

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

J. Egger¹, K. Kostoglou¹, G. R. Müller-Putz^{1,2}

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

²BioTechMed Graz, Austria

E-mail: gernot.mueller@tugraz.at

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

INTRODUCTION

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

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

MATERIALS AND METHODS

A. EEG recordings throughout the day and night

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

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

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

B. Processing of recordings

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

C. Analysis of MRCPs

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

D. Classification of gestures

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

E. Analysis of temporal changes in classification

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

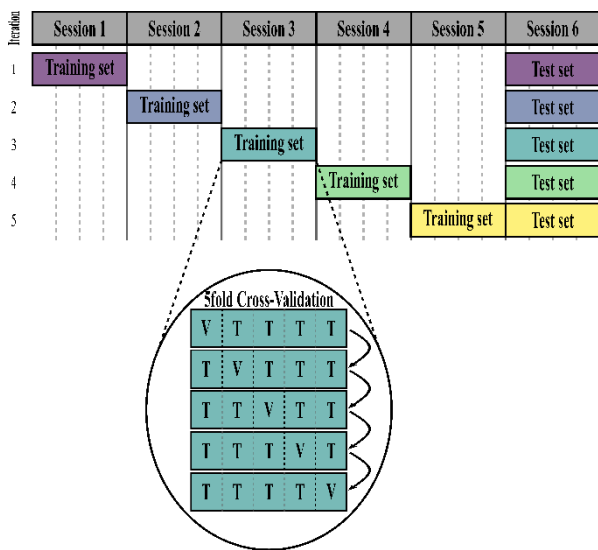


Figure 1: Classification procedure of single classifiers tested on session 6. Additionally, the trial based 5-fold cross-validation procedure for the single session results on session 3 is depicted.

F. Comparison between adaptive and unrevised classification

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

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

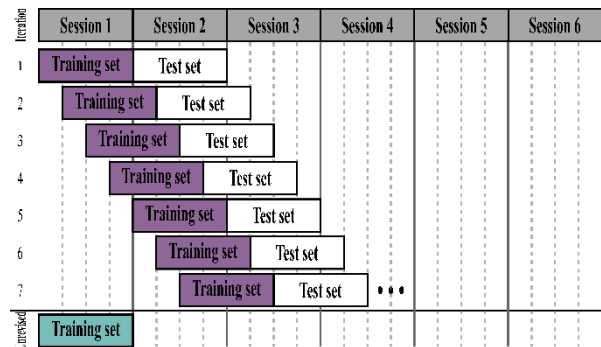


Figure 2: Adaptive (violet) and unrevised (turquoise) classification approach. As an example, only the first seven iterations of the adaptive classifiers are depicted. The test sets were the same for both classification approaches.

RESULTS

A. Analysis of MRCPs

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

B. Analysis of temporal changes in classification

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

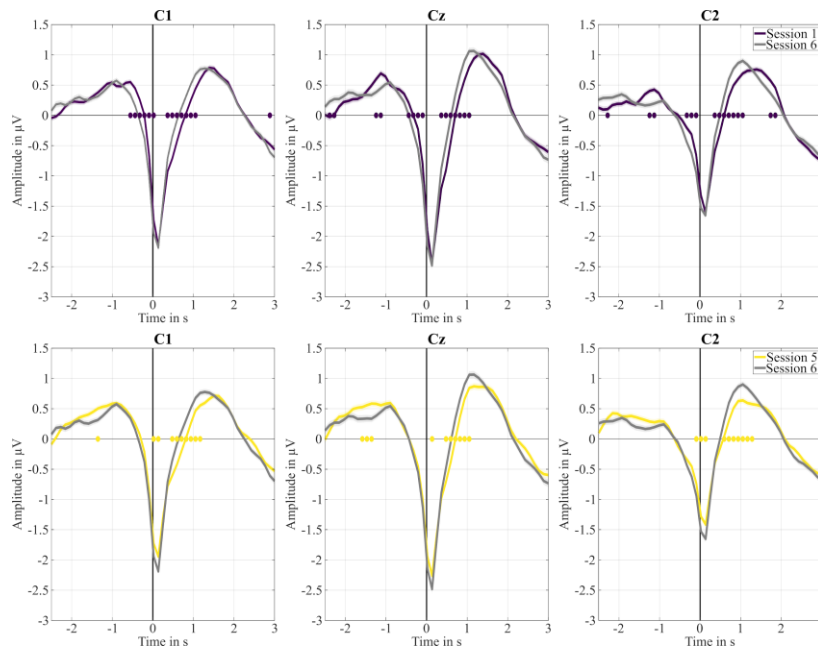


Figure 3: Average MRCPs across all participants for sessions 1 and 6 (top panel) and sessions 5 and 6 (bottom panel). The movement onset occurred at $t=0$ s. Statistically significant differences ($p < 0.05$) between sessions at each time point within a trial are indicated with color-coded dots on the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

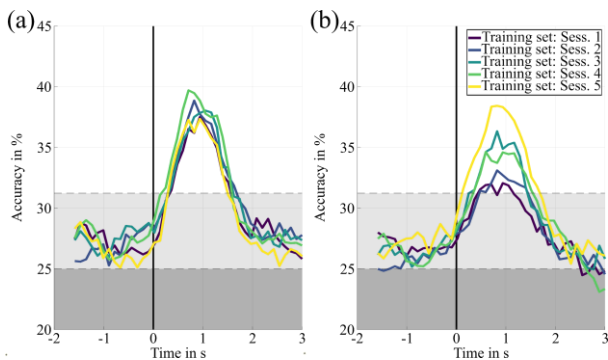


Figure 4: Classification results of different sessions. (a) Single session results. (b) Results of classification when tested on session 6. Indicated by the horizontal dashed lines are the theoretical chance level (25%) and the level of statistical significance (31.25%) as estimated using a permutation-based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

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

C. Comparison between adaptive and unrevised classification

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

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

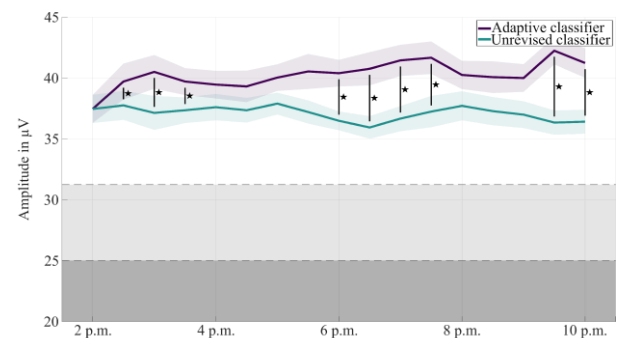


Figure 5: Comparison of the maximum classification accuracies obtained from both the adaptive (violet) and unrevised decoders (turquoise) shifted along the time axis. Depicted are the averages across participants (\pm standard error). The horizontal dashed line at 25 % indicates the theoretical chance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

DISCUSSION

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

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

A. Analysis of MRCPs

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

B. Analysis of temporal changes in classification

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

C. Comparison between adaptive and unrevised classification

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

CONCLUSION

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

ACKNOWLEDGEMENTS

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

REFERENCES

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

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