls were numbered 1 to 15 and contained 5 to 8 dots around the number (Fig. 1, right). In the "dynamic" conditions (SMB, FA, CT) balls moved linearly on the screen at ~230 px/s (6.8°/s) speed, changing their movement direction in a natural way when hitting each other or the screen edges.

Balls with randomly varying numbers of dots were generated as 15 different images, separately for each block of four runs. Each image was assigned to each ball number at the start of a run.

In a run, the participants were asked to complete one of the following tasks:

Select moving balls (SMB) – select 15 balls one by one according to their numbers in ascending order and then (only for subjects 7-14) in descending order (see details about selection below).

Select static balls (SSB) – select 15 balls one by one as in the SMB condition, but in this case, the balls were fixed at their initial positions.

Find accelerated (FA) – find the ball that was moving faster than the other balls. This target ball, randomly chosen, accelerated once after 15 to 20 s from the block start by 1.3˚/s (19% of the basic speed). Participants were advised to announce the number of the ball that appeared to be the fast one only when they made sure that it was actually faster than the others.

Counting task (CT) – find, one by one, five objects with a specified number of dots and summarize the objects' numbers. In the CT condition, participants were asked, five times in a row, to find a ball with the number of dots specified by the experimenter (this number could be from 5 to 8, chosen randomly each time). After they had heard the number, they had to find an appropriate ball and summarize its labeling number with previous the number(s) of the previous ball(s).

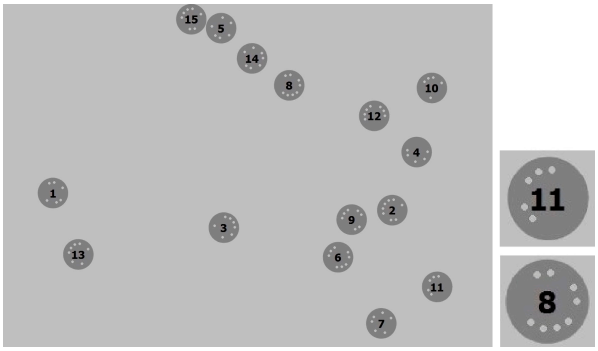


Figure 1: A screenshot of a typical experimental layout (**left**), and examples of "balls" with different amount and positioning of dots (**right**).

The SMB and SSB tasks modeled intentional gaze selection, while FA and CT tasks served to collect data on smooth pursuit eye movements without intention to make a selection. The SMB task was similar to the task used in [12] where gaze based selection was compared to a mouse-based one. However, in the current study the number of balls and their speed were set at lower levels in all tasks, otherwise, FA and CT tasks appeared to be too difficult.

At the beginning of the run, all the balls were dark grey. In the SMB and SSB conditions, a selected ball was highlighted with a green color. Ball color returned to dark grey when another ball was selected. Participants were asked to select the 15 balls in ascending order (all participants) and then in descending order (only participants 7-14) as soon as possible. In the FA and CT conditions, the balls that were considered selected by the online selection algorithm were not highlighted.

Online Gaze-Based Selection: In all conditions, the median distance from the gaze position to each ball's center was calculated in a moving window of ~867 ms length. If this measure computed for a ball did not exceed 55 pixels (in other words, the distance from the gaze position to this ball's center did not exceed 55 px longer than half of the length of this window) and was the smallest among all balls, this ball was selected. Some samples from the eye tracker (sampling rate 90 Hz) were discarded from time to time to maintain synchrony with the monitor (its refresh rate was 75 Hz). Using a high-speed video camera (240 fps) we found out that a delay between a saccade start and the start of a gaze-controlled cursor moving was approximately 140 ms. To improve synchronization between the gaze and

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-53

balls coordinate streams, the latter was intentionally delayed by the same value. The eye trackers' built-in alignment filter was not applied. With a few exceptions, the selection algorithm followed one described in [12].

Algorithm triggering time was recorded using a photosensor attached to the lower left edge of the screen, where a circle not visible to a participant changed its brightness as selection algorithm triggered (in the SMB and SSB conditions, it was in the same video frame when the visual feedback was given). The signal from the sensor was recorded synchronously with the EEG by the amplifier. In addition, the ID of the selected ball was recorded regardless of its highlighting.

The visual task and the selection algorithm were implemented in C++ and QML with the use of the Qt framework.

Procedure: Participants were seated in an adjustable chair in front of an office table with the monitor. To ensure head position stabilization a chin rest was used.

The experiment consisted of six blocks, each containing all four tasks (conditions). The order of conditions was the same over the blocks for a single participant but counterbalanced in the group of participants. Conditions where balls were intentionally selected with gaze (SMB, SSB) were not allowed to immediately follow each other. Each condition in a block included one run of ~2 min duration.

Before the start of the first block, the experimenter showed the moving balls and in detail explained the tasks. The first block was considered as a practice, and the participants were allowed to ask any questions during the tasks. The data from the subsequent five blocks were used for analysis. Before each block, native 7-points calibration of the eye tracker was used. Also, calibration was monitored by the experimenter during SMB and SSB conditions. If the participant felt discomfort at those conditions, re-calibration was run and the condition was overwritten. Within a single block, participants were asked to avoid movements to prevent calibration distortion. After completing one block, the participants were allowed to rest for a few minutes. 10 minutes break came after the third block.

In the end of an experiment, participants were asked to rate how difficult was performing a task by putting a mark on four visual analog scales (VAS), one per condition. Left end of the scale was labeled "very easy" and the right end was labeled "very difficult".

Data Analysis: In SMB and SSB conditions we considered only selections in right order (balls #1, #2, … #15, #14, … #1).

Averaging of pursuit-related EEG was triggered using a sensor (photodiode) placed in the bottom left screen corner, which was activated (not visible to the participant) as the algorithm triggered. In SMB and SSB conditions, this was followed with a ball highlighting. In FA and CT conditions, highlighting didn't happen, but the data about fixations still recorded.

Preprocessing: All EEG analysis was made offline using the data collected during the experiments. For individual trial analysis, EEG data were filtered with lowpass Butterworth 4th order filter with cutoff at 30 Hz and then cut into epochs around the time of interest, including the time of online gaze-based selection algorithm triggering and the time of gaze pursuit initial fixation onset. The latter was found using gaze velocity estimated from the EOG prior to the algorithm triggering (our *Tobii* eye tracker license did not allow for gaze coordinate offline analysis). Epochs (trials) were discarded from analysis if either of the following conditions were met:

1) Pursuit duration was shorter than 450 ms or exceeded 1200 ms
2) The trial was not in the correct selection sequence (for SMB and SSB)
3) The trial resulted in the selection of the accelerated ball (for FA)
4) Speech sounds were detected between fixation onset and algorithm triggering (for participants 11 and 14)
5) Amplitude range in any EEG or EOG signal exceeded 150 μV during gaze pursuit

The resulting dataset was used for the EEG data analysis and classification.

Signal preprocessing, segmentation and visualization was performed using MATLAB (*MathWorks*, USA). Function *topoplot* from the EEGLAB package [13] was used to create topographical amplitude maps.

Classification: For training, the SMB and SSB data were joined together and the second class consisted of the CT data, while during testing the SMB and SSB data were considered separately, and the second class again was CT. Features were amplitudes from the time interval starting -400..0 ms relative to online gaze-based selection triggering. Average over -100..0 ms interval was subtracted from the same channel data in each trial, i.e., was used as a baseline. Note that a saccade typically preceded the start of pursuit, thus we could not use the pre-pursuit EEG as the baseline in single-trial processing without special measures to remove the EOG, while such measures could introduce data distortions that are difficult to control.

We used Keras implementation of the convolutional neural network EEGNet [14] (https://github.com/vlawhern/arl-eegmodels) under Python 3.6. The procedure consisted of 4-fold cross-validation (80% of the whole dataset) and testing on the holdout set (20% of the whole dataset). Target to non-target class ratio was preserved when forming the test subset. For each fold, we saved the best model after the optimal number of training iterations with respect to performance on the validation set of each fold, and for each trial from the testing set, we averaged predicted probabilities from these four models. We used hyperparameters optimized for the case of classifying the EEG accompanying static object selection by gaze in our study [15] (see it also for other details of the classification procedure). Since the EBCI design makes possible the use of classifiers with relatively low sensitivity while maintaining high specificity, but the balance between them might be dependent of the use

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-53

case [3, 5, 6], we found more relevant to present the test results using the ROC AUC measure which is not related to the choice of the classifier threshold.

RESULTS

In the responses to the questionnaire, the FA condition was ranked as the most "difficult" (M±SD 71.8±17.6), while for the CT condition the difficulty score was 46.3±22.7 (the score could be between 0 and 100). Only two participants perceived CT as more difficult than FA. The intentional selection conditions, SMB and SSB, were scored as less difficult: 19.2±21.3 and 10.2±9.1, respectively (the difference from the CT condition was significant, with p=0.33 in both cases, according to Wilcoxon matched pairs test).

According to visual inspection of the averaged EEGs with and without rejection of epochs containing speech signal, they did not substantially differ. For further EEG analysis and classification, we excluded data obtained from two participants: #12, who often did not follow instructions correctly enough, and #13, who had issues with several EEG electrodes. The number of EEG epochs obtained in each condition for the remaining 12 participants is shown in Table 1.

Fig. 2 shows the grand average EEG amplitude at the electrode POz. In all conditions, a similar EEG pattern was observed about 0.5-0.6 s prior to the selection algorithm triggering, i.e., near the beginning of the pursuit (the largest positive peak could be the lambda wave, well known in the EEG accompanying static gaze fixations). In SMB and SSB, a slow negative wave was developing up to the time of the online gaze-based selection algorithm triggering; soon after it, the amplitude abruptly went in the opposite direction (evidently, in response to the visual feedback). The time course of EEG amplitude in the FA condition was quite similar to the SMB and SSB conditions along most time before the selection algorithm triggering, while in the CT condition only very light deviation to negative direction was observed.

The EEG amplitude head maps (Fig. 2, bottom) demonstrated a pattern similar to previously reported for the intentional static object selection with gaze [5] in the SMB and SSB conditions, and its weaker variant in the FA condition, but nothing similar in the CT conditions.

Mean EEG amplitude at POz over the last 100 ms before selection triggering varied significantly across conditions, as determined by one-way ANOVA ($F(3,43) = 6.97$, $p = 0.0006$). A Tukey post hoc test revealed that POz was significantly more negative in SMB and SSB compared to CT (p=0.003 and p=0.001, respectively). Amplitude in FA showed no significant difference when compared to any other condition.

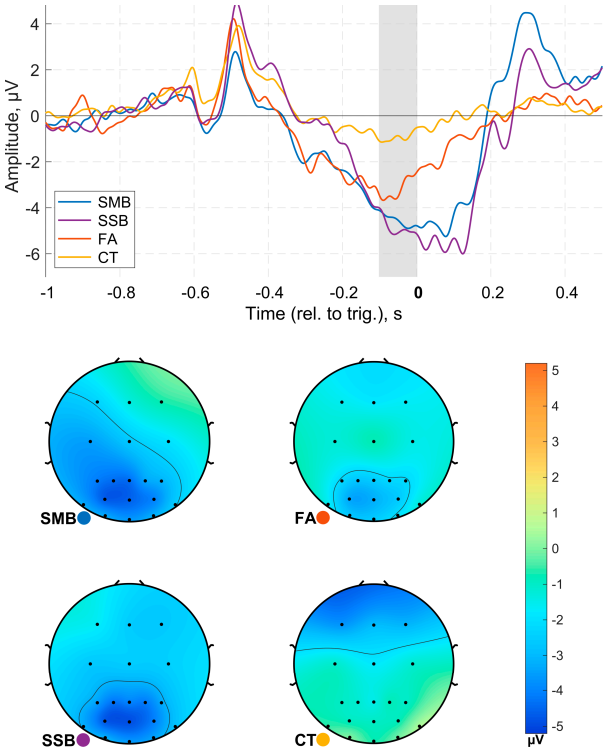


Figure 2: (**Top**) The grand average EEG amplitude at electrode POz relative to gaze-based selection algorithm triggering (N=12, baseline -1200...-700 ms). Feedback was provided to the participants at zero time in SMB and SSB but not in CT and FA conditions. (**Bottom**) Topographical maps of the EEG amplitude averaged over the last 100 ms before triggering.

Fig. 3 demonstrates the variability of the individual EEG data and shows (in its right part) the EEG dynamics with a more precise synchronization with the pursuit start. As in the studies of static object selection with gaze [5, 6], the E-wave was evident in the intentional selection conditions in all participants. Improvement of synchronization with the pursuit start tended to improve tracking the E-wave, although the difference between the two types of averaging was not large. In some participants (here, S09 and S10) the EEG dynamics in the pre-triggering time interval in the CT condition was similar to the SMB and SSB. Visual comparison of the averaged posterior EEG and EOG waveforms revealed no correspondence between them.

Classification results are presented in Table 1. We did not obtain enough data in the FA condition, so the classifier was trained and tested on both intentional conditions only against the CT condition (for training, the SMB and SSB subsets were collapsed). Note that intentional moving object selection with smooth pursuit eye movements (SMB) was classified almost as good as the selection of static objects with gaze fixations (SSB).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-53

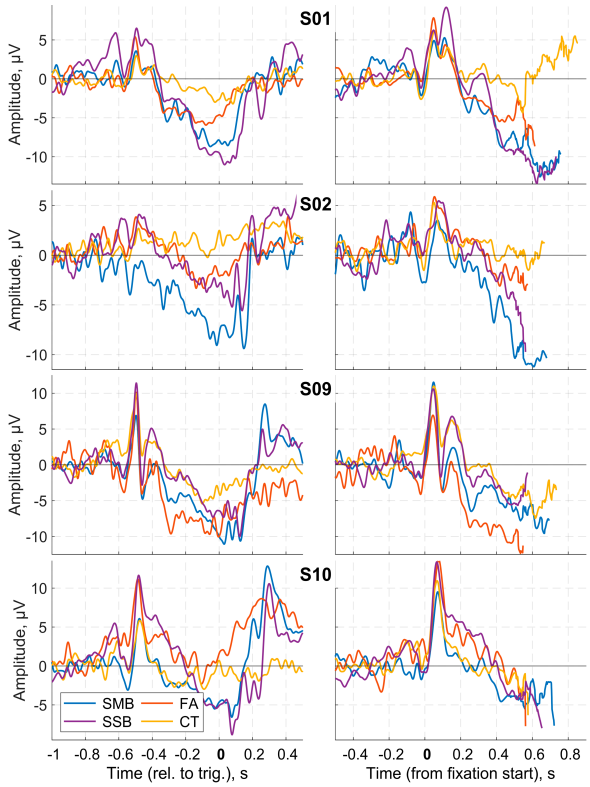


Figure 3: Averaged EEG amplitude at POz for two participants with good (S01, S02) and weak (S09, S10) separation of the SMB and SSB conditions from the CT condition: (**Left**) Synchronized (zero time) with gaze-based selection algorithm triggering (baseline -1200...-700 ms). (**Right**) Synchronized (zero time) with smooth pursuit start and truncated at online algorithm triggering (averaged curves truncated where n of trials was < 25; baseline -700 to -200 ms).

DISCUSSION

This study, with the preceding preliminary studies [10, 11], demonstrated for the first time that intentional selection of a moving object with smooth pursuit eye movements can be accompanied by a posterior EEG negativity, similar to the phenomenon observed earlier in the selection of static objects with gaze fixations.

In our preliminary studies in small groups of participants who selected moving objects with their gaze [10, 11] we found it difficult to find appropriate control condition in which moving objects are often pursued but without expecting any upcoming event. Moreover, when the participants were instructed just to pursue a moving object for a certain time without any other task, a similar negative wave also developed, likely due to mental counting time. In the current study, pursuit under the instruction to look for a faster ball also evoked a similar EEG dynamics, which could be related to waiting for a good opportunity to compare the speed of the pursued ball with the speed of some other ball(s) (a common strategy to solve this task). Surprisingly, even the less expectation-loaded condition, the CT (counting the dots on the moving balls and adding up the number of the ball to the previous sum), was accompanied with a similar deviation of the averaged EEG in some participants. Thus, cautions are due concerning the specific relation of the marker to intentional selection. Nevertheless, in the majority of participants and in the grand average (Fig. 2) the difference between the intentional selection and this control condition was clear.

Table 1: Number of epochs per condition used for EEG analysis and classification, and the classification results.

Subject	N of epochs			ROC AUC	
	SMB	SSB	CT	SMB vs CT	SSB vs CT
1	124	148	149	0.59	0.69
2	47	62	99	0.50	0.46
3	62	62	112	0.70	0.65
4	40	34	49	0.89	0.86
5	58	62	90	0.45	0.53
6	60	64	121	0.74	0.74
7	184	165	140	0.76	0.70
8	155	161	139	0.74	0.77
9	123	123	155	0.68	0.76
10	122	128	66	0.69	0.85
11	123	130	61	0.89	0.89
14	118	128	74	0.71	0.67
M	101.3	105.6	104.6	0.69	0.71
SD	46.4	45.6	36.8	0.13	0.13

Another concern could be inflation of the classification results due to the EEG contamination by the EOG. However, we used only features from the pursuit time interval. While it was accompanied by certain EOG dynamics, the longest possible distance during the time interval used for feature extraction could be 3° and in the majority of trials the distance was much lower, so the EOG contribution to EEG could not be strong. Removing the frontal electrodes from the analysis did not lead to a substantial decrease in ROC AUC, demonstrating that the EOG contribution could not strongly support classification. Nevertheless, additional study will be needed to quantify possible effects from the EOG or other sources not specific to the intentional selection that may inflate the classification results. Moreover, cleaning data from EOG might even help to improve classification.

Unlike in the studies of static object selection with gaze fixations [5, 6], the task in the current study was not free from visual search, and the pursuit could often start just after identification of the target object. Thus, the P300 wave that provides the most useful features for classification in the case of visual search [9] could also contribute to classification in our case. However, we did not observe a strong P300 wave in the averaged waveforms (Fig. 2, 3; as the first large positive peak was likely the lambda wave, the P300 peak could be expected 200 ms after it or later), so its contribution to classification could not be high.

We used an advanced classifier that demonstrated impressive performance in several BCI paradigms [14]

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-53

and also in the EBCI paradigm, i.e. in the classification of intentional vs. spontaneous gaze fixations [15]. The hyperparameters used in this study, however, were tuned for a different setup. It is possible that their more specifical tuning would provide better classification results. It is also possible that other improvements can enhance the interface design and the data processing pipeline, making the EBCI performance sufficiently good at least for certain applications requiring selection in a dynamically changing visual environment.

CONCLUSION

The results obtained in this study suggest that the EEG negative wave that is developing when a gaze-based input user expects feedback from the interface might be robust enough to serve, after further improvement of the methodology, as the basis of hybrid eye-brain-computer interfaces applied for selection in dynamically changing visual environment.

ACKNOWLEDGMENT

This study was supported in part by the Russian Science Foundation, grant 18-19-00593 (development, adaptation and application of techniques for data preprocessing and classification).

REFERENCES

[1] Walter WG. Expectancy waves and intention waves in the human brain and their application to the direct cerebral control of machines. Electroenceph. Clin. Neurophysiol. 1966; 21: 616

[2] Ihme K, Zander TO. What you expect is what you get? Potential use of contingent negative variation for passive BCI systems in gaze-based HCI, in Proc. ACII'11, 2011, 447-456

[3] Protzak J, Ihme K, Zander TO. A passive brain-computer interface for supporting gaze-based human-machine interaction, in Proc. UAHCI'13, 2013, 662-671

[4] Jacob RJK. The use of eye movements in human-computer interaction techniques: what you look at is what you get. ACM Trans. Inf. Syst. 1991; 9(2): 152-169

[5] Shishkin SL, Nuzhdin YO, Svirin EP, Trofimov AG, Fedorova AA, Kozyrskiy BL, Velichkovsky BM. EEG negativity in fixations used for gaze-based control: Toward converting intentions into actions with an Eye-Brain-Computer Interface. Front. Neurosci. 2016; 10: 528

[6] Nuzhdin YO, Shishkin SL, Fedorova AA, Kozyrskiy BL, Medyntsev AA, Svirin EP, Korsun OV, Dubynin IA, Trofimov AG, Velichkovsky BM. Passive detection of feedback expectation: Towards fluent hybrid eye-brain-computer interfaces, in Proc. 7th Graz BCI Conf., 2017, 361-366

[7] Vidal M, Bulling A, Gellersen H. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets, in ACM UbiComp'13, 2013, 439-448

[8] Ganin IP, Shishkin SL, Kaplan AY. A P300-based brain-computer interface with stimuli on moving objects: Four-session single-trial and triple-trial tests with a game-like task design. PloS ONE. 2013; 8(10): e77755

[9] Uščumlić M, Blankertz B. Active visual search in non-stationary scenes: coping with temporal variability and uncertainty. J. Neural Eng. 2016; 13(1): 016015

[11] Zhao DG, Isachenko AV, Melnichuk EV, Kozyrskiy BL, Shishkin SL. EEG potentials related to moving object selection with gaze: a possible basis for more flexible eye-brain-computer interfaces. Opera Medica et Physiologica. 2018; 4(S1): 109-110

[12] Zhao D, Isachenko I, Melnichuk E, Shishkin S. Eye-brain-computer interfacing with smooth pursuit eye movements, in Proc. 4th Int. Conf. BCI: Science and Practice, Samara, 2018, 21

[13] Isachenko AV, Zhao DG, Melnichuk EV, Dubynin IA, Velichkovsky BM, Shishkin SL. The pursuing gaze beats mouse in non-pop-out target selection, in 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, 3518-3523

[14] 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

[15] 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

[16] Kozyrskiy BL, Ovchinnikova AO, Moskalenko AD, Velichkovsky BM, Shishkin SL. Classification of the gaze fixations in the eye-brain-computer interface paradigm with a compact convolutional neural network. Procedia Computer Science. 2018; 145: 293-299

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-54

PARTIAL BRAIN MODEL FOR REAL-TIME CLASSIFICATION OF RGB VISUAL STIMULI: A BRAIN MAPPING APPROACH TO BCI

A.F. Soler-Guevara[1], E. Giraldo[2], and M. Molinas[1]

[1]Norwegian University of Science and Technology, Trondheim, Norway
[2]Universidad Tecnológica de Pereira, Pereira, Colombia

E-mail: andres.f.soler.guevara@ntnu.no

ABSTRACT: Brain-computer interface (BCI) applications are characterized by real-time feature extraction and classification. Thus, brain activity reconstruction is not generally performed due to the long computation times required to estimate brain activity. Here, we present a method for applying brain mapping solutions to BCI applications, based on a reduced model of the brain covering the occipital region using only four local electrodes (P3, P4, O1, and O2) for the classification of red, green, and blue (RGB) visual stimuli. We obtained a classification accuracy of 100% within feasible estimation times for BCI applications using an event-related potential (ERP) dataset. We first validated the occipital-zone model with a test using synthetic EEG data to evaluate its performance for local brain mapping with a small number of electrodes. In a second test, we used the validated partial model to map the occipital and part of the parietal lobes for RGB classification with outstanding results.

INTRODUCTION

Brain-computer interface systems (BCI) use brain activity to control external devices, improving the capacity of the interaction of people with disabilities or allowing a user to generate a command to actuate a device directly from the brain. BCI systems based on electroencephalograms (EEGs), use the electrical signals from electrodes in the scalp as a non-invasive technique to obtain neural activity information corresponding to human activity or external stimuli. Then, a feature extraction process from the data allows the classification of several human activities to generate a command that can actuate devices. Thus, the time between data acquisition and generation of the command is critical for developing real-time applications. The brain-mapping approach has not been considered for BCI applications due to the computational time required to process large head models and estimate a solution.

Brain mapping solutions provide an estimation of the cortical activity from the information measured with electrodes in the scalp. Such an estimation is known as the inverse problem solution from EEG. It has, however, the characteristics of being ill-posed and ill-conditioned, due to the reduced number of electrodes (hundreds), limited available information, and the high quantity of points of activity or distributed sources (thousands) to be estimated. Several methods have been presented in the literature to overcome such conditions, but brain mapping is a time-consuming technique and the quality of the solutions is highly dependent on the number of electrodes. Thus, BCI systems are rare and have not been widely studied. However, in [1] the authors present a simplified EEG inverse solution through the use of a small quantity of dipoles in the Broadmann areas, where the dimension of the inverse problem is reduced, opening the possibility to use this type of representation in BCI systems due to the reduced computational time. In [2] the feasibility of combining the source of information and electrode information was studied using a right-hand imaginary task. This study paved the way to the possibility of improving BCI solutions with brain mapping. However, in both studies, a full-brain model was considered and the solutions were not focused on specific areas of the brain, as we are proposing.

Here, we investigated a method for estimating brain activity using a partial brain model, based on a specific zone of the brain, the occipital zone and part of the parietal region, and fewer electrodes around this zone. Moreover, we present a BCI application for classifying RGB colors, obtaining a remarkable improvement over the classification results shown in [3], using the same database for classifying the three colors in real-time. This method paves the way for online BCI applications involving brain mapping techniques.

MATERIALS AND METHODS

EEG Forward model:
The equation which relates the signals measured in the scalp with the activity inside the brain is known as the EEG forward equation, and can be represented as shown below in 1

$$y_k = Mx_k + \varepsilon_k \tag{1}$$

with $\boldsymbol{y}_k \in \mathbb{R}^{d \times N}$ being the observed EEG measurements and containing the signals of d electrodes and N the total

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-54

number of samples, where the subscript k is the sample associated with the time instant $t_k = kh$ which represents the $k-$th sample of the EEG recording, where $k = 1, ..., N$ and h the sample time. The term $x_k \in \mathbb{R}^{n \times N}$ represents the source activity inside the brain that produces the measured electrical impulses. In a head model based on distributed sources, n is the number of sources considered in the brain. The matrix $M \in \mathbb{R}^{d \times n}$ is the lead field matrix or sensor matrix, which relates the measured EEG y_k to the neural activity x_k; and represents the current flow from each source to each electrode. In addition, white noise is considered in the measurements and is represented by ε_k with a mean of zero and C_ε covariance.

EEG Inverse problem solutions:
The EEG inverse problem has several characteristics that are difficult to manage -the neural activity reconstruction is an ill-conditioned and ill-posed problem- where the solution is highly noise dependent and an infinite number of estimations of the activity $\hat{x}_k$ can explain the same measurement in the EEG y_k. These issues can be overcome by several potential solutions found in the state-of-the-art, based on constrained minimum norm regularization, such as minimum norm estimation (MNE) [4], low-resolution tomography (LORETA) [5] and and the iterative regularization algorithm based on the L2 norm (IRA L2) [6],and probabilistic frameworks, such as the dynamic multi-model source localization method (DYNAMO) [7] and multiple sparse priors (MSP) [8]. We selected two methods to estimate the underlying neural activity, MSP and IRA L2, of which the approaches are briefly presented below. The MSP method assumes a priori that the activity x_k follows a normal distribution with zero mean and covariance C_x. Based on this assumption, the expected operator is applied to estimate the activity.

$$\hat{x}_k = E[p(x_k|y_k)] \tag{2}$$

After applying the Bayes theorem and further mathematical manipulations, the estimation is performed by:

$$\hat{x}_k = C_x M^T (C_\varepsilon + M C_x M^T)^{-1} y_k \tag{3}$$

where the problem is centered on estimating the covariance matrices C_ε and C_x. On the other hand, the IRA L2 method is based on a constrained minimum norm solution adding spatial and temporal constraints, such as the following cost function:

$$J = ||Mx_k - y_k||_2^2 + \rho_k^2 ||x_k||_2^2 + \lambda_k^2 ||x_k - x_{k-1}||_2^2 \tag{4}$$

where the solution involves the temporal behavior of the sources and their location, involving the terms ρ_k and λ_k, which are the regularization parameters related to the spatial distribution of the sources and the temporal evolution of the activity. The estimation can be performed using the equation below:

$$\hat{x}_k = (M^T M + \rho_k^2 I + \lambda_k^2 I)^{-1} (M^T y_k + \lambda_k^2 x_{k-1}) \tag{5}$$

Zone-Based Partial Brain Model:
The lead field matrix M represents the current flow from each of n sources to each of d electrodes. This matrix usually involves the conductivity of the skull, skin and grey matter as in [9], which is obtained to relate the neural activity x_k to the measurements in the scalp y_k as well as possible. The formation of the matrix is developed by numerical methods, such as the finite element method (FEM) and the head model considers the position of the electrodes and establishes the position of the distributed sources. For the proposed zone-based partial brain model, the model for a specific zone is created by applying a spatial filter for evaluating which sources belong to the target zone. Then, the columns corresponding to the sources in the target zone are selected from the gain matrix M and appended in a new temporal matrix $M_{rs} \in \mathbb{R}^{d \times rs}$, with rs being the number of reduced sources. At this point of the reduction, we work under the first assumption that *all the measured activity in the electrodes can be found in the target zone*, an assumption that is not true, because the activity can be generated in another zone of the brain and projected in the target zone. However, a second assumption is considered, in which *the activity registered in an electrode mostly represents activity in the neighbor spaces around it*. Under this assumption, the rows corresponding to the nearest electrodes in the zone can be selected from the matrix M_{rs}, and appended to form a lead field matrix of the target zone $M_{rz} \in \mathbb{R}^{rd \times rs}$ with rd being the number of reduced electrodes. This assumption reduces the effects of the first and allows us to create a hypothesis in which *the activity in a reduced zone, can be mapped with a reduced number of electrodes around such a zone*. Here, the forward reduced problem can be represented by the following equation:

$$y_{rz_k} = M_{rz} x_{rz_k} + \varepsilon_k \tag{6}$$

where y_{rz_k} represents the measurements in the reduced set of electrodes and x_{rz_k} the activity in the sources that belong to the target zone.

EXPERIMENTAL SETUP

We performed two experiments to evaluate the presented hypothesis and explore the use of the zone-based partial model in a BCI application. We used a database presented and used in [3] for RGB color classification, which contains the information of the electrodes P3, P4, O1, and O2. The tests are described below:

Test 1: Mapping the Occipital region:
We evaluate the hypothesis by simulating the activity of one random source in the aforementioned region and calculating the EEG with 60 electrodes (Fpz, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-54

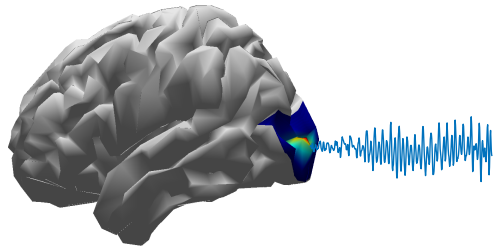


Figure 1: Simulated activity in the occipital region using the full brain model.

FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, and Iz) according to the standard 10-10 layout, using equation 1 involving the standard head model presented in [9], in which a mean head model from a 152 participant MRI study was computed using the FEM method, considering the position of 231 electrodes. The zone model was computed for the occipital and part of the parietal regions. Fig.1 shows an example of simulated activity in the target zone using the full brain forward equation. The simulated activity corresponds to normal activity based on a continuum brain model of the neuronal activity presented in [10], which was used in [6] and [11] performing a sampling using the model in Eq.7, in which the activity is represented in discrete time.

$$\frac{1}{\gamma^2}\frac{d^2\boldsymbol{x}(t)}{dt^2}+\frac{2}{\gamma}\frac{d\boldsymbol{x}(t)}{dt}=c_1\boldsymbol{x}(t)+c_2\boldsymbol{x}(t-t_0)+c_3\boldsymbol{x}^2(t)+c_4\boldsymbol{x}^3(t)+\eta \tag{7}$$

where the constants C_i with $i = 1,2,3,4$ represent the instantaneous feedback due to nearby neurons and delayed feedback via an extra-cortical loop, t_0 is a time delay, and γ is a characteristic decay rate of the field activity $\boldsymbol{x}(t)$. η is random white noise and considered to be an external stimulus that perturbs the brain states. With the activity x_k equation 1 was computed to obtain an EEG. Then, noise was added in the electrodes using a signal-noise-ratio (SNR) of 50db, after which, the brain mapping was performed in two ways: the first, using the full head model with 2000 distributed sources and 60 electrodes, and the second, using the set with the occipital model with 622 distributed sources, and the P3, P4, O1, and O2 electrodes as the EEG dataset in [3]. In both cases, the MSP and IRA L2 methods were applied to estimate neural activity. The overall process followed is summarized in the Fig.2

Five thousand trials of random activity were generated. The error measurements to evaluate the performance of the algorithms with the zone-based partial brain model and full brain model were: the localization error of the estimated source and the relative error between the simulated and estimated activity, using the following expressions:

$$RelE = \frac{||\hat{x}_k - x_k||_2^2}{||x_k||_2^2} \tag{8}$$

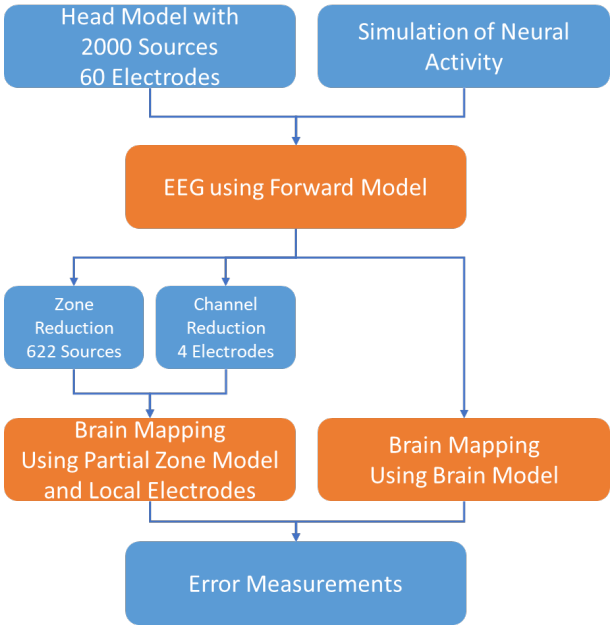


Figure 2: Test 1: Flowchart explaining the zone-based partial brain model and its validation procedure.

$$LocE = ||\hat{P}_x - P_{\hat{x}}||_2 \tag{9}$$

where P_x represents the 3D position in a coordinated space of the simulated activity and $P_{\hat{x}}$ the position of the maximum amplitude source in the estimated activity.

Test 2: RGB ERP classification using the occipital region-based partial brain model:

A dataset of seven subjects exposed to red, green, and blue visual stimuli was used in this experiment, as presented in [3]. The event-related potentials (ERP) were taken with a g.tec g.MOBIlab+ portable device sampled at 256 Hz, for which each color was randomly presented 60 times, resulting in 60 trials of each color for each subject. These data were used to perform brain mapping using an occipital lobe and part of parietal lobe partial brain model with 622 distributed sources, involving the four aforementioned electrodes. The inverse solutions were performed using the MSP and IRA L2 methods to obtain the source space. The mapped source space was divided into 64 and 81 sub-regions, using the mean of the sources involved in each one of the sub-regions as features for classification. The data were normalized and then classified using the naive Bayes and decision tree classifiers techniques with 10-fold cross-validation.

The procedure followed for the test, feature extraction and classification of the activity are summarized in Fig. 3 and how the division was performed for the 81 features is depicted in Fig. 4 depicts . As shown in the images, there were no sources involved in the upper corners. Thus, these features were removed and not considered for the classification process.

An evaluation of computational time performing brain mapping, feature extraction, and classification are part of

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-54

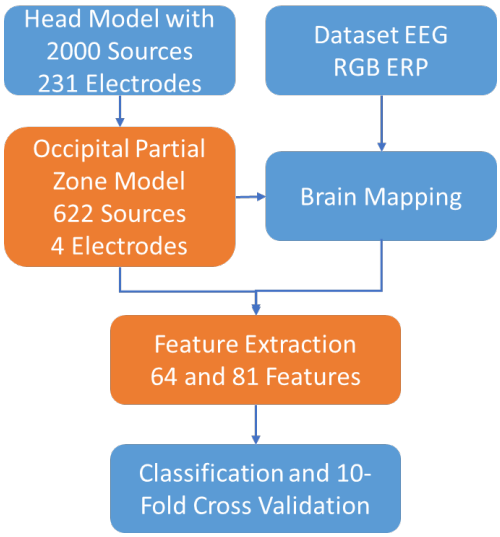


Figure 3: Test 2: Flowchart explaining the RGB ERP classification procedure.

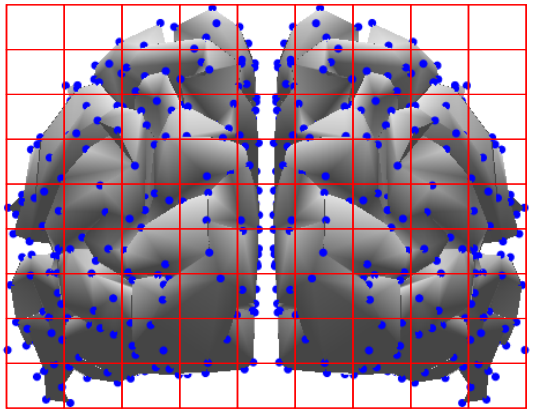


Figure 4: Occipital Zone division for feature extraction.

the study. To measure the execution times and run the algorithms, a computer with the following characteristics was used: 16GB RAM, Core i7-4790 processor, OS windows 10 64-bit, and Matlab® 2016b.

RESULTS AND DISCUSSION

Test 1 Results: The results from the first test are shown in Figs. 5, 6 and 7, in which the Fig. 5 shows the mean of the localization error and the relative error, measured between the simulated and estimated sources with the two applied methods for the reconstruction with 60 electrodes using the full brain model and four electrodes using the occipital model. As expected, the localization errors were higher if the zone model was used. The the error using the zone model was approximately 23mm, whereas the mean separation between was approximately 13 mm. This can be explained by the loss of information due to the reduction in the number of electrodes.

The localization error can be reduced by increasing the number of sources in the brain model. Nonetheless, the activity was clearly identifiable in all the cases (6).

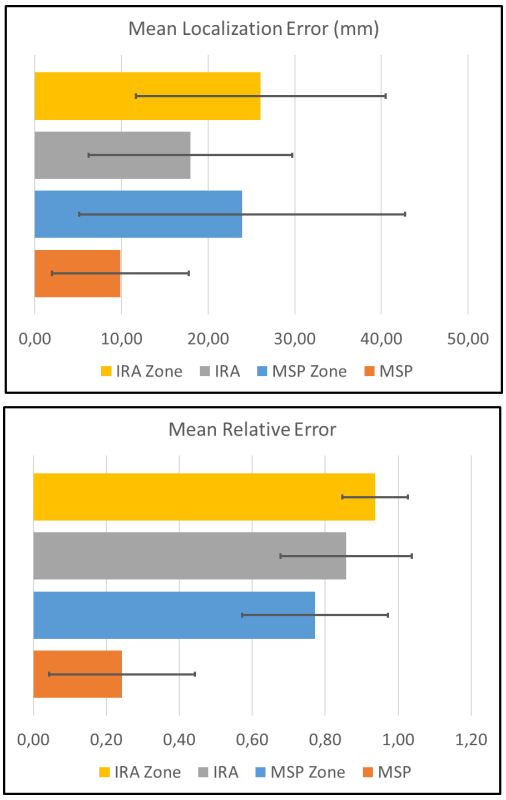


Figure 5: Mean localization error (Top) and mean relative rrror (Bottom) across the 5000 trials with the proposed zone-based partial brain model and MSP and IRA L2 methods.

MSP with the full brain model achieved the lowest relative error, because MSP gives the best results in experiments with one source estimation, as explained in [12]. However, the partial model with MSP obtained a lower error value than IRA L2 method with the full brain model. In addition, the activity with the partial and full brain models showed a highly similar pattern, with similar amplitude, using MSP (Fig. 7). The activity with IRA L2 showed a simmilar pattern but the amplitude was attenuated, explainning why the relative error with IRA L2 was higher in the two cases (Fig. 7).

A reconstruction between the simulated activity and the solutions with the full brain and partial occipital brain model for visual comparison, performed with the MSP method, showed the simulated and reconstructed activity to be highly similar with a only small error in source localization (Fig. 6). The simulated activity and each of the reconstructions results obtained using MSP and IRA L2 are shown in Fig. 7. The MSP reconstruction with the brain model overlapped the simulated activity. The other reconstructions followed a pattern similar to the original, with the MSP reconstruction using the zone model showing a smaller difderence in amplitude than for the other reconstructions.

The differences in amplitude between IRA L2 reconstruction and the simulated source can be explained by the fact that the IRA L2 solutions are smooth, which means that the source amplitude of the true source is represented by a combination of amplitudes of several

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-54

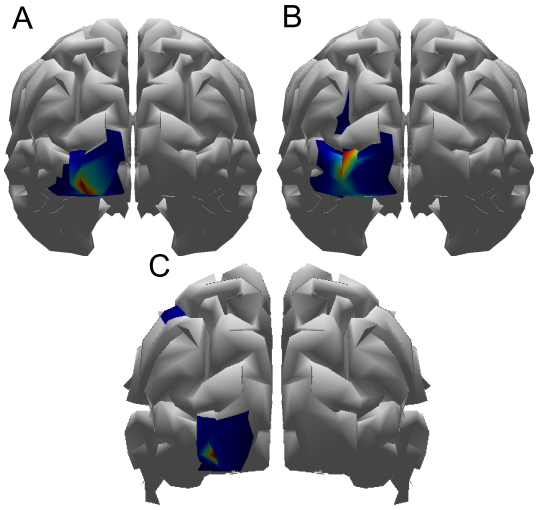


Figure 6: Simulated activity (A), estimated activity using the full brain model with MSP (B) and estimated activity using the zone model with MSP (C).

neighbors sources in the reconstruction and the source with the maximum mean activity in time is used for calculating the relative error.

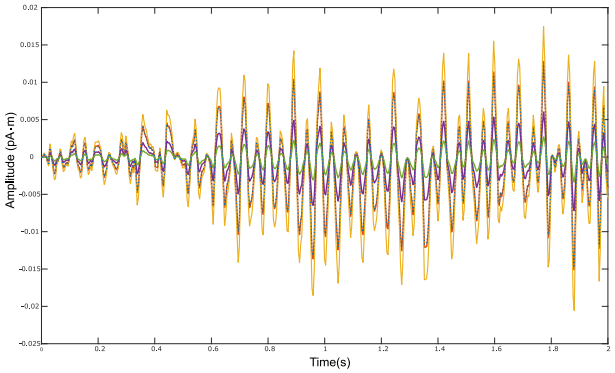


Figure 7: Simulated activity (blue dots), MSP (orange) and IRA L2 (purple) reconstructions using the full brain model and MSP (yellow) and IRA L2 (green) reconstruction using the zone model.

Test 2: RGB Classification Results: The classification results of the second test are summarized in 1, in which the naive Bayes classifier achieved and accuracy of 99.6% and 100% with 64 and 81 features, respectively. The decision tree classifier achieved a 100% for both number of features. These results shown that each of the RGB colors can be uniquely represented as a combination of activity in terms of localization and intensity as shown in Fig. 8 in which the mean of the 64 features for the seven subjects and 420 trials is presented for each color.

Table 1: Classification accuracy with the two classifiers, and feature numbers for RGB visual stimuli.

Feature Numbers	Naive Bayes	Decision Tree
64	99,6% ± 0,3	100% ± 0
81	100% ± 0	100% ± 0

The computational time for brain mapping, feature extraction, and classification using the two inverse solutions for a three-second window of the trial is shown in Table 2. In summary, a solution that includes the aforementioned steps can be performed in approximately 150 ms for the longest combination, using MSP for brain mapping and naive Bayes for classification, and 45 ms for the shortest combination, using IRA L2 and the decision-tree for classification. These results show that the methodology proposed here is feasible for BCI implementation and general online applications.

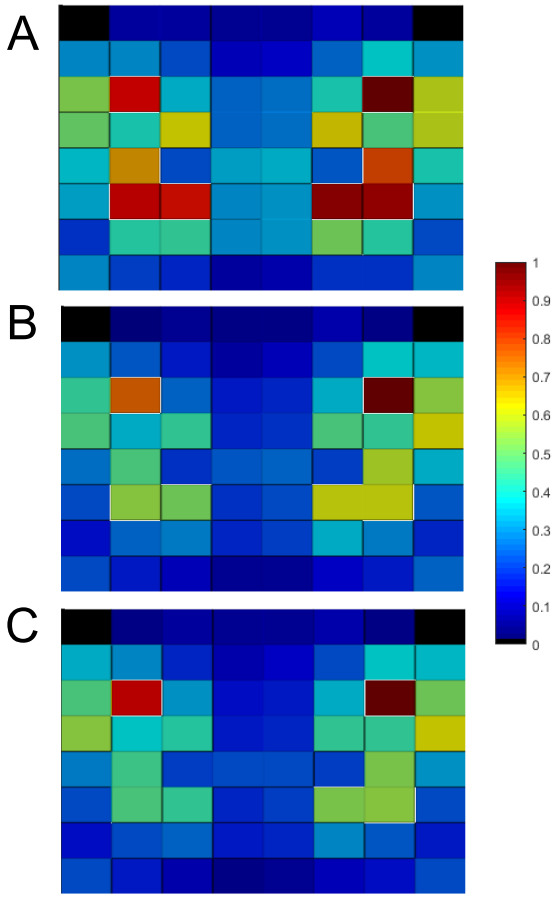


Figure 8: Mean of 64 features for the seven subjects for red (A), green (B) and blue (C).

Table 2: Computational times for brain mapping, feature extraction, and classification steps in the RGB classification.

	Time (ms)	
Process /Methods	MSP zone	IRA L2 zone
Brain Mapping	101,8 ±5,2	27,3 ±3,2
Feature Extraction	12,1 ±0,2	
Naive Bayes Classification	35,1 ±0,2	
Decision Tree Classification	4,9 ±0,1	

CONCLUSIONS AND FUTURE WORK

We performed experiments with the zone-based partial brain model methodology using synthetic EEG and visual ERP. We studied the behavior of the estimations using a reduced number of local electrodes. Based on the

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-54

accuracy of the results presented herein, activity can be mapped using a zone-model and electrodes around this area for the given stimuli.

Our classification results showed high accuracy in relatively short computational times, suitable for BCI applications. These encouraging results show promise for the extension of the same methodology to other brain zones, depending on the neuro-paradigm studied. Some immediate applications could be the classification of motor cortex activity in motor or imaginary motor tasks, diverse visual and auditory stimuli, and emotions.

The signature of the RGB colors can be represented as a combination of intensities of activities in the sources located around the visual cortex in the occipital lobe. This test can be extended for the evaluation of more colors and their differentiation in the cortical space.

The next step in this research will be the experimental demonstration of the uniqueness of the RGB signatures in device actuation. An ON/OFF actuation command will be sent to actuate a device using the signatures obtained from RGB stimuli on-line with the proposed method. The partial brain model approach will be further evaluated for various neural activities to assess the real-time capabilities of the method for a wider set of human activities and external stimuli.

ACKNOWLEDGMENTS

This work was funded by the Norwegian University of Science and Technology NTNU project "David and Goliath: single-channel EEG unravels its power through adaptive signal analysis" where a concept of FlexEEG is being developed for scanning the activity inside the brain with a reduced number of electrodes using a moving robotic headset over the scalp.

REFERENCES

[1] Duque-Muñoz L., Vargas F., López J. D. Simplified EEG inverse solution for BCI real-time implementation. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016, 4051–4054.

[2] Ahn Minkyu, Hong Jun Hee, Jun Sung Chan. Feasibility of approaches combining sensor and source features in brain–computer interface. Journal of Neuroscience Methods. 2012;204(1):168 –178.

[3] Rasheed Saim, Marini Daniele. Classification of EEG Signals Produced by RGB Colour Stimuli. Journal of Biomedical Engineering and Medical Imaging. 2015;2(5):56.

[4] Hyder R, Kamel N, Tang T B, Bornot J. Brain source localization techniques: Evaluation study using simulated EEG data. 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES). 2014 (;December):942–947.

[5] Pascual-Marqui R D, Michel C.M., Lehmann D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of Psychophysiology. 1994;18(1):49–65.

[6] Giraldo-Suarez E., Martinez-Vargas J. D., Castellanos-Dominguez G. Reconstruction of Neural Activity from EEG Data Using Dynamic Spatiotemporal Constraints. International Journal of Neural Systems. 2016;26(07):1650026.

[7] Antelis Javier M., Minguez Javier. DYNAMO: Concurrent dynamic multi-model source localization method for EEG and/or MEG. Journal of Neuroscience Methods. 2013;212(1):28 –42.

[8] Friston Karl et al. Multiple sparse priors for the M/EEG inverse problem. NeuroImage. 2008;39(3):1104–1120.

[9] Huang Yu, Parra Lucas C., Haufe Stefan. The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting. NeuroImage. 2016;140:150 –162.

[10] Kim J. W., Shin H.-B., Robinson P. a. Compact continuum brain model for human electroencephalogram. Proceedings of SPIE. 2007;6802:68020T–68020T–8.

[11] Soler Andrés Felipe, Muñoz-Gutiérrez Pablo Andrés, Giraldo Eduardo. Regularized State Observers for Source Activity Estimation. In: Brain Informatics. Ed. by Wang Shouyi et al. Cham: Springer International Publishing, 2018, 195–204.

[12] Soler A. F., Giraldo E. Non-Linear Iterative Regularization Algorithm for Neural Activity Reconstruction from EEG Data. In: 2018 IEEE ANDESCON. 2018, 1–4.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-55

CORRELATIONS BETWEEN THE LATERALITY COEFFICIENT AND FUNCTIONAL SCALES IN STROKE PATIENTS

M. Sebastián-Romagosa[1,3], R. Ortner[1], J. Dinarès-Ferran[1], C. Guger[2]

[1] g.tec medical engineering Spain SL, Barcelona, Spain
[2] g.tec medical engineering GmbH, Graz, Austria
[3]Neuroscience Institute, Universitat Autònoma de Barcelona, Spain

E-mail: sebastian@gtec.at

ABSTRACT: Motor impairments are the most common and incapacitating consequences for the stroke survivors. As of today, the effects of this pathology in the central nervous system as well as the brain proprieties to restore the function are still not fully understood. The conventional physical therapy techniques are limited and sometimes have an innocuous effect for non-cooperative or strongly impaired patients who only can receive passive movement treatments. Brain Computer Interface (BCI) systems are adding new possibilities for the stroke patient's rehabilitation, helping the patients in the relearning process of lost movements, and inducing neuroplastic changes in the affected motor cortex. The electrical brain signals can provide valuable information about the brain functions, thence the BCI systems can process these signals to understand what is happening in each situation. The event-related synchronization and even-related desynchronization (ERD/ERS) calculated with the brain signals during the motor imagery tasks, could be related with the functional state of the stroke patients. The Laterality Coefficient (LC) is a parameter calculated using the ERD/ERS changes in the mu wave. Twenty-six stroke patients with hemiparesis in the upper limb have been enrolled on this study and performed 25 sessions of BCI therapy. All of them performed assessment sessions before and after the therapy. The results showed significant correlation between the LC and functional scales like the Fugl-Meyer Assessment (FMA) or Box and Block Test (BBT). The findings of this experiment suggest that the LC parameter could be a good biomarker for the functional state of stroke patients.

INTRODUCTION

Stroke is one of the most prevalent pathologies around the world, with severe effects to the motor and sensory system that hinder the daily living activities. The major part of the stroke patients needs a long rehabilitation process to overcome the hemiplegia and adapt again to the environment. The conventional rehabilitation techniques have a roof effect to get a complete degree of rehabilitation. New technologies like the Brain Computer Interfaces (BCI) are important tools to improve the functional results of the rehabilitation process. The BCI systems are able to measure the brain activation and to generate a control signal for external devices in real-time [1], [2]. After the stroke, the brain signals do not follow a normal activation, usually the affected cortex presents less excitability due to the change in the cortical representation areas and other physiological alterations on the nervous tissue [3], [4].

However, BCI systems can help the stroke survivor to relearn the lost movements, using EEG signals during Motor Imagery (MI) exercises [5]. The detected brain oscillations can be used to move a virtual reality avatar or trigger a functional electrical stimulator device to reproduce the imagined movement with the paretic limb (e.g. [6], [7]). This way it provides the patient a closed loop feedback to ease the motor learning process.

During the MI tasks the patient should concentrate on performing an indicated movement mentally. At this moment typical brain waves appear in the EEG. During MI, the contralateral motor cortex produces a desynchronization (event-related desynchronization or ERD) of motor neurons, showing a decrease in the EEG amplitude in the frequency of 8-13 Hz (mu frequency rhythm). When the imagery period is finished, the contralateral motor cortex restores the synchronization state (event-related synchronization or ERS) and increases again the amplitude of the EEG ([3], [4], [8], [9]).

Considering the stroke patients do not present normal brain signals, the ERD and the ERS patterns could be atypical as well. Kaiser et al. [4] investigated the relation between these patterns versus the patient's functional state and spasticity, using a new parameter, the Laterality Coefficient (LC). For functional assessment they used the European Stroke Scale (ESS), the Medical Research Council (MRC) and the Modified Ashworth Scale (MAS). The LC presented significant correlations with the MRC scale and MAS. The findings of Kaiser and colleagues showed that strong ERD patterns on the contralesional hemisphere are related to a high degree of impairment [4].

The objective of this study is to find correlations between the LC parameter in alpha and beta band, calculated using the ERD/ERS patterns, with other functional scales like the Fugl-Meyer Assessment (FMA).

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-55

MATERIALS AND METHODS

Study design: Twenty-six stroke patients with upper extremity hemiparesis were recruited for this study. All these patients have been classified in four groups based on their stroke diagnosis: Cortical, Subcortical, Cortical + Subcortical and Unknown. The inclusion criteria were: i) able to understand written and spoken instructions, ii) residual hemiparesis, iii) the stroke occurred at least four days before the beginning of the study, iv) Functional restriction in the upper extremities, v) stable neurological status, vi) willing to participate in the study and to understand and sign the informed consent, vii) have the opportunity to attend meetings.
All patients have completed between 23 and 25 sessions of BCI therapy, two sessions per week. Two assessment visits have been performed by an expert clinician before and after the therapy to track the therapy effect in the functional patient state. The Pre1 assessment is performed 1 month before starting the therapy, and Pre2 assessment is performed just before the therapy starts. Post1 is performed just after the last session, and Post2 is performed one month after the last session.
Table 1 shows the used scales for the assessments. In the first column appears the scales name, the second column is the short name of each scale, the column number three shows a short description of each scale and the last column presents the worst and best possible score. For the Fahn Tremor Rating Scale (FTRS) and BBT we have assessed both hands. For the BBT, the patient is asked to move as many blocks as possible from one box to the contralateral box in less than 1 minute. In the case that the patient cannot move any block, the final score would be 0.
The FMA scale has two different parts; the first part (up to 66 points) is for the motor assessment, and the other part is for the assessment of the sensation (up to 12 points).

BCI System: The BCI system used on this study is recoveriX® (g.tec medical engineering GmbH, Austria). The recoveriX system combines the visual feedback using a virtual reality avatar with a proprioceptive feedback using functional electrical stimulation (FES).
Every patient performed 25 sessions of BCI training. The patient was seated in a comfortable chair with the arms on the table. In front of the patient was a computer screen, showing two hands in virtual reality. The total time of one session was about 60 minutes, including preparation and cleaning. Every session was composed by up to 3 runs of 80 trials, depending of the patient's fatigue. Patients wore EEG caps with 16 active electrodes (g.LADYbird or g.Scarabeo, g.tec medical engineering GmbH). The electrode positions were according to the international 10/10 system (extended 10/20 system): FC5, FC1, FCz, FC2, FC6, C5 C3, C1, Cz, C2, C4, C6, Cp5, Cp1, Cp2, Cp6. A reference electrode was placed on the right earlobe and a ground electrode at position of FPz.

Two FES electrodes were placed on the skin over wrist extensors of the left and right forearms. The stimulation

Assessment scales				
Scale name	***Short name***	***Description***	***Score***	
			Worst	***Best***
Barthel Index	BI	Daily living activities	0	100
Fahn Tremor Rating Scale	FTRS[a]	Degree of tremor	12	0
Modified Ashworth Scale	MAS	Spasticity	4	0
Box and Block Test	BBT[a]	Grasp	Block's number	
Fugl-Meyer Assessment	FMA_motor	Motor function on upper limb	0	66
	FMA_sens	Sensation of Upper extremity	0	12

Table 1: Scales used in the assessment visits

parameters (g.Estim FES, g.tec medical engineering GmbH, Austria) were adjusted for each patient and session individually, to find the optimal passive movement without pain.
The frequency was set to 50 Hz, the pulse-width to 300µs.
Then, the therapist increased the current amplitude until the optimal stimulation point was observed.
The sequence of trials (motor tasks) was specified by the recoveriX software in pseudo random order. One single motor task is depicted in Fig. 1.
The patients first heard an attention beep. Two seconds later, an animated arrow with spotlight to the expected hand for motor imagery indicated the task of each trial with an auditory instruction saying either "left" or "right" in the patient's mother tongue. The patient started to imagine the movement and recoveriX processed the EEG using the features from a Common Spatial Patter (CSP) filter and using a Linear Discriminant Analysis (LDA) classifier to infer which hand the patient is imagining. If recoveriX detected the appearance of the expected hand side, FES and avatar feedback were activated during the feedback phase. Feedback was otherwise deactivated. Updating the feedback was carried out five times per second. The animated forearm movement in avatar simultaneously performed the similar wrist dorsiflexion as produced by FES. The full recoveriX system is described in Figure 2.
Both hands are trained, the patient should learn the strategy used with the healthy hand to move the affected one. This is a key point for a correct embodiment and activate the motor learning process.

Laterality coefficient analysis: The EEG raw data recorded during the recoveriX sessions has been used to calculate the LC parameter.
The LC coefficient (1) is calculated for each session twice: one time for trials of MI of the paretic (p) hand and another time for the trials of the healthy (h) hand.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-55

Figure 1: Timing of one trial.

Figure 2: Components of the recoveriX system.

$$LC_{p/h} = (C\text{-}I) / (C+I) \quad (1)$$

Where C and I refer to the contralateral and ipsilateral values of the ERD/ERS patterns during the MI. C and I are calculated following these steps: 1) band filtering (8-13 Hz or 13-30Hz) of the EEG signal on the C3 and C4 electrodes. 2) Artifact rejection. 3) Laplacian derivation using the surrounding electrodes. 4) Calculate ERD/ERS patterns according to [9]. 5) Summation of all ERD/ERS values from second 2 until the end of the ERD map (second 8). And 6) apply the formula to obtain the LC coefficients.

RESULTS

Participant baseline information: The mean age of the participants was 61.5 years (±12.8), the maximum age was 86 years, and the minimum was 33 years old. The mean time since the stroke was 4.2 years (±4.8), the maximum time since stroke was 24 years, and the minimum 10 months. In terms of kind of stroke; fourteen patients had a subcortical stroke, one had a cortical stroke, five a mixed cortical+subcortical and for six of them the kind of stroke is not clear. From the total number of patients, eight of them presented a hemiparesis on the right side, and eighteen on the left side.

Functional scales: The FMA mean before the therapy was 23.08 points, with an SD of ±16.99 points. The highest possible FMA score is 66. The BI mean was 78.46 (±21.45) points, the mean FTRS of the healthy hand was 0.53 (±2.00) points, the one of the paretic hand 8.98 (±4.83) points. The mean of the MAS scale of the wrist was 1.76 (±1.34) points, the MAS of the fingers 2.11 (±1.12) points. The mean of the BBT of the healthy hand was 56.67 (±14.38) boxes, and the same scale with the paretic hand was 4.96 (±11.90) boxes.

LC variance: Fig. 3 shows the variance of the LC parameter in alpha and beta band. In both bands, the LC of healthy hand (LCh) is strongly related with the results of the LC of paretic hand (LCp).

LC during BCI therapy: Fig. 4 shows the delta of the LC parameter calculated in the alpha band during the BCI therapy. Typically, the LC delta in both hands goes to values near 0 through the therapy.
The LCh and LCp delta are a mean of the delta of each participant. Two of these participants only performed 22 sessions of BCI therapy, for external problems no

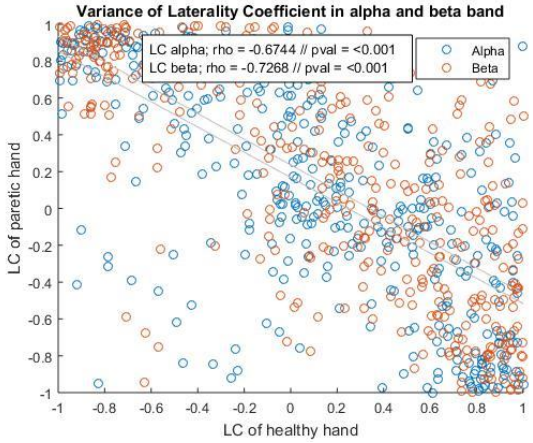


Figure 3. Variance of LC in alpha and beta band

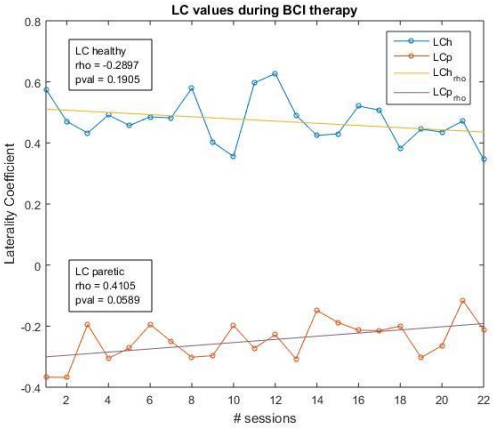


Figure 4. LC parameter in alpha band during BCI therapy

related with the study. For this reason, the Fig.4 only shows the LC behavior between session 1 to session 22.

Correlation with the functional scales: Statistical analysis was performed using MATLAB R2015a. The Kolmogorov-Smirnov Test showed that this data does not follow a normal distribution. Hence, for the statistical analysis we have used a non-parametrical method, the Spearman Test.
No significant correlations have been found between the functional scales and the LC of the beta band.
The average of the LC calculated during these 25 sessions in the healthy hand (LCh), shows significant correlations with the FTRS score of the paretic hand, also with the BBT score of the paretic hand, and the FMA of the motor part and the FMA of the sensation part.
The average of the LC calculated during the therapy in the paretic hand (LCp) shows correlations with the BBT of the paretic hand, and the different parts of the FMA scale.

Laterality Coefficient of the healthy hand: Fig. 5.A shows a moderate relation between the tremor degree of the paretic hand using the FTRS, and the LC parameter of the healthy hand. This correlation is present in all the assessment (pre and post) less on Pre2 assessment, where the p-value is near to 0.05.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-55

Finally, there are significant correlations between this parameter and the FMA scale. On the Fig. 5.B appears the relationship between the LCh and the FMA_motor. Fig.5.C shows the relation of this parameter and the sensation part of the FMA scale.

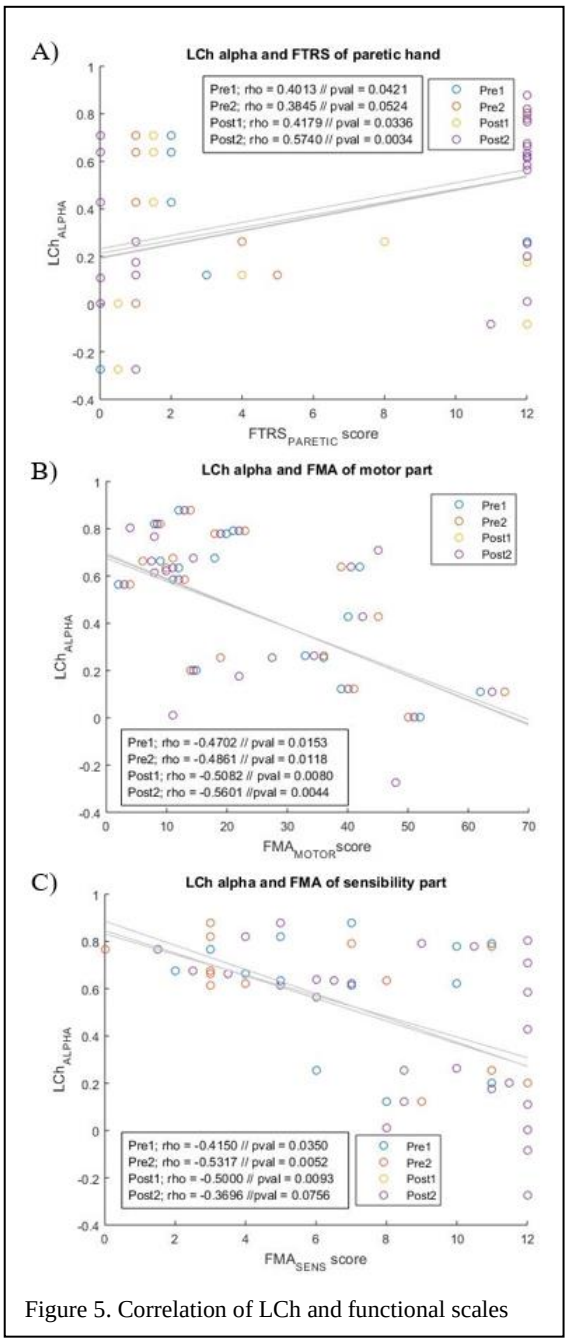


Figure 5. Correlation of LCh and functional scales

Laterality Coefficient of the paretic hand: The BBT on the paretic hand, also present a significant relation with the LCp, but in this case the correlation is present in all the assessment except on Pre2 (Fig. 6.A).

In Fig. 6.B appears the correlation between LCp and the FMA_motor. The correlation is only significant in the two post assessment visits, in the case of the other non-significant cases (Pre1 and Pre2), the p-value is near to 0.05 points.

Finally, in Fig. 6.C appears the correlation with the sensation part of the FMA scale.

The correlations present coherence amongst them. The correlation coefficients express that the high levels of functionality are related with LCp values near to 0, and the low functional levels are related with very negative

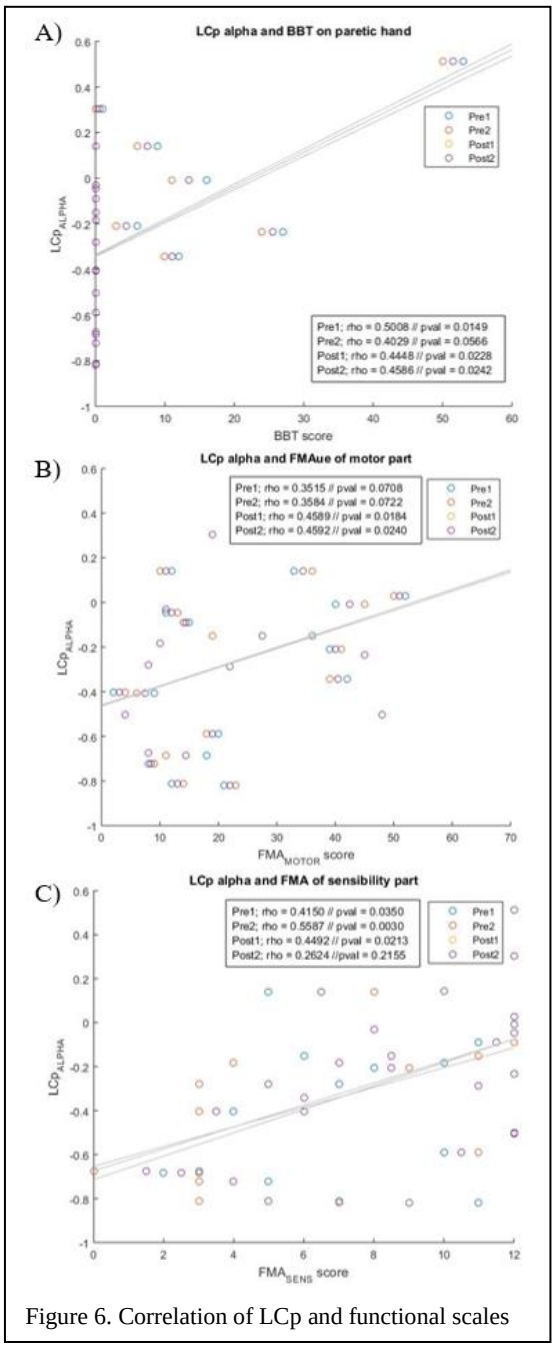


Figure 6. Correlation of LCp and functional scales

LCp values. Furthermore, the LCh have the opposed sign on the correlation coefficient with the scales.

In terms of the used functional scales, the high scores are related with values near to 0 in LCh and LCp, and the low scores with values near to 1 of LCh, and -1 in LCp. The FTRS is a special case of this typical positive or negative trend related with the LCh or LCp, because on the FTRS the high score is related with high degree of tremor.

DISCUSSION

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-55

The objective of this study was to find correlations between the LC parameter in the alpha band, calculated using the ERD/ERS patterns with the functional state of stroke patients. For this, we analyzed 26 stroke patients who performed 25 sessions of therapy with BCI system.
Usually the EEG parameters present high variability, but this is not the case for the LC parameter, as Fig.3 shows.
The LCh in alpha band shows significant correlations with the tremor degree, with the global functionality of the upper extremity and with the sensation part of the FMA. In the other hand, the LCp in alpha band shows a marked correlation with the grasp functionality (Fig. 6.A), with the global motor function in the upper extremity (Fig. 6.B) and the sensation degree (Fig.6.C).
The general rule that can be applied to all these correlations is: LC values near to 0 points are related with high functional degree. LCh values near to 1 and LCp values near to -1 are related with poor functional degree.
The first important result to point out is that our significant results of the LC against the MAS are not similar to the results presented by Kaiser et al. The different kind of stroke patient, or the sample size could explain this.
Another important finding is the correlation with the FMA motor score. The FMA is a very extended scale, used to evaluate the patient's functional state. FMA has been validated many times by many researchers, and the correlations between this scale with EEG features are not common. This correlation is especially interesting because it could mean that the quantification of the cortical activation, using the LC parameter is related to the peripherical motor performance. This relation is present in the affected hemisphere and also in the healthy hemisphere. The healthy hemisphere is not related directly to the motor activity of the paretic side, but for the LC calculation it is necessary using and compare the signals of both hemispheres. This is a reason why the LCh are important values for the assessment of the paretic side. Even though the sample size in our study is too small to give conclusive results, it is worth to point out the significance of this finding.
And last but not least, the LC alpha also presented a strong correlation with the FMA sensation scale part.
The superficial sensitivity and the proprioception are essential players on the BCI systems. The patients should feel as much as possible the feedback that the system provides for a correct closed loop interaction. Only if a correct synchronization between the intention of movement and the real feeling of this movement is provided the motor learning process is optimal [10]. This is only possible with BCI, and this is the greatest limitation of the conventional therapy techniques like the mirror therapy.
The other used scales of this study did not present significant correlations with the LC parameter. Again, the sample size of our study could be a limitation to find such correlations.
Concerning the LC of the beta band, it shows only some isolated significant correlations with the scales.
Further studies with more patients will be needed to confirm these correlations and to find out how useful the LC parameter is in the daily clinical practice.

CONCLUSION

The results of this study suggest that the LC parameter, calculated using the ERD/ERS of the stroke patients could be related with the Fugl-Meyer Assessment scale. This study opens the door to find more correlations between the EEG parameter with the patient's functional state.

ACKNOWLEDGMENT

We appreciate the collaboration of the Ministry of Business and Knowledge of the Government of Catalonia that partially supported this study (ACCIO RD15-1-0020 project and the Industrial Doctorates Plan). This study was also supported by the H2020-ESCEL Project Astonish (692470-1), the EC SME Phase 2 project recoveriX, and the MSCA-RISE grant Progait (agreement No 778043).

REFERENCES

[1] D. J. McFarland and J. R. Wolpaw, "EEG-based brain–computer interfaces," *Curr. Opin. Biomed. Eng.*, vol. 4, pp. 194–200, Dec. 2017.
[2] B. H. Dobkin, "Brain-computer interface technology as a tool to augment plasticity and outcomes for neurological rehabilitation: BCI for rehabilitation," *J. Physiol.*, vol. 579, no. 3, pp. 637–642, Mar. 2007.
[3] C. Neuper, M. Wörtz, and G. Pfurtscheller, "ERD/ERS patterns reflecting sensorimotor activation and deactivation," *Prog. Brain Res.*, vol. 159, pp. 211–222, 2006.
[4] V. Kaiser, I. Daly, F. Pichiorri, D. Mattia, G. R. Muller-Putz, and C. Neuper, "Relationship Between Electrical Brain Responses to Motor Imagery and Motor Impairment in Stroke," *Stroke*, vol. 43, no. 10, pp. 2735–2740, Oct. 2012.
[5] M. A. Cervera *et al.*, "Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis," *Ann. Clin. Transl. Neurol.*, vol. 5, no. 5, pp. 651–663, May 2018.
[6] W. Cho *et al.*, "Hemiparetic Stroke Rehabilitation Using Avatar and Electrical Stimulation Based on Non-invasive Brain Computer Interface," *Int. J. Phys. Med. Rehabil.*, vol. 05, no. 04, 2017.
[7] D. C. Irimia *et al.*, "Brain-Computer Interfaces With Multi-Sensory Feedback for Stroke Rehabilitation: A Case Study: BCI for Stroke Rehabilitation," *Artif. Organs*, vol. 41, no. 11, pp. E178–E184, Nov. 2017.
[8] G. Pfurtscheller and A. Aranibar, "Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movements," *Electroen Clin Neuro*, vol. 46, pp. 138–146, 1979.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-55

[9] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller, "Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data," *Clin. Neurophysiol.*, vol. 113, no. 1, pp. 43–47, 2002.

[10] A. Ramos-Murguialday *et al.*, "Brain-machine interface in chronic stroke rehabilitation: A controlled study: BMI in Chronic Stroke," *Ann. Neurol.*, vol. 74, no. 1, pp. 100–108, Jul. 2013.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-56

LONG-TERM HOME USE OF A FULLY IMPLANTED BCI FOR COMMUNICATION: VISUAL AND AUDITORY SPELLING

E.J. Aarnoutse, S. Leinders, Z.V. Freudenburg, B.H. Van der Vijgh, E.G.M. Pels, M.P. Branco, N.F. Ramsey, M.J. Vansteensel

UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, Utrecht, The Netherlands

E-mail: e.j.aarnoutse@umcutrecht.nl

ABSTRACT: We report a case study where a user provided with a fully implantable Brain-Computer Interface (BCI) has been using the BCI at home for almost three years for communication. At the start of participation ECoG electrodes were placed subdurally over motor cortex and signals are amplified and transmitted to a tablet with BCI software. This software processes and translates the brain signal to control AAC software. To select letters on the tablet, the user generates clicks by attempting hand movement. Since the start of home use, the participant used the BCI on average during 11 sessions per month with an average duration of 21.2 hours per month. Use of the BCI correlates with situations in which other means of communication are more difficult.
An auditory spelling paradigm was investigated for future use. Auditory spelling was reliable, but speed was lower than vision-based spelling.
This report emphasizes that a BCI has an added value to a user.

INTRODUCTION

Brain-Computer Interfaces (BCIs) that restore communication promise greater autonomy and independence for people with locked-in syndrome (LIS). BCIs will give them the possibility to self-initiate communication with family, friends and caregivers.
Over the last decades BCIs for communication have developed into systems with increasingly better performance, with users and researchers gaining more experience with BCIs. However, the proof of the pudding is in the eating, i.e. a BCI can only be perceived as an added value for a user if the user is actually using the BCI at home.
Abandonment of assistive technology is a concern for augmentative and alternative communication (AAC) technology. This is true for BCIs as well. A study of home use by people with LIS due to late stage Amyotrophic Lateral Sclerosis (ALS) with EEG based BCIs [1] showed that not all participants kept their BCI at home for future use, even though reliability of the BCI was high. The use of this BCI was discontinued for several reasons, including preference for other methods of communication. In this study 24% of potential users could not use vision-based BCIs, however, the vast majority of BCIs rely on visual feedback. Auditory BCIs do exist, but communication is slower than with visual BCIs [2].
Here we report a case study of a user with an implanted BCI who has been using the system at home for almost three years. The system is fully implantable BCI and wirelessly interfaces with a tablet with speech synthesis for communication [3]. The system was designed to be used at home by the user without help from experts, with 24/7 functionality. Due to progression of the disease over the three years, other means of communication became more difficult and use of the BCI increased. In addition, on request of the user and caregivers, auditory BCI was investigated for future use to mitigate increasing problems with closing the eyelids.

MATERIALS AND METHODS

Participant: The user is a woman with late stage ALS, who was diagnosed with the disease in 2008 and required positive-pressure mechanical ventilation through a tracheostomy. At the start of the study she had a score of 2 on a scale of 40 on the Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS). She gave informed consent in September 2015, when she was 58 years old. The study was approved by the MREC Utrecht and was conducted in accordance with the Declaration of Helsinki (2013). The study protocol allows for yearly extension of participation, which the participant did for three consecutive years.
Her methods of communication in September 2015 were rudimentary speech, an AAC device controlled by eye-movements with spelling software (Tobii, Communicator-5, Tobii Dynavox) and eye-blinks for answering yes/no to closed questions. Since then her ability to use these means of communication declined.

Implanted BCI: In October 2015 the participant was implanted with a BCI system as the first participant in the Utrecht Neuroprosthesis study.
The Utrecht Neuroprosthesis records ECoG signal from two subdural electrodes strips (Resume II®, Medtronic, 4 electrodes each, 4mm diameter, 1cm distance, off label use) which were implanted through burr holes (1cm diameter) over the hand area of the left motor

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-56

cortex and over the left dorsolateral prefrontal cortex. The target areas were determined by an fMRI session several weeks prior to implant, where the participant performed motor and counting backwards localizer tasks. The strips were connected subcutaneously to an amplifier/transmitter device (Activa® PC+S, Medtronic, off label use), that was placed subcutaneously under the clavicle. ECoG data is streamed through an antenna to a receiver which is connected to a Microsoft Surface 3 tablet computer running BCI software derived from the BCI2000 software platform.

BCI Control: The participant attempts to move her hand to generate a click. Bandpass filtered data with center frequencies of 20Hz and 80Hz are streamed 5 times per second to the tablet where the BCI software filters the data and translates the data into a click.
The BCI software converts this click to a keypress signal in Communicator-5 AAC software (Tobii Dynavox). A matrix of letters (and a few icons for e.g. backspace and pronunciation of a written word or sentence) is presented and a moving rectangle highlights the letters one by one. Several schemes are used to speed up spelling, e.g. selecting letters in two steps (rows, columns) or three steps and predicted words are presented as an option to select. The rectangle moves at a fixed scan rate of once per 2.4 s. Recently, the speed is increased, the scan rate is now once per 2.0 s. The user decides to pronounce by selecting the pronounce icon.
In the auditory version the letters are announced one by one, by a different voice than when the spelling result is pronounced. Similar to the visual interface, letters can be selected by making a click. The scan rate is reduced to allow for announcing letters to once every 3 or 4 s.

BCI data: Frequency and duration of home use is logged, but no ECoG data is analysed to preserve privacy. Therefore, we rely on user feedback for what the BCI is used for.
Data from auditory BCI was recorded during research visits and is available to the researchers. The communication software does not allow for precise logging. The task was to copy-spell words of 3-5 letters. Accuracy is defined by (TP+TN)/(TP+TN+FP+FN). No home use of the auditory paradigm was performed yet.

RESULTS

The participant used the eye-tracker as the main means of communication for free spelling, but she reported before implant that the performance of eye-tracking was unsatisfactory outside her home due to different light conditions. Prior to eye-tracker use she preferred her voice, but unfortunately, she lost the ability to use her voice within months after giving informed consent. In autumn 2018 progression of the disease reached her orbicularis oculi muscle, causing difficulties to close her eyelid, and drying out of the cornea. With this, the use of the eye-tracker became painful.

Since May 2016, home use, outside of research visits, was on average 1270 minutes per month, which equals 21.2 hours (Fig. 1). Frequency of BCI use however was closely correlated with seasonal frequency of outdoor activities and progression of the disease. The user reports that the BCI is used mostly for spelling, and secondly to alert the caregiver.

Figure 1. Summary of home use of the BCI. The number of sessions per month since the start of home use in May 2016 until January 2019 is shown in the upper panel. The total number of minutes where the BCI was used is depicted in the lower panel. The duration of home use in the periods May-June 2016 and February-October 2017 are estimated, because logging turned out to be unreliable.
Note that in winter periods use was lower than during summer, correlated with outside use of the BCI. In the first month of home use (May 2016) many short sessions were performed as home practice. The increase of BCI use since August 2018 corresponds to decline of control over the orbicularis oculi muscle.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-56

Spelling with an auditory paradigm was performed in 12 runs. In 9 runs the word was spelled correctly, 1 run was aborted for technical reasons. In all runs, corrections were allowed. Total accuracy over 12 runs was 90.2%. On average 1.3 correct characters were spelled per minute.

DISCUSSION

In this report the user has continued use of the BCI until present, almost three years after she gained reliable control of the BCI and the system was left at her home. Commitment to use an implanted BCI is arguably larger than with a non-invasive BCI, given the required surgery. However, BCI use was mainly associated with failure to use other available methods for communication. The user preferred spelling with the eye-tracker over spelling with the BCI, because of speed of communication. In situations where the eye-tracker was less reliable (outside her home or when problems with her eyelid caused dryness of her eyes), BCI was her primary (and increasingly only) method of communication.
In this study feedback proved to be of quintessential importance to improve on the design of this BCI and BCIs in general. The user requested a function to alert the caregiver, which gave her an improved feeling of safety [4]. She also requested an auditory feedback mode, anticipating loss of control over her eyelids. Continued participation in the study is voluntary and not required for home use. Provisions for post-trial access to intervention, as required in the declaration of Helsinki, are made in this study. The BCI software in this study is developed compliant to the requirements of medical software.

Use of the BCI using an auditory paradigm showed that this compromised speed. The speed of auditory EEG based BCI is reported to be approximately 1 character per min [2]. It has to be noted that the user had ample experience with the BCI with a visual paradigm. It is unknown whether the same speed and accuracy would have been reached without prior visual feedback. These results emphasize a clear need to improve on auditory paradigms.

CONCLUSION

The implanted BCI system proves to be valuable to the user. It functions as a fail-safe method for communication when other AAC technologies are not (or no longer) accessible or available.
Auditory BCI was successful, although at a reduced but encouraging speed, and may give the user reassurance that communication remains possible when control over eyelids fails.

ACKNOWLEDGEMENTS

This study was funded by grants from the Dutch Technology Foundation STW (grant UGT7685), European Research Council (ERC-Advanced ‘iConnect’ project, grant ADV 320708) and the NIDCD of the National Institutes of Health (award number U01DC016686).
We thank our participant for all her feedback, courage, motivation, and hospitality.

REFERENCES

Boldface names denote co–first authors

[1] Wolpaw JR, Bedlack RS, Reda DJ, Ringer RJ, Banks PG, Vaughan TM et al. (2018) Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis. Neurology 17;91(3):e258-e267.
[2] Schreuder M, Rost T, Tangermann M. Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. Front Neurosci 2011;5:112.
[3] **Vansteensel, MJ**, **Pels, EGM**, Bleichner, MG, Branco, MP, Denison, T, **Aarnoutse, EJ**, et al. (2016). Fully Implanted Brain–Computer Interface in a Locked-In Patient with ALS. New England Journal of Medicine, 375(21), 2060–2066.
[4] Leinders, S, Pels, EGM , Vansteensel, MJ, Branco, MP, Freudenburg ZV, van den Boom, MA et al. (2017) Using A One-Dimensional Control Signal For Two Different Output Commands In An Implanted BCI. Proceedings of the 7th Graz Brain-Computer Interface Conference 2017, DOI: 10.3217/978-3-85125-533-1-50

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-57

A SCORE BASED METHOD FOR P300 COLLABORATIVE BCI

L. Bianchi[1], F. Gambardella [1], C. Liti [1], V. Piccialli[1]

[1]Department of Civil Engineering and Computer Science, University of Rome Tor Vergata, Rome, Italy

E-mail: chiara.liti@uniroma2.it

ABSTRACT: Group decision-making is the process where two or more people are engaged in generating a solution for a given problem. In the last decade, researchers started exploiting collaborative Brain-Computer Interfaces to enhance group performance. Various methods have been proposed to integrate EEG data of multiple users showing the improvement in group decision-making over single-user BCIs and non-BCI systems. In this study, we investigate four EEG integration strategies: EEG averaging across participants, the standard majority voting rule and two weighted voting system. For each approach, we evaluate three different scenarios varying the number of iterations necessary to perform a single selection. In all cases, it is possible to exceed 90% of accuracy with at least one collaborative BCI.

INTRODUCTION

Group (or Collective, or Collaborative) decision-making is the process where two or more people are engaged in generating a solution for a given problem [1]. Combination of sensing and cognition capabilities allow a group to make better decisions than single individuals [2]. Nevertheless, group decisions can be negatively affected by several factors, such as lacking of time, sharing of information, group and leadership style and communication biases [2, 3]. In the last decade, researchers started to use Brain-Computer Interfaces (BCIs) to enhance group decision-making. BCIs allow people to interact with the environment without requiring any peripheral muscle activity to complete the interaction [4]. Brain signals are acquired and processed by a computer to identify a particular type of neural process called event-related potential (ERP), which is the brain response resulting after specific sensory or cognitive events. ERP-based BCIs use an oddball paradigm to elicit the ERP components: the user has to focus on 'target' (rare) stimuli which are inserted in a stream of 'non-target' (frequent) stimuli. As target and non-target stimuli elicit different responses, they can be distinguished and exploited by the BCI. Single-user BCIs are widely exploited for clinical purposes, most of them aiming at restoring communication capabilities to severely disable people [5]. Instead, a collaborative Brain-Computer Interface (cBCI) is a system designed for integrating brain signals from a group of users for improving a decision-making process.

In the last decade, various approaches to integrate EEG signals have been proposed. For example, single-trial ERPs can be averaged across group members and then processed as a single-user BCI. Alternatively, neural features can be inferred from the EEG data of each user and concatenated afterwards to build a feature vector for the group, which is then passed to a single classifier. Finally, the output of several single-user BCI can be combined by means of a voting system to compute the group decision [6]. In [7], these approaches have been applied to the EEG data collected from 20 subjects in a movement-planning experiment. In the voting method, a SVM classifier was trained for each subject. The classification output was then weighted according to each user's training accuracy. All the three cBCIs outperformed the single-user BCIs. Moreover, the voting strategy turned out to be the optimal method for collaborative EEG-based classification. In [8] and [9] the same strategy has been applied to a visual target detection task and a visual Go/NoGo task, respectively. The output of the single-user SVMs has been used as the input for a second-layer SVM. In [10], the authors integrated the EEG of 20 individuals engaged in the discrimination among pictures of cars and faces, using various voting decision rules for combining information across user. The advantages of a cBCI has been evaluated also in [11], where data recorded using a P300 speller paradigm have been analyzed showing that combining data from users led to an improved accuracy with respect to fusing data from the same participant over time. In [3], a completely different weighted majority rule has been introduced. The authors developed a hybrid cBCI that does not predict the user decision but combines neural signals and response times to determine the decision confidence of each user and then weights their behavioral responses accordingly to produce the group decision. This hybrid cBCI was evaluated on several tasks, such as visual matching [3], visual search with simple shapes [12], visual search with realistic stimuli [2].
In literature, cBCIs have been tested on several visual tasks showing their reliability. Moreover, various studies suggest that voting methods are often optimal for collaborative EEG-based classification, especially when the scores of the single classifier (instead of the predicted class) are used for the integration [6]. In this work we propose a voting method for a cBCI that exploits several information behind a standard ERPs stimulation

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-57

paradigm to achieve a global group decision.

MATERIALS AND METHODS

Datasets: In this study, we test our voting strategy on a healthy-subject dataset recorded using a P300-Speller paradigm. The dataset can be downloaded from the BNCI Horizon 2020 database data09 (Dataset 9: Covert and overt ERP-based BCI (009-2014)). In a standard P300-Speller paradigm [13], cues are organized in a 6×6 matrix. Given a character (also referred as a trial) to select, each row and column of the matrix flash every 250ms in a pseudo-random order. A single flash is called stimulus. A block of twelve different stimuli, six rows and six columns, constitutes a stimulation sequence (or an iteration). Due to the low signal-to-noise ratio (SNR) of EEG signals, several iterations have to be carried out and then averaged in order to perform a single selection step. We analyzed data recorded from 10 participants that had to select a total of eighteen characters. Each selection consists of eight iterations. In summary, for each participant, we analyzed a total of 1728 stimuli (18 trials $\times$ 8 iterations $\times$ (6 columns + 6 rows)). The characters to select as well as the used pseudo-random stimulation sequences and timings are the same for all participants. Therefore it is possible to test offline the benefits of combining this EEG data to perform a group decision.

Pre-Processing and Classification: For each participant, we analyze six pairs of training and test sets. Pairs differ in terms of the number of training trials. More in detail, the number of training characters ranges from one to six. As a result, the number of test trials varies from seventeen to twelve. We use a linear Support Vector Machine (SVM) [14] to classify the response of each stimulus. This classifier discriminate brain responses by means of a separating hyperplane, that is built on the basis of the training data, and it is defined as:

$$f(x) = w^T x + b \tag{1}$$

where w is the vector containing classification weights and b is the bias term. In (1) the right-hand side is called decision value. Its absolute value represents a measure of the distance of the sample point x from the separating hyperplane. In a typical P300-Speller [13], based on the assumption that the P300 is elicited by one of the six rows/columns stimuli, the target class is assigned to the stimulus matching the maximum decision value for the respective rows, as well as for the columns. The predicted character is identified as the intersection of the predicted row and column in the matrix [15]. In this work we use the *Decision Weighted Function (DWF),* introduced in [16] to classify the brain responses. Consider, without loss of generality, the rows stimuli. The computed decision values are normalized by dividing for the norm of vector w and sorted in decreasing order. We assign a score to each stimulus based on both the sign of their decision values and their position with respect to the separating hyperplane. We set a distance-threshold (t) equal to the median of the normalized decision values. This threshold represents an overall measure of the stimuli's distribution with respect to the separating hyperplane. We assign a score to each stimulus based on both the sign of their decision values and their position with respect to t. The score s_i^j is defined for redstimulus i at the iteration j as:

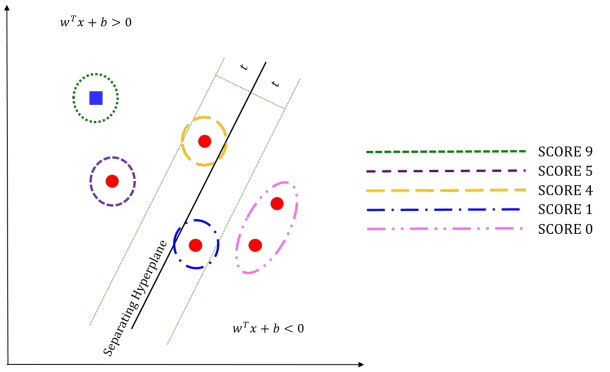


Figure 1: Score assignment procedure. Consider rows stimuli. The stimulus having the maximum decision value (the blue square) gets the maximum score, that is equal to 9 (obtained as round(1.5×6). The stimuli that fulfill the second condition in (2) get a score equal to 5 (obtained as round(0.85×6)). A score equal to 4 (obtained as round(0.6×6)) is assigned to stimuli fulfilling the third condition. Stimuli with negative decision values and that fulfill condition number four get a score equal to 1 (obtained as round(0.1×6)), otherwise they get a score equal to 0.

$$s_i^j = \begin{cases} 1.5 \times n, & \text{if } d_i^j = \max_i(d_i^j) \\ 0.85 \times n, & \text{if } d_i^j \geq 0 \wedge |d_i^j| \geq t \\ 0.65 \times n, & \text{if } d_i^j \geq 0 \wedge |d_i^j| < t \\ 0.1 \times n, & \text{if } d_i^j < 0 \wedge |d_i^j| < t \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

For each stimulus, the assigned scores are summed up iteration by iteration. We assign the target class to the stimulus having the highest total score at the last iteration. Whenever the maximum score at the last iteration corresponds to more than one stimulus, a suitable rule for breaking the ties is applied looking to the scores at the previous iterations. The same procedure is applied over columns stimuli. The predicted character is identified as the intersection of the predicted row and column in the matrix. Figure 1 shows an example of score assignment procedure given the distribution of the decision values assigned to six stimuli (think as an example to six rows of P300 Speller matrix).

Collaborative BCI: Four different collaborative decision making approaches are evaluated.

EEG Grand-Averaging (Avg_EEG): For each stimulus, the EEG responses are averaged across participants. The obtained responses are then classified using the DWF.

Majority (Maj): For each participant a single classification task is performed. According to a standard majority voting system, the target class is assigned to the character chosen by the majority of the participants. Ties are ran-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-57

domly broken by a flip of coin.

Accuracy-weighted voting rule (AW): For each participant a single classification task is performed. Let y_i be the character chosen by the participant i among the j available characters. The accuracy-weighted voting rule is defined as follows:

$$y^* = max_j \sum_i w_i y_i \quad (3)$$

where w_i is the weight of the i-*th* participant. The weight w_i is given by the training accuracy of the i-th participant. In case of parity, a random selection is performed among the contenders.

Confidence-weighted voting rule (CW): A classification task is performed for each participant. The confidence-weighted voting rule is defined as in (3), but the weight w_i is defined as follows: let t be the target trial, and let s_t^i be the total score assigned to it by DWF, the confidence of the i-th subject is computed as

$$w_i = \frac{\sum_{t=1}^{n} s_t^i}{nS}$$

where n is the number of training trials and S is the maximum achievable score. Note that $w_i \in [0,1]$.

In case of ties, a random decision is made.

We also investigate whether a cBCI approach can be applied in order to reduce the number of iterations necessary to select a single character and, therefore to speed up the selection rate. Indeed, in BCI systems, to perform a single selection step, several iterations are carried out and averaged in order to improve the EEG signal-to-noise ratio. Typically, a larger number of iterations implies higher accuracy but lower selection rate. We then evaluate three different scenarios: in the first scenario, we consider each iteration as a single selection step. As a result, let n be the number of characters to copy-spell and let t the number of iteration we perform $n \times t$ selection steps. In the second scenario we consider two iterations as a single selection step, therefore we perform $n \times t/2$ selection steps. In the third scenario we evaluate four iterations at a time, thus performing $n \times t/4$ selection steps.

RESULTS

Table 1 shows the accuracy (i.e., the percentage of correctly classified characters) over the test set for both single-user and collaborative BCIs using all the available iterations (i.e., considering a block of eight iteration as a single selection step). The single-user classification accuracy varies according to the number of characters used to train the model. It is not possible to identify a common trend across the participants. Six training characters allow participants 01, 02, 05, 07 and 10 reaching the 100% of accuracy. For users 02, 05 and 10 the same result can be obtained also with a number of training characters ranging from three to six. Three or at most four training characters are necessary to achieve the 100% of accuracy for subjects 04 and 09. Subjects 03, 06 and 08 reach their best result using five and four training characters, respectively. Note that for the single-user BCIs in some cases increasing the number of training characters can lead to a decay of the performance due to overfitting. Except for the accuracy-weighted voting rule with two training characters, all the evaluated cBCIs allow to get the 100% of accuracy with any number of training characters.

Models	Number of training trials					
	1	2	3	4	5	6
01	29.41	81.25	86.67	78.57	84.61	100
02	47.06	62.5	93.33	92.86	100	100
03	23.53	43.75	60.0	50.0	61.54	58.33
04	64.71	68.75	100	100	92.31	91.67
05	94.12	87.5	100	100	100	100
06	52.94	50.0	80.0	85.71	84.61	83.33
07	23.53	43.75	86.67	57.14	76.92	100
08	47.06	37.5	73.33	85.71	76.92	75.0
09	82.35	93.75	93.33	100	92.31	91.67
10	94.12	93.75	100	100	100	100
Avg_EEG	100	100	100	100	100	100
AW	100	93.75	100	100	100	100
Maj	100	100	100	100	100	100
CW	100	100	100	100	100	100

Table 1: Single-user and collaborative BCIs test accuracy using all the available iteration with training characters ranging from 1 to 6

Figure 2 shows the cBCI accuracy over the test set considering each iteration as a single selection step varying the number of training characters. None of the cBCIs allow to reach the 100% of accuracy. All the cBCIs reach a high accuracy ($> 85\%$) using only three training characters. This implies that these systems can be efficiently used to communicate.

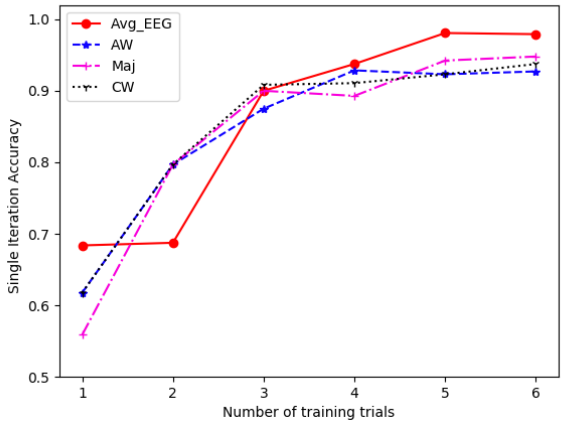


Figure 2: Comparison of the different cBCIs considering each iteration as a single selection step with training characters ranging from 1 to 6

Figure 3 depicts the cBCI accuracy over the test set considering each pair of iteration as a single selection step varying the number of training characters. Except for the Avg_EEG approach, all the cBCIs guarantee an accuracy

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-57

$>$ 90% using only two training characters. The grand-average of the EEG signals across the users allows reaching 100% of accuracy using only four training characters. Figure 4 shows the cBCI accuracy over the test set look-

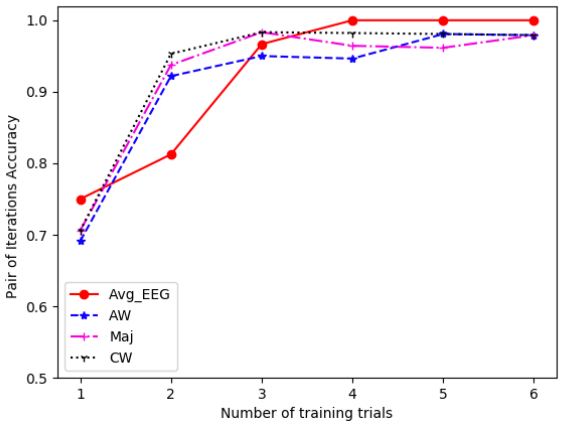


Figure 3: Comparison of the different cBCIs considering each pair of iteration as a single selection step with training characters ranging from 1 to 6

ing upon a block of four iterations as a single selection step and varying the number of training characters. All the cBCIs allow reaching an accuracy $>$ 90% using only one training character. Moreover, three training trials are sufficient to get 100% of accuracy with any collaborative approach.

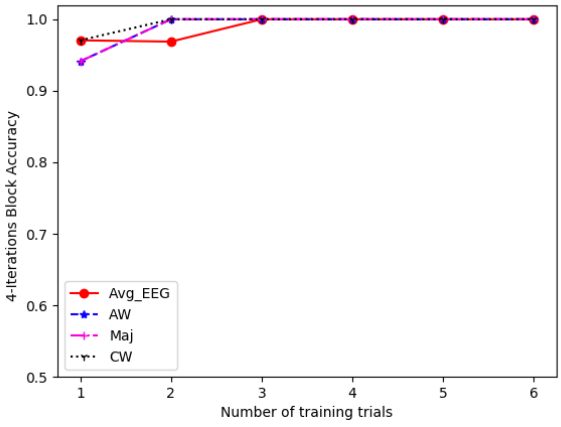


Figure 4: Comparison of the different cBCIs looking upon a block of four iterations as a single selection step with training characters ranging from 1 to 6

DISCUSSION AND CONCLUSION

In the last decade, researchers started exploiting collaborative BCIs to enhance group performance. Various methods have been proposed to integrate EEG data of multiple users showing the improvement in group decision-making over single-user BCIs and non-BCI systems. In this study, we investigated four EEG integration strategies: EEG averaging across participants, the standard majority voting rule and two weighted voting system. All the evaluated cBCIs outperform single-user BCIs confirming the trend found in the collaborative BCI literature. Moreover considering all the available iterations per selection step, it is possible to obtain 100% of accuracy with just one training character. The possibility to reduce the number of training characters preserving the test accuracy not only reduces the training computational effort but also corroborates the reliability of the evaluated cBCIs. A high number of iterations typically means higher accuracy but lower selection rate. In this work, we also investigated whether it was possible to reduce the number of iterations necessary to make a group decision preserving the 100% of accuracy. We thus evaluated three different scenarios choosing three different values for the number of iterations necessary to perform a single step selection: one iteration, pairs of iterations and blocks of four iterations. Note that this implies that we could speed up the test phase of a factor of 8, 4 or 2 depending on the chosen number of repetitions. In all the examined scenarios, it is possible to exceed 90% of accuracy with at least one collaborative BCI. Thus the systems can be efficiently used for communication. More in details, at least four iterations per selection step are necessary with one training character. With two or three training characters two iterations per selection step are enough. From four training characters on, a single iteration is enough. The minimum number of training characters and iterations that allow reaching 100% of accuracy is two and four respectively. Our results corroborate the cBCIs stability and indicate that it is possible to choose the best scenario depending on the application, in other terms based on the desired trade-off between accuracy and communication speed.

References

[1] Desanctis Gerardine, Gallupe R Brent. A foundation for the study of group decision support systems. Management science. 1987;33(5):589–609.
[2] Valeriani Davide, Cinel Caterina, Poli Riccardo. Group Augmentation in Realistic Visual-Search Decisions via a Hybrid Brain-Computer Interface. Scientific reports. 2017;7(1):7772.
[3] Poli Riccardo, Valeriani Davide, Cinel Caterina. Collaborative brain-computer interface for aiding decision-making. PloS one. 2014;9(7):e102693.
[4] Wolpaw Jonathan, Wolpaw Elizabeth Winter. Brain-computer interfaces: principles and practice. OUP USA, 2012.
[5] Nijboer Femke et al. A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. Clinical neurophysiology. 2008;119(8):1909–1916.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-57

[6] Cinel Caterina, Valeriani Davide, Poli Riccardo. Neurotechnologies for Human Cognitive Augmentation: Current State of the Art and Future Prospects. Frontiers in Human Neuroscience. 2019;13(13).
[7] Wang Yijun, Jung Tzyy-Ping. A collaborative brain-computer interface for improving human performance. PloS one. 2011;6(5):e20422.
[8] Yuan Peng et al. A collaborative brain-computer interface for accelerating human decision making. In: International Conference on Universal Access in Human-Computer Interaction. Springer. 2013, 672–681.
[9] Yuan Peng et al. Study on an online collaborative BCI to accelerate response to visual targets. In: Proceedings of 34nd IEEE EMBS Conference. 2012.
[10] Eckstein Miguel P et al. Neural decoding of collective wisdom with multi-brain computing. NeuroImage. 2012;59(1):94–108.
[11] Cecotti Hubert, Rivet Bertrand. Subject combination and electrode selection in cooperative brain-computer interface based on event related potentials. Brain sciences. 2014;4(2):335–355.
[12] Valeriani Davide, Poli Riccardo, Cinel Caterina. Enhancement of group perception via a collaborative brain–computer interface. IEEE Transactions on Biomedical Engineering. 2017;64(6):1238–1248.
[13] Sellers E. W., Kubler A., Donchin E. Brain-computer interface research at the University of South Florida Cognitive Psychophysiology Laboratory: the P300 Speller. IEEE Trans Neural Syst Rehabil Eng. 2006;14(2):221–224.
[14] Fan Rong-En et al. LIBLINEAR: A library for large linear classification. Journal of machine learning research. 2008;9(Aug):1871–1874.
[15] Krusienski Dean J et al. A comparison of classification techniques for the P300 Speller. Journal of neural engineering. 2006;3(4):299.
[16] Bianchi Luigi, Liti Chiara, Piccialli Veronica. A new early stopping method for P300 spellers. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2019;under revision.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-58

DEFINING BRAIN-COMPUTER INTERFACES: A HUMAN-COMPUTER INTERACTION PERSPECTIVE

H. Si-Mohammed[1], G. Casiez[2], F. Argelaguet[1], N. Roussel[3], A. Lécuyer[1]

[1] Inria, Univ Rennes, CNRS, IRISA, Rennes, France
[2] Univ Lille, CRIStAL, Villeneuve d'Ascq, France
[3] Inria, Bordeaux, France
E-mail: hakim.si-mohammed@inria.fr
gery.casiez@univ-lille.fr
anatole.lecuyer@inria.fr

ABSTRACT: Regardless of the term used to designate them, Brain-Computer Interfaces are "Interfaces" between a user and a computer in the broad sense of the term. This paper aims to discuss how BCIs have been defined in the literature from the day the term was introduced by Jacques Vidal. From a Human-Computer Interaction perspective, we propose a new definition of Brain-Computer Interfaces as "*any artificial system that transforms brain activity into input of a computer process*". As they are interfaces, their definition should not include the finality and objective of the system they are used to interact with. To illustrate this, we compare BCIs with other widely used Human-Computer Interfaces, and draw analogies in their conception and purpose.

INTRODUCTION

In 1973 Jacques Vidal introduced the term Brain-Computer Interface [1]. The "BCI" project as it was imagined back then, was meant to exploit the electrical activity arising from the brain to control a computer program. Even though other terms can be found in the literature to designate BCIs: Brain-Machine Interfaces (BMI), Direct Brain Interfaces (DBI), Direct Neural Interfaces (DNI) or Brain-Interfaces (BI). It is admitted that they all designate the same thing [2]: an interface between a brain and a computer (in the broad sense of the term). This is even noticeable from the fact that most of them share the "I" of "Interface", a term commonly used in Human-Computer Interaction, the field that aims to design, evaluate and implement interactive systems [3].

The main goal of this paper is to stimulate discussion around the terminology of BCIs from a Human-Computer Interaction point of view by discussing the current definitions of BCIs, by highlighting the definition of an "interface" and by setting up an analogy between BCIs and other popular computer interfaces, leading to a new definition of BCI.

CURRENT DEFINITIONS OF BRAIN-COMPUTER INTERFACES

The most widely recognized definition of a BCI was proposed by Wolpaw et al. in 2002: *"A Brain-Computer Interface is a communication system in which messages or commands that an individual sends to the external world do not pass through the brain's normal output pathways of peripheral nerves and muscles"* [4]. The main goal of a BCI would be then, to allow users who may be suffering from "locked in" syndrome or "paralysis" to communicate and to express their wishes to caregivers, or to operate word processing programs and neuroprostheses. The very core idea of a BCI would thus be to provide the brain with *"a new, non-muscular communication and control channel for conveying messages and commands to the external world"*. This brings up a core element of a BCI which is the "non-muscular" nature. A BCI derives its input solely from the brain activity.

A few years later, Blankertz stated that a BCI is *"a new augmented communication system that translates human intentions reflected by suitable brain signals, into a control signal for an output device such as a computer application or a neuroprosthesis"* [5]. An interesting feature of this definition is the concept of translation. In order to interact using a BCI, the brain activity has to be transformed into commands or messages.

According to Daly and Wolpaw [6], a BCI system *"enables a new real-time interaction between the user and the outside world"* specifying that the signals extracted from brain activity *"are translated into an output"* from which the user receives feedback that affects his brain activity. This definition highlights the real-time component of a BCI system and introduces the notion of closed-loop between the user and the system through the feedback.

Aggregating the previously mentioned features, Grainmann et al. [7] stated 4 criteria under which a system can be called a BCI: (1) A BCI system has to acquire its inputs *directly* from the brain activity; (2) The signal has to be processed and translated in *real-time*; (3) The user must obtain *feedback* from his activity; (4) The user has to send *intentional* commands

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-58

to the system. The last criterion though, has been qualified by a newer classification of BCIs proposed by Zander and Kothe [8] who distinguish between *active, reactive* and *passive* BCIs depending on the endogenous and intentional nature of the interaction. Today, passive BCIs should also be considered as BCIs.
All these definitions across time have helped better understand and better design Brain-Computer Interfaces. They are remarkably complementary and perfectly describe what most of the BCI systems are able to achieve today, in terms of communication and control. However, most of these definitions go beyond what can etymologically fall into the definition of an interface.
From an HCI perspective, considering the usage of the term "interface", strictly speaking, the term Brain-Computer Interface should only be applied to describe the intermediary hardware and software components between the brain and the interactive system.

A NEW DEFINITION OF A BRAIN-COMPUTER INTERFACE

In the Oxford dictionary, the word "interface" defines "*A point where two systems, subjects, organizations, etc. meet and interact*" [9]. In the context of computing, it defines "*A device or program enabling a user to communicate with a computer*". From both of these definitions, it clearly appears that the notion of interface only makes sense when considering the items it helps to bridge. In HCI, these two items are the "user" and the "interactive system".
A system is said to be interactive if it depends on unpredictable inputs incoming from an external environment that it does not control [10]. In HCI, the unpredictable inputs are the user commands.
In this context, an interface is the set of hardware and software means by which the user communicates with the interactive system. The interface comprises the input device (the hardware), the algorithms and methods to process the outcomes of the input device, and the presentation mechanisms to render the feedback. The final objective of the system though, does not belong in the boundaries of the interface. A computer mouse for example, is made of hardware parts, comprising the plastic box, the motion sensor and the microcontroller, for measuring the user movements. It also requires a transfer function for translating these movements into the motion of a pointer displayed on screen, and the presentation of graphical elements (i.e. buttons) so that the user can designate elements to interact with.
We believe that any BCI comprises the same set of components as any user-computer interface. Hence, arising from an HCI perspective, we propose to define a Brain-Computer Interface as:

"*any artificial system that transforms brain activity into input of a computer process".*

COMPARATIVE STUDY: A COMPUTER MOUSE VS. AN SSVEP-BASED BCI

Brain-Computer Interfaces have a lot in common with other Human-Computer Interfaces. Both aim to allow the communication between a user and a computer system. Moreover, even in terms of design and components, BCIs and HCI interfaces often share the same high-level architecture.
In this section, in order to explain our point of view, we choose to illustrate the similarities between a Steady-State Visual Evoked Potential (SSVEP) based BCI, based on EEG, and a computer mouse. SSVEP is a reactive paradigm based on the property that the brain, when the subject is focused on a periodic visual stimuli of frequency *f*, reacts with an increase in its activity at the same frequency *f*. In other words, it means it is possible to determine the frequency of periodic stimulation the user is focused on. Interested readers may refer to [11] for more details.
The choice of the SSVEP and a computer mouse is arbitrary, and the same analogy could be drawn with other instances. Figure 1 illustrates the different components of this analogy.

Hardware components: The first thing that comes to users' mind when they hear "computer mouse", is the plastic-made body. Similarly, the first thing that comes to mind when evoking a BCI is the EEG cap itself with its electrodes. In fact, both of them correspond to the entry point of the interface for the user. The second component of the mouse is usually a digital camera, the microcontroller and the LED lens embedded in the body which are responsible of acquiring the user's hand movements. In the case of an EEG BCI, it is the amplifier that provides one time-dependent signal per measured electrode.

Software components: In both of the computer mouse and the BCI, the software components hold an important part. In the mouse, the direction, amplitude and speed of the hand movement are translated into new coordinates for the cursor on the screen using a transfer function. These movements can typically be predicted by Fitts' law [12]. In a BCI, the amplified EEG signal is processed and classified into a mental state. Depending on the mental state, and depending of the position of the cursor when clicking, a particular command is determined from the interface and sent to the interactive system.

Presentation: In the cases of the mouse and an SSVEP BCI, a graphical presentation is primordial. For the SSVEP BCI, it is essential to have external stimuli under the form of flickering targets, in order to infer different commands from the BCI. This is what makes SSVEP a reactive paradigm. For the mouse, on the other hand, it is necessary to present buttons or specific areas on the screen so that the user knows where to click. In the same way as clicking on different buttons generate different commands, focusing on targets flickering at different frequencies will generate different commands.
This analogy between a mouse and a SSVEP BCI do

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-58

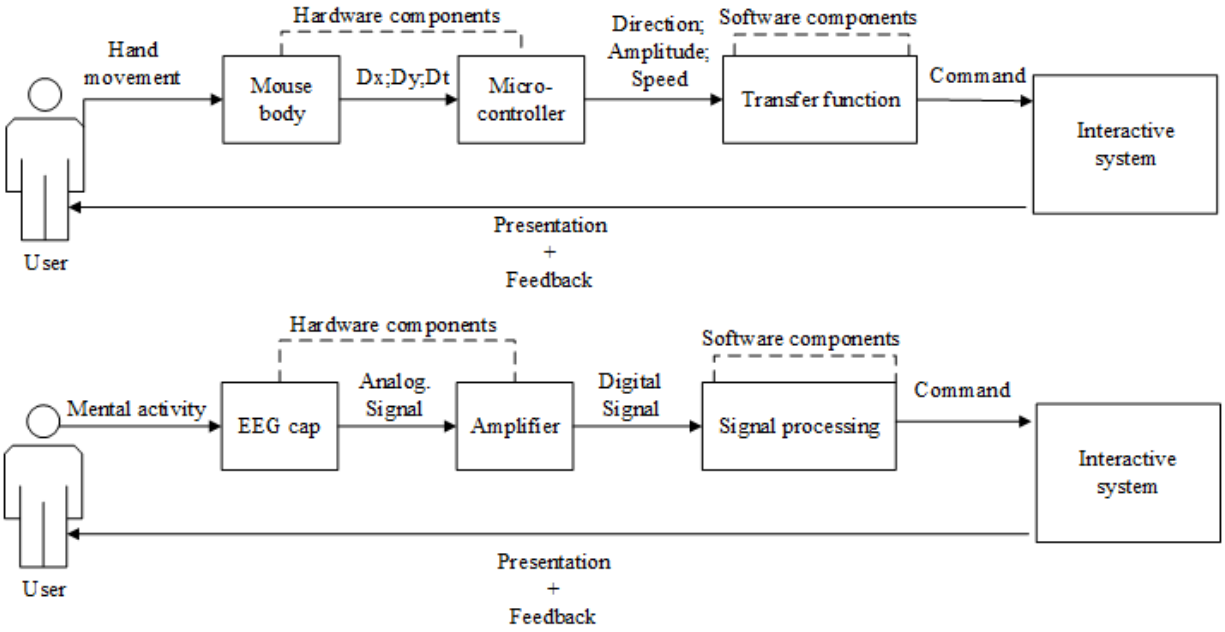


Figure 1: Illustration of the components comprised in a user interface. (Top) A Brain-Computer Interface comprises an acquisition equipment, a signal amplifier and all the methods used to transform the signal into a computer input. (Bottom) A computer mouse is made of a hardware body, a microcontroller and a transfer function.

apply to other reactive paradigms. Yet, it is possible to draw the same kind of comparison for active or passive BCIs. In the case of Motor Imagery for instance, we can easily make an analogy with a joystick or a game controller. When the subject imagines left or right-hand movement, the appropriate command is sent to the interactive system, without the need for a graphical presentation, even though a feedback mechanism can help.

For passive BCIs, it is possible to draw a comparison with proactive and transparent interfaces. A good example of this would be a smart-home interface, where the user's behaviors and locations are transparently monitored to infer commands to send to the smart-home. Adapting a system to the user's mental state can somehow be seen as automatically turning on the light when the user moves into the room.

These comparisons and analogies have to be seen as what they are meant to be: illustrative examples. They are not intended to be exhaustive nor universal truth, but they serve to illustrate that a Brain-Computer Interface can fundamentally be seen as a Human-Computer Interface.

DISCUSSION

There have been several definitions of Brain-Computer Interfaces since Jacques Vidal introduced the name. Today, research and innovation in the field of BCIs is very transversal. Neuroscientists, computer-scientists, medical doctors and engineers have different perspectives on what Brain-Computer Interfaces are.

The definition we proposed, present the advantage of being very inclusive because it does not comprise the finality of the interactive system.

Instead, we consider "*any*" system, that unlike muscles is "*artificial*" and falls in the proposed definition as a BCI. The purpose of this system is to "*transform*" the brain activity into "*inputs*" that a "*computer process*" can exploit.

A similar reflection has been conducted by Jeunet et al. [13] when comparing Neurofeedback and Motor Imagery. They highlighted that both Neurofeedback (NF) and MI-BCI users have to learn how to regulate their neurophysiological activity, sometimes with similar features, through given feedback but the final objective is different. While MI-BCI consists in producing a specific EEG pattern to send a command, the goal of NF is to learn how to generate the specific pattern.

If one wants to designate the whole interaction between a user and an interactive system using a BCI to achieve a particular role, one should use the term Brain-Computer Interaction.

With our new definition of Brain-Computer Interfaces (BCI), any artificial system that involves the exploitation of brain signal, can be considered as using a BCI, instead of being the BCI itself.

CONCLUSION

In this discussion paper, our goal was to explore the different definitions of Brain-Computer Interfaces. All these definitions are very complementary, reflect the transversality of the field, and illustrate what it is possible to achieve with BCIs today. In an attempt to aggregate these definitions, in order to be as inclusive of new and incoming types of BCIs as possible, we proposed a new definition motivated by the etymology of the term, and with our perspective from the Human-

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-58

Computer Interaction domain. We illustrated and grounded our point of view through an analogy between BCIs and more widely used interfaces.
We believe that this new definition could support the future discussion about what BCIs are, and that it could constitute a first step towards the design of Brain-Computer Interfaces as interaction medias.

REFERENCES

[1] J. J. Vidal, "Toward direct brain-computer communication," *Annu. Rev. Biophys. Bioeng.*, vol. 2, no. 1, pp. 157–180, 1973.

[2] S. G. Mason, A. Bashashati, M. Fatourechi, K. F. Navarro, and G. E. Birch, "A comprehensive survey of brain interface technology designs," *Ann. Biomed. Eng.*, vol. 35, no. 2, pp. 137–169, 2007.

[3] T. T. Hewett *et al.*, *ACM SIGCHI curricula for human-computer interaction*. ACM, 1992.

[4] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–91, 2002.

[5] B. Blankertz *et al.*, "The Berlin Brain-Computer Interface: EEG-based communication without subject training," *IEEE Trans. neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 147–152, 2006.

[6] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *Lancet Neurol.*, vol. 7, no. 11, pp. 1032–1043, 2008.

[7] B. Graimann, B. Allison, and G. Pfurtscheller, "Brain–computer interfaces: A gentle introduction," in *Brain-Computer Interfaces*, Springer, 2009, pp. 1–27.

[8] 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. 25005, 2011.

[9] Interface. (2019). In: *Oxford University Press*. [online] Available at: https://en.oxforddictionaries.com/definition/interface [Accessed 20 Feb. 2019].

[10] D. Goldin, S. A. Smolka, and P. Wegner, *Interactive computation: The new paradigm*. Springer Science & Business Media, 2006.

[11] F.-B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki, "Steady-state visually evoked potentials: focus on essential paradigms and future perspectives," *Prog. Neurobiol.*, vol. 90, no. 4, pp. 418–438, 2010.

[12] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement.," *J. Exp. Psychol.*, vol. 47, no. 6, p. 381, 1954.

[13] C. Jeunet, F. Lotte, J.-M. Batail, P. Philip, and J.-A. M. Franchi, "Using recent BCI literature to deepen our understanding of clinical neurofeedback: A short review," *Neuroscience*, 2018.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-59

INVESTIGATION OF NEEDS AND CHARACTERISTICS OF END-USERS, FOR A FUTURE INCLUSION OF BCIS IN AT-CENTERS

A. Riccio[1], F. Schettini[2], E. Giraldi[1]
F. Cappalonga[3], L. Pelagalli[4], F. Cincotti[1-3] D. Mattia[1]

[1] Neuroelectrical Imaging and BCI Lab, IRCCS Fondazione Santa Lucia, Rome, Italy
[2] Servizio Ausilioteca per Riabilitazione Assistita con Tecnologia (SARA-t), Fondazione Santa Lucia (IRCCS), Rome, Italy
[3] Department of Computer, Control, and Management Engineering Antonio Ruberti, "Sapienza" University of Rome, Rome, Italy
[4] Dipartimento di Ingegneria Elettrica e dell'Informazione "Maurizio Scarano", Università degli Studi di Cassino e del Lazio Meridionale

E-mail: a.riccio@hsantalucia.it

ABSTRACT Brain-Computer Interface as an alternative channel to access Assistive Technologies for communication disorders is not currently available in the AT-centers portfolio. A step forward such availability would consist in BCIs integration with existing AT inputs thus, resulting in a personalized hybrid BCI-based communication device.

The overall aim of the study is to generate profiles of patients that would potentially use the BCI as an additional/alternative channel for AT-access.

In an AT-center, we have started to systematically screen patients with different degree of disability. Currently, 10 patients have been screened in relation to their needs by means of the Individual Prioritised Problems Assessment (IPPA) and the matching AT solutions.

Preliminary results of IPPA showed that the most common problems to be solved with AT were "reading/writing" (N=8), "communication" (N=7) and "phone access" (N=6). AT solutions were mostly characterized by input devices (touch screen, eye tracker) controlling a customized user interface.

As a next step, participants will be screened for P3-based-BCI control.

INTRODUCTION

Communication is a basic need and right of every human being. Many common brain disorders or neuromuscular diseases lead to severe and complex impairments of communication and interaction abilities. These disabilities could create social isolation and dependency, having a considerable impact on a person's quality of life (QoL). Assistive technologies (ATs) could restore basic communication capabilities and improve interaction in people with complex communication needs, alleviating dependency and supporting the participation in society. This would significantly improve their own and their caregivers' QoL.

AT was defined as "an umbrella term indicating any product or technology-based service that enables people of all ages with activity limitations in their daily life, education, work or leisure" [1]. The definition includes both "mainstream" technologies (general-purpose technologies) and "assistive" technologies (purposely designed for people with disabilities), whose assembly varies case-by-case, depending on the individual characteristics of the person, the activities he/she is intended to perform, and the physical and human context where he/she lives [1]. Current ATs provide a powerful array of communication, information, organization, and social networking options for individuals with complex communication needs [2].

Several AT solutions are available for end-users to improve accessibility to communication and environmental control technologies: those are based on various technological approaches and interaction modalities (e.g. eye-trackers, adapted joysticks, speech recognition …) which are available in the portfolio of the AT-centers. AT-centers are AT service delivery, where AT experts with various backgrounds (such as therapists, psychologists, engineers, medical doctors) are responsible for the selection of AT solutions and for their customization to users' needs and their motor, sensory and cognitive impairment (disabilities). Furthermore, AT experts are responsible for training the users in the AT daily-life usage.

Brain-Computer Interfaces (BCIs) measure signals related to specific brain activity and translate them into outputs to control external devices for a range of applications such as communication, environmental control, movement control and motor rehabilitation [3]. It was widely demonstrated that BCI can provide people with communication disorders with an AT, restoring their interaction with the environment [4, 5, 6]. Thus, the primary motivation of BCI research in this field has

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-59

been the reestablishment of communication and facilitation of daily life activities for people with communication and interaction disabilities, due to many common disorders such as neurodegenerative diseases (amyotrophic lateral sclerosis-ALS, spinal muscular atrophy –SMA), spinal cord injury (SCI), acquired brain injury (ABI).

Currently, BCIs are not available in the portfolio of AT-centers for a full deployment to end-users even though BCI technology could improve the inclusiveness of AT solutions. A step forward for the inclusion of BCIs in AT-centers is their integration with existing (available in the market) assistive or mainstream technologies. This integration would result in a hybrid BCI-based communication device [7] that will allow more end-users to use standard means of communication integrating a BCI with their residual muscular activity [4]: the end-users will be able to switch to BCI channel when the muscular one is fatigued or weak (e.g. touch screen, buttons, eye tracker, joysticks); alternatively, he/she will use them as complementary channels.

The first step to bridge the translational gap between BCI development and end-users, by incorporating BCI technology into an AT device for everyday communication and interaction, consists in the clear definition of the users' characteristics and their relation with the abilities to control a BCI. Indeed, the development of prototypic BCIs and ATs, as well as the definition of their specifications, often take place in research laboratories, whit exiguous contact with end-users. The involvement of AT centers in the development of innovative devices and in their customization and validation could bridge this gap.

ALS patients were often considered as the target population of BCI technologies, because of the neurodegenerative characteristics of the disease [8]: it leads to a progressive muscular paralysis causing a loss of communication and interaction ability. However, users attending AT centers are affected by many different diseases with many different etiologies, and complex sensorial, physic, functional and cognitive disabilities.

Here we present a study aimed at investigating the characteristics of patients attending an AT-center (needing an AT for communication and environmental interaction), who could take advantage from BCI introduction in AT-centers as an additional/alternative AT channel. Ten patients were involved in this preliminary study: they underwent multidisciplinary evaluation and AT training, performed in the AT-center of Fondazione Santa Lucia, Roma, to identify the problem (related to communication and environmental interaction) that they wished to solve with the AT and also to identify the best AT solution to solve such problems. Such users will be screened in a following session for their abilities to control a P3-based BCI.

We will analyze the clinical, functional and neurophysiological features influencing such ability. Furthermore, during the BCI control evaluation users' eye movements will be recorded, in order to evaluate if they influence the BCI classification and how the two control channels (eyes and brain) can be used complementary and/or alternatively as part of an AT solution.

MATERIALS AND METHODS

Participants. Ten patients (43.3 ± 9.9 years old, 2 men) with different diagnosis were enrolled in the study: 1 participants with traumatic brain injury (TBI), 3 with amyotrophic lateral sclerosis (ALS), 1 with Friedreich's ataxia, 1 with autosomal dominant leukodystrophy, 2 with hemorrhagic stroke, 1 with encephalitis, 1 with multiple sclerosis (MS). All patients had undergone a multidisciplinary evaluation and an AT training in the AT-center (SARA-t) in Fondazione Santa Lucia, IRCCS, Rome, because they were limited in (at least) one aspect related to interpersonal communication and/or interaction with digital technologies (smartphone, PC, Tablet).

Participants were recruited through the AT-center SARA-t, where the study was conducted.

Patients with other concomitant neurological or psychiatric disorders, any impediment in the acquisition of electroencephalography (EEG) data from the scalp (e.g. wounds, dermatitis), severe concomitant pathologies (fever, infections, metabolic disorders, and severe heart failure), global cognitive impairment, aphasia, or episodes of reflex epilepsy were excluded from the study.

Participant's functional disability was assessed by means of the "World Health Organization Disability Assessment Schedule 2.0 (Whodas, 12 items version [9]). Whodas investigates the functioning level in six life domains with 2 items each: understanding and communicating, getting around, self-care, getting along with people, life activities (i.e., household, work, and/or school activities), and participation in society. Each item asked the participant to rate how much difficulty, (from 1="none" to 5="extreme or cannot do") he or she has had in specific areas of functioning during the past 30 days. 12 items Whodas 2.0 scores range from 0 ("no disability") to 100 ("full disability"). Mean Whodas scores was 67.08 ± 15.6 (range from 45.83 to 85.42). Patient's demographic and clinical information are reported in Table 1.

Table 1: Demographic and clinical information

Participant	Age	Gender	Diagnosis
P1	46	F	ALS
P2	30	F	Encephalitis
P3	58	M	Hemorrhagic stroke
P4	47	F	Hemorrhagic stroke
P5	48	F	ALS
P6	33	F	Friedreich's ataxia
P7	30	F	TBI
P8	50	F	Leukodystrophy
P9	53	F	ALS

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-59

Participant	Age	Gender	Diagnosis
P10	38	M	MS

Protocol. The protocol consists of two parts, *i)* need assessment protocol (results presented here) and *ii)* Brain-Computer Interface protocol (presented here and to be performed in a following study). The study was approved by the Independent Ethics Committee of Fondazione Santa Lucia, IRCCS, Rome.

Needs assessment protocol. Patients were administered with the Individual Prioritized Problems Assessment (IPPA [10]) questionnaire during the multidisciplinary evaluation. They were asked to list (a maximum of seven) problems that they wished/expected to improve/solve with the AT. Moreover, they had to score the perceived importance (1 = "not important at all" to 5 = "most important") and the difficult (1 = "no difficulty at all" to 5 = "too much difficulty to perform the activity at all") associated to each problem. After the needs assessment participants were involved in the AT training. The training was aimed at the identification of the AT solution matching user's need and their motor, sensorial and cognitive characteristics. Table 1 lists the problems reported by participants and the AT solution used by each participant. Patients were proposed to and agreed to participate in the BCI session, which will be performed in the next weeks.

Brain-computer Interface protocol. Scalp potential will be acquired by means of a 16-channel amplifier (g.USBamp, g.tec, Austria) from 16 active electrodes (g.Ladybird, G.tec, Austria) placed according to 10-10 international standard (Fz, Cz, Pz, Oz, P3, P4, PO7, PO8, F3, F4, FCz, C3, C3, CP3, CPz, and CP4 (right ear lobe reference; left mastoid ground). Signals will be digitized at 256Hz. Stimulus paradigm and online delivery will be managed by means of the BCI2000 framework [11]. A P3-speller [12] interface (5 by 6 matrix of alphabetic items) will be displayed full screen, placed approximately at eye level and at a distance of 60 cm.
In the calibration phase (i.e., no feedback on performance), subjects will be asked to focus on 15 items forming 3 predefined words (3 runs; 5 items for each run). The target to focus on will be shown to the participant by a single flash, after which rows and columns will randomly be intensified for 125ms, with an inter stimulus interval (ISI) of 125ms (Stimulus Onset Asynchrony, SOA, 250ms). A stepwise linear discriminant analysis (SWLDA) will be applied offline to determine the classifier coefficients [13] for the testing phase. During the testing phase (i.e., provision of feedback on performance), participants will have to spell four predefined (copy mode) words (4 runs; 5 items for each run 20 characters in total).
The Tobii Technology 4C eye-tracker will be used to collect eye-gaze data. The 4C is a binocular, IR eye tracker that samples at a frequency of 90 Hz. The eye-tracker estimates the user's gaze or point of focus on the monitor in both the vertical and horizontal axis. The initial calibration will be performed using Tobii's calibration software Gaze Point Bundle (version 2.0.8). This software allows also for mouse pointer control. During EEG recording mouse pointer will be not visible and its x and y-coordinates will be recorded and synchronized with EEG signal as a BCI2000's state. The monitor used in this experiment has a resolution of 1280 × 1024 (width × height) pixels and is 17″ along the diagonal.

RESULTS

Needs assessment protocol. The mean total IPPA Score was 19.58 (± 3.04 SD, min = 15, max = 25). Fifty-eight problems in total were identified, with an average of 5.8 problem for participant (SD = ±1.9; min = 1, max = 7). Problems identified were grouped in the following 10 categories:
1) *Reading/Writing*, reported by 8 patients; 2) *Communication*, reported by 7 patients; 3) *Phone access* reported by 6 patients; 4) *TV access*, 5 patients; 5) *Social Network access*, reported by 5 patients; 6) *PC access*, reported by 4 patients; 7) *Listen to music*, 4 patients; 8) *Relationship/Social life*, 3 patients; 9) *Turn on/off the light*, 1 patient. The AT solutions matching with the problems that users wished to solve with the AT solutions and with their motor, sensorial and cognitive characteristics are reported in the following. The touch screen was used as input by 4 patients, the eye tracker was used by 3 patients and the head tracker was the AT commercial input used by 2 patients. One patient used mainstream accessibility settings to improve control accuracy in controlling the PC and 5 patients accessed to the device by mean of a customized user interface.
Problems pointed out from and AT solutions identified for each patient are reported in Table 2.

Brain-Computer Interface protocol. Significant relationships between patients clinical (e.g. aetiology, onset, score in the functional scales, lesion...) and neurophysiological characteristics (e.g. ERPs amplitude and latency) and BCI control performance will be reported. Influence of the number of fixations on the target and off the target, and influence of the mean duration of single on-target fixation (measured by means of the eye tracker) on BCI control will be investigated and reported.

Table 2: problems identified with IPPA and AT solutions

Participant	IPPA	AT solution
P1	Phone Access	Customized
	TV Access	User Interface

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-59

	PC Access Reading/Writing Communication	Input device: eye tracker to control mouse cursor and mechanical switch to perform clicks
P2	Social Network Phone Access Reading/Writing Music TV Access	Customized User Interface Input device: touch screen
P3	Communication	Commercial text-to-speech application Input device: Touch screen
P4	Relationship/Social Life Communication Reading/Writing Music Social Network TV Access	Accessibility settings to improve accuracy using PC
P5	Relationship/Social life Communication Reading/Writing Music	Customized User Interface Input device: eye tracker
P6	Communication Relationship/Social Life Music Social Network Phone Access	Customized User Interface Input device: eye tracker to control mouse cursor and mechanical switch to perform clicks
P7	Communication Reading/Writing Social Network	Commercial text-to-speech application and accessibility settings Input device: Touch screen
P8	Phone Access PC Access Reading/writing TV Access Social Network	Accessibility settings to facilitate the access to tablet Input device: Touch screen
P9	Reading/writing Phone Access Communication PC Access	Customized User Interface Input device: head tracker to control mouse cursor and mechanical switch to perform clicks
P10	Phone Access Reading/Writing PC Access TV Access Turn on/off Light	Head tracker

DISCUSSION

The assumption of this study is the need and potential benefit to include BCIs in AT-centers as an alternative/additional input channel. Indeed, despite the demonstration that BCI can provide people with communication disorders with an AT, such technology is currently absent in the AT-centers portfolio. We propose, as a step forward to such inclusion, BCI integration with existing AT as a hybridization of the BCI-based communication device.

To meet this aim, we propose a screening of the ability to control a P3-speller BCI, in a group of participants with various diseases considering their needs and their clinical and neurophysiological characteristics. Participants involved were attending the AT-center SARA-t during the period of neurorehabilitation that was taking place in Fondazione Santa Lucia. They needed an AT for communication to support daily interaction with communication partners, to access to mainstream communication devices (tablet, smartphone, personal computer) and also to support the process of neurorehabilitation.

Preliminary results on their need (IPPA) and the matching AT solutions showed their need of an AT to improving reading and writing abilities (8 participants), communication abilities (7 participants) and smartphone access possibility (6 participants). AT solutions were identified as matching participants needs and their motor, sensorial and cognitive (dis)abilities and were various input devices (e.g. touch screen, buttons, eye tracker) mostly controlling a customized user interface.

As a next step, participants will be screened for their ability to control a P3-based BCI (P3-Speller) recording eye movements during the BCI session.

Aim of this second part of the protocol is to investigate the relationship between participants' clinical and neurophysiological characteristics and BCI control. This will allow the definition of various profiles of BCI accessing to different versions of personalized hybrid BCI for communication. Such personalization will match end users' motor, sensorial and cognitive characteristics.

The hybridization of the BCI based communication device by mean of BCI integration with existing AT would lead to the concept of a highly personalized (hybrid) BCI customized for each user and merged with input channels specific for each user. This would follow the patient in the course of neurorehabilitation consequently improving their QoL and also the quality of rehabilitation and will open the way for multicentric studies to be performed in different AT-centers.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-59

CONCLUSION

In the present study, involving 10 patients with different degree of disabilities, we reported preliminary data about the screening of needs of potential end-users of a hybrid-BCI device for communication and we also reported the matching AT solutions. The overall aim was to generate profiles of patients that would potentially use the BCI as an additional/alternative channel for AT-access.
Results showed that a range of input channels customized on the basis of patients motor, sensorial and cognitive (dis)abilities was used to solve/improve problems related to reading and writing abilities, communication ability and smartphone access ability.
In the next step, their performance in controlling a P3-based BCI for communication will be investigated and the relationship with the user's characteristics (among witch eye movement's peculiarities) will be established.
We consider this as an important step for the integration of BCI with daily/commercial AT devices, for the consequent development of a personalized hybrid BCI device for communication and for BCI inclusion AT-centers portfolio.

ACKNOWLEDGMENT

This work was partially supported by the Promobilia Foundation (Grant Ref. 17137).

REFERENCES

[1] Andrich R, Mathiassen N-E, Hoogerwerf E-J, Gelderblom GJ. Service delivery systems for assistive technology in Europe: An AAATE/EASTIN position paper. Technol Disabil 2013;25:127–146.
[2] Abbott C, Brown D, Evett L, Standen P. Emerging issues and current trends in assistive technology use 2007-2010: practising, assisting and enabling learning for all. Disabil Rehabil Assist Technol 2014;9:453–62.
[3] Wolpaw J, Wolpaw EW. Brain-computer interfaces: principles and practice. OUP USA; 2012.
[4] Riccio A, Holz EM, Aricò P, Leotta F, Aloise F, Desideri L, et al. Hybrid P300-based brain-computer interface to improve usability for people with severe motor disability: electromyographic signals for error correction during a spelling task. Arch Phys Med Rehabil 2015;96:S54–S61.
[5] Schettini F, Riccio A, Simione L, Liberati G, Caruso M, Frasca V, et al. Assistive device with conventional, alternative, and brain-computer interface inputs to enhance interaction with the environment for people with amyotrophic lateral sclerosis: a feasibility and usability study. Arch Phys Med Rehabil 2015;96:S46–S53.
[6] Zickler C, Riccio A, Leotta F, Hillian-Tress S, Halder S, Holz E, et al. A brain-computer interface as input channel for a standard assistive technology software. Clin EEG Neurosci 2011;42:236–244.
[7] Müller GR, Breitwieser C, Cincotti F, Leeb R, Schreuder M, Leotta F, et al. Tools for brain-computer interaction: a general concept for a hybrid BCI. Front Neuroinformatics 2011;5:30.
[8] Riccio A, Schettini F, Simione L, Pizzimenti A, Inghilleri M, Olivetti-Belardinelli M, et al. On the Relationship Between Attention Processing and P300-Based Brain Computer Interface Control in Amyotrophic Lateral Sclerosis. Front Hum Neurosci 2018;12:165..
[9] Üstün TB, Kostanjsek N, Chatterji S, Rehm J. Measuring health and disability: Manual for WHO disability assessment schedule WHODAS 2.0. World Health Organization; 2010.
[10] Wessels R, Persson J, Lorentsen Ø, Andrich R, Ferrario M, Oortwijn W, et al. IPPA: Individually prioritised problem assessment. Technol Disabil 2002;14:141–145.
[11] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans Biomed Eng 2004;51:1034–43.
[12] 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.
[13] Krusienski DJ, Sellers EW, Cabestaing F, Bayoudh S, McFarland DJ, Vaughan TM, et al. A comparison of classification techniques for the P300 Speller. J Neural Eng 2006;3:299–305.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-60

BCI THERAPIES TRIGGERING SENSORY FEEDBACK FOR MOTOR REHABILITATION AFTER STROKE: A SYSTEMATIC REVIEW

Claudia Bigoni[1,2]& Friedhelm C. Hummel[1,2,3]

[1] Defitech Chair of Clinical Neuroengineering, Swiss Federal Institute of Technology (EPFL), 1202 Geneva and [2] Swiss Federal Institute of Technology (EPFL Valais), 1951 Sion, Switzerland
[3] Clinical Neuroscience, University of Geneva Medical School, 1202 Geneva, Switzerland

E-mail: claudia.bigoni@epfl.ch

ABSTRACT: In the last two decades, brain-computer interfaces (BCI) triggering external devices to perform movements for upper-limb rehabilitation after stroke, have proven to be successful. However, the literature is quite heterogeneous in terms of targeted patient population and study protocols. In this systematic review, we aim at identifying those patient characteristics and protocol features that might best explain the variance observed in treatment-induced motor recovery. Using the data from 15 studies and a total of 168 patients with a BCI-based intervention, patients in the sub-acute phase and mildly impaired patients showed significantly stronger improvement with the BCI-based intervention. Furthermore, receiving conventional therapy additionally to the BCI-based intervention leads to a significantly larger improvement. In summary, BCI-based neurorehabilitation combined with conventional therapy might induce a synergistic effect leading to stronger functional recovery. Larger patient samples in studies and more information on individuals characteristics and protocol features would be desirable in order to build predictive models for motor rehabilitation towards personalization of interventions.

INTRODUCTION

In the last two decades, the use of rehabilitation technologies for motor recovery after stroke has increased [1]. Among these, BCI is a very promising technology. Its goal is to re-create contingency between voluntary brain activation in the motor cortex, movement induction and afferent peripheral sensory feedback, often given by an exoskeleton or functional electrical stimulation (FES). This contingent activation of the efferent (motor) and afferent (sensory) part of movements is thought to be one of the main neurophysiological mechanisms BCI-based therapy relies on. This leads to neuroplastic changes following Hebbian learning principles supporting reorganization and the recovery process. This intended contingency is achieved with the following framework: brain activity is continuously recorded and fed to a decoding algorithm trained to differentiate between specific brain states (e.g. rest vs. motor intention); whenever such signal is detected, an external device/machine is triggered to perform a motor action and provide a well-timed sensory feedback contingent to brain engagement. BCI is still a relatively new technology for stroke rehabilitation and the current literature is mainly based on proof-of-principle clinical trials targeted to provide first evidence for feasibility and efficacy of this intervention in comparison to more standard rehabilitation treatments. Meta-analyses on the topic [2,3] have shown that BCI therapies result in significantly stronger motor improvement compared to various controls (i.e. standard therapy [4,5], external device alone [4,6,7], sham BCI where feedback is given randomly, but with the same external device [8,9,10]). Certainly, these reviews and the respective trials highlight the potential of BCI-based rehabilitation for motor recovery after stroke. Although a positive effect of BCI-based rehabilitation has been suggested, the BCI protocols used are not standardized and are quite heterogeneous; the same applies to the targeted patient population in the different trials. Therefore, the present review addresses whether specific intervention parameters relate better to stronger motor recovery in experiments using BCI-based therapy.
Considering all the studies (controlled and non-controlled) that recruited stroke patients for motor rehabilitation using BCIs, large differences in outcome measures, protocol design, as well as individual patient characteristics are apparent. Main differences are found e.g., in total intervention duration, external device used, performed movements, and in terms of individual patient characteristics, impairment at baseline and stroke onset. The goal of our systematic review is to determine, within BCI studies, how the degree of motor recovery is related to these features.

MATERIALS AND METHODS

Articles retrieval strategy and inclusion criteria: Within this review, protocols and patient characteristics which could be potential parameters for predicting motor recovery in stroke patients were evaluated. The retrieval of relevant papers used the following keywords in Google Scholar research: brain-computer/machine/robotic interface, exoskeleton/orthosis/robot, functional/neuromuscular electrical stimulation, upper-limb, stroke, rehabilitation. In addition, we looked through the references of the retrieved papers and reviews on the topic. Papers

Published by Verlag der TU Graz
Graz University of Technology

Table 1 –Characteristics of selected studies

Reference	Sample	Age(y)	Stoke onset (months)	Motor impairment (baseline)	Brain instruction	Sensory feedback provider	Standard therapy (hours)	BCI therapy (hours)	Additional feedbacks	Outcomes	Movement[1]
Biasiucci et al., 2018 [9]	14	56.35±9.55	56.35±9.55	Moderate-to-severe	AM	FES	7.5	10	Therapist	FMA	Hand extension
Ramos-Murguialday et al., 2013 [8]	16	49.3 ± 12.5	49.3 ± 12.5	Severe	AM	EXO	20	20	None	FMA	Reach and grasp
Pichiorri et al., 2015 [15]	14	64.1 ± 8.4	64.1 ± 8.4	Moderate-to-severe	MI	EXO	36	12	Therapist	FMA	Fingers extension
Li et al., 2014 [7]	7	66.3±4.53	66.3±4.53	Severe	MI	FES	40	24	Visual and auditory	FMA, ARAT	Hand extension
Sullivan et al., 2015 [18]	6	57.5±7.25	51.5±38.27	Mild	AM	FES	0	7.5	None	FMA	Elbow flexion
Bundy et al., 2017 [19]	10	58.6±9.72	73.6±98.86	Moderate-to-severe	MI	EXO	0	40.2	Visual	ARAT	3-finger pinch
Várkuti et al., 2013 [20]	6	40.17±13.23	15.78±12.33	Mild-to-moderate	MI	EXO	0	24	None	FMA	Shoulder elbow flexion
Frolov et al., 2017 [10]	55	58.0 [48.0; 65.0]	8.0 [4.0; 13.0]	Mild-to-severe	MI	EXO	0	5	Visual	FMA, ARAT	Hand extension
Tabering et al., 2018 [14]	8	61.25±18..96	36.75±24.18	Severe	MI	FES	0	60	None	FMA	Hand extension
Prasad et al., 2017 [16]	4	-	chronic	Mild-to-severe	MI	EXO	0	12	Visual	ARAT	Hand extension
Caria et al., 2011 [21]	1	67	14	Severe	MI	EXO	40	40	Visual	FMA	Fingers flexion
Mukaino et al., 2015 [22]	1	38	14	Moderate	AM	FES	6.6	8.3	None	FMA	Fingers extension
Belardinelli et al., 2017 [23]	8	57.25±11.02	75.5±36.78	Severe	MI	FES	0	7	None	FMA	Fingers extension
Remisk et al., 2018 [11]	14[2]	60.43±13.4	42.64±50.06	No restriction	AM	FES	0	30	Visual	ARAT	Fingers extension
Ibáñez et al., 2017 [17]	4	54.25±11.84	48±9.78	Mild	AM	FES	0	8	None	FMA	Reach and grasp

[1] If extension or flexion, it means also movement in the other direction
[2] The actual sample number was 19, however 5 patients were recruited with already the maximal score (57 in the ARAT) and remained stable. Therefore, they were not included since they did not have any window for recovery

Published by Verlag der TU Graz
Graz University of Technology

DOI: 10.3217/978-3-85125-682-6-60

 DOI: 10.3217/978-3-85125-682-6-60

published until the end of 2018 were considered. Moreover, only studies that reported clinical motor assessment were included. In particular, for this preliminary review, we only accepted studies using the Fugl-Meyer Assessment (FMA) or the Action Research Arm Test (ARAT) since they are the most widely used scales and because a correlation between the two has already been extensively confirmed [11,12]. Given the objective of this review, we could only include articles reporting the motor improvement for each patient (i.e. if only the average recovery of a group was given, the study was not considered). If the study was a randomized trial, we only considered patients in the experimental therapy. Finally, due to this "individuality" feature we had no limitation in sample size. Whenever an eligible article was lacking data, we tried to contact the authors multiple times.

Analyses methods: In terms of study protocol, we looked at six characteristics. (1) The device giving the sensory feedback and (2) the type of movement produced. (3) The type of instruction to calibrate and exploit the BCI was also considered, specifically we divided into motor imagery (with no inner subdivision, as most studies explicitly say they used the kinesthetic type) and active motor intention ("try to move the arm/hand"). (4) The additional feedbacks received (e.g. visual, auditory, tongue stimulation). (5) Overall intervention therapy and (6) if any, additional standard therapy. When possible, some characteristics of the patients were also taken into account: motor impairment at baseline and the time after stroke in months.
In these analyses the dependent variable is always the motor improvement. For each binary comparison, we computed non-parametric tests for the means, namely the Mann-Whitney test and the effect size with the *Cohen's d* [13]. For non-binary features, such as the intervention duration, we looked at the correlation. Along the article, we will refer to motor improvement as the difference in motor score between the end and the beginning of the study. We will call proportional motor improvement the motor improvement normalized by the baseline. All the values are normalized by the maximal achievable score. This choice was made since some studies used a modified FMA [8,14] and because the maxima of the ARAT and the FMA differ.

RESULTS

Search results: The pruned research gave overall 30 eligible papers. Of these, 6 were redundant reports, 1 used both motor imagery and attempt the movement as instruction, 3 used no motor scale, 2 did not assess motor abilities with FMA nor ARAT and 3 did not report the motor scores for each subject. This left us with 15 papers.

Overall characteristics of protocols and samples: A summary of study protocols and overall subject characteristics for each of the 15 articles can be found in Table 1. In total, we included 168 patients who underwent non-invasive BCI therapy that triggered sensory feedback. For 120 of them, the device used was an exoskeleton. The remaining received above motor-threshold FES. Additional feedbacks ranged in terms of visual response (6 studies), the therapist voice (2 studies) or tongue stimulation (1 study). In some cases, more than one additional feedback was given. Overall intervention time ranged between 5 and 64 hours (μ=17.13 σ=16.76). Some studies also involved conventional therapy (if present, it was between 7.50 and 40 hours, sometimes in higher amount compared to the experimental therapy [7,15]). In terms of personal characteristics, we could not retrieve the age and sex for most of the time. However, 83 subjects (chronic patients) had a stroke more than 6 months before participating in the experiment (μ=53.88, σ=53.03 months) and 26 were sub-acute (μ=2.83, σ=1.06 months). For two studies [10,16] this parameter was not given. In terms of severity of impairment, almost half of the subjects were severe (FMA<=20 and ARAT<=20) and the remaining moderate-to-mild (here we refer to mild as less impaired than moderate).

Patient characteristics: For patient individual characteristics only stroke onset and impairment at baseline was investigated, due to lack of data for other parameters (e.g. age, lesion side). In Fig. 1 the difference between binarized groups of subject is displayed: chronic vs. acute/sub-acute, severe vs. moderate/mild and the combination of the former. Patients in the sub-acute phase showed significantly better recovery compared to the other groups (Fig. 1b), the same can be observed for patients who are rather mildly impaired before starting the experiment (Fig. 1a). When combining the two groups (chronic and severe vs. sub-acute and moderately impaired), the effect size is increasing.

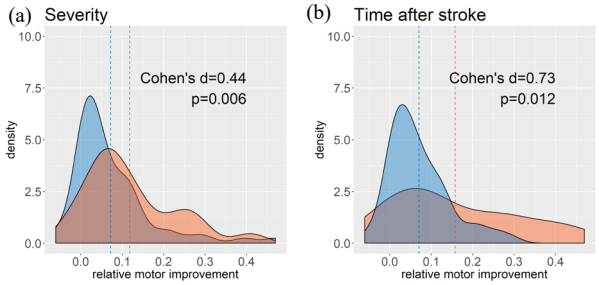


Figure 1 – Density plots for comparison between patients' characteristics and their motor improvement, normalized by the maximum score. We consider their impairment at baseline (severe or not) and the chronicity (stroke occurred more than 6 months before study). **(a)** Impairment at baseline; **(b)** chronicity. In blue: severe and chronic, in red: moderate–to-mild impairment and sub-acute patients respectively. The dashed lines represent the mean of each contribution.

Protocol features: One source of variance for the motor improvement comes from the subjects' characteristics, another from the design of the experimental protocol. The 15 studies considered all differed in terms of intervention time, duration (if any) of an additional conventional therapy, type of instruction to command the BCI and the device exploited for sensory feedback. The type of movement triggered by the BCI also changed among studies. We divided them into proximal (shoulder, arm and elbow), distal (hand and wrist) and movements combining distal and proximal parts (e.g. reach and grasp). As motor improvement also

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-60

significantly differed according to the baseline, we decided to use proportional improvement of motor functions as dependent variable to take into account baseline differences and give stronger importance to improvements in severe patients. We acknowledge that there are also other computational approaches that could have been used [15]. Due to space limitations detailed technical aspects related to BCI and EEG recordings were not addressed in the review; also, thorough information about these aspects is often not available. Nevertheless, it is obvious that the processing pipelines used, as well as the classifiers, were quite heterogeneous. The targeted brain signal was mainly the sensorimotor rhythm, but recorded from different areas; moreover, the pre-processing pipeline varied and the decoding algorithm ranged from support vector machine to common-spatial patterns, from naïve Bayes and Gaussian classifiers to independent component analysis.

Firstly, instructions given to subjects to command the BCI were addressed. In most papers, they were asked to perform motor imagery (MI) and in fewer to try and make the movement that the device would have then actually completed. Patients using MI, had a better recovery compared to the other group with a moderate, but not significant effect size (d=0.3). Attempting the movement was never tried in studies that included non-severe subjects to avoid the possibility of some actively initiated movements for only a subgroup. However, the results did not change when we only looked at severe patients. This may be due to the fact that proportional improvement rather than motor improvement was used.

No significant difference in recovery was observed when comparing the devices used, nor the number of additional feedback provided.

Interesting results can be observed when looking at the proportional improvement according to the movement performed. All the movements were quite simple (e.g., hand opening, elbow flexion) and only in two study more complex ones (i.e. reaching and grasp [8,17]) were attempted. For these specific cases, complexity goes along with an action involving distal and proximal parts. Fig. 2 shows significant larger improvement when the latter types are performed compared to movements involving either part.

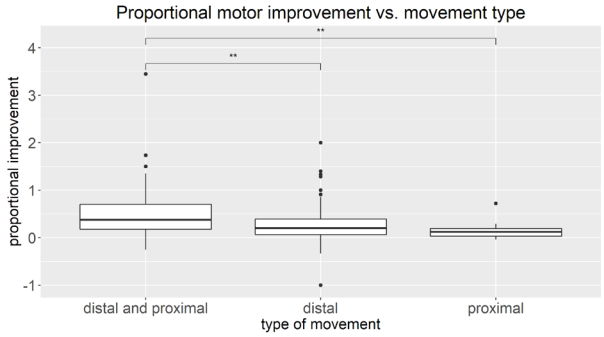


Figure 2 – Proportional motor improvement according to type of movement trained with either device. There is a significant difference in recovery when a movement involving both distal and proximal parts of the upper-limb is used compared to movements involving only one part. ** p<=0.001

Other significant results arose when studying the explicit use of conventional therapy in parallel with the experimental one. Initially, we grouped the subjects according to whether or not they received additional standard therapy. The effect size was quite strong (d=0.5), and the two distributions significantly different (p<0.001) with the group receiving conventional therapy showing larger improvement. Secondly, we stratified patients into more groups according the total amount of standard therapy received and reported the results in Fig. 3a. Examining correlations between the total amount of therapy received and the respective proportional improvement, we found r=0.34, p<0.001. However, no significant correlation was found when looking at the total amount of experimental therapy. Finally, we summed the total hours of therapy (experimental with BCI and standard) and found a strong correlation (r=0.24, p<0.01). Boxplots can be seen in Fig. 3b.

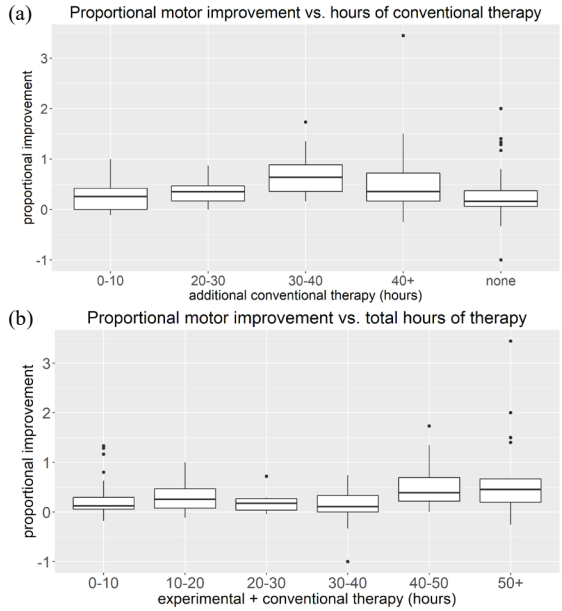


Figure 3 – Proportional motor improvement compared to overall hours of therapy received. **(a)** Additional conventional therapy; **(b)** sum of interventional and conventional therapy. If not explicitly specified in the study protocol, standard therapy was considered as none.

DISCUSSION

This systematic review aimed at looking at possible relationships between patient and study protocol characteristics and motor improvement of the upper-limb. The target population was stroke patients who managed to control a BCI to trigger an external device for supporting the movement and delivering sensory feedback; specifically, we looked at studies with FES or exoskeletons.

Stroke onset and severity at baseline: chronicity and severity of impairment at baseline were found to be both important features. As expected, the group of sub-acute patients generally improved more compared to the chronic ones. This is probably due to the "high-neuroplasticity window" that opens for around 8-12

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-60

weeks right after the stroke incident in which also the brain naturally starts to re-organize [24]. During this period re-learning processes are more efficient and efficacious [25]. Overall, this result is in agreement with recent meta-analyses on BCI rehabilitation [2]. We also observed that starting at a lower level of motor capability leads to smaller improvement, even though the range for improvement is much higher. Considering that often larger impairment results from larger lesions, it seems reasonable that the capability for reorganization, plastic changes and re-learning might be limited.

Motor imagery vs. attempted movement: Among the protocol design features, we were expecting to see larger improvement when the instruction to command the BCI was to try to attempt the movement rather than MI, as this should try to re-activate the "normal" motor pathways. Yet, there is a great overlap between MI and real-movement pathways. This correspondence has been proven both with fMRI studies [26] and in terms of physics law preservation in MI, such as the Fitt's law of timing [27]. In this analysis, it seems that MI leads to a larger, although not statistically significant, recovery. A clear limitation here is that the number of patients for the two groups is very different, being the group of MI almost double in size compared to the other. Furthermore, the group attempting to do the movement is composed only of severe patients, who generally have lower performance. Nonetheless, when focusing only on hemiparetic stroke subjects, the results do not differ much. Therefore, a feature to look at, rather than the instruction given, is the BCI classifier and the accuracy obtained. However, because the primary outcome of all the included studies was motor recovery, the classifier algorithm exploited was seldom reported in detail. Moreover, it was seen that performance, whenever reported, had been computed in different manners (i.e. different proportions between true/false positive/negatives), making comparisons difficult.

Movement type: We found significantly stronger motor recovery when the repeated movement contained joints in both the proximal and the distal part of the upper-limb. Moreover, there was a slight trend for which distal movements perform better than proximal ones, though the difference is not statistically significant. Overall, findings from this section suggest that doing more complex movements, and possibly different types of movement during the BCI therapy, may lead to better motor improvement [28]; additive effects may be obtained with functional movements [5]. At this regard, we speculate that more movements can be achieved by combining an orthosis with FES. Such a union has already been tested in other studies, but rarely triggered by a BCI [29].

Therapy quality or quantity?: In one out of three reviewed studies, for a total of 53 patients, conventional therapy was provided in addition to the experimental one. From the current analyses, patients who received it showed significantly better improvement, with a strong effect size. Furthermore, the amount of training hours significantly correlated with the degree of recovery. A similar correlation was observed when looking at the total amount of therapy, of either type, during the study period. On the contrary, no correlation was found between the total hours of the intervention and motor recovery. It must be pointed out that except for [19], the hours of intervention and additional therapy were kept constant for patients belonging to the same study. Taken together, one could summarize that the addition of standard therapy (1) increases the total amount of therapy and (2) although not working on the synchrony between brain intentions and feedbacks, helps the patient doing more types of movements and especially functional ones. Two randomized clinical trials [4,5] evaluated the difference between BCI and conventional rehabilitation; unfortunately, due to the lack of individual data, they were not included in this review. Nonetheless, in [5] the BCI group received additional therapy, having 50% more overall time of therapy with respect to those receiving only standard. Differently, in [4] the control with standard therapy received an overall same amount of therapy. In both studies, the experimental group improved significantly better than their control and when reported, kept this improvement in the follow-up measurement. To more deeply investigate the role of standard therapy and how a BCI performs compared to it, a randomized clinical trial should be designed in which the total duration for the BCI and the conventional therapy are exactly the same, as well as the performed movements.

Limitations: In this systematic review only studies providing individual data for motor assessment were included. Therefore, the overall sample size was limited. Moreover, the restriction on the motor scales, led to discard some studies. Indeed, adding other motor scales would be relevant. In terms of characteristics, we here only looked at two patients' personal features: motor score at baseline and stroke onset. In further analyses, it would be interesting to add more parameters such as age, lesion site, lesion size or hemispheric dominance and handedness. Due to space limitation technical aspects of BCIs used in rehabilitation settings were also not in detailed scope of this review. To make BCI-based interventions more comparable and standardize them it is of crucial importance that the details about classifiers and analytical pipelines are provided and the different approaches compared. As a brief overview aspects relevant for the classifiers are both the type of pre-processed signal fed to the decoder and the algorithm itself. Furthermore, related to the neurophysiological mechanisms targeted, the location side from which the signal is read is also important and may have an effect on the motor improvement.
Finally, within the present review each feature impacting on the effect of the intervention was addressed separately, however in upcoming meta-analyses it is essential to combine the different aspects and features to develop a strong predictive model.

Recommendations: Limiting factors in these analyses were related to the absence of some features, especially for each individual. A more detailed reporting of

Published by Verlag der TU Graz
Graz University of Technology

individual therapy, BCI-based features, patients' characteristics and treatment durations will also help to better compare studies and develop standards and optimized interventions to maximize the treatment effects. In terms of BCI-therapy, it is important to develop devices which support multiple, functional, longer, and more complex movements. However, at this concern, we acknowledge that multi-class discrimination with non-invasive techniques may be strenuous. Therefore, developments towards hybrid BCI or invasive techniques might pave the way towards these goals.

CONCLUSION

In the present review, we identified five features that show relevance for driving motor recovery of the upper-limb in stroke patients in BCI rehabilitation. In terms of patient characteristics, being in the sub-acute phase and mildly impaired support functional improvement within this treatment strategy. Regarding protocol features, we observed the importance of adding conventional therapy that can be related to both longer rehabilitation duration and more variety and functionality of executed movements.

E-mail address:
claudia.bigoni@epfl.ch

Postal address:
Claudia Bigoni
CNP and BMI, EPFL
Campus Biotech, 9, Chemin des Mines, 1202 Geneva

REFERENCES

[1] Raffin E, Hummel FC. Restoring Motor Functions after Stroke: Multiple Approaches and Opportunities. The Neuroscientist 2018. 24(4):400-416.
[2] Cervera MA, Soekadar SR, Ushiba J, Millán Jdr, Liu M, Birbaumer N, et al. Brain-Computer Interfaces for Post-Stroke Motor Rehabilitation: A Meta-Analysis. Annals of Clinical and Translational Neurology 2018. 5(5): 651–663.
[3] Lopez E, Sarasola A, Irastoza N, Birbaumer N, Ramos A. Brain-Machine Interfaces for Rehabilitation in Stroke: A Review. Neurorehabilitation 2018. 43:77-97.
[4] Ang KK, Guan C, Puha KS, Wang C, Zhou L, Tang KY, et al. Brain-Computer Interface-Based Robotic End Effector System for Wrist and Hand Rehabilitation: Results of a Three-Armed Randomized Controlled Trial for Chronic Stroke. Frontiers in Neuroeng. 2014.. 7.
[5] Kim T, Seongsik K, And Lee B. Effects of Action Observational Training Plus Brain-Computer Interface-Based Functional Electrical Stimulation On Paretic Arm Motor Recovery in Patient with Stroke: A Randomized Controlled Trial: Effects of AOT Plus BCI-FES On Arm Motor Recovery. Occupational Therapy International 2016.23(1):39-47
[6] Ang KK, Chua KSG, Phua KS, Whang C, Chin ZY, Kuah CWK, et al. A Randomized Controlled Trial on EEG-Based Motor Imagery Brain-Computer Interface Robotic Rehabilitation for Stroke. Clinical EEG and Neuroscience 2015. 46(4):310-20.
[7] Li M, Liu Y, Wu Y, Liu S, Jia J, Zhang L. Neurophysiological Substrates of Stroke Patients with Motor Imagery-Based Brain-Computer Interface Training. International Journal of Neuroscience 2014. 124(6):403-15.
[8] Ramos-Murguialday A, Broetz D, Rea M, Läer L, Yilmaz Ö, Brasil F, et al. Brain-Machine Interface in Chronic Stroke Rehabilitation: A Controlled Study: BMI in Chronic Stroke. Annals of Neurology 2013. 74(1): 100–108.
[9] Biasiucci A, Leeb R, Iturrate I, Perdikis S, Al-Kohdairy A, Corbet T, et al. Brain-Acutated Functional Electrical Stimulation Elicits Lasting Arm Motor Recovery After Stroke. Nature Communications 2018. 9(2421).
[10] Frolov AA, Mokienko, Lyukmanov R, Biryukova E, Kotov S, Turbina L et al. Post-Stroke Rehabilitation Training with a Motor-Imagery-Based Brain-Computer Interface (BCI)-Controlled Hand Exoskeleton: A Randomized Controlled Multicenter Trial. Frontiers in Neuroscience 2017. 11(400).
[11] Remisk AB, Dodd K, Willimas L, Thoma J, Jacobson T, Allen JD, et al. Behavioural Outcomes Following Brain-Computer Interface Intervention for Upper Extremity Rehabilitation in Stroke: A Randomized Controlled Trial. Frontiers Neuroscience 2018. 12(752).
[12] Rabadi MH, Rabadi FM. Comparison of the Action Research Arm Test and the Fugl-Meyer Assessment as Measures of Upper-Extremity Motor Weakness After Stroke. Arch Phys Med Rehabil 2006. 87(7): 962-6.
[13] Cohen, J. Statistical Power Analysis for the Behavioural Sciences. Routledge. *New York, NY* 1988.
[14] Tabering CB, Lopez CA, Carrere LC, Spaich EG, Ballario CH. Neurorehabilitation Therapy of Patients with Severe Stroke Based on Functional Electrical Stimulation Commanded by a Brain Computer Interface. Journal of Rehabilitation and Assistive Technologies Engineering 2017. 5:1-12.
[15] Pichiorri F, Morone G, Petti M, Toppi J, Pisotta I, Molinari M, Paolucci S, et al. Brain-Computer Interface Boosts Motor Imagery Practice During Stroke Recovery: BCI and Motor Imagery. Annals of Neurology 2015. 77(5): 851–865.
[16] Prasad G, Chowdhury A, Raza H, Meena YK, Dutta A. Combining Mental and Physical Practice Using Brain Computer Interface Driven Hand Exoskelton for Personalized Post-Stroke Neuro-Rehabilitation. 8th Annual Translational Medicine (TMED8) Conference 2017
[17] Ibáñez J, Monge-Pereira E, Molina-Rueda F, Serano JI, Castillo Mdm Cuesta-Gómez A, et al. Low Latency Estimation of Motor Intentions to Assist Reaching Movements Along Multiple Sessions in Chronic Stroke Patients: A Feasibility Study. Frontiers in Neuroscience 2017. 11(126)
[18] Sullivan J. L., Bhagat N. A., Yozbatiran N, Paranjape R, Losey CG, Grossman RG, et al., Improving Robotic Stroke Rehabilitation by Incorporating Neural Intent Detection: Preliminary Results from a Clinical Trial. IEEE International Conference on Rehabilitation Robotics 2017: 122-127
[19] Bundy DT, Souders L, Baranyai K, Leonard L, Schalk G, Coker R, et al. Controlesional Brain-Computer Interface Control of a Powered Exoskeleton For Motor Recovery In Chronic Stroke Survivors. Stroke. 2017 48(17):1908-1915
[20] Várkuti B, Guan C, Pan Y, Ang KK, Kuah CWK., Chua K, et al. Resting State Changes in Functional Connectivity Correlate with Movement Recovery for BCI and Robot-Assisted Upper-Extremity Training after Stroke. Neurorehabilitation and Neural Repair 2013 27(1):53-62
[21] Caria A, Weber C, Brötz D, Ramos A, Ticini LF, Gharabaghi A, et al. Chronic Stroke Recovery After Combined BCI Training and Physiotherapy: Case Report. Psychophysiology 2011 48:578-582
[22] Mukaino M, Ono T, Shindo K, Fujiwara T, Ota T, Kimura A, et al. Efficacy of Brain-Computer Interface-Driven Neuromuscular Electrical Stimulation for Chronic Paresis after Stroke. Journal Rehabilitation Medicine 2014 46: 378-382
[23] Belardinelli P, Laer L, Ortiz E, Braun C, Gharabaghi A. Plasticity of Premotor Cortico-Muscular Coherence in Severely Impaired Stroke Patients with Hand Paralysis. Neuroimage: Clinical 2017 14:726-733
[24] Kwakkel G, Kollen BJ. Predicting Activities after Stroke: What is Clinically Relevant? International Journal of Stroke 2013. 8(1): 25-32
[25] Langhorne P, Bernhardt J, Kwakkel G. Stroke Rehabilitation. The Lancet 2011. 377(9778): 1693-1702.
[26] Gerardin E, Sirigu A, Lehéricy S, Poline JB, Gaymard B, Marsault C, et al. Partially Overlapping Neural Networks for Real and Imagined Hand Movements. Cerebral Cortex 2000. 10:1093-1104
[27] Lotze M and Halsband U. Motor Imagery. Journal of Physiology. 2006 Paris 99: 386-395.
[28] Ramos A, Curado MR, Broetz D, Yilmaz O, Brasil FL, Liberati G, et al. Brain-Machine Interface in Chronic Stroke: Randomized Trial Long-Term Follow-Up. Neurorehabilitation and Neural Repair. 2019.
[29] Grimm F, Walter A, Spüler M, Naros G,Rosenstiel W, And Gharabaghi A. Hybrid Neuroprosthesis For The Upper-Limb: Combining Brain-Controlled Neuromuscular Stimulation With Multi-Joint Arm Exoskeleton. 2016 10(367).

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-61

TOWARDS EEG-BASED SIGNALS CLASSIFICATION OF RGB COLOR-BASED STIMULI

Sara Åsly, Monika Gilde, Luis Alfredo Moctezuma, Marta Molinas

Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway

E-mail: sarahaa@stud.ntnu.no

ABSTRACT: This research looks at the possibility to actuate devices by looking at primary colors, which is thought to be especially useful for impaired individuals having restricted motor control. Such a brain-computer interface (BCI) requires reliable detection of color related features captured in electroencephalograph (EEG) data. This paper presents analytic and empirical signal analysis methods for analyzing EEG signals, motivated by the search for features directly related to the color perception in the human brain. Methods used are Fourier transform (FT) and short time Fourier transform (STFT). Empirical mode decomposition (EMD) is used to extract information used for feature extraction. Classification accuracies are tested using the machine learning algorithms: random forest (RF), support vector machine (SVM), k-nearest neighbors (kNN), decision tree (DT) and naive Bayes (NB). Using data from 7 subjects, a general model classifies RGB with 0.37, while the best subject specific model achieves an accuracy of 0.58. The classification accuracy between gray and any one of the RGB colors is 0.98 with NB. These results are encouraging and can be improved by further exploring features and classification techniques.

INTRODUCTION

Electroencephalographic signals (EEG) represent the electrical activity in the brain. By placing electrodes on the scalp, one can record these signals. One electrode records the cumulative electrical activity of neurons. EEG signals are non-stationary, time-dependent, and because of cumulative electrical activity, most likely multicomponent signals [1]. Also, non-invasive EEG signals have a small amplitude and are extremely noisy. These properties are but a few of the reasons raw EEG signals do not provide useful information alone, and dedicated signal analysis is therefore required to extract relevant information contained within the signal.

Choosing a suitable signal analysis method is a crucial step in the process of finding useful information in EEG data. In general, no particular method will provide the best results. Choice of signal analysis tool depends for instance on the characteristics of the signal and the aim of the experiment.

The goal of most EEG experiments is to classify signals produced by specific brain activity correctly. A feature is an individual measurable property of the process being observed [2], and any recorded EEG activity includes many different features [3]. Researchers, therefore, search for a limited amount of features that can differentiate signals with certainty.

The process of selecting only a subset of variables in the input which can efficiently describe the data is called feature selection. Feature selection decreases the effect of noise, irrelevant or redundant variables are reduced, and the predictor performance improved [2][4].

Researchers have explored techniques to predict which color a subject is looking at using different indirect approaches such as analyzing psychological and emotional response to color [5][6]

Recently, classification of EEG signals produced by random visual exposure to primary colors was presented in [7]. Independent component analysis (ICA) was used to remove artifacts. Event-related spectral perturbations (ERSP) were used as features for a support vector machine (SVM), and the highest classification accuracy was 0.97, more information at [3]. In general, empirical mode decomposition (EMD) for feature extraction from color related EEG signals has proven to be successful in several studies [8].

A neural signature of the unique hues (red, yellow, green, and blue) were discovered 230 ms after stimulus onset at a post-perceptual stage of visual processing [9]. This interesting study uses ERPs (recorded neural activity time-locked to an event) evoked in the response to different hues.

In this paper, analytic and empirical signal analysis methods are investigated in order to evaluate their ability to reveal color specific patterns in EEG signals produced by exposure to primary colors. EMD is used as basis for feature extraction. Identifying a set of features for color identification in EEG signals would enable less complex machine-learning based models, reducing the computational time for real-time color identification. A reliable real-time classification of EEG signals produced by looking at a color could enable physically disabled people with cognitive functions to control their environment. For instance, a user can open and close doors by looking at colored signs.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-61

METHODS AND MATERIALS

Fast fourier transform (FFT):
Information is often contained in the frequencies of a signal. A signal is transformed from the time domain to the frequency domain with the Fourier transform (FT). The discrete Fourier transform (DFT) [10] is defined as

$$F_n = \sum_{k=0}^{N-1} f_k \cdot e^{\left(-\frac{2\pi i n k}{N}\right)}$$

For faster computation, FFT is often used. FFT computes the DFT of a signal. For a signal of length N, DFT needs $2N^2$ computations, while the FFT uses only $2N \cdot log(N)$. A significant drawback of the FT is the loss of time characteristics and is therefore not suitable for interpreting time-dependent signals. Methods based on the time-frequency domain have been developed for feature extraction in non-stationary signals.

Short time Fourier transform (STFT):
STFT preserves information about time domain by windowing the signal around a particular instant in time and calculating the local FT for each time window. The information obtained from the STFT is presented in a spectrogram. Spectrograms show how the spectral density of a signal varies with time, giving the information about the quantity of the frequency, and at what time this frequency is present.
STFT is limited due to windowing of the signal, which causes a trade-off between time precision and frequency resolution. Frequency resolution must be sacrificed to detect an event precisely in time, and vice verse. This trade-off between time and frequency resolution makes it essential to choose an appropriate window size to optimize both time and frequency [11].

Empirical mode decomposition (EMD):
EMD is a well-known technique used to analyze non-stationary and non-linear data [12]. EMD does not make assumptions regarding stationary or linearity of data, which motivates it's use for analyzing EEG data [8]. In contrast to FFT and STFT, EMD is data-driven, based on the assumption that a signal consists of several intrinsic mode functions (IMFs), that must satisfy two basic conditions:

- Number of zero-crossings must equal or differ by one compared with number of extrema in the signal.
- The mean value of the upper and lower envelope of the signal must be equal to zero at any point.

The EMD algorithm finds all the IMFs trough a process called *Sifting*. The calculation of the IMFs given a signal $x(t)$ are done as follows [12]:

1. Identify all extrema (maxima and minima) in $x(t)$
2. Interpolate between minima and maxima, generating the upper and lower envelope; e_{upper} and e_{lower}
3. Determine the local mean as $a(t) = \frac{e_{upper}+e_{lower}}{2}$
4. Extract the mean from the signal; $h_1(t) = x(t) - a(t)$
5. Decide whether it is an IMF or not based on two basic conditions for IMFs mentioned above
6. Repeat step 1 to 4 until an IMF is obtained.
7. Subtract the IMF from the original signal
8. Repeat steps 1-6 until there are no IMFs left to extract, the last extraction resulting in a residue

The decomposition is complete when the sum of the IMFs and the residue is negligible.

Feature extraction and classification:
The main method used for feature extraction and classification is based on the work presented in [13]. The feature extraction stage for each electrode consists on the computation of energy (instantaneous and teager energy) and fractal (Petrosian and Higuchi fractal dimension) features, but additionally, in this paper, a set of statistical values (min, max, mean, median, variance, standard deviation, kurtosis, skew) are computed for each channel. This procedure is illustrated in Fig. 1. Lastly, supervised machine-learning based models were created using 10-folds cross validation using the accuracy metric. The machine-learning based algorithms used are, random forest(RF), support vector machine (SVM), k-nearest neighbors (kNN), decision tree (DT) and naive Bayes (NB).

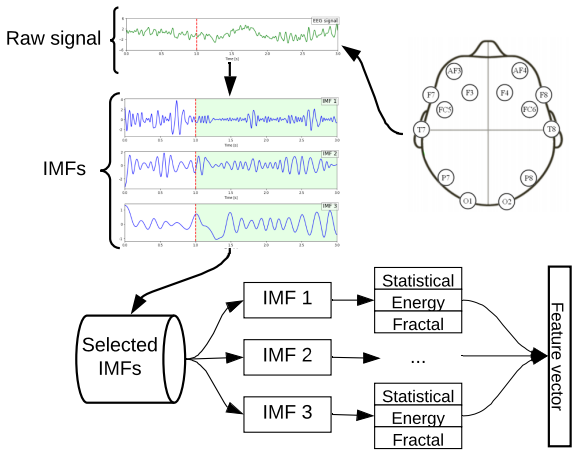


Figure 1: Flowchart illustrating the feature extraction procedure using EMD. The procedure is the same for each channel

Dataset description:
The dataset consist of EEG signals that were recorded from P1, P2, O1 and O2 channels according to the 10-20 international system using the BCI200 with g.tec's MOBIlab portable device with a sample rate of 256 Hz [7].
The dataset consist of EEG signals from 7 Subjects while watching RGB colors, each color was presented randomly 60 times to each subject. Signals were band pass filtered from $0.1 - 30$Hz. To reduce the effect of abnormal values, signals crossing $\pm 60\mu V$ were removed. In addition, some trails were excluded due to electromyogram- (EMG) and electrooculogram (EOG) artifacts. The final dataset used in this paper consist of 52 trails for each color, in order to obtain a balanced dataset Next, the data was re-organized in 3 seconds long "epochs" (768 data points). One epoch contains samples

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-61

from all channels where the subject is looking at gray for one second, followed by two seconds of looking at one of the RGB colors. The colored light switched on at $t = 1s$ in all the following results.

RESULTS

Signal analysis:

Fig. 2 shows a single EEG signal, where the red vertical line indicates the moment of color exposure, and green background illustrates when green light is continuously on.

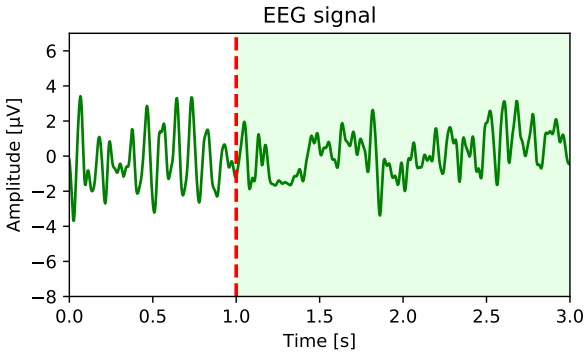


Figure 2: Example of one EEG signal where green light is switched on at $t = 1$, from channel P1

As an example of FFT from the above EEG signal is presented in Fig. 3. A larger magnitude at lower frequencies $0-10$Hz than $10-33$Hz is observed. Presence of higher magnitudes in $0-12$Hz is expected. These frequency bands corresponds to Delta, Theta and Alpha rhythms in the brain [14]. It is reasonable to believe that the person in the experiment was in a day dreaming / relaxed / wide awake state during data collection, and FFT satisfies the expectation. There are no frequencies above 33Hz, confirming successful prepossessing. As frequencies with lower magnitude are by eye inspection uniformly spread, it is difficult to draw further conclusions about frequencies in the EEG signal based on FFT.

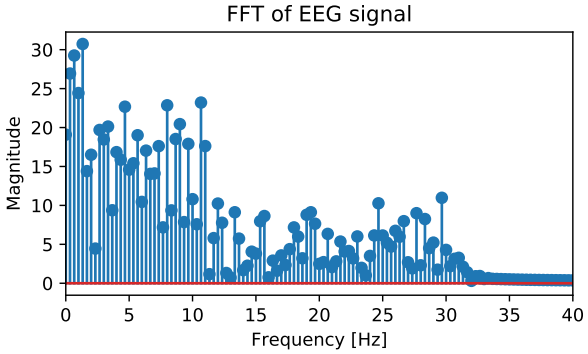


Figure 3: FFT of grand average EEG signal (green)

Since the aspect of time is lost in FFT, STFT was applied to investigate possible changes of frequencies over the given time period. A STFT with a "Hanning" window size of 200 samples, overlap off $200-10$ samples and sampling frequency of 256 Hz was used to produce the spectrogram in Fig.4

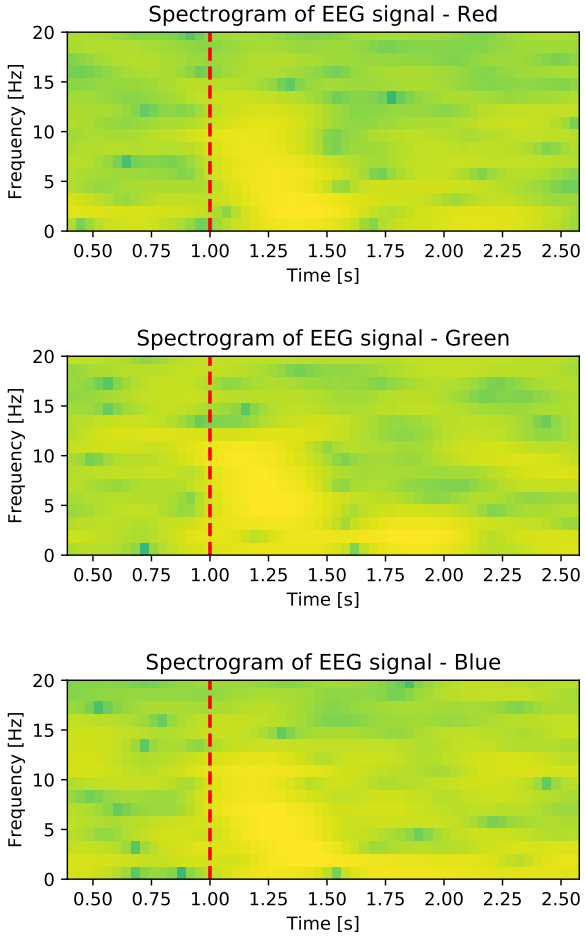


Figure 4: Spectrogram of grand average EEG signal for RGB

The spectrogram represents the grand average for RGB respectively. Despite apparent prevalence of noise, there is an amplitude increase in $2-12Hz$ for all colors, and for green there is also a notable amplitude increase for $0-5Hz$ in the time frame $1-2s$. Hence, averaging data reveals a change caused by visual stimuli from gray to RGB colors $200-300$ms after exposure. However, is clear from their overlap that frequency alone is not sufficient to separate three colors. In addition, there is no lasting change in frequency, even though all subjects are continuously looking at color from from $t = 1s$ to $t = 3s$. Information gain from STFT is limited, and doubtfully sufficient to reveal a signal feature specific for each of the colors.

For this reason, EMD algorithm was applied on the signal, and after 10 siftings, the residual fulfills IMF requirements discussed in the methodology section. Fig. 5 shows an example of the 5 IMFs and the residual for color green.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-61

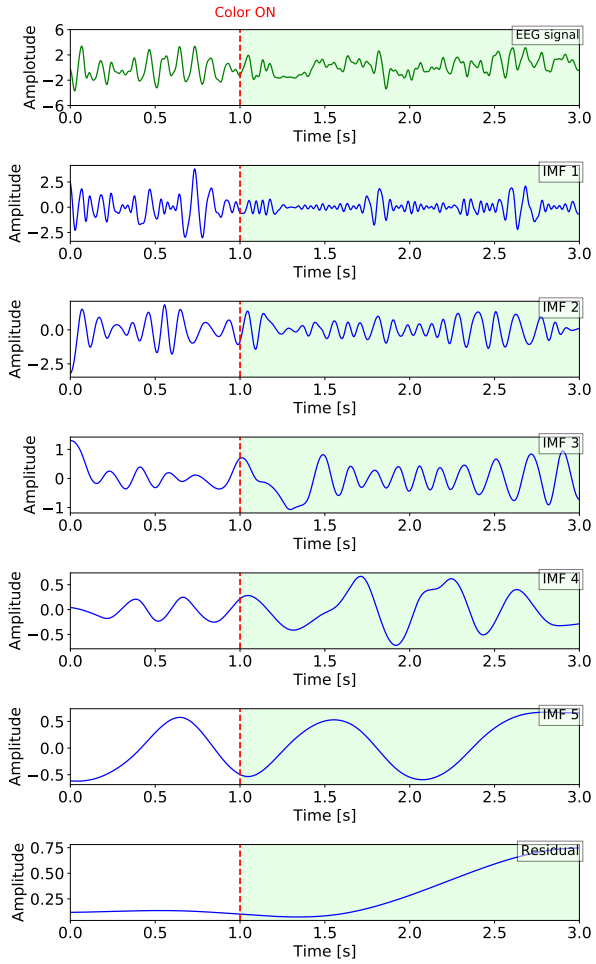


Figure 5: Original EEG signal, extracted IMFs and the residual. Green background represents green light is continuously on

EMD does not use windows. Using windows in analysis of the signal would force the ends to zero, and therefore mask end effects. The end effect problem has not been taken into account in this paper. In Fig. 6 a spectrogram of each of the IMFs are plotted. EMD successfully extracts the highest frequency components in the first IMFs. IMF1 reveals slight increase in magnitude for all frequencies at $t \approx 1.5$. This might be related to color exposure or change of mental state for the person in the experiment. Extracted IMFs can be representing physical properties of the process from which the signal is obtained. However, the problem of mode mixing in EMD caused by presence of adjacent frequencies will cause loss of meaningful information in the IMFs. A new method for separating closely spaced spectral tones using EMD is presented in [15][16], and could be implemented to improve results.

Neither spectrograms nor IMFs reveal distinct color dependent frequency or amplitude related characteristic by visual inspection.

Classification:

To test if machine-learning models can classify RGB colors from EEG signals using features based on EMD, the following experiments are proposed:

1. Classify RGB colors from gray color
2. Classification of red, green and blue colors for each subject
3. Classification of red, green and blue considering the EEG signals from all the subjects

The aim of the first experiment is to provide experimental information about the performance of the method and to check if there are feature that can separate these two classes (gray or RGB colors).

In the second experiment mentioned, the classifier consist of three classes (red, green and blue) with the aim to check if using the proposed method is possible to differentiate between them. It can be the second step for a real implementation of a BCI based on RGB colors. Since the first step can identify when a RGB color is presented and then recognize the specific color. Following this aim is important to check the feasibility of a general model for the second experiment, that is why, the last experiment consist of the same experiment but considering the EEG signals from all the Subject to create the classifier.

For all experiments, the procedure described for *Feature extraction and classification* is used. Accuracy metric after 10-fold cross-validation is presented. All the classifiers are tested with different kernels, number of neighbours or deph depending of each one and the best parameters are automatically selected. Unless otherwise stated, default parameters of scikit-learn classifiers are used [17]. Note that the chance level for the first experiment is 0.5 of accuracy, and for experiment 2 and 3 it is 0.33

Experiment 1; gray vs RGB:

For a possible real-time application, it will be important to clearly distinguish if the subject is looking at nothing in particular, or decisively looking at a color. To simulate "nothing in particular", gray color is used. The complexity of such differentiation was investigated by first classifying if subjects were looking at gray or an RGB color. An event-related potential (ERP) (P300) is expected approximately 300*ms* after presentation of an infrequent stimulus. The part of the signal where the subject is exposed to the color will therefore contain the P300 component, and it can easily be distinguished from a signal not containing an ERP. Therefore, classification removing data points between $t = 1-2$ was investigated, and referred to as "data excluded". Results for gray vs color classification are presented in Tab. 1.

Table 1: Accuracy (Acc) obtained for the first experiment using all the features (feats.), the statistical features (S.) and only one statistical feature (the mean).

	Acc		Acc, data excluded	
All feat.	0.99	RF depth 5	0.87	RF depth 5
S. feat.	0.92	linear SVM	0.89	RF depth 6
Mean	0.91	8-NN	0.92	4-NN

Surprisingly, when excluding data from $t = 1-2$, the accuracy only decreases accuracy by 0.12 using all features. An interesting finding is a 0.92 accuracy when using data

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-61

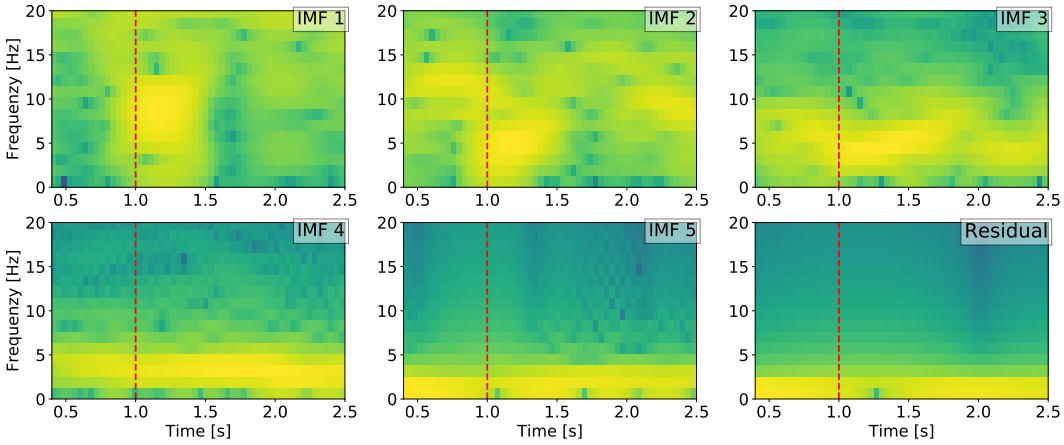


Figure 6: Spectrogram of original EEG signal, each IMF and residual

without ERP, and only one feature; the mean. These results yields a promising first step towards a less complex real-time application for separating between gray and RGB colors.

Second experiment; classification of red, green and blue color:

Firstly, a model including data from all seven subjects were developed, reaching an accuracy of 0.37 using a Gaussian distributed NB. Limited amount of data and individual differences are believed to impair the result, and hence subject specific models were developed. No classifier alone performed better for all subjects; but rather different classified yield better result dependent on the subject. There were in particular one subject that consistently obtained higher accuracy, when testing with all classifiers: 0.58 of accuracy using NB, 0.51 using linear SVM, 0.47 with 6-NN, 0.53 using DT, and finally 0.57 using RF with depth 4. On the other hand, another subject model classified at chance level. Tab. 2 summarizes accuracies of the RGB models considering each subject separately.

Table 2: Accuracy (Acc) reached for the second experiment, classifying red, green and blue colors considering each subject separately

	Accuracy	Classifier
Average	0.42	Depends on the subject
Max	0.58	NB

The mean accuracy for the subject model is found by finding the maximum accuracy for each subject individually, and then performing the mean of these. The best performing classification algorithm differs dependent on the subject, and hence no algorithm in particular can be preferred. The maximum accuracy is the highest individual accuracy obtained for one subject.

Interestingly, the accuracy increase when including only one feature - the mean. A possible explanation can be that redundant features forms the model, due to very limited source data.

CONCLUSIONS AND DISCUSSION

In this paper, several methods have been explored in order to check if there exist features that can be useful to describe the EEG data while the subject is looking at gray or RGB colors, and also considering RGB separately. In the signal analysis step, FFT, STFT and EMD were investigated.

The FFT indicates that the subject was in relaxed and/or in an awake state during the data collection, which indicates a realistic dataset. The STFT successfully identify the P300 when averaging all the data for each color.

The EMD method decomposed the original signals from each channel into several IMFs. Since the IMFs alone do not provide any information, they are analyzed further with STFT for visual inspection, and later used as basis for feature extraction.

None of the methods yields a lasting unique frequency marker sought after for RGB, however, there where clear frequency modulations detected in the spectrogram of each IMF. The frequency modulation after color exposure is confirmed with a successful classification of gray and RGB color with 0.99 of accuracy.

Accuracies from the second experiment, classifying RGB considering all subjects together yields incomplete or poor results, considering the chance level of 0.33 for the 3 classes, and with the best accuracy of 0.37 using NB. The highest classification of RGB on individual subject level was obtained using NB with and accuracy of 0.58. It can be concluded that color classification suffers from subject dependencies. Though NB yields highest accuracy in the classifications, it should not be concluded as a general preference for RGB classification algorithm.

These results indicate the feasibility of using the method for feature extraction and experimental evidence of differences between RGB colors EEG-based. Considering the results obtained in this paper and the experiments

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-61

proposed, it is reasonable to assume that improving the feature extraction stage with a subject tailored system, the accuracy can improve, which will be tested in future works.

REFERENCES

[1] Cohen L. What is a multicomponent signal? Proc. ICASSP. 1992;5:113 –116 vol.5.
[2] Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering. 2014;40(No.1):16–28.
[3] Alharbi ET, Rasheed S, Buhari SM. Feature selection algorithm for evoked EEG signal due to RGB colors. 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2016:1503–1520.
[4] Guyon I, Elisseeff A. Special issue on variable and feature selection. An Introduction to Variable and Feature Selection. 2003;3:1157–1182.
[5] Yoto A, Katsuura T, Iwanaga K, Shimomura Y. Effects of object color stimuli on human brain activities in perception and attention referred to EEG alpha band response. Journal of physiological anthropology. 2007;26:373–9.
[6] Zhang H, Tang Z. To judge what color the subject watched by color effect on brain activity. IJCSNS International Journal of Computer Science and Network Security. 2011;11(No.2):80–83.
[7] Rasheed S, Marini D. Classification of EEG signals produced by RGB colour stimuli. Journal of Biomedical Engineering and Medical Imaging. 2015;vol. 2(no. 5):p. 56.
[8] Riaz F, Hassan A, Rehman S, Niazi IK, Dremstrup K. Emd-based temporal and spectral features for the classification of EEG signals using supervised learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2016;24:28–35.
[9] Forder L, Bosten J, He X, Franklin A. A neural signature of the unique hues. Scientific Reports. 2017;7.
[10] Proakis JG, Manolakis DK. Digital signal processing (. 2014).
[11] Allen J. Short term spectral analysis, synthesis, and modification by discrete fourier transform. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1977;25(3):235–238.
[12] Huang NE, Shen Z, Long SR, *et al.* The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 1998;454(1971):903–995.
[13] Moctezuma L, Molinas M. EEG-based subjects identification based on biometrics of imagined speech using emd. Dec. 2018, 458–467.
[14] Tan D, Nijholt A. Brain-computer interfaces: Applying our minds to human-computer interaction. 2010.
[15] Fosso OB, Molinas M. Method for mode mixing separation in empirical mode decomposition. arXiv preprint arXiv:1709.05547. 2017.
[16] Gao Y, Ge G, Sheng Z, Sang E. Analysis and solution to the mode mixing phenomenon in emd. 2008 Congress on Image and Signal Processing. 2008.
[17] Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;12:2825–2830.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-62

CLASSIFICATION OF IMAGINED SPOKEN WORD-PAIRS USING CONVOLUTIONAL NEURAL NETWORKS

C. Cooney[1], A. Korik[1], R. Folli[2], D. Coyle[1]

[1] Intelligent Systems Research Centre, Ulster University, Derry, UK
[2] Institute for Research in Social Sciences, Ulster University, Jordanstown, UK

E-mail: cooney-c@ulster.ac.uk

ABSTRACT: Imagined speech is gaining traction as a communicative paradigm for brain-computer-interfaces (BCI), as a growing body of research indicates the potential for decoding speech processes directly from the brain. The development of this type of direct-speech BCI has primarily considered feature extraction and machine learning approaches typical to BCI decoding. Here, we consider the potential of deep learning as a possible alternative to traditional BCI methodologies in relation to imagined speech EEG decoding. Two different convolutional neural networks (CNN) were trained on multiple imagined speech word-pairs, and their performance compared to a baseline linear discriminant analysis (LDA) classifier trained using filterbank common spatial patterns (FBCSP) features. Classifiers were trained using nested cross-validation to enable hyper-parameter optimization. Results obtained showed that the CNNs outperformed the FBCSP with average accuracies of 62.37% and 60.88% vs. 57.80% ($p<0.005$).

INTRODUCTION

A direct-speech brain-computer interface (DS-BCI) is one in which a user's imagined speech is harnessed as the mode of communication between themselves and a computer, or interlocutor [1]. Imagined speech is the internal pronunciation of words or sentences, which does not result in any audible output [2]. Imagined speech has received relatively little attention from BCI researchers in comparison with more common paradigms such as motor imagery (MI) or steady-state visually-evoked potentials (SSVEP) (see [3] for a review). However, a DS-BCI does offer the possibility of a more naturalistic form of communication and must therefore be considered an important field of study within the BCI community. Both invasive and non-invasive approaches to data acquisition have been applied to the recording of imagined speech, primarily through electrocorticography (ECoG) [4] and electroencephalography (EEG) [5]. In this study, we focus specifically on the decoding of imagined speech from EEG recordings.

Approaches to imagined speech decoding have typically employed traditional BCI feature extraction and classification algorithms. Among the features used to represent imagined speech from EEG are autoregressive coefficients [6], common-spatial patterns [7] and spectrotemporal features [8]. More recently, Mel Frequency Cepstral Coefficients (MFCC) [9], [10] and Riemannian manifold features [5] have enabled imagined speech classification.

Several traditional machine learning approaches have been applied to the task of decoding imagined speech from EEG. These include support vector machines (SVM) [9], [11], Linear Discriminant Analysis (LDA) [6], [12], Naïve Bayes [12], k-Nearest Neighbors [10] and Random Forests [13]. Of these, the SVM has been the most-often utilized classification method, resulting in accuracies of 71.3% [7] and 69.3% [14] in binary tasks. However, to-date no combined feature extraction and classification method has proven itself to be the dominant approach. For this reason, research into a deep learning approach to imagined speech classification is a logical undertaking. Deep learning has been enormously successful across fields such as computer vision [15] and automatic speech recognition [16]. More recently, it has been successfully applied to BCI in relation to MI [17] and SSVEP [18] but its application to imagined speech has been relatively sporadic [19]. Of the deep learning approaches available, convolutional neural networks (CNN) have been the most heavily-utilized in relation to BCI/EEG. Among many others, the applicability of CNNs has been demonstrated for automated screening of depression [20], and prediction of drivers' cognitive performance [21]. For a review into deep learning analysis of EEG, see [22].

Here we evaluate the performance of two CNNs tasked with decoding imagined speech from EEG. The data used consisted of fifteen word-pairs extracted from a dataset containing six Spanish words produced with imagined speech. The performance of the CNNs are rated in comparison with a regularized-LDA (rLDA) trained on FBCSP features. A nested approach to cross-validation is implemented to facilitate parameter-optimization and improve the robustness of results. The results obtained show that the CNNs perform significantly better than the rLDA classifier, and that the performance of the deep CNN was significantly better than that of the shallow CNN.

MATERIALS AND METHODS

The methodology implemented in order to classify imagined speech production from EEG signals is

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-62

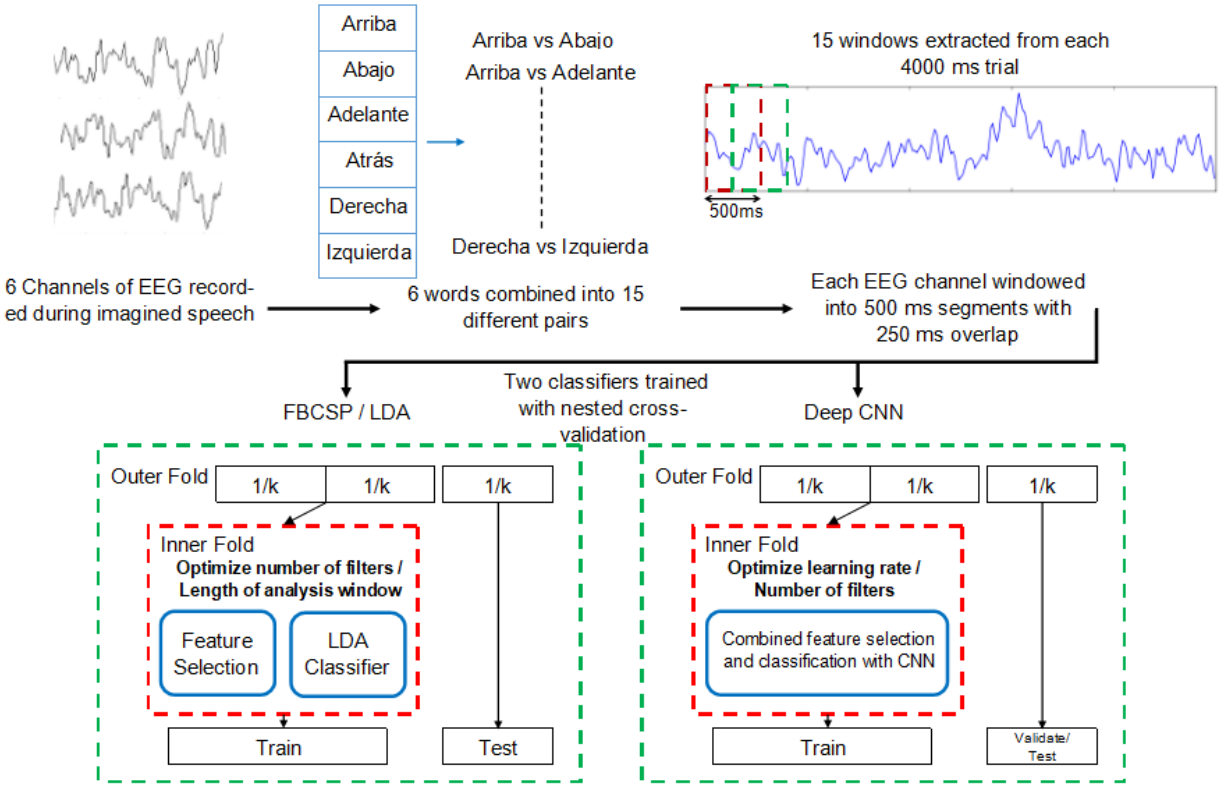


Figure 1Depiction of the methodology followed for this study

depicted in Figure 1 and described throughout the remainder of this section, beginning with the dataset.

Dataset: The dataset used for this research was recorded at the offices of the Laboratorio de Ingeniería en Rehabilitación e Investigaciones Neuromusculares y Sensoriales (LIRINS) in the Faculty of Engineering at the National University of Entre Ríos (UNER) by Pressel Coretto et al. [23]. EEG signals were recorded while 15 subjects performed overt and imagined speech tasks corresponding to the production of Spanish words and vowels. Only the EEG associated with imagined word production was analysed for this study. Thus, the EEG data used consisted of those trials recorded while participants imagined the production of six Spanish words: "arriba", "abajo", "derecha", "izquierda",

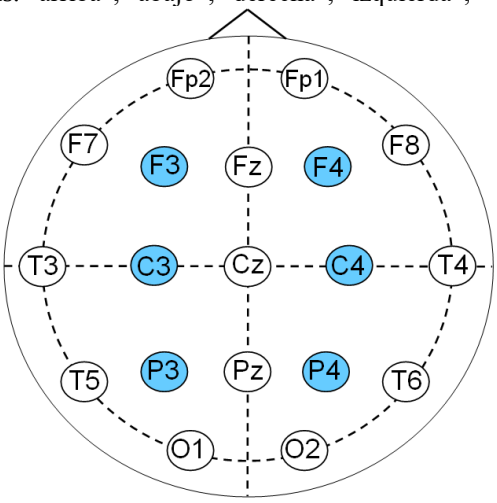


Figure 2 The EEG montage used to acquire data.

"adelante" and "atrás" (corresponding to the English words up, down, left, right, backward and forward). These words were selected as commands a user might make when interacting with a BCI. The experimental protocol for the imagined words tasks required participants to imagine speaking one of the prompted words at three audibly-cued time-points during the 4-second trial-period. Prior to the trial-period, stimuli were presented in both visual and auditory form, showing each subject the word for 2 seconds before being removed during the trial. EEG signals were recorded using a six-channel system, sampled at 1024 Hz. Electrodes were positioned according to the 10-20 international system over F3, F4, C3, C4, P3 and P4 (Figure 2).

Preprocessing: The original dataset was filtered between 2 Hz and 40 Hz using a finite impulse response bandpass filter [23], so no further filtering was applied for this study. Artefact detection and removal were implemented using Independent Component Analysis with Hessian approximation preconditioning [24].

Data splitting: In order to facilitate the analysis of multiple binary classifiers, all possible pairs of words were extracted from the dataset (Figure 1B). This resulted in 15 different pairs of imagined words for binary classification (e.g. *arriba* vs. *abajo*). The number of trials per class varied across subjects, with a maximum of 51. However, all pairs were balanced prior to training. Due to the high computational load associated with the nested cross-validation scheme employed for this study (see below), a 500 ms segment was extracted from each 4000 ms trial (Figure 1C) to act as the classification window. The selection of this window was based on the description of the experimental protocol described in

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-62

[23], in which three audible clicks directed participants when to imagine speaking. Therefore, a window was extracted to encompass the first of these periods of imagined speech production, about the 1-second mark of the overall trial. Concretely, this was the 500 ms segment between 750 ms and 1250 ms of the overall trial window.

Classification methods: Three distinct methods of classification were applied to the imagined words EEG data. The first of these methods, was the use of FBCSP features to train a rLDA (Figure 1D). The rLDA classifier is a regularized version of the LDA algorithm [25], which reduces the dispersion of the eigenvalues of the sample covariance matrix when a diverging dimension *p* is large. It has been employed elsewhere as a classification method for EEG signals [26], and is used here to provide a baseline reference for the performance of the CNNs. Unlike CNNs, the rLDA requires separate feature engineering and classification, and thus the type of features must be selected prior to training. FBCSP is a widely-used feature extraction algorithm across multiple BCI paradigms [27]. Linear combinations of the EEG channels are computed to enhance discrimination of band power features between classes. FBCSP has been proven successful in MI tasks, including as the winner in several EEG decoding competitions (e.g. [28]). Its proven results as a decoding algorithm in BCI, and the fact that there is no clear benchmark specifically for imagined speech, has led to selection of FBCSP as a reasonable baseline in this study.

The second classifier tested was a deep CNN designed by Schirrmeister et al. [29] specifically for EEG decoding applications (Figure 1E). The network architecture is based on similar CNNs used in computer vision [30] and is constructed to extract a wide range of features from the EEG signals. Figure 3 depicts the composition of the deep CNN. The input block of the CNN consists of two convolutional layers, one to perform convolution over time and one for spatial filtering. This first block also contains batch normalization, a non-linear activation function and a mean-pooling layer. Following this are three identical convolution blocks, each containing dropout, convolution, batch normalization, non-linear activation and mean-pooling. Finally, the output consists of a dense softmax layer for classification. The second CNN has been constructed with a shallow architecture and designed to decode band power features from EEG [29]. The shallow CNN is constructed of the same series of layers featured in the input block to the deep CNN (Figure 3), followed by dropout and a softmax classification layer. Here, we set dropout to 0.1 and selected the leaky rectified linear units activation function to add non-linearity into the two networks. Both CNNs used the ADAM optimizer [31] and the cross entropy loss function, and were allowed to train for 60 epochs with a *patience* of 30. A batch size of 64 was used. The CNNs were implemented in PyTorch [32], using the Braindecode repository [29].

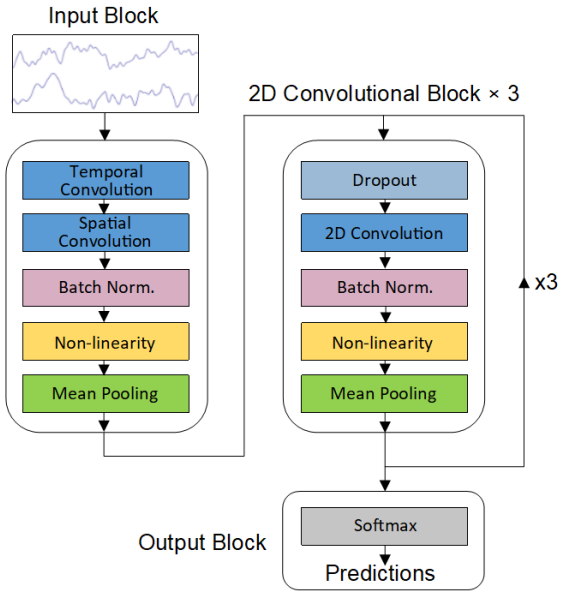


Figure 3 Deep CNN architecture designed by [22].

Nested cross-validation: A nested approach to cross-validation has been applied to training and testing of both the rLDA and the CNNs, with only small differences implemented when required by the respective classifiers (Figure 1D and E). Although not typically employed in deep learning contexts, nested-cross validation is utilized here to improve the robustness of results and to facilitate hyper-parameter optimization. Based on principles described in [33], the data are first split into 4-folds, one of which is retained in the outer-fold. An inner fold is then instantiated using the remaining 3 folds. The combined inner-fold is then re-split into 4 folds, with each fold iteratively acting as the test-set. The inner-fold facilitates training and testing of the two classifiers using each possible combination of hyper-parameters. The hyper-parameter combination with the best average accuracy across all inner-folds is then used to train the classifier on the entire inner-fold data. The outer-fold is then used as the test-set, or in the case of the CNNs, both validation- and test-sets. The classification accuracy is reported as the average accuracy across all 4 outer folds.

Hyper-parameters: Two hyper-parameters were selected for optimization with each classifier (Table 1). In the case of the rLDA, hyper-parameters required in the computation of FBCSP features were used. The first of these is the number of selected spatial filter pairs (1,3,4,5). The second hyper-parameter used here was the mutual information quantization level, with the values considered being 6, 8, 10 and 12. Hyper-parameters selected for the deep CNN were learning-rate and the number of filters implemented in the final convolutional layer (Table 1). The four learning-rates evaluated were

Table 1 Hyper-parameters optimized using nested cross-validation.

	Hyper-parameter 1	Hyper-parameter 2
FBCSP	# spatial filters: (1,3,4,5)	mutual information: (6,8,10,12)
Deep CNN	learning–rate: (1,0.1,0.01,0.001)	# final layer filters: (100,500,1000,1500)
Shallow CNN	learning-rate: (1,0.1,0.01,0.001)	# spatial filters: (20,40,60,80)

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-62

1.0, 0.1, 0.01 and 0.001 and the number of filters used in the final layer of convolution was 100, 500, 1000 and 1500. The same learning-rate range was used for the shallow CNN but the second parameter considered was the number of spatial filters (Table 1).

Statistics: Wilcoxon signed-rank tests were used to determine whether or not differences between the classifiers were statistically significant.

RESULTS

Here we report classification accuracies for each subject in the cohort and for each word-pair used to train the classifiers. Cross-subject classification accuracies are presented in Figure 4. Here, the highest classification accuracy obtained was 65.67%, achieved by subject 13 with the CNN. The shallow CNN showed similar peak performance with 65.28% average accuracy for subject 8. Results obtained by the baseline rLDA and the CNNs are significantly above chance accuracy (50%) for all word-pairs across all subjects. The average classification accuracies for the word-pairs when trained on the rLDA and the two CNNs were 57.80%, 62.37% and 60.88% respectively (Figure 5). The Wilcoxon signed rank tests determined that the greater performance of both the CNNs across word-pairs was significant in comparison with the FBCSP ($p<0.005$). The greater performance of the deep CNN was also significant in relation to the shallow network ($p<0.05$). Accuracies for the different word-pairs do not deviate substantially from the mean for any of the combinations (Figure 5). The highest average classification accuracy for a single word-pair was 64.55%, achieved by the *abajo vs derecha* pair, using the deep CNN (Figure 5). The highest single-subject accuracy obtained for a single word-pair was 78.33%, achieved by subject 11, also for the *abajo vs derecha* pair with the deep CNN.

The number of spatial filters used for FBCSP feature extraction was selected by the nested cross-validation as 5 (Table 2), although the difference between selecting 5, 4 or 3 was minimal. A mutual information coefficient of 8 was most often selected for optimization. In the case of

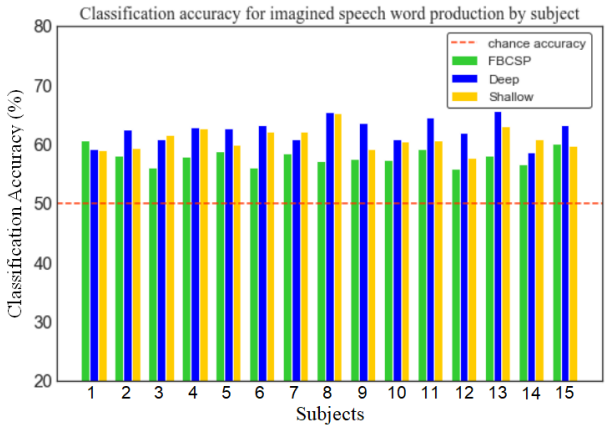


Figure 4 Subject classification accuracies for rLDA and CNNs.

Table 2 Hyper-parameters selected with nested cross-validation.

	Hyper-parameter 1	Hyper-parameter 2
FBCSP	# of filters = 5	mutual info. = 8
Deep	lr = 0.001	# of filters = 1000
Shallow	lr = 0.001	# of filters = 20

the CNNs, the hyper-parameter optimization selected a learning-rate of 0.001 more often than any of the other options. 1000 filters were selected for the final analysis of the convolution layer of the deep network and 20 were selected for spatial convolution in the shallow CNN.

DISCUSSION

The results presented here support the assertion that employing deep learning methodologies to the task of decoding imagined speech from EEG is a reasonable undertaking. For each subject, and for each word-pair, the CNNs outperformed the FBCSP-trained rLDA, achieving accuracies significantly above chance in each case. Despite indicating promise, the results also show that this level of performance is not yet close to what would be required of a functional DS-BCI. However, the greater overall performance of the deep architecture in relation to the shallow CNN does indicate the potential of deep learning for imagined speech decoding. Hyper-parameter optimization through nested cross-validation enabled selection of parameters most appropriate to the current task. Here, we determined that 5 spatial filters and a mutual information coefficient of 8, resulted in greater performance. Of the CNNs, 0.001 was the optimum learning-rate for use with the ADAM optimizer. The number of filters selected for the final convolutional layer was 1000. While this is greater than the number in the original paper [29], it provided the best accuracies here.

A weakness of the present study is the selection of a single 500 ms classification segment from the 4000 ms trial window. Although this approach was followed in the interests of computational efficiency, it is likely that a sliding-window would have improved overall classification performance. Furthermore, the number of trials per class was quite small, ranging from 39 to 51 for a single word. This relatively small volume of data constrains the classifiers' ability to infer classes by recognizing common patterns. Clearly, if deep learning is going to become a useful decoding approach for DS-BCI, larger datasets are required.

Interestingly, the results presented in Figure 5 do not suggest any significant differences in the effects of the linguistic content of the word-pairs. This may be a direct result of the choice of words used for this study. However, we agree with views expressed elsewhere [1], [2] that neurolinguistics research into imagined speech can aid the design of experiments in future research.

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-62

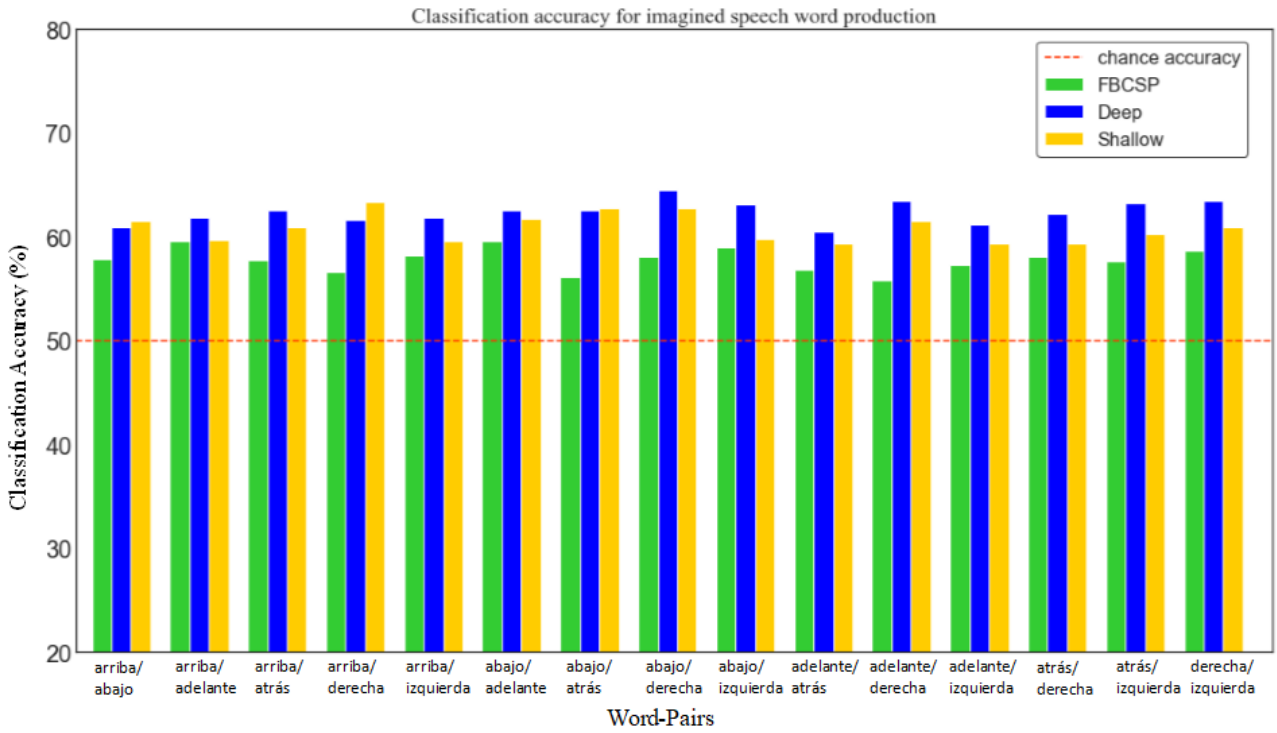


Figure 5 Classification accuracies for imagined speech production, by word-pair

CONCLUSION

In this study, we trained three different types of classifier with the purpose of decoding imagined speech from EEG. A rLDA using FBCSP features, and two CNNs, were trained on a 500 ms classification window extracted from trials where subjects imagined speaking Spanish words. 15 word-pairs were extracted from the dataset to enable multiple binary classifications. Nested cross-validation was employed to facilitate hyper-parameter optimization during training.
Results showed that the CNNs performed significantly better than the rLDA classifier with average accuracies of 62.37% and 60.88% vs. 57.80%. The differences observed between the two CNNs were significant, with the deeper network performing better. Results also indicated that differences in the accuracies obtained between the different word-pairs were not significant. The results suggest that, while further work is required in the field, deep learning is a realistic decoding methodology for imagined speech EEG.

ACKNOWLEDGEMENTS

The authors acknowledge the dataset's authors for the data used in this study. This work is supported by a Northern Ireland Department for the Economy studentship.

REFERENCES

[1] O. Iljina *et al.*, "Neurolinguistic and machine-learning perspectives on direct speech BCIs for restoration of naturalistic communication," *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 186–199, 2017.

[2] C. Cooney, R. Folli, and D. Coyle, "Neurolinguistics Research Advancing Development of a Direct-Speech Brain-Computer Interface," *IScience*, vol. 8, pp. 103–125, 2018.

[3] R. A. Ramadan and A. V. Vasilakos, "Brain computer interface: control signals review," *Neurocomputing*, vol. 223, no. October 2016, pp. 26–44, 2017.

[4] S. Martin *et al.*, "Word pair classification during imagined speech using direct brain recordings," *Sci. Rep.*, vol. 6, no. 1, 2016.

[5] C. H. Nguyen, G. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian Manifold features," *J. Neural Eng.*, 2017.

[6] Y. Song and F. Sepulveda, "Classifying speech related vs. idle state towards onset detection in brain-computer interfaces overt, inhibited overt, and covert speech sound production vs. idle state," *IEEE 2014 Biomed. Circuits Syst. Conf. BioCAS 2014 - Proc.*, pp. 568–571, 2014.

[7] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Networks*, vol. 22, no. 9, pp. 1334–1339, 2009.

[8] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2015–Augus, pp. 992–996, 2015.

[9] C. Cooney, R. Folli, and D. Coyle, "Mel Frequency Cepstral Coefficients Enhance

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-62

Imagined Speech Decoding Accuracy from EEG," *ISSC 2018 Irish Signals Syst. Confernece, 2018*, 2018.

[10] N. Hashim, A. Ali, and W.-N. Mohd-Isa, "Word-Based Classification of Imagined Speech Using EEG," in *Computational Science and Technology*, 2018, pp. 195–204.

[11] T. Kim, J. Lee, H. Choi, H. Lee, I. Y. Kim, and D. P. Jang, "Meaning based covert speech classification for brain-computer interface based on electroencephalography," *Int. IEEE/EMBS Conf. Neural Eng. NER*, pp. 53–56, 2013.

[12] X. Chi, J. B. Hagedorn, D. Schoonover, and M. D. Zmura, "EEG-Based Discrimination of Imagined Speech Phonemes," *Int. J. Bioelectromagn.*, vol. 13, no. 4, pp. 201–206, 2011.

[13] E. F. González-Castañeda, A. A. Torres-García, C. A. Reyes-García, and L. Villaseñor-Pineda, "Sonification and textification: Proposing methods for classifying unspoken words from EEG signals," *Biomed. Signal Process. Control*, vol. 37, pp. 82–91, 2017.

[14] A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Online EEG Classification of Covert Speech for Brain–Computer Interfacing," *Int. J. Neural Syst.*, vol. 27, no. 8, p. 1750033, 2017.

[15] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," pp. 1–13, 2016.

[16] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, no. 6, pp. 6645–6649, 2013.

[17] B. E. Olivas-padilla and M. I. Chacon-murguia, "Classification of multiple motor imagery using deep convolutional neural networks and spatial filters," *Appl. Soft Comput. J.*, vol. 75, pp. 461–472, 2019.

[18] N. Waytowich *et al.*, "Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials," *J. Neural Eng.*, vol. 15, no. 066031, 2018.

[19] A. Rezazadeh Sereshkeh, R. Trott, A. Bricout, and T. Chau, "EEG Classification of Covert Speech Using Regularized Neural Networks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 12, pp. 2292–2300, 2017.

[20] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated EEG-based screening of depression using deep convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 161, pp. 103–113, 2018.

[21] M. Hajinoroozi, Z. Mao, T. P. Jung, C. T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Process. Image Commun.*, vol. 47, pp. 549–555, 2016.

[22] Y. Roy and A. Gramfort, "Deep learning - based electroencephalography analysis : a systematic review," *arXiv Prepr. arXiv1901.05498v2*, 2019.

[23] G. A. Pressel Coretto, I. E. Gareis, and H. L. Rufiner, "Open access database of EEG signals recorded during imagined speech," p. 1016002, 2017.

[24] P. Ablin, J. Cardoso, and A. Gramfort, "Faster independent component analysis by preconditioning with Hessian approximations arXiv : 1706 . 08171v3 [stat . ML] 8 Sep 2017," pp. 1–23, 2017.

[25] J. H. Friedman, "Regularized Discriminant Analysis," *J. Am. Stat. Assoc.*, vol. 84, no. 405, pp. 165–175, 1988.

[26] E. Neto, F. Biessmann, H. Aurlien, H. Nordby, and T. Eichele, "Regularized Linear Discriminant Analysis of EEG Features in Dementia Patients," vol. 8, no. November, pp. 1–10, 2016.

[27] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface," *Proc. Int. Jt. Conf. Neural Networks*, pp. 2390–2397, 2008.

[28] M. Tangermann *et al.*, "Review of the BCI competition IV," vol. 6, no. July, pp. 1–31, 2012.

[29] 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.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[31] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *Comput. Res. Repos.*, 2017.

[32] A. Paszke, S. Gross, S. Chintala, and G. Chanan, "Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration." 2017.

[33] A. Korik, R. Sosnik, N. Siddique, and D. Coyle, "Decoding imagined 3D hand movement trajectories from EEG: Evidence to support the use of mu, beta, and low gamma oscillations," *Front. Neurosci.*, vol. 12, no. MAR, pp. 1–16, 2018.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-63

A SELF-PACED P300 SPELLER WITH IMPROVED TYPING SPEED USING CONTINUOUS STIMULUS PRESENTATIONS

W. Speier[1,2], C. Arnold[1,2], N. Chandravadia[3], D. Roberts[4], S. Pendekanti[3], A. Marttini[3], N. Pouratian[3,4,5,6]

[1] Department of Radiology, University of California, Los Angele, United States
[2] Medical Imaging and Informatics Group, University of California, Los Angeles, United States
[3] Neuroscience Interdepartmental Program, University of California, Los Angeles, United States
[4] Department of Neurosurgery, University of California, Los Angeles, United States
[5] Department of Bioengineering, University of California, Los Angeles, United States
[6] Brain Research Institute, University of California, Los Angeles, United States

E-mail: Speier@ucla.edu

ABSTRACT: Communication is a critical human function that can be severely compromised in patients with neurological diseases such as amyotrophic lateral sclerosis (ALS). The P300 speller is a brain-computer interface (BCI) device that restores communication in these patients by detecting evoked responses in subjects' electroencephalography signals. One of the bottlenecks of these systems is the pause after character selections. This pause has been necessary for the P300 speller because it signals users that a character selection has been made and gives them time to transition to the next character. If this pause is too long, the system is slowed down unnecessarily. If it is too short, stimuli for the next character begin before the user is ready. We propose a system that does away with the pause entirely and continually flashes stimuli. We employ a joint model that determines the target characters as well as the transition times so that users can change between characters at their own pace. A preliminary study on eight subjects showed a selection rate of 16.35 characters/minute and an average accuracy of 94.85%, both significant improvements over performance in an equivalent system with standard flashing. These results suggest that the P300 speller could be improved by implementing a continuous flashing paradigm.

INTRODUCTION

Communication is a critical human function that can be severely compromised in patients with neurological diseases such as amyotrophic lateral sclerosis (ALS). Brain-computer interfaces (BCI) provide a unique opportunity to restore communication in severe cases where traditional augmentative and alternative communication devices are unusable due to lack of motor control. The P300 speller is one BCI system that provides this communication ability by detecting evoked responses after flashing characters on a visual display [1]. Limits to the signal to noise ratio require multiple stimuli before making a selection, leading to slow typing rates. Several projects have improved performance by incorporating optimizations such as varying the dimensions of the character matrix [2], [3]; optimizing system parameters [4], [5]; and employing various signal-processing methods [6], [7].

Recent work has involved the incorporating of language models into the classifier [8]. This movement in BCI research integrates knowledge about the domain of natural language to improve classification, similar to methods used in other domains such as speech recognition [9]. Several BCI studies have shown incremental improvements in system speed and accuracy using n-gram language models, first using Naïve Bayes [10], [11] and later using a partially observable Markov decision process [12] and a hidden Markov model [13]. Recently, a particle filter (PF) algorithm made the use of more complicated language models possible, which was shown to have superior results [14].

One area that language models have had particular impact is the ability to provide prior probabilities for dynamic stopping [8]. These methods compute the probability distribution over all characters after each stimulus and select that character after the probability of a character reaches a threshold [11], [15]. This method has the potential to drastically improve typing speed as selections are made once the system is confident in a selection rather than continuing until a set number of stimuli is reached. It also allows the system to spend more time on selections where the confidence is lower so that it can collect more information rather than forcing a selection.

A P300 speller with continuous stimulus presentations is an extension of the dynamic stopping paradigm in that it allows subjects to move at their own pace rather than forcing them to wait a predefined period of time between character selections. It also allows them to take more time if they are unable to find a character or are unsure of what they want to say next. It accomplishes this task by

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-63

computing the joint probability distribution over possible target characters and transition times between characters after every stimulus presentation.

In this study, we propose a novel P300 speller system that incorporates continuous stimulus presentations. This system was incorporated into the BCI2000 [16] framework using a previously published particle filtering classifier [14]. A pilot study of eight healthy subjects was conducted to compare typing performance using this paradigm to the traditional P300 speller system.

MATERIALS AND METHODS

Data Collection All data collection was performed using g.tec amplifiers, active EEG electrodes, and electrode cap (Guger Technologies, Graz, Austria); sampled at 256 Hz; referenced to the left ear; grounded to AFz; and filtered using a passband of 0.1 – 60 Hz. Additional artifact detection (e.g., eye blinks) was not performed as it was left to the classifier to determine whether a signal contained a valid ERP. The electrode montage consisted of a previously reported set of 32 electrodes [5]. The subjects for the online study consisted of eight healthy volunteers with normal or corrected to normal vision between the ages of 20 and 35. The system used a 6 x 6 character grid, famous faces stimuli [17], row and column flashes, and a stimulus onset asynchrony of 125 ms. Using the standard interface, a 3.5-second gap was included between characters to allow subjects time to find the next character in the sequence.

To validate our methods, we implemented the continuous P300 speller in BCI2000 [16]. In a pilot study, eight subjects used the continuous speller to type out the sentence: "THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG." Eye tracking was not used, so training sessions consisted of two copy spelling sessions using the traditional P300 speller. When using the online speller, subjects were instructed to focus on each character for approximately three seconds (about two complete sets of flashes) before moving to the next character. Feedback was turned off to avoid distraction.

Continuous Speller The continuous speller formulation is similar to the traditional p300 speller with dynamic stopping [11]. After each flash, the probability distribution across the set of characters is estimated based on a set of observation probabilities and transition probabilities based on a language model. The main difference is the introduction of a variable d_t that represents the amount of time the subject will remain at the current state, x_t.

$$p(x_t, d_t|y_t, x_{0:t-1}, d_{t-1}) \propto p(x_t, d_t|x_{0:t-1}, d_{t-1})p(y_t|x_t)$$

In this model, d_t is reduced after each flash until it reaches zero. At that point, the transition probability is determined by the language model as in the traditional speller.

$$p(x_t, d_t|x_{0:t-1}, d_{t-1}) = \begin{cases} p(x_t|x_{0:t-1})p(d_t|x_t) & d_{t-1} = 0 \\ \delta_{x_{t-1}}^{x_t}\delta_{d_t}^{d_t+1} & d_{t-1} > 0 \end{cases}$$

A Gaussian distribution was used to estimate the time taken to transition between characters, $p(d_t|x_t)$. Initially, this distribution was set empirically at a mean of one second and a standard deviation of 0.5 seconds (with a minimum of zero seconds). To further tailor this distribution, expectation maximization was used to find the distribution for each subject in an unsupervised manner. This process was similar to the methods used in previous studies that trained the P300 speller with unlabeled data [18], [19]. In this version, the empirical distribution was used to find preliminary labels for and transition times for all characters. These labels were then used to find more accurate parameters for the transition distribution. Iteration between these two steps continued until the distribution stabilized.

Because it will always take time for the subject to find new characters in the grid, an additional state needs to be made for when the subject is transitioning between letters. During this time, the stimulus responses will look different from those when the subject's attention is on a character. The observation probability distribution needs to take this into account.

$$p(y_t|x_t) = \begin{cases} f(y_t|\mu_a, \sigma_a^2) & x_t \in A_t \\ f(y_t|\mu_{trans}, \sigma_{trans}^2) & x_t = x_{trans} \\ f(y_t|\mu_n, \sigma_n^2) & otherwise \end{cases}$$

Language Model: A language model is used to determine the transition probabilities, $p(x_t|x_{0:t-1})$. This probability can be simplified using the nth-order Markov assumption to create a n-gram model [11], [20]. While n-gram models are able to capture character patterns, they allow for strings that are not valid words on the language. A probabilistic automaton (PA) creates a stronger prior by creating states for every substring that starts a word in the corpus. Thus, the word "the" would result in four states: "t", "th", "the", and the start state which corresponds to a blank string. Each state then links to every state the represents a superstring that is one character longer. Thus, the state "th" will link to the states "the" and "tha" (Fig. 1).

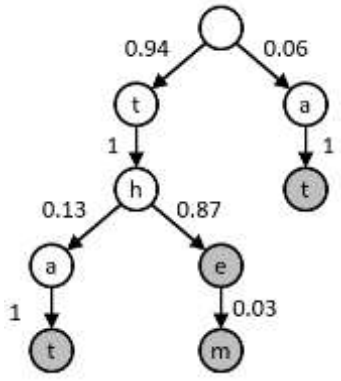


Figure 1: Example language model containing only the words at, that, the, and them.

Transition probabilities are determined by the relative frequencies of words starting with the states' substrings in the Brown English language corpus [21].

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-63

Particle Filter: As more sophisticated language models are used, the ability to fully represent the probability distribution over possible output sequences becomes impractical. Particle filtering (PF) is a method for estimating this distribution by creating a set of realities (called particles) and projecting them through the model based on the observed data [22], [23]. Each of these particles contains a reference to a state in the model, a history of previous states, and an amount of time that the particle is going to remain in the current state. The distribution of states occupied by these particles represents an estimation of the probability distribution.

When the system begins, a set of P particles is generated and each is associated with the root node of the language model. At the start of a new character, samples are drawn from the proposal distribution defined by the transition probabilities from the previous state.

$$x_t^{(L)} \sim p\left(x_t \middle| x_{t-1}^{(L)}\right)$$

The time that the particle will stay in that state is drawn from a distribution representing how long the subject is expected to spend looking at that character. After each stimulus response, the probability weight is computed for each of the particles.

$$\widehat{w}_t^{(L)} \propto \widehat{w}_{t-1}^{(L)} p\left(y_t \middle| x_t^{(L)}\right)$$

The weights are then normalized and the probability of possible output strings is found by summing the weights of all particles that correspond to that string. When feedback is enabled, the string with the highest probability is then displayed to the user. The effective number of particles is then computed.

$$P_{eff} = \frac{1}{\sum_i \left(\widehat{w}_t^{(L)}\right)^2}$$

If the effective number falls below a threshold, P_{thresh}, a new set of particles are drawn from the particle distribution.

After each stimulus, the amount of time for a given particle to remain in a state is decremented. Once that counter reaches zero, the particle transitions to a new state in the language model based on the model transition probabilities $p(x_t|x_{0:t-1})$.

Evaluation: Performance of the system is evaluated in terms of the speed and accuracy of typing characters. Selection rate (SR) is the average number of characters selected per minute and accuracy (ACC) is the percentage of those characters that match the target sentence. Because of the tradeoff between speed and accuracy, performance is also evaluated in terms of information transfer rate (ITR), which measures the bits of information conveyed through the output message divided by time. ITR has been shown to be an imperfect measure of BCI communication due to assumptions about the uniform probability across characters and the independence of selections [24], [25]. However, it remains a standard metric used in the BCI field and can compare relative performance on identical sequences. Calculation of ITR starts by computing the number of bits of information contained per symbol in the output sequence.

$$B = \log(N) + ACC\log(ACC) + (1 - ACC)\log\left(\frac{1 - ACC}{N - 1}\right)$$

Where N is the number of characters in the grid (36). ITR can then be found by multiplying the by SR. Significance was tested using Wilcoxon rank-sum tests.

RESULTS

Even though subjects were all instructed to type at the same speed (three seconds per letter), each subject typed at a slightly different pace (Fig. 2).

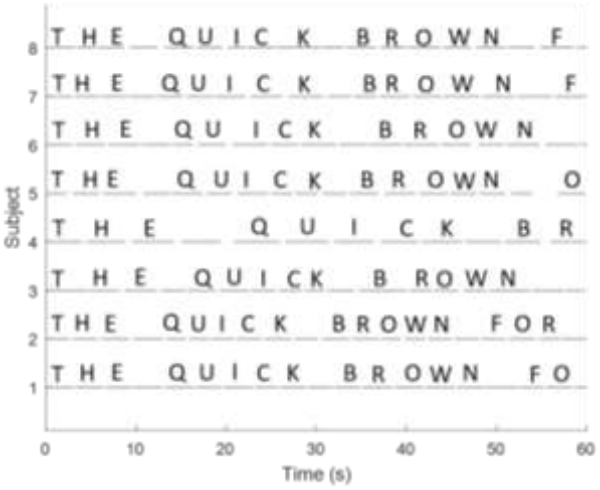


Figure 2: Output for each subject during the first 60 seconds of use of the continuous speller. Each subject was attempting to spell "THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG."

Table 1. Selection rate (SR), typing accuracy (ACC) and information transfer rate (ITR) for the traditional P300 speller and the continuous speller (CS).

	SR (characters/minute)		ACC (%)		ITR (bits/minute)	
Subject	P300	CS	P300	CS	P300	CS
1	14.06	17.79	93.33	100.00	62.92	89.89
2	13.81	18.18	95.00	90.91	63.91	77.50
3	13.77	16.51	100.00	91.11	71.21	75.74
4	7.58	11.94	75.00	100.00	23.33	61.72
5	13.43	16.28	95.00	90.70	62.13	69.12
6	12.63	16.31	100.00	95.35	65.30	72.64
7	13.69	17.45	65.00	97.73	33.42	85.47
8	13.04	17.09	75.00	93.02	40.13	75.98
Mean	12.71	16.35	89.05	94.85	54.60	76.01

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-63

Overall, subjects types an average of 16.35 characters/minute using the continuous speller, which was a significant improvement over the 12.71 characters/minute achieved using the traditional P300 speller (p=0.004, Tab. 1). When using a generic model for dwell and transition times, the average selection accuracy for the continuous speller (85.4%) was lower than the P300 speller (89.05%), although the difference was not statistically significant (p=0.42). The resulting ITR for the continuous speller (64.5 bits/minute) was higher than that of the P300 speller (54.6 bits/minute), although the difference was not statistically significant (p=0.13). Once subject-specific transition distributions were learned, the accuracy rose to 94.9%, resulting in an average ITR (76.0 bits/minute) that was significantly higher than that of the P300 speller (p=0.004).

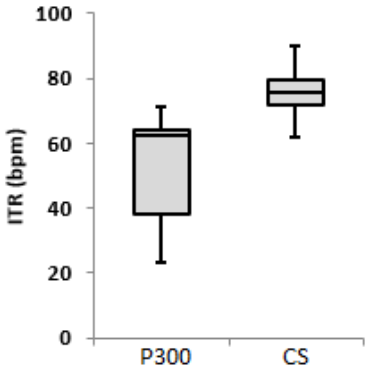


Figure 3: Boxplots of ITR values using the standard P300 and the continuous speller (CS).

In general, subjects' gaze times were close to the suggested value of three seconds (Tab. 2). One subject (subject 4), however, spent over one second longer per letter on average. The other subjects' averages were between 2.93 and 3.32 seconds. The average time spent transitioning between characters was also fairly consistent across the subjects with an average of 0.36 seconds.

Table 2: Average gaze and transition times

Subject	Gaze Time (s)	Transition Time (s)
1	3.28 (0.58)	0.14 (0.06)
2	2.93 (0.65)	0.31 (0.23)
3	2.98 (0.75)	0.35 (0.26)
4	4.39 (0.73)	0.59 (0.28)
5	3.13 (0.70)	0.41 (0.31)
6	3.32 (0.61)	0.32 (0.21)
7	3.06 (0.61)	0.32 (0.27)
8	3.07 (0.63)	0.43 (0.31)
Mean	3.27 (0.47)	0.36 (0.13)

DISCUSSION

The average typing speed achieved using the continuous speller was 16.35 characters/minute, which was significantly faster than the 12.71 characters/minute achieved by the same subjects using the traditional P300 speller. The bit rate for the P300 speller was similar to values achieve in previous studies using similar systems [14].

Only one subject failed to achieve a typing speed greater than 16 characters/minute. While this subject was still able to type at a speed faster than she did using the standard P300 speller, it was substantially slower than the other subjects in this study, taking over a minute longer to type the sentence than any other subject. This difference is likely because she dwelled longer on each character than the suggestion of three seconds.

The results achieved using continuous spelling demonstrate that this system has the potential to outperform existing ERP-based BCI systems. However, the results from this study are likely far from optimal, as most of the parameters were not optimized for this system. For instance, the flashing rate and ISI values used were optimized for the traditional p300 speller, and do not necessarily reflect the best configuration for this system. We also used the row/column flashing paradigm as it is the standard stimulus paradigm for the P300 speller. There are various other paradigms that have been introduced, including the checkerboard [2], combinatorial [3], and asynchronous [26] paradigms, which could improve typing performance in this system as well. Finally, the strategy of looking at each character for approximately three seconds was chosen empirically as it gave approximately four positive stimuli for each character, which we felt would be sufficient to make an accurate classification. It is possible that less information is needed, which could allow the system to achieve even greater speeds.

This study was conducted using healthy volunteers who did not have the same constraints as "locked-in" patients, such as restrictions on eye gaze. While a similar P300 speller system was previously tested in the ALS population [27], it is unclear whether the continuous flashing will be more difficult and therefore offset the gains seen by applying continuous flashing. The healthy subjects in this study generally had no problems with the additional cognitive task, and therefore appreciated the added speed that continuous flashing afforded. However, it is possible that his additional task will make the system more taxing for ALS patients, which could make it less practical despite the performance increase. Future studies in the ALS population should be conducted to determine how these results in healthy subjects translate to affected population. If continuous flashing is a hindrance to some subjects, those subjects could continue to use the traditional P300 speller with the continuous flashing paradigm included as an option for those subjects who would benefit.

The results presented in this study are promising, but they represent offline performance which does not include several factors that occur in an online implementation. For instance, offline systems do not include feedback to

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-63

the user, which can provide additional motivation or allow the user to adjust their strategy. Particularly in this system, feedback can be difficult to implement because the user does not have specific built in to allow for checking past input. Also, the fact that the system optimizes over multiple characters at once means that the system will likely make changes to the displayed text, which can distract the user's focus from the current character. The lack of feedback can make the task more difficult, however, as subjects can lose track of where they are in the target word or phrase. They are also unable to adjust their strategy such as slowing down if the system is unable to make accurate selections.

CONCLUSION

Overall, continuous stimulus presentation allowed subjects to type an average of 16.35 characters/minute with an accuracy of 94.85%, resulting in an average ITR of 76.01 bits/minute, all significantly higher than the values achieved using the standard P300 speller. Future work involves optimizing system parameters for the continuous flashing paradigm, and implementing feedback in an online system.

REFERENCES

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

REFERENCES

[1] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials.," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–23, Dec. 1988.

[2] G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. E. Frye, C. K. Hauser, N. E. Schwartz, T. M. Vaughan, J. R. Wolpaw, and E. W. Sellers, "A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns," *Clin. Neurophysiol.*, vol. 121, no. 7, pp. 1109–1120, Jul. 2010.

[3] J. Jin, P. Horki, C. Brunner, X. Wang, C. Neuper, and G. Pfurtscheller, "A new P300 stimulus presentation pattern for EEG-based spelling systems.," *Biomed. Tech. (Berl).*, vol. 55, no. 4, pp. 203–10, Aug. 2010.

[4] D. J. McFarland, W. A. Sarnacki, G. Townsend, T. Vaughan, and J. R. Wolpaw, "The P300-based brain-computer interface (BCI): Effects of stimulus rate," *Clin. Neurophysiol.*, vol. 122, no. 4, pp. 731–737, Apr. 2011.

[5] J. Lu, W. Speier, X. Hu, and N. Pouratian, "The effects of stimulus timing features on P300 speller performance," *Clin. Neurophysiol.*, vol. 124, no. 2, pp. 306–314, Feb. 2013.

[6] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI Competition 2003--Data set IIb: support vector machines for the P300 speller paradigm.," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1073–6, Jun. 2004.

[7] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, "BCI Competition 2003 - Data Set IIb: Enhancing P300 Wave Detection Using ICA-Based Subspace Projections for BCI Applications," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1067–1072, 2004.

[8] W. Speier, C. Arnold, and N. Pouratian, "Integrating language models into classifiers for BCI communication: a review," *J. Neural Eng.*, vol. 13, no. 3, p. 31002, 2016.

[9] F. Jelinek, *Statistical methods for speech recognition*. MIT press, 1998.

[10] P.-J. Kindermans, H. Verschore, and B. Schrauwen, "A Unified Probabilistic Approach to Improve Spelling in an Event-Related Potential-Based Brain–Computer Interface," *Biomed. Eng. IEEE Trans.*, vol. 60, no. 10, pp. 2696–2705, 2013.

[11] W. Speier, C. Arnold, J. Lu, R. K. Taira, and N. Pouratian, "Natural language processing with dynamic classification improves P300 speller accuracy and bit rate," *J. Neural Eng.*, vol. 9, no. 1, p. 16004, Feb. 2011.

[12] J. Park and K.-E. Kim, "A POMDP approach to optimizing P300 speller BCI paradigm," *Neural Syst. Rehabil. Eng. IEEE Trans.*, vol. 20, no. 4, pp. 584–594, 2012.

[13] W. Speier, C. Arnold, J. Lu, A. Deshpande, and N. Pouratian, "Integrating language information with a hidden markov model to improve communication rate in the P300 speller," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 678–684, 2014.

[14] W. Speier, C. W. Arnold, A. Deshpande, J. Knall, and N. Pouratian, "Incorporating advanced language models into the P300 speller using particle filtering.," *J. Neural Eng.*, vol. 12, no. 4, p. 46018, Aug. 2015.

[15] H. Serby, E. Yom-Tov, and G. F. Inbar, "An improved P300-based brain-computer interface.," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 1, pp. 89–98, Mar. 2005.

[16] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system.," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–43, Jun. 2004.

[17] T. Kaufmann, S. M. Schulz, A. Köblitz, G. Renner, C. Wessig, and A. Kübler, "Face stimuli effectively prevent brain-computer interface inefficiency in patients with neurodegenerative disease.," *Clin. Neurophysiol.*, vol. 124, no. 5, pp. 893–900, May 2013.

[18] W. Speier, J. Knall, and N. Pouratian,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-63

"Unsupervised training of brain-computer interface systems using expectation maximization," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, 2013, pp. 707–710.

[19] P.-J. Kindermans, M. Schreuder, B. Schrauwen, K.-R. Müller, and M. Tangermann, "True Zero-Training Brain-Computer Interfacing–An Online Study," *PLoS One*, vol. 9, no. 7, p. e102504, 2014.

[20] C. D. Manning and H. Schütze, *Foundations of statistical natural language processing*. MIT Press, 1999.

[21] W. N. Francis and H. Kucera, *Brown Corpus Manual*. Providence, RI: Dept of Linguistics, Brown University, 1979.

[22] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *Signal Process. IEEE Trans.*, vol. 50, no. 2, pp. 174–188, 2002.

[23] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, 1993, vol. 140, no. 2, pp. 107–113.

[24] P. Yuan, X. Gao, B. Allison, Y. Wang, G. Bin, and S. Gao, "A study of the existing problems of estimating the information transfer rate in online brain–computer interfaces," *J. Neural Eng.*, vol. 10, no. 2, p. 26014, 2013.

[25] W. Speier, C. Arnold, and N. Pouratian, "Evaluating True BCI Communication Rate through Mutual Information and Language Models.," *PLoS One*, vol. 8, no. 10, p. e78432, Jan. 2013.

[26] G. Townsend and V. Platsko, "Pushing the P300-based brain–computer interface beyond 100 bpm: extending performance guided constraints into the temporal domain," *J. Neural Eng.*, vol. 13, no. 2, p. 26024, 2016.

[27] W. Speier, N. Chandravadia, D. Roberts, S. Pendekanti, and N. Pouratian, "Online BCI Typing using Language Model Classifiers by ALS Patients in their Homes," *Brain-Computer Interfaces*, vol. 4, no. 1–2, pp. 1–8, Nov. 2016.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-64

MOVEMENT RELATED CORTICAL POTENTIALS DURING HAND OPENING AND CLOSING IN ALS PATIENTS

Strahinja Dosen[1], Natalie Mrachacz-Kersting[1]

[1] Center for Sensory-Motor Interaction, Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Denmark

E-mail: sdosen@hst.aau.dk and nm@hst.aau.dk

ABSTRACT: Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease, which leads to the progressive loss of muscle control. In Denmark, approximately 150 people are diagnosed with ALS each year. The ensuing burden on related health care costs increases substantially as patients lose the ability to perform tasks of everyday living. Brain-computer interface (BCI) systems provide an approach to allow these patients some interaction with their environment. In this ongoing study we investigate whether the movement-related cortical potential (MRCP) extracted from electroencephalography (EEG) signals can be used for BCI control of an assistive glove in ALS patients. To this aim, the BCI needs to detect the intention of the subject to open and close the hand. The MRCP is a slow negative drift that commences 2 seconds prior to movement onset and contains features that differ between different types of movement, making it an ideal signal modality for multi-dimensional BCI control. Preliminary analysis from three ALS patients reveal a classification accuracy above 85% for the best channel between movement and rest conditions. However, classification accuracy between movement types (hand opening/closing), was lower (>65% for the best channel). Combining MRCP detection, with methods that allow using a single input (brain switch) to select multiple commands, may thus be a viable solution for this patient group.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a debilitating progressive neurological disease. In Denmark, more than 150 new ALS patients are diagnosed per year. Symptoms vary from patient to patient, but may be manifested by reduced muscle strength, muscle wasting, fatigue, increased muscle tension, convulsions, muscle soreness, dysfunction, speech impairment, and difficulty breathing; several patients also present with behavioral changes and Frontotemporal Dementia. In Denmark, approximately 20% of ALS patients end up with respiratory therapy due to respiratory problems [1]. ALS has both personal as well as societal consequences. The economic impact of ALS at the community level is high [2], a large proportion being due to the care costs when patients can no longer perform everyday activities like personal care due to muscle weakness [3]. Therefore, it is of great interest to provide these patients with assistive systems that would facilitate and/or prolong functional independence, and thereby decrease the need for constant care.

The focus of the present study is on the assistance of hand grasping function. Several methods can be used to assist grasping function in paralyzed patients (e.g., functional electrical stimulation – FES and exoskeletons) [4]. For example, rigid exoskeletons are efficient in restoring movements but they are also bulky and esthetically unappealing. FES can be delivered using a compact hardware but the movements are difficult to control precisely and selectively. Recently, a number of soft exoskeletons have been proposed [5], [6]. A typical solution comprises a textile glove that is actuated using a network of tendons pulled by a motor placed somewhere on the body (e.g., around the waist). These systems combine good controllability and compact design, and therefore, represent an attractive solution to assist highly disabled patients.

The overall aim of the present project is to develop a brain-computer interface (BCI) [7], [8] that can be used to control a soft hand exoskeleton in ALS patients. The envisioned BCI will be based on detection and classification of movement related cortical potentials (MRCPs) [9]. The MRCP is a characteristic modulation of brain potentials comprising a negative deflection that anticipates the movement, followed by a positive rebound. The MRCP has been selected because it does not require training, it is present during motor execution and imagination, and it has been successfully detected in different patient populations [10], [11], including ALS [12]. In order to control a soft glove, the user should be able to generate opening and closing commands via BCI (Fig. 1). Previously, the MRCPs related to foot dorsiflexion were successfully detected and used for online control of a foot orthosis [13]. In addition, several studies have used MRCPs to detect and classify hand motions, including different grasp types [14], and movement speed and force [15]. In a recent study [16], low-frequency time domain features within the bandwidth of the MRCP were used to detect and classify natural reach-to-grasp movements. However, these studies were performed in healthy subjects, except [15] which was conducted in stroke patients.

In the present study, electroencephalography (EEG) signals during hand opening and closing were recorded

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-64

in three patients suffering from ALS. The patients were recruited at different stages of ALS, as assessed by the ALS functional rating scale [17]. The aim was to preliminarily assess the feasibility of detecting the patient intention to perform the two hand movements (closing vs. rest, opening vs. rest) as well as to discriminate between them (closing vs. opening).

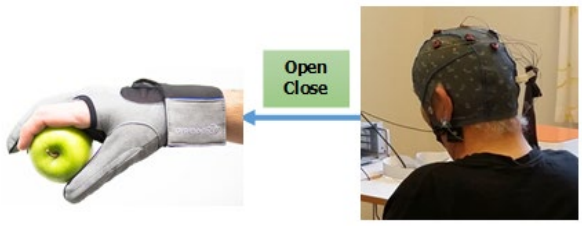


Figure 1: The envisioned BCI system for restoring grasping using a soft exoskeleton glove

MATERIALS AND METHODS

Patients

Three patients with ALS participated to the recording. All signed an informed consent form approved by the local ethical committee. The patients were recruited by a neurologist, who also determined the functional score using the ALS functional rating scale [18]. The scores for the patients numbered 13, 14 and 23 were 0, 0, and 3.6, respectively. All recordings were performed at the patient's home.

Signal acquisition

Non-invasive electroencephalographic (EEG) recordings were obtained using the g.USBamp (g.tec, AU) and g.GAMMAcap equipped with nine active electrodes (g.LADYbird). The electrodes were arranged according to the standard 10-20 system over the motor area of the arm/hand (channels: F3, FC1, FC5, Cz, C3, C7, CP1, CP5, P3) contralaterally to the dominant hand of the patient. Electromyography (EMG) was recorded from the hand flexor and extensor muscles using a bipolar configuration (Ag/AgCl electrodes (AMBU Neuroline, US)). The ground for the EMG recording was separate from the EEG ground. The sampling frequency was set to 1200 Hz for all signals and no filtering was activated in the amplifier.

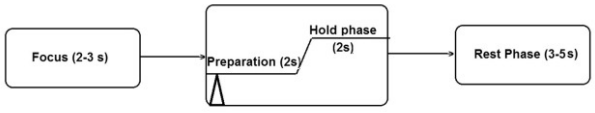


Figure 2: The cue-based data collection paradigm comprising focus, preparation, hold and rest phase. The protocol is explained in the text.

The patients were seated in a chair in front of a table. Since two of the patients had no residual hand function, EMG could not be implemented to indicate movement

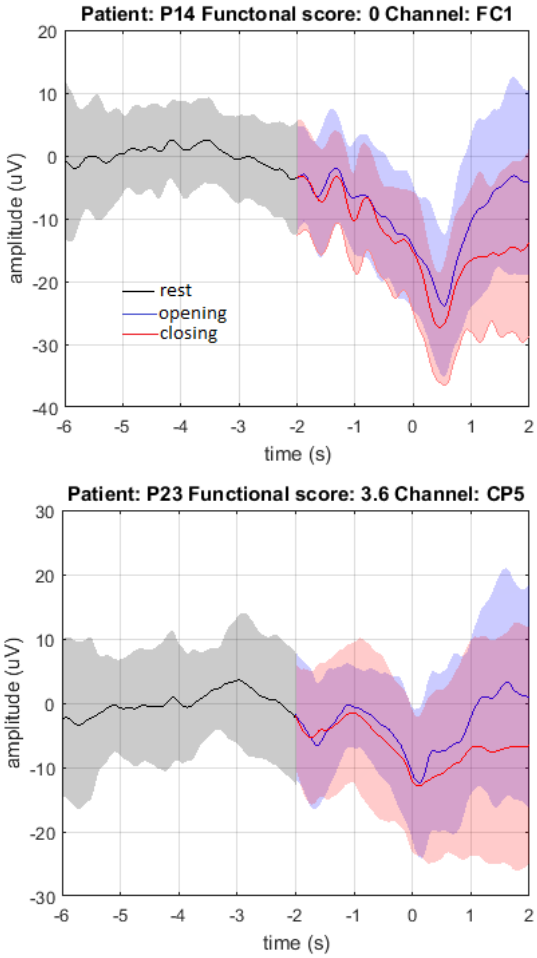


Figure 3: The traces (mean ± standard deviation) for rest (black line), MRCP during closing (red line) and opening (blue line) for patient P14 (top) and P23 (bottom). Movement onset is at 0 s. Two representative channels are depicted (FC1 and CP5).

onset and thus to extract the MRCP of each trial. Patients were thus presented with a visual cue that was displayed on a computer screen following a predefined experimental paradigm (Fig. 2). In the focus phase, the patient was asked to focus on the middle of the screen. In the preparation phase, the movement to be executed was indicated by a text message (open/close) and a triangular cursor started moving across the screen. When the cursor arrived at the middle of the screen, it instantly jumped, indicating the moment when the patient should perform the movement. The patient was instructed to hold the movement, until the cursor disappeared from the screen (hold phase). This was followed by a resting phase. In each recording block, 15 hand opening and 15 hand closing movements were collected in randomized order, and two blocks were recorded in the experimental session, hence 30 movements in total in each class.

Signal processing

The EMG signals were bandpass filtered using a second order Butterworth filter with a cut off frequency at 10 Hz (movement artifacts) and 3 Hz (linear envelope). The

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-64

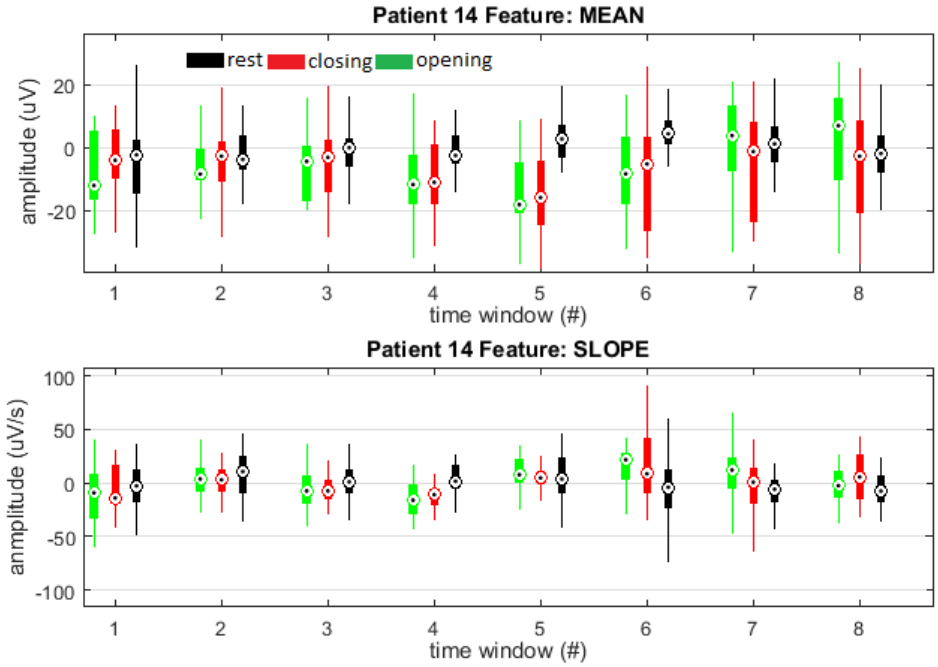


Figure 4: Distribution of time domain features extracted from the epochs shown in Fig. 3 (Patient 14). The boxplots indicate median, interquartile range and min/max values for mean (top) and slope (bottom) in each time window for rest (black), closing (red) and opening (green).

EEG signals were bandpass filtered between 0.01 and 3 Hz (zero-phase Butterworth 2nd order) which is the bandwidth of the MRCP. Segmenting the epochs was challenging due to noise from the electrical/medical equipment (e.g., patient respirator) and low signal levels (e.g., weak or no EMG). Therefore, a custom-made Matlab application was developed and used to manually inspect each trial and to discard those that were corrupted. In all patients, there was at least 20 good trials in each movement class. The onset of the trial was identified either manually based on the recorded EMG or based on the cue to move.
The selected "good" trials were then segmented to extract the MRCP and rest. The interval from -2 s to 2 s, with respect to the movement onset (0 s), was considered as an MRCP and the interval between -6 s and -2 s was assumed to be the resting state. The MRCPs were baseline corrected by subtracting the mean of the resting state.

Classification

For the purpose of classification, the time domain features were computed from the epochs (MRCP and rest). Most studies that rely on MRCP consider only the phase preceding the movement (negative deflection). This is done to minimize the detection delay and provide timely feedback using electrical stimulation following the paradigm of Hebbian learning [19]. In the context of the present project, however (Fig. 1), the overall reliability of detection and classification is more important than the delay since the focus is on robust control. Therefore, the entire epoch was considered, including both pre and post movement phases (i.e., deflection and rebound of the MRCP).
Each epoch was divided in windows (500 ms) and the mean and the slope of the signal were computed within each window, resulting in 16 features per epoch. Linear discriminant analysis was used for classification between the pairs of classes (open vs. rest, close vs. rest, and open vs. close). The classification was tested for each individual channel. Due to the low number of trials, the classifier was validated using the leave-one out cross validation scheme. The classification accuracy was adopted as the outcome measure.

RESULTS

The average traces of the rest and MRCP epochs recorded in two patients with substantially different functional scores are shown in Fig. 3. In both cases, the MRCP is prominent for both movements, and there is a difference between the two MRCPs, particularly during the rebound phase. However, there is also a significant variability and overlap between the MRCPs corresponding to opening and closing.
The distribution of the time domain features for the same data as in Fig. 3 top (patient 14, channel FC1) is depicted in Fig. 4. Both mean and slope features exhibit clear differences between movement and rest, and this reflects the morphology of the MRCP. The mean in the windows 4, 5 and 6 indicate the negative deflection of the MRCP. The slope in the windows 4 and 6 exhibit the change in the sign, which corresponds to the downward (negative slope) and then upward (positive slope) trend of the MRCP. The difference in the features between the two

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-64

Table 1. Classification results

Subject	Classes	Classification accuracy (%)								
		F3	FC1	FC5	Cz	C3	C7	CP1	CP5	P3
Patient 14	O/C	67	76	43	74	44	48	53	49	59
	O/R	90	91	73	94	90	83	86	80	79
	C/R	93	99	75	97	91	82	88	79	84
Patient 13	O/C	65	51	65	47	51	49	47	49	51
	O/R	88	88	77	78	85	78	88	83	77
	C/R	78	77	75	86	89	69	80	75	80
Patient 23	O/C	59	53	46	50	49	54	60	72	51
	O/R	78	77	78	84	86	83	81	78	86
	C/R	78	81	73	89	86	84	95	80	78

C, O and R stand for closing, opening and rest

movements is however less visible, i.e., nevertheless, there seems to be a difference in the mean in time windows 7 and 8 and in the slope in time windows 6 and 7, reflecting the faster and stronger rebound of the MRCP associated to hand opening (see also Fig. 3).

The channel-wise classification success rates obtained using the leave one out cross validation scheme are shown in Table 1. The grayed cells indicate the channels with the highest classification success rate.

DISCUSSION

In the present study, EEG data were collected during hand opening and closing movements from three ALS patients, and the MRCP profiles were extracted. An important conclusion from the collected profiles is that a pronounced MRCP can be observed even in highly disabled ALS patients (patient 13 and 14). The profiles resemble, in shape as well as in depth, those that were collected in a more functioning patient (patient 23). This is in line with the results reported in [12] where they compared the MRCPs of ALS patients to those of healthy subjects and found no significant difference in the peak negativity.

A preliminary classification of the collected patterns has demonstrated that they have sufficient discriminative information to be correctly classified. Both opening and closing could be differentiated from rest with high accuracy. Classifying between opening and closing was however a substantially more challenging task, as can be seen from the low classification accuracies for many channels (O/C in Table 1). An important insight is that the classification should likely consider the post movement period, since this is the phase where the two profiles somewhat diverge (Fig 3).

The present test is a first and simple evaluation of the feasibility of discriminating between the patterns. The next step is to implement classification using a sliding window during pseudo online (offline data) and online applications. This represents a more difficult task and it is to be expected that the accuracies will be lower than those reported here (Table 1). In addition, instead of classifying on each channel individually, they could be combined using a spatial filter [20] and classification could be performed on the surrogate channel. Finally, for the offline analysis in the present study, the corrupted trials were eliminated manually. During an online application, an automatic artifact rejection scheme needs to be implemented [21], [22].

Nevertheless, the obtained results are encouraging for detection (movement versus rest), which is the most important command for the online system. Even if direct classification (closing versus opening) is revealed insufficiently reliable, the lack of this input can be circumvented using state machine and/or recently presented electrotactile menus [15]. Both of these methods allow using a single input (brain switch) to select multiple commands.

REFERENCES

[1] P. Dreyer, C. K. Lorenzen, L. Schou, and M. Felding, “Survival in ALS with home mechanical ventilation non-invasively and invasively: A 15-year cohort study in west Denmark,” *Amyotroph. Lateral Scler. Front. Degener.*, vol. 15, no. 1–2, pp. 62–67, Mar. 2014.

[2] M. Gladman and L. Zinman, “The economic impact of amyotrophic lateral sclerosis: a systematic review,” *Expert Rev. Pharmacoecon. Outcomes Res.*, vol. 15, no. 3, pp. 439–450, May 2015.

[3] M. Obermann and M. Lyon, “Financial cost of amyotrophic lateral sclerosis: A case study,” *Amyotroph. Lateral Scler. Front. Degener.*, vol. 16, no. 1–2, pp. 54–57, Mar. 2015.

[4] P. Maciejasz, J. Eschweiler, K. Gerlach-hahn, A. Jansen-troy, and S. Leonhardt, “A survey on robotic devices for upper limb rehabilitation.,” *J. Neuroeng. Rehabil.*, vol. 11, no. 1, p. 3, Jan. 2014.

[5] S. Biggar and W. Yao, “Design and Evaluation of a Soft and Wearable Robotic Glove for Hand Rehabilitation,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 10, pp. 1071–1080, Oct. 2016.

[6] L. Cappello *et al.*, “Assisting hand function after spinal cord injury with a fabric-based soft robotic glove,” *J. Neuroeng. Rehabil.*, vol. 15, no. 1, p. 59, Dec. 2018.

[7] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-64

Computer Interfaces, a Review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, Jan. 2012.

[8] R. Rupp, S. C. Kleih, R. Leeb, J. del R. Millan, A. Kübler, and G. R. Müller-Putz, "Brain–Computer Interfaces and Assistive Technology," 2014, pp. 7–38.

[9] 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 Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 288–96, Feb. 2014.

[10] S. Aliakbaryhosseinabadi *et al.*, "Influence of attention alternation on movement-related cortical potentials in healthy individuals and stroke patients," *Clin. Neurophysiol.*, vol. 128, no. 1, pp. 165–175, Jan. 2017.

[11] R. Xu *et al.*, "Movement-related cortical potentials in paraplegic patients: abnormal patterns and considerations for BCI-rehabilitation," *Front. Neuroeng.*, vol. 7, Aug. 2014.

[12] Y. Gu, D. Farina, A. R. Murguialday, K. Dremstrup, and N. Birbaumer, "Comparison of movement related cortical potential in healthy people and amyotrophic lateral sclerosis patients," *Front. Neurosci.*, vol. 7, 2013.

[13] Ren Xu *et al.*, "A Closed-Loop Brain–Computer Interface Triggering an Active Ankle–Foot Orthosis for Inducing Cortical Neural Plasticity," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2092–2101, Jul. 2014.

[14] M. Jochumsen, I. K. Niazi, K. Dremstrup, and E. N. Kamavuako, "Detecting and classifying three different hand movement types through electroencephalography recordings for neurorehabilitation," *Med. Biol. Eng. Comput.*, vol. 54, no. 10, pp. 1491–1501, Oct. 2016.

[15] M. Jochumsen, I. Khan Niazi, D. Taylor, D. Farina, and K. Dremstrup, "Detecting and classifying movement-related cortical potentials associated with hand movements in healthy subjects and stroke patients from single-electrode, single-trial EEG," *J. Neural Eng.*, vol. 12, no. 5, p. 056013, Oct. 2015.

[16] 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, Feb. 2018.

[17] "The Amyotrophic Lateral Sclerosis Functional Rating Scale. Assessment of activities of daily living in patients with amyotrophic lateral sclerosis. The ALS CNTF treatment study (ACTS) phase I-II Study Group.," *Arch. Neurol.*, vol. 53, no. 2, pp. 141–7, Feb. 1996.

[18] J. M. Cedarbaum *et al.*, "The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III).," *J. Neurol. Sci.*, vol. 169, no. 1–2, pp. 13–21, Oct. 1999.

[19] N. Mrachacz-Kersting *et al.*, "The effect of type of afferent feedback timed with motor imagery on the induction of cortical plasticity," *Brain Res.*, vol. 1674, pp. 91–100, Nov. 2017.

[20] M. Jochumsen, I. K. Niazi, N. Mrachacz-Kersting, N. Jiang, D. Farina, and K. Dremstrup, "Comparison of spatial filters and features for the detection and classification of movement-related cortical potentials in healthy individuals and stroke patients," *J. Neural Eng.*, vol. 12, no. 5, p. 056003, Oct. 2015.

[21] Sheng-Hsiou Hsu, T. Mullen, Tzyy-Ping Jung, and G. Cauwenberghs, "Online recursive independent component analysis for real-time source separation of high-density EEG," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 3845–3848.

[22] F. Karimi, J. Kofman, N. Mrachacz-Kersting, D. Farina, and N. Jiang, "Detection of Movement Related Cortical Potentials from EEG Using Constrained ICA for Brain-Computer Interface Applications," *Front. Neurosci.*, vol. 11, Jun. 2017.

Published by Verlag der TU Graz
Graz University of Technology

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-65

TOWARDS ANSWERING QUESTIONS IN DISORDERS OF CONSCIOUSNESS AND LOCKED-IN SYDROME WITH A SMR-BCI

Natalie Dayan[1], Alain Bigirimana[1], Alison McCann[2], Jacqueline Stow[2], Jacinta McElligott[2], Aine Carroll[2], Damien Coyle[1]

[1] Intelligent Systems Research Centre, Ulster University, Northern Ireland
[2] National Rehabilitation Hospital, Dún Laoghaire, Dublin, Republic of Ireland

E-mail: n.dayan@ulster.ac.uk

ABSTRACT: Sensorimotor rhythm-based brain-computer interfaces (SMR-BCI) may enable patients with prolonged disorders of consciousness (PDoC) or severe physical impairment to learn to intentionally modulate motor cortical neural oscillations. SMR-BCI could mitigate the need for movement-dependent behavioural responses, hence providing diagnostic information and/or communication strategies. Here, an SMR-BCI was evaluated in a three-staged protocol for PDoC. Stage I assessed awareness and capacity to modulate brain activity intentionally. Stage II facilitated SMR-BCI learning via stereo-auditory feedback training. Stage III tested use of SMR-BCI to answer closed categorized yes/no questions. Out of 14 patients with PDoC and locked in syndrome (LIS), eight patients showed capacity to modulate brain activity during stage I and thus participated in stage II. For practical reasons only five of these patients completed stage III. Two able-bodied participants were enrolled for benchmarking. Five of the eight participants performed significantly greater than chance level in 50-100% of runs ($p<0.05$). Average top run performance accuracy correlated with diagnoses category. Participants across the PDoC spectrum showed capacity to engage with SMR-BCI to answer closed questions.

INTRODUCTION

A gold standard assessment tool for Prolonged Disorders of Consciousness (PDoC) is yet to be realized, and communication strategies are difficult to establish. Consciousness requires arousability and awareness. Patients with PDoC have altered states of consciousness whereby, unresponsive wakefulness syndrome (UWS) is defined by clear signs of arousal but absence of awareness; minimally conscious state (MCS) is defined by preserved arousal level and distinguishable yet shifting signs of awareness. An individual with locked-in syndrome (LIS) is both conscious and aware, yet unable to communicate verbally or move. In LIS, usually blinking and vertical eye movements are retained and occasionally used to communicate [1], [2].

Standardization of PDoC clinical evaluation has been established through response scales such as the Coma Recovery Scale-revised (CRS-R) or Wessex Head Injury Matrix (WHIM). The CRS-R Scale is composed of six sub scales testing: audition, vision, motor, oromotor/ verbal, yes-no communication and arousal [3]. The WHIM is a 62-item hierarchical scale of defined behaviours that are considered to be sequentially more advanced [4]. These assessments are intended to decipher discrimination and localization from reflexive behaviours, and degree of patient interaction with environment, to establish the state of consciousness. However, since the introduction of several behavioural scales (including the aforementioned) as recommended by the Royal College of Physicians National Clinical Guidelines (RCP NCG) [5], misdiagnosis rates are still reportedly ~15-40% indicating an enduring unmet need for better assessment protocols [6], [7].

Applying sensorimotor rhythm (SMR)-brain-computer interfaces (BCI) to PDoC may augment clinical evidence supporting diagnoses and/or increase response reliability as a movement-independent communication channel. The primary sensorimotor cortex consists of topographic mapping dedicated to sensory and motor processing of anatomical divisions of the body. The SMR denotes localized frequencies in the μ (8–13 Hz) or β (15–30 Hz) range of electroencephalography (EEG) recorded across the sensorimotor cortices [8]. μ-rhythm decreases/ desynchronization is observed contralateral to left/right hand motor imagery (MI), similar to preparation or execution of movement. Classification of different motor imageries through SMR-BCI could facilitate discriminatory choice-making, independently of motor pathways, yet dependent on purposeful modulation of the motor cortex. Based on the premise that a PDoC patient is able to achieve above chance performance accuracy (AC) in SMR-BCI, it may be inferred that the individual has intact short-term memory in order to recall instructions, an ability to remain attentive for periods, and some degree of awareness. Cruse et al. [9], [10] showed 19% (three out of 16) UWS and 22% (five out of 23) MCS patients were able to perform command following via imagining squeezing their right hand or moving their toes in a single session. Coyle et al. showed that patients with PDoC can modulate visual and auditory feedback when learning to control an SMR-BCI and pilot data showing response to questions [11]–[13].

Here, an SMR-BCI protocol is evaluated in PDoC to further evidence its potential to assess awareness, and to develop an understanding of the influence multisession, SMR stereo-auditory feedback training in preparation

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-65

for patients to engage with a Q&A system, whereby imagined movements are used to answer closed questions with known answers. The Q&A paradigm is derived from three main influences: a BCI-functional near-infrared spectroscopy study in amyotrophic lateral sclerosis [14]; the Montreal Cognitive Assessment (MOCA) [15]; and annex 1a of the NCG - Operational evaluation of parameters for demonstrating consistent functional communication using autobiographical and situational questions [5].

MATERIALS AND METHODS

The study involved two able-bodied (AB) participants (as a benchmark) and 14 patients: eight with unresponsive wakefulness syndrome (UWS), three with minimally conscious state (MCS), and three with locked-in syndrome (LIS). Two participants with UWS were included in previous studies: [13], [16], [17]. The study was approved by National Rehabilitation Hospital of Ireland and carried out in accordance with the Declaration of Helsinki. Proxy-consent was given by participants' families. Trials were conducted in patient homes, care homes and hospitals in the Rep. of Ireland. EEG was recorded from 14 channels, Fp1, Fp2, F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, PO7, Oz, PO8 (g.Nautilaus amplifier with active electrodes (g.tec Medical Engineering, Austria)) at a sampling rate of 250Hz. The reference electrode was fixed on the right earlobe and the ground electrode was mounted on the forehead. The data were resampled at 125Hz. Bad channels were identified and removed via spectrum and kurtosis thresholding functions from an EEGLAB toolbox [18]. The number of channels removed varied from 0-4 channels. Data recorded were visually inspected for significant artefacts (e.g., eye-blinks). Trials with strong artefacts in most of the electrodes were removed.

Stage I (Session 1) entailed a block design assessment. Participants were asked to imagine one movement per block, cued with an auditory tone circa every 8s (6 blocks, 15 trials/block). In Stage II, following assessment, real-time stereo-auditory feedback was given as broadband (pink) noise or music samples (see [17] for details), over 5-10 sessions of 1-4 runs (60 trials/run, randomized equal number per class) cued with voice command e.g., "left", "feet" or "right" to matching ear via earphones: cue at 3s, feedback at 4-7s, followed by a "relax" cue. Feedback was modulated by continually varying the sound's azimuthal position between ±90° via imagined movement. Stage III, following training with auditory feedback, involved 4-6 question-answer runs (over 2-5 sessions) of 48 closed yes-no questions. Instructions were repeated at the start of each run and participants were asked to respond yes or no with respective hand/feet imagery. 96 unique closed questions were asked in total and were repeated across runs. Four question categories were evaluated: biographical, situational, basic logic, and numbers and letters. The questions and statements posed were adapted from the MOCA and NCG for PDOC [5], [15]. "Yes" questions had semantically similar "no" questions e.g., "You are 33 years old" vs "You are 47 years old". Recordings of family members reading questions were played back to participants in a timed paradigm. Familiar voices were recorded in order to encourage participants engagement through self-relevant stimuli [19], [20]. A CRS-R and WHIM assessment was conducted each day BCI sessions took place with UWS and MCS participants.

BCI setup: Throughout each stage of the experimental paradigm (i.e. assessment, stereo-auditory feedback and Q&A) the cue occurred at 3s and the window of interest was 3s prior and 5s post cue. During Q&A the cue was the end of each question, which lasted no longer than 7s. Event related offline peak accuracy (AC) was compared to random accuracy (RA) (class labels for the trials in the test sets were permuted randomly) and offline pre-cue (baseline) accuracy, respectively. The AC during the task execution period should be significantly higher than RA and the pre-cue accuracy (baseline period). Online single-trial accuracies were computed too.

Offline analysis was conducted through a filter bank common spatial patterns (FBCSP) framework, detailed in [21], to train a classifier to be applied in the auditory feedback runs. In this FBCSP framework, the EEG data are filtered into 9 frequency bands as shown in Fig. 1, and common spatial pattern (CSP) features are extracted from each band on a 2s sliding window. The features

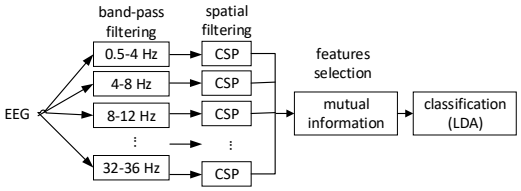


Fig. 1: The FBCSP-based framework.

from all the frequency bands are concatenated together, and then between 4 and 12 features are selected using mutual information. The best individual feature mutual information was used as detailed in [21].

The parameters to be optimized in the FBCSP setup used are the number of features selected and CSP filters pairs (between 1 to 4 pairs). These parameters are optimized in a nested cross-validation (CV): 6-fold-CV with an inner 5-folds-CV. A 2s sliding window was used (with an 80% overlap) in the task related period of the trial. At each window's position the parameters are optimized in the inner 5 folds cross-validation. An LDA, trained on the inner folds' data (i.e., training set of the outer fold) with the optimized parameters, is used to classify the outer fold test set. For each outer fold, the classifier with highest accuracy from different window positions is then used to classify the fold's test set at each time point of the trial segment. The resulting 6 time-courses of accuracy are averaged to get the time-course of cross-validation accuracy (CVAC). The parameters for the FBCSP differ across the 6 folds so the fold with highest accuracy is used to determine the parameters to be applied online. Using these parameters,

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-65

the final classifier to be deployed in the online feedback setup is trained at a 2s window positioned at the time of CVAC peak.

A permutation test was used to evaluate if the AC during the task execution is significantly higher than RA with a 95% confidence interval. The RA is computed by repeating the 6-fold cross-validation 100 times, and each time the trials' labels are randomized. This leads to 100 time-courses of random CVAC corresponding to 100 permutations. At each time-point of non-random CVAC, the probability that the classification accuracy is achieved by chance is computed using expression (1) as in [22]:

$$p = \frac{|\{D' \in \hat{D} : ac(D') \geq ac(D)\}| + 1}{n + 1} \quad (1)$$

where, $\hat{D}$ is a set of n-randomized versions *D'* of the original data *D*, and *ac(D)* is the accuracy achieved with the non-randomized data *D*. The computed *p* is the probability that given the permuted data, we can achieve accuracy level that is higher or equal to the accuracy achieved with non-permuted data. The Null hypothesis that classification accuracy was achieved by chance is rejected for $p < 0.05$. The *p*-value at each time point is computed enabling assessment of the time-course of CVAC significance.

Online feedback setup: For online processing during feedback runs, at each sample point, a distance is computed from the classifier's learned weights vector, distance referred to as time-varying signed distance (TSD) [23] [24]. The TSD value at a given time point *t* during nth trial is given by expression in (2). The distance's sign indicates the classifier's output label and its magnitude measures the classification confidence. The magnitude of the TSD indicates how the direction and amplitude of the audio feedback (panning to the right or left ear). The performance in online auditory feedback runs is given by the percentage of the trials with TSD's sign correctly matching the trial's task (class).

$$tsd_t^{(n)} = w^T \bar{\omega}_t^{(n)} + a_0 \quad (2)$$

where w^T and a_0 are the slope and bias of the discriminant hyperplane, respectively, of our trained LDA, $\bar{\omega}_t^{(n)}$ is the features vector at the time point *t* of the n^{th} trial. The tsd_t is debiased by subtracting the mean of *tsd* for the past 30-35s.

RESULTS

Six patients were withdrawn from the study after stage I (assessment stage) as AC was not significantly different to baseline (inclusion criteria for stage II). This withdrawal was not necessarily based on an inability to respond to command: it was difficult to acquire clean data from two participants with large frontal and bifrontal craniectomies and data contained noise as a result of persistent movement artefacts. The remaining patients completed study stage II: three UWS, two MCS, three LIS (3 Female). Time since condition onset varied between 11 months and 16 years (median= 3 years). For practical reasons, Stage III was only completed by one UWS, one MCS and three LIS. Clinical data are provided in Tab. 1. The three-staged paradigm was validated on two AB participants, whom achieved average AC of 77% and 84% across all session types. Every run's peak AC was significantly greater than baseline accuracy and RA ($p<0.05$), aside from three runs whereby peak AC was not significantly different from baseline accuracy. AB participants showed across session improvement, especially 1AB with AC increasing to 98%, but then decreasing to ~78% for Q&A runs (refer to Fig. 4). 3UWS and 5MCS both achieved top ACs of 69% and 77% during Q&A runs. Every participant was able to achieve above 70% accuracy during at least 1 run (refer to Tab. 1), with the top average run AC correlated with severity of diagnosis i.e., ascending from VS, MCS, LIS through to AB participants (Refer to Fig. 3). Across patients, AC did not progressively increase as a function of number of training sessions, and the top AC run did not necessarily appear in the latter half of total runs completed (Refer to Fig. 4). In five participants 50-100% of run ACs were significantly greater than RA ($p<0.05$). 53%, 57%, 56% and 100% of runs were significantly different from RA in 3UWS, 5MCS, 6LIS,

Tab. 1: Patient demographics, overall CRS-R and WHIM average scores, and top run performance accuracy with corresponding performance accuracy at 2 seconds during baseline period. UWS-unresponsive wakefulness syndrome, MCS-minimally conscious state, LIS-locked-in syndrome, AB-able-bodied, baseline 2s (A), (F), (Q) represent baseline for Assessment, feedback and Q&A runs respectively. WHIM score reported is the total number of behaviors observed.

ID	Sex	Age	Type of injury	Time since onset (months)	Av CRS-R	Av WHIM	BCI Top run Performance Accuracy (%)					
							Baseline 2s (A)	Assess	Baseline 2s (F)	Feedback	Baseline 2s (Q)	Q&A
1 UWS	M	34	Non-traumatic	192	5	4	55	63	62	72	-	-
2 UWS	M	34	Non-traumatic	103	3	3	61	61	57	76	-	-
3 UWS	M	29	Traumatic	74	5	4	58	69	47	73	54	69
4 MCS	M	49	Non-traumatic	23	11	16	53	66	52	73	-	-
5 MCS	F	56	Traumatic	35	18	17	67	80	57	80	52	77
6 LIS	F	34	Non-traumatic	11	-	-	52	60	50	88	52	81
7 LIS	M	28	Traumatic	25	-	-	71	73	58	75	58	67
8 LIS	F	27	Non-traumatic	36	-	-	70	78	40	68	56	88
1AB	M	20	-	-	-	-	61	69	45	98	42	79
2AB	M	23	-	-	-	-	65	78	52	88	50	79
Average of BCI Top run Performance Accuracy (%):							**61**	**68**	**52**	**79**	**52**	**77**

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-65

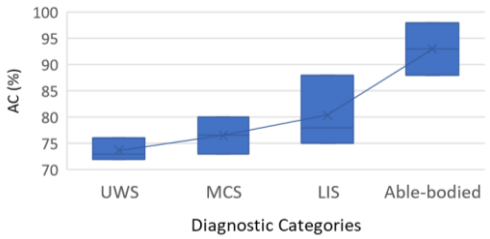


Fig. 3: Whisker-box plot of average top run performance accuracy per participant diagnostic category. X's demarcate the mean, central line is the median, and inclusive median quartile calculation is displayed. Performance is shown to increase across levels of awareness.

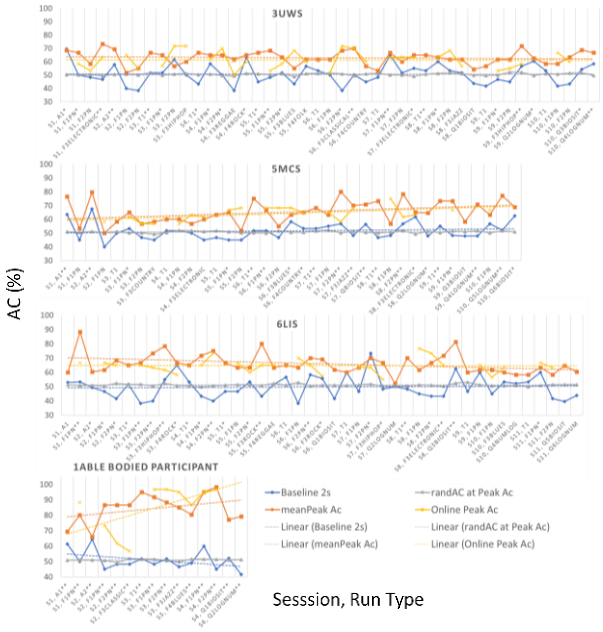


Fig. 4: Run accuracies of best patients and best able-bodied participant across sessions. ss=session, a=assessment run, t = training run, f= feedback run, pn = pink noise, q= q&a run biosit = biographical and situational, lognum = logic and numbers/letters. $p<0.05$ signified by*, $p<0.01$ signified by **(Wilcoxon signed ranks test).

1AB respectively. For Q&A, five participants presented significant peak ACs relative to RA ($p<0.05$) in 50-100% of runs apart from 6,7LIS (refer to Fig. 4). However, Peak AC was significantly different from baseline AC in ≥50% of Q&A runs for four participants 6,8LIS and AB1,2 ($p<0.05$). Across participants, on average, baseline AC was significantly different from RA in 7% of runs.

Fig. 5 illustrates the time courses of the top AC runs across participants from each group/condition: 3UWS, 5MCS, 6LIS and 1AB, with select corresponding event-related desynchronization (ERDS) plots. Across these runs, AC is at chance level at the start of the trial (cue at 3s) and increases (deviating from chance level) as the participant executes the task. Peak AC during feedback runs are maximal and have a similar time course for 1AB and 6LIS. The maps show the power change with respect to the baseline (pre-cue period of 0.2-3s). For the MI tasks, activation is expected in electrodes placed

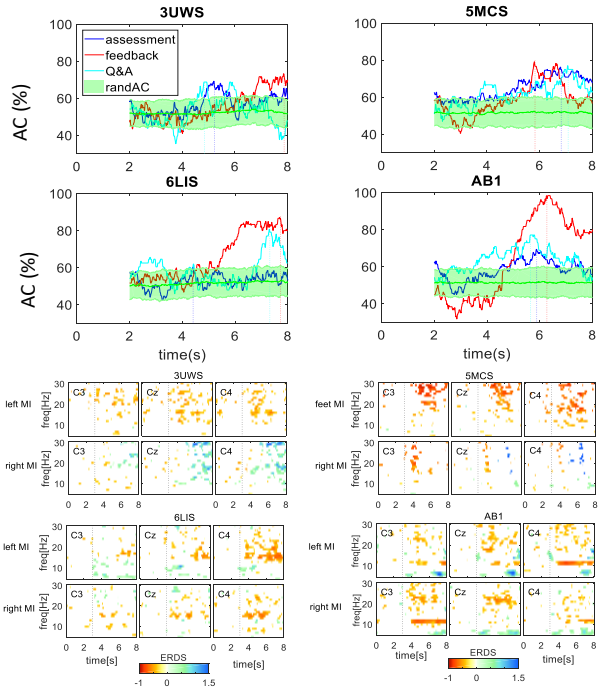


Fig. 5: Example time courses of top performance accuracies (AC) of assessment, feedback and Q&A on best participants per diagnostic group (top 4 panels), green and shaded area indicates mean and variation in accuracy from randomly permutated trials. One example corresponding ERDS map per participant (bottom 4 panels). MI = Motor imagery, left, right refers to hand.

around the motor cortex (C3, Cz, and C4). The ERDS map for 1AB shows clear contralateral activation indicated by mu (8-12Hz) rhythm ERD in the electrodes mounted around the motor cortex (C3 for right MI and C4 for left MI). Patients present task differences for ERD/S plots, however these do not conform entirely to typical observations expected for MI.

Average CRS-R and WHIM for UWS and MCS patients are shown in Tab. 1. As expected, there is strong positive correlation between CRS-R and WHIM scores ($r=0.88$, $p<0.0001$). Average BCI AC during assessment sessions for UWS + MCS was shown to have a positive, yet insignificant, correlation to the average sessional CRS-R ($r=0.4$, $p>0.05$) and WHIM scores ($r=0.4$, $p>0.05$) (2 tailed Spearman's rank correlation). A weak relationship was observed in comparing all average session ACs for UWS+MCS to CRS-R ($r=0.18$, $p>0.05$) and WHIM scores ($r=0.09$, $p>0.05$) (2 tailed Spearman's rank correlation). A Spearman's rank correlation (2-tailed) with average AC for MCS patient, showed a positive correlation to CRS-R scores, $r=0.42$, with a tendency towards significance, $p=0.057$; and an insignificant correlation to WHIM total behavior scores at the participant level $r=0.07$, $p>0.05$.

DISCUSSION

We sought to determine if AC could be used to provide indication of awareness in a 1-2 sessions of assessment (stage I) and whether this corresponded with

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-65

conventional scales (e.g. CRS-R and WHIM). We observed that 3/8 UWS, 2/3 MCS and 3/3 LIS participants were capable of modulating brain activity through SMR strategies during stage I. Both assessment and feedback average ACs in MCS participants, were found to weakly positively correlate with CRS-R scores, and there was no correlation with WHIM scores. UWS participants demonstrated significant above chance AC during multiple runs, which conflicts with their UWS diagnoses. An insignificant correlation between UWS+MCS participants and behavioral scores was found, indicating SMR-BCI may provide supplementary or corroborative diagnostic information in PDoC. These results further demonstrate that EEG-based SMR-BCI provides evidence of awareness not detected by standard behavioral tests in UWS. Some analytical hurdles have been reported concerning block-design [25], [26], however this was necessary for the assessment stage in order for the cue to be demarcated by a tone rather than a word describing the type of MI. This eliminates the likelihood of the response being automatic/unconscious [27] and requires short term memory of the instruction given at the start of the block.

In Stage II, during multisession stereo-auditory feedback training runs, most participant ACs were significantly greater than RA, indicating cohort-wide engagement. However, across patients, progressive AC increase as a function of number of completed runs was not observed. $\geqslant$70% accuracies were not consistently reported (AC of 70% being viewed as the lower limit for ability to communicate effectively with a BCI [28]). AC variation across runs and sessions may be influenced by many factors such as patient motivation Proper patient positioning may encourage arousal/ minimize involuntary movements or persistent involuntary hypertonicity that may be induced by frustration due to miscommunication, particularly in LIS. Other studies reported that after post learning during early training, patients AC stabilizes within the first 10–20 training sessions [29]. This is in line with across session performance observations observed here. Individuals were trained over the period of a few weeks, in some cases months due to various interruptions, which may have impacted performance. Our dataset is consistent with other studies in terms of patients obtaining higher inter-run/session and inter-individual variability relative to AB participants [30], [31]. In stage III, ability to encode yes/no responses through motor imageries to closed questions was assessed. The feedback sessions were implemented to encourage SMR learning prior to the more complex Q&A. All diagnostic groups were able to respond at above chance levels in at least one Q&A run demonstrating feasibility of adoption by this patient cohort. High intersession variability is also demonstrated in Q&A runs indicating the importance of multiple sessions, as recommended with behavioural scale assessments by the RCP NCG for PDoC [5]. It is yet to be established whether consistent, sufficient accuracy can be achieved across the patient cohort to enable communication and further sessions with these patients will be conducted. Given availability of more data, AC as a factor of question category could isolate different cognitive deficits in relation to knowledge of self and environment, logic and numbers/letters. It would be interesting to test further iterations of the paradigm e.g. tailoring the response time window. Here, this was set to 5s and might not have been sufficient for some participants. Response time was constrained as AC is hinged on a trade-off between duration/complexity of task/keeping patient fatigue to a minimum and maximizing the amount of trials/answers collected.

PDoC is a challenging patient group to evaluate due to tendencies for; heterogeneity in aetiology and potential neural atrophy and cortical remapping; muscle spasms, seizures, fluctuating arousal, ease of exhaustion; limited memory capacity; medication affecting vigilance e.g. muscle relaxants, anti-epileptics and anxiolytics; and suboptimal EEG recording due to ocular, respiratory and muscular artefacts. EEG quality may have been affected by presence of nutritional life supporting systems or other equipment where private rooms away from other hospital equipment e.g., airbeds was not possible.

In future, it would be ideal to analyse the relationship between AC and type of injury, time since altered consciousness onset, time of day of session. A further investigation might be to add a third “I don’t know” class reflected by another MI to maximize separability for yes/no classes. The existing paradigm assumes the user will polarize their response when the answer is unclear. Having three classes would increase the cognitive load, nonetheless other groups have also experimented with providing options beyond yes/no in PDoC, e.g. a 4-choice auditory oddball EEG-BCI paradigm based on a P300 response [32].

CONCLUSION

This study showed demonstrable feasibility of an initial assessment of SMR engagement; multisession auditory feedback to train SMR-BCI control; and an SMR-BCI Q&A system in PDoC and LIS. Adaptation of the paradigm in order to maximize the number of runs where 70% AC is reached prior to commencing Q&A is crucial to effective adoption by patients with a PDoC. This is the first targeted group of this patient cohort and further training is necessary to progress to open Q&A sessions.

Acknowledgments: We are indebted to the patients and their families, who enabled this research.

REFERENCES

[1] S. Laureys, M. Boly, G. Moonen, and P. Maquet, “Coma,” *Encyclopedia of Neuroscience*. Elsevier Ltd., p. vol. 2, pp. 1133–1142, 2009.

[2] G. Bauer, F. Gerstenbrand, and E. Rumpl, “Varieties of the locked-in syndrome,” *J. Neurol.*, vol. 221, no. 2, pp. 77–91, Aug. 1979.

[3] J. T. Giacino, K. Kalmar, and J. Whyte, “The

Published by Verlag der TU Graz
Graz University of Technology

JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility.," *Arch. Phys. Med. Rehabil.*, vol. 85, no. 12, pp. 2020–9, Dec. 2004.

[4] A. Shiel, S. A. Horn, B. A. Wilson, M. J. Watson, M. J. Campbell, and D. L. McLellan, "The Wessex Head Injury Matrix (WHIM) main scale: a preliminary report on a scale to assess and monitor patient recovery after severe head injury," *Clin. Rehabil.*, vol. 14, no. 4, pp. 408–416, Aug. 2000.

[5] Royal College of Physicians, "Prolonged disorders of consciousness: national clinical guidelines | RCP London." RCP, London, 2013.

[6] C. Schnakers *et al.*, "Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment.," *BMC Neurol.*, vol. 9, p. 35, Jul. 2009.

[7] R. T. Seel *et al.*, "Assessment Scales for Disorders of Consciousness: Evidence-Based Recommendations for Clinical Practice and Research," *Arch. Phys. Med. Rehabil.*, vol. 91, no. 12, pp. 1795–1813, Dec. 2010.

[8] E. Niedermeyer and F. H. Lopes da Silva, Electroencephalography : basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins, 2005.

[9] D. Cruse *et al.*, "Bedside detection of awareness in the vegetative state: a cohort study.," *Lancet (London, England)*, vol. 378, no. 9809, pp. 2088–94, Dec. 2011.

[10] D. Cruse *et al.*, "Relationship between etiology and covert cognition in the minimally conscious state.," *Neurology*, vol. 78, no. 11, pp. 816–22, Mar. 2012.

[11] D. Coyle, A. Carroll, J. Stow, A. McCann, A. Ally, and J. McElligott, "Enabling Control in the Minimally Conscious State in a Single Session with a Three Channel BCI," *1st Int. Decod. Work.*, no. April, pp. 1–4, 2012.

[12] D. Coyle, Á. Carroll, J. Stow, K. Mccreadie, and J. Mcelligott, "Visual and Stereo Audio Sensorimotor Rhythm Feedback in the Minimally Conscious State," *Proc. Fifth Int. Brain-Computer Interface Meet.*, 2013.

[13] D. Coyle, N. Dayan, J. Stow, J. McElligott, and A. Carroll, "Answering questions in Prolonged disorders of consciousness with a brain-computer interface," in *Seventh International BCI Meeting*, 2018.

[14] U. Chaudhary, B. Xia, S. Silvoni, L. G. Cohen, and N. Birbaumer, "Brain–Computer Interface–Based Communication in the Completely Locked-In State," *PLOS Biol.*, vol. 15, no. 1, p. e1002593, Jan. 2017.

[15] E. de Guise *et al.*, "The Montreal Cognitive Assessment in Persons with Traumatic Brain Injury," *Appl. Neuropsychol. Adult*, vol. 21, no. 2, pp. 128–135, Apr. 2014.

[16] K. A. McCreadie, D. H. Coyle, and G. Prasad, "Learning to modulate sensorimotor rhythms with stereo auditory feedback for a brain-computer interface," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, vol. 2012, pp. 6711–6714.

[17] D. Coyle, J. Stow, K. McCreadie, J. McElligott, and Á. Carroll, "Sensorimotor Modulation Assessment and Brain-Computer Interface Training in Disorders of Consciousness," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 3, pp. S62–S70, Mar. 2015.

[18] 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, 2004.

[19] F. Perrin, M. Castro, B. Tillmann, and J. Luauté, "Promoting the use of personally relevant stimuli for investigating patients with disorders of consciousness," *Front. Psychol.*, vol. 6, p. 1102, Jul. 2015.

[20] A. M. Kempny *et al.*, "Patients with a severe prolonged Disorder of Consciousness can show classical EEG responses to their own name compared with others' names," *NeuroImage Clin.*, vol. 19, pp. 311–319, Jan. 2018.

[21] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Front. Neurosci.*, vol. 6, no. MAR, pp. 1–9, 2012.

[22] M. Ojala and G. C. Garriga, "Permutation Tests for Studying Classifier Performance," *J. Mach. Learn. Res.*, vol. 11, no. Jun, pp. 1833–1863, 2010.

[23] D. Coyle *et al.*, "Action Games, Motor Imagery, and Control Strategies: Toward a Multi-button Controller," in *Handbook of Digital Games and Entertainment Technologies*, no. June, M. Cavazza and R. M. Young, Eds. 2016, pp. 1–16.

[24] G. Pfurtscheller, C. Neuper, A. Schlögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–25, Sep. 1998.

[25] A. M. Goldfine, J. C. Bardin, Q. Noirhomme, J. J. Fins, N. D. Schiff, and J. D. Victor, "Reanalysis of 'Bedside detection of awareness in the vegetative state: a cohort study,'" *Lancet*, vol. 381, no. 9863, pp. 289–291, Jan. 2013.

[26] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage*, vol. 56, no. 2, pp. 387–399, May 2011.

[27] P. Nachev and M. Husain, "Comment on "Detecting awareness in the vegetative state".," *Science*, vol. 315, no. 5816, pp. 1221; author reply 1221, Mar. 2007.

[28] A. Kübler and N. Birbaumer, "Brain–computer interfaces and communication in paralysis: Extinction of goal directed thinking in completely paralysed patients?," *Clin. Neurophysiol.*, vol. 119, no. 11, pp. 2658–2666, Nov. 2008.

[29] A. Kübler, N. Neumann, B. Wilhelm, T. Hinterberger, and N. Birbaumer, "Predictability of Brain-Computer Communication," *J. Psychophysiol.*, vol. 18, no. 2/3, pp. 121–129, Jan. 2004.

[30] D. Lulé *et al.*, "Probing command following in patients with disorders of consciousness using a brain–

Published by Verlag der TU Graz
Graz University of Technology

 DOI: 10.3217/978-3-85125-682-6-65

computer interface," *Clin. Neurophysiol.*, vol. 124, no. 1, pp. 101–106, Jan. 2013.

[31] A. M. Goldfine, J. D. Victor, M. M. Conte, J. C. Bardin, and N. D. Schiff, "Determination of awareness in patients with severe brain injury using EEG power spectral analysis," *Clin. Neurophysiol.*, vol. 122, no. 11, pp. 2157–2168, Nov. 2011.

[32] D. Lulé *et al.*, "Probing command following in patients with disorders of consciousness using a brain–computer interface," *Clin. Neurophysiol.*, vol. 124, no. 1, pp. 101–106, Jan. 2013.

Published by Verlag der TU Graz
Graz University of Technology