Proceedings of the
9th Graz Brain-Computer Interface Conference 2024
10.3217/978-3-99161-014-4-068

VALIDATING NEUROPHYSIOLOGICAL PREDICTORS OF BCI PERFORMANCE ON A LARGE OPEN SOURCE DATASET

D.Trocellier[1,2], B. N'Kaoua[3], F. Lotte[1,2]

[1] Inria center at the University of Bordeaux, Talence, France
[2] LabRI (CNRS - Univ. Bordeaux - Bordeaux INP), Talence, France
[3] Université de Bordeaux, Bordeaux, France

E-mail: david.trocellier@inria.fr

ABSTRACT

Brain-computer interfaces (BCI) are systems that process brain activity to decode specific commands from it such as motor imagery patterns generated when users imagine movements. Despite the growing interest in BCI, they present significant challenges, notably in decoding distinct neural patterns, due to considerable variability across and within users. The literature showed that various predictors were correlated with subject's BCI performance. Among these indicators, neurophysiological predictors appeared to be the most effective, although studies generally involved small samples and results were not always replicated, thus questioning their reliability. In our study, we used a large dataset with 85 subjects to analyse the relationship between different predictors identified in the literature and BCI performance. Our findings reveal that only four of the six predictors tested could be replicated on this dataset. These results underscore the necessity of validating literature findings to ensure the reliability and applicability of such predictors.

INTRODUCTION

Brain-computer interfaces (BCI) are devices that measure brain activity, typically through electroencephalography (EEG), to extract specific commands for computer input. A particularly widespread paradigm of BCI is motor magery (MI), which involves decoding EEG patterns characteristic of movement imagination (typically signals from 8-13 Hz located over the motor and sensory motor cortices). MI leads to patterns similar to motor execution, characterized by an event-related desynchronization (ERD), i.e, a diminution of the amplitude, within the mu band (8-13 Hz) and beta band (13-30 Hz) of the contralateral sensory motor cortex, followed by an event-related synchronization (ERS), i.e an augmentation of the amplitude of the signal within the beta band after the imagined movement ends. A primary limitation of current MI-BCI technology is decoding accuracy. Producing clear neurophysiological signals that can be decoded by existing classification algorithms is not a competence that all BCI users have. In their article, Blankertz et al. [1] demonstrate that using a common spatial pattern (CSP) with a linear discriminant analysis (LDA) classifier, the average BCI accuracy is 74.4 % ± 16.5 %, with significant variability among participants, ranging from perfect classification (100 %) to performance equivalent to chance level (50 %). More recently, Dreyer et al. [2] published a large database of 87 first-time BCI users performing MI-BCI, reporting a mean accuracy of 63.53 % with a large variability of performances (std = 17.61 %). This variability may be due to users' inability to produce clear and distinguishable patterns that are strong enough to be classified by current algorithms. It is considered that for effective BCI control, performance should exceed 70 % [3]: users below this threshold are deemed "BCI illiterate" or the BCI "BCI deficient". Understanding the parameters explaining differences in user control of such devices has been an important research question for the past 20 years. A better comprehension of those predictors is essential for developing better BCIs, e.g., to later identify the best BCI type for each user or to create BCIs that consider those predictors in their design and into classification algorithms. The literature identifies a broad spectrum of predictors that can be categorized into four main groups: personality traits, cognitive profiles, demographic factors, and neurophysiological patterns [4]. Traits are "stable and enduring, caused by internal circumstances" whereas mental states, as defined by Chaplin et al. [5], are "temporary, brief, and caused by external circumstances". Demographic characteristics correspond to personal characteristics (age, gender, etc.), habits, and environment-related factors. Neurophysiological predictors are predictors from the EEG signal during MI tasks, pre-cue MI tasks, or during a resting state, serving as markers of the user's mental states, such as attention [6] [7], fatigue [8], or initial capacities for producing the pattern to be decoded [1].

Among personality traits, cognitive profiles and demographic characteristics correlated with BCI performance. Jeunet et al. [4] identified 3 major elements : user relationship with technology, attentional capacities, and spatial abilities. More recently, Leeuwis et al. [9] re-evaluated these predictors in an experiment with 55 subjects and found that MI-BCI performance was significantly correlated with vividness of visual imagery, and the personality traits of orderliness and autonomy. How-

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-068

ever, Benaroch et al. suggested that personality traits, cognitive profiles and demographic characteristic might not be sufficient to predict MI-BCI performance using statistical models [10]. They conducted a follow-up experiment incorporating into the models neurophysiological predictors measured during a two-minute baseline at rest [11], and measured their predictive capabilities. They found significant predictability with neurophysiological predictors ($p<0.01$) while traits and demographic information led to a predictability that was not better than chance level ($p=0.88$).

The most robust neurophysiological predictor of BCI performance is the sensory motor rhythm (SMR) predictor, with a correlation coefficient (R) of 0.53 validated with 80 participants, proposed by [1]. This predictor is computed from a 2-minute EEG recording during an open-eye baseline period, wherein the participant is not engaged in MI tasks associated with BCI activities. The SMR-predictor encapsulates the participant's capability to modulate their SMR. The effectiveness of predictions based on the SMR was confirmed by subsequent studies [11, 12], which also highlights the efficacy of SMR-based predictors. Specifically, Tzdaka et al. [11] introduced additional predictors based on the estimation of the power spectral density (PSD) during the open-eye baseline. With 56 participants, they showed that the mean performance during the session was significantly correlated with the number of frequency peaks necessary to model the PSD at electrodes C3 and C4 ($R = 0.351$) and with the temporal variance in the amplitude of the peak within the mu band during baseline ($R = -0.477$). This also highlights the efficacy of baseline SMR-based predictors of MI-BCI performances.

Other predictors have also been correlated with BCI performance. Grosse-Wentrup et al. identified in a dataset of 10 participants a correlation between the high gamma (55-85 Hz) rhythm in centro-parietal and frontal regions during the MI-BCI task and the SMR quality score, a metric of BCI classification performance ($r = 0.0786$, $p = 9.998 \times 10^{-5}$) [13]. Here, the gamma rhythm was thought to be related to attentional networks active during MI-BCI tasks [7]. Foong et al. demonstrated that for stroke patients (n=11), the mean relative beta power (12-30 Hz) before the cue across all trials was correlated with session performance [8]. This correlation was present for the relative beta power in the frontal (F3,Fz,F4)($r = 0.251$, $p = 0.0005$) and central (C3,Cz,C4)($r = 0.181$, $p = 0.0130$) brain regions but not in parietal-occipital regions ($r = 0.033$, $p = 0.6486$). It is important to note that this study made the hypothesis (not empirically verified) that relative beta power was a neural correlate of fatigue. Ahn et al. [14] divided the users in two different groups based on their performances during MI-BCI training, and have shown that during the baseline the efficient group has significantly higher θ and lower α than the inefficient one. Based on these results, they proposed a predictor of BCI performance, named PPfactor, that combined θ, α, β and γ power which had a strong correlation with BCI performance ($R = 0.59$) in their 61 subject dataset. Interestingly, although no significant correlations were found between BCI classification performance and either β or γ power, they were nonetheless included in the PPfactor equation.

Despite the extensive literature in the field, identifying strong and robust predictors of MI-BCI remains a significant hurdle. While the SMR predictor efficacy has been replicated in other datasets [11, 12], it is not the case for the other predictors. Botrel et al. have highlighted the issue, where numerous studies introduce new potential psychological or neurophysiological predictors of performance, yet often fail to replicate the findings of earlier research [15]. Moreover, several of these predictors were identified on small data sets, which can question their reliability. There is thus a pressing need to consolidate these predictors within a single study, on a different data set than the one on which they were identified, to assess their validity and reliability comprehensively. Therefore, in this paper, we aim to evaluate these diverse neurophysiological predictors together using a large (n=85) open source dataset to attempt to confirm their correlations with BCI performance.

This paper is organised as follows. In the Materials and Methods section, we present the dataset used and the methodology for extracting the neurophysiological predictors. The Results Section analyses the correlation between BCI performance and the neurophysiological predictors extracted. The Discussion Section finally compares those results with the one obtained by the original articles.

MATERIALS AND METHODS

In order to validate those different neurophysiological predictors on a large and open source data set, we need to first extract those predictors from the dataset and then correlate them with BCI performance. In this section, we will first briefly present the dataset, then we will detail the 6 different neurophysiological predictors and how to compute them, lastly we will develop the statistical analysis that we performed.

Dataset: The dataset used in this analysis is sourced from an open-access EEG database with 87 participants, collected during a single day of MI-BCI experiments [2]. The experimental protocol was organized into a single session of motor imagery, divided into six runs. Before the first runs the participants were asked to perform two three minutes baseline recordings, one with open eyes, the other with closed eyes, where the participants were asked to fix a cross and relax. Participants were then required to engage in 20 motor imagery tasks (trials) for each hand per run. The first two runs corresponded to a calibration phase, where the feedback provided does not reflect the participants' actual motor imagery performance, but was a sham feedback. Based on the EEG data from these two first runs, a linear classifier employing three pairs of CSP spatial filters and a LDA classifier is

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-068

trained. In the subsequent four runs, participants received real-time feedback, visually represented by a horizontal bar whose length varied in proportion to the accuracy of the classifier predictions. The goal for participants was to maximize the length of this bar through their motor imagery efforts. In this analysis, we used epoch-wise accuracy (EAcc) as classification performance metrics. This metrics is expressed in ratio of correctly classified epochs compared to the total number of classification. We divided each MI trial in 1 seconds epoch with 1/16 overlap from 0.5 to 4.5 seconds after the visual cue.
Two participants, identified as A40 and A59, were excluded from the statistical analysis due to missing trials that could potentially impact the data analysis pipeline in subsequent steps.

Neurophysiological predictors:
In this paper we focused on six distinct predictors, four of them are extracted during the two-minute eyes-open (OE) baseline, one during the motor imagery (MI) trial, and one across MI trials. The extraction of these neurophysiological predictors was conducted as delineated in their respective foundational studies. The following subsections detail the extraction process for each predictor.
The SMR-predictor [1] is computed from a 2-minute EEG recording, taken during an open-eye baseline period when the user is relaxing. The signal is filtered between 4 and 40 Hz using a Laplacian filter around the C3 and C4 electrodes. The SMR-predictor is computed based on an estimation of the power spectral density (PSD) composed by the sum of two functions, as shown in Equation 1. The function g_1, given in Equation 2, models the noise floor of the signal, while g_2, outlined in Equation 3, estimates the peaks in the μ (8-15 Hz) (with μ_1 and σ_1) and β (15-30 Hz) (with μ_2 and σ_2) bands in order to best estimate the PSD during the baseline. The SMR predictor is defined as the maximum difference within the mu band between the noise floor and the estimated peak g_2. From the PSD estimation, Tzdaka et al. [11] found two other predictors correlated with BCI performance: **The sum of the number of peaks** in C3 and C4 for each PSD reconstruction g (either 0, 1, or 2 for each channel) and **the variance across time of the peak frequency of the SMR predictor during the baseline**. For the latter, the two-minute baseline signal is divided into 10-second epochs with a 3-second overlap. For each epoch, the PSD of the signal is estimated with g. The variance is computed across all the estimations of the SMR predictor for C3 and C4, and then averaged between the two sensors.

$$g(f;\lambda,\mu,\sigma,k) = g_1(f;\lambda,k) + g_2(f;\mu,\sigma,k) \quad (1)$$

$$g_1(f;\lambda,k) = k_1 + \frac{k_2}{f^{\lambda}} \quad (2)$$

$$g_2(f;\mu,\sigma,k) = k_3\phi(f;\mu_1,\sigma_1) + k_4\phi(f;\mu_2,\sigma_2) \quad (3)$$

where $k = (k1;k2;k3;k4) \in \mathbb{R}^4$; $\lambda \in \mathbb{R}$ correspond to the steepness of the noise floor and $\phi(.,\mu,\sigma)$ indicates the probability density function of a normal distribution with mean μ and standard deviation σ.

The PPfactor is calculated using data from a two-minute baseline period with open eyes [14]. During this phase, after applying a common average reference and a notch filter centred at 50 Hz to mitigate power line interference, the power spectrum of the EEG signal at electrodes C3 and C4 is analysed across various frequency bands:θ (4-8 Hz) α (8-13 Hz), β (13-30 Hz) and γ (30-70 Hz). The power values for each band are then normalized by the total power across the entire spectrum (4-70 Hz) to account for individual differences in overall brain activity levels. The PPfactor is subsequently calculated using Equation 4

$$PPfactor = \frac{\alpha+\beta}{\theta+\gamma} \quad (4)$$

The high gamma predictor is a marker of attention, extracted during the MI task [7]. During the entire epoch of the MI task, only the electrodes from the Frontal, Central, Central Parietal, and Occipital regions are considered. The EEG signal from these regions is filtered between 55 and 85 Hz to isolate the gamma band. The signal is then processed as follows: The filtered EEG signal is squared to calculate the power of the signal at each time point within the epoch. The mean power is computed by averaging the squared signal over all time stamps within the epoch. The power values are then log-transformed to normalize the distribution. Finally, the mean across all selected channels is computed to obtain a single value representing the mean log power for the gamma band during the MI task. It is important to note that the signal used for classification of MI is filtered between 8 and 30 Hz. Therefore the gamma band is not used as a classification feature.
The Relative beta power at rest is derived from a 3-second pre-MI task EEG segment, without explicit rest instructions, using frontal electrodes (F3, Fz, F4) [8]. As presented in equation 5 it is computed from the power ω in the broad band (4-50 Hz) and the power β in the beta band (12-30 Hz) and is thought to reflect the level of fatigue of the user during the task.

$$RelativeBeta = 10 * \log_{10}(\frac{\beta}{\omega}) \quad (5)$$

Statistical analysis: We conducted correlations between the predictors and the offline classification performances. Our approach varied depending if the predictors is compared with the overall performance during the session or if this predictor is compared with the evolution of performance during the session. In the first case we performed Spearman correlation and in the second multi repeated correlation analysis [16]. The detailed analysis is described below:
- For predictors obtained during the baseline phase, we explored the correlation between each subject's overall performance in a single session and the predictor. We used Spearman correlation tests for this analysis. Spearman correlation was chosen for its ability to capture

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-068

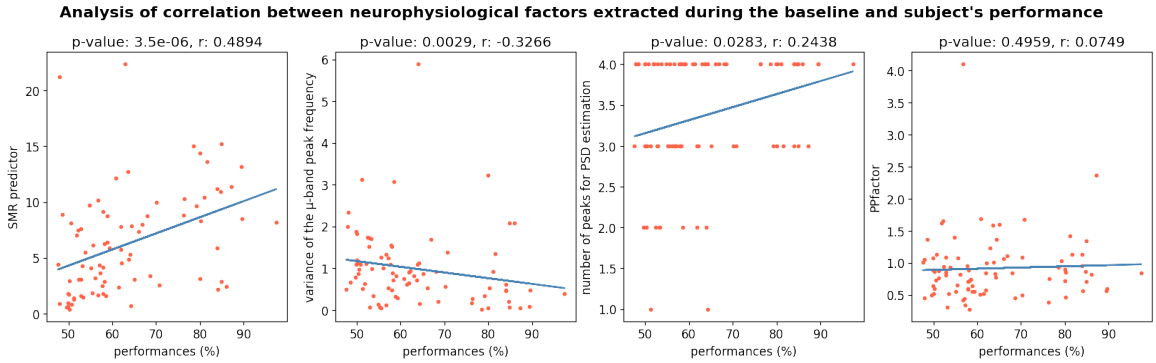

Figure 1: Correlation between subjects' BCI performance and neurophysiological predictors extracted during baseline. The scatter plots from left to right represent: SMR-predictor, the number of PSD peaks in the sensorimotor channels C3 and C4, the variance in the peak of μ-band frequency during baseline, and the PPfactor. Each plot displays individual subject performances against the respective neurophysiological predictor, with the fitted linear regression line illustrating the trend. Significance levels (p-values) and correlation coefficients (r) between each predictor's and the performance are indicated based on Spearman correlation.

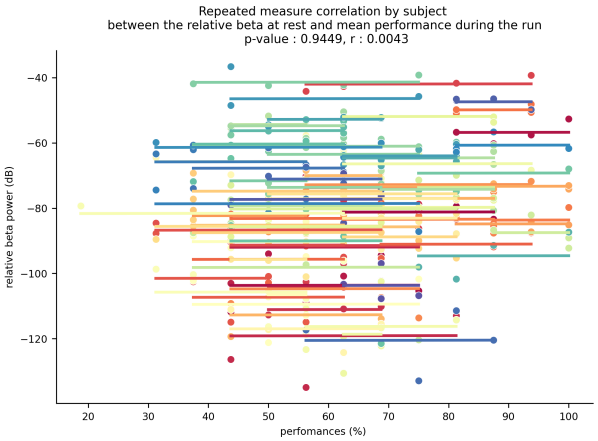

Figure 2: repeated measure correlation analysis between individual subjects' BCI performances and their average relative beta power during rest phases across four runs (n=85 subjects). The lines correspond to the regression for each subject where the slope is fixed and correspond to the repeated measure correlation coefficient (r = 0,0043).

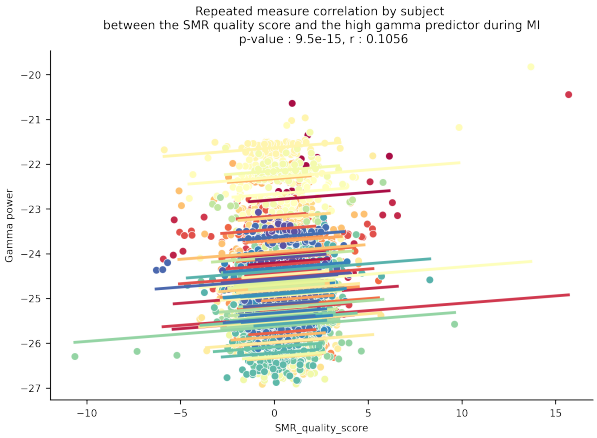

Figure 3: repeated measure correlation analysis between SMR quality score and the high gamma predictor power during the 40 trials of MI (n=85 subjects). The lines correspond to the regression for each subject where the slope is fixed and correspond to the repeated measure correlation coefficient (r = 0,1056).

monotonic relationships without assuming linearity, for possibly being more robust to outliers and for reflecting better neurophysiological dynamics [17].

- For relative beta power at rest, we assessed the multi repeated correlation for each subject [16] between performance during the run and the mean relative beta power during resting phases before the MI trials. Repeated measures correlation is a statistical technique for determining the common within-individual association for paired measures assessed for multiple individuals, in this case the mean performance and the relative beta power at rest. Note that the original paper [8] found correlations between beta power at rest and BCI performance across sessions while we assess it across trials for a single session.

- For high gamma predictor, we investigated the correlation between the SMR quality score and high gamma predictor during MI. The SMR quality score, as defined in the work of Grosse-Wentrup and colleagues [7], is the output of the LDA classifier (negative for left-hand MI and positive for right-hand MI) multiplied by the sign of the actual class. Consequently, correct trials yield a positive SMR quality score, whereas incorrect ones result in a negative score. It is essential to recognize that the magnitude of the LDA output (positive or negative) directly relates to the model's confidence in its classification, indicating that higher positive SMR quality scores denote accurate predictions made with high confidence, and higher negative scores denote incorrect predictions made with similar confidence. Correlations were conducted with repeated measures correlation per subject to appropriately analyse these relationships.

RESULTS

In the analysis of neurophysiological predictors extracted during the baseline phase, we investigated the correlation between each predictor and the subjects' performance during the entire session. Figure 1 illustrates the Spear-

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-068

man correlation coefficients with the linear regression being also added to illustrate the correlation. The most significant predictor of performance was found to be the SMR predictor, which exhibited a positive and significant Spearman correlation ($\rho = 0.49$, $p < 0.01$). The second most significant predictor was the variance of the peak frequency necessary to reconstruct the PSD, as defined in Equation 3. This predictor showed a significant negative Spearman correlation ($\rho = -0.33$, $p = 0.0029$). The third predictor evaluated was the number of peaks required to model the PSD during the baseline, demonstrating significant Spearman correlation ($\rho = 0.24$, $p = 0.028$). The PPfactor did not exhibit any significant correlation with performance ($\rho = 0.07$, $p = 0.50$).

The analyse of the correlation between the mean relative beta power at rest and the corresponding performance, depicted in Figure 2, reveals no significant relationship ($r = 0.0043$, $p = 0.94$). In contrast, the correlation between gamma predictor and SMR quality scores, as shown in Figure 3, is weak but very statistically significant correlation ($r = 0.1056$, $p = 9 \times 10^{-15}$).

DISCUSSION

The identification of robust and reliable neurophysiological predictors of BCI performance is a challenging and promising research issue. In this study, we successfully validated only four out of six proposed neurophysiological predictors using a large dataset comprising 85 subjects. Among these validated predictors, three were extracted during the baseline and demonstrated correlations with the mean session performance. Interestingly, the high-gamma predictor showed a correlation with the SMR quality score, which allows a trial-wise analysis to understand performance variations. In contrast, the predictive values of PPfactor and relative resting beta were not replicated in this dataset.Two important differences between our study and the initial study showing a correlation with relative beta [8] could explain these discrepancies. Firstly, that predictor was identified in stroke patients, a group often affected by fatigue [18]. Since relative beta power was used as a neurophysiological indicator of fatigue, it may have greater predictive value in this patient population rather than in the general population. Secondly, in their research, the repeated measure correlation was conducted across multiple sessions, while in our study, it was assessed across runs within a single session with only four runs per subject. Therefore, relative beta power may serve as a better indicator of overall fatigue rather than reflecting the evolution of fatigue within a session. Further analysis of this predictors will be necessary to confirm this relationship. Regarding the PPfactor, its lack of reproducibility cannot be attributed to differences in the experimental setup. Previous studies attempting to replicate this predictor only found significant correlations in one out of two datasets they tested [19]. Notably, when they applied a Laplacian filter around the C3 and C4 electrodes before extracting the PPfactor, this enhanced the correlation between this predictor and performance metrics as compared to using and C3 and C4 without Laplacian filters. Therefore, incorporating this filtering step into the PPfactor computation may be essential for enhancing the predictor's significance.
While previous studies have suggested that demographical information and personality traits [10, 15] may not be reliable predictors of perfomance, our study show the reliability and the reproducibility of most neurophysiological predictors.

CONCLUSION

In this paper, we compiled various studies that identify neurophysiological predictors of MI-BCI performance across and within different subjects. We noted that these predictors have predominantly been explored in small datasets and lacked widespread replication on independent data sets. Therefore, we sought to replicate the predictive value of six distinct indicators within a large (n=85), open source dataset. Of these, we managed to replicate the results of 4 of them: the SMR predictor, the number of peaks necessary for estimating the PSD during the baseline, the variability across time in the peak frequency within the mu band during baseline, and the high gamma predictor during the trial. However, the PPfactor and the relative beta power at rest were not successfully replicated. This study reinforces the significance of the SMR predictor as a robust indicator of BCI performance and highlights the critical need for replication of results reported in the BCI literature.
Future research needs to explore the influence of classification algorithm on predictors of BCI classification performance and validate that the predictors that we validate in this article are reliable predictors independently of the classifier used.

REFERENCES

[1] B. Blankertz *et al.*, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, 2010.

[2] P. Dreyer, A. Roc, L. Pillette, S. Rimbert, and F. Lotte, "A large EEG database with users' profile information for motor imagery brain-computer interface research," *Scientific Data*, 1 2023.

[3] B. Z. Allison and C. Neuper, "Could Anyone Use a BCI?" In *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*, ser. Human-Computer Interaction Series, D. S. Tan and A. Nijholt, Eds., London: Springer, 2010.

[4] C. Jeunet, B. N'Kaoua, and F. Lotte, "Chapter 1 - Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates," in *Progress in Brain Research*, ser. Brain-Computer Interfaces:

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-068

Lab Experiments to Real-World Applications, D. Coyle, Ed., Elsevier, 2016.

[5] W. F. Chaplin, O. P. John, and L. R. Goldberg, "Conceptions of states and traits: Dimensional attributes with ideals as prototypes," *Journal of Personality and Social Psychology*, 1988.

[6] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu, "The predictive role of pre-cue EEG rhythms on MI-based BCI classification performance," *Journal of Neuroscience Methods*, 2014.

[7] M. Grosse-Wentrup and B. Schölkopf, "High gamma-power predicts performance in sensorimotor-rhythm brain–computer interfaces," *Journal of Neural Engineering*, 2012.

[8] R. Foong *et al.*, "Assessment of the Efficacy of EEG-Based MI-BCI With Visual Feedback and EEG Correlates of Mental Fatigue for Upper-Limb Stroke Rehabilitation," *IEEE Transactions on Biomedical Engineering*, 2020.

[9] N. Leeuwis, A. Paas, and M. Alimardani, *Psychological and Cognitive Factors in Motor Imagery Brain Computer Interfaces*, 2021.

[10] C. Benaroch, C. Jeunet, and F. Lotte, "Are users' traits informative enough to predict/explain their mental-imagery based BCI performances ?," presented at the 8th Graz Brain-Computer Interface Conference 2019, 2019.

[11] E. Tzdaka, C. Benaroch, C. Jeunet, and F. Lotte, "Assessing The Relevance Of Neurophysiological Patterns To Predict Motor Imagery-based BCI Users' Performance," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020.

[12] C. Sannelli, C. Vidaurre, K.-R. Müller, and B. Blankertz, "A large scale screening study with a SMR-based BCI: Categorization of BCI users and differences in their SMR activity," *PLOS ONE*, 2019.

[13] M. Grosse-Wentrup, B. Schölkopf, and J. Hill, "Causal influence of gamma oscillations on the sensorimotor rhythm," *NeuroImage*, Multivariate Decoding and Brain Reading, 2011.

[14] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, "High Theta and Low Alpha Powers May Be Indicative of BCI-Illiteracy in Motor Imagery," *PLOS ONE*, 2013.

[15] L. Botrel and A. Kübler, "Reliable predictors of SMR BCI performance — Do they exist?" In *2018 6th International Conference on Brain-Computer Interface (BCI)*, 2018.

[16] J. Z. Bakdash and L. R. Marusich, "Repeated Measures Correlation," *Frontiers in Psychology*, 2017.

[17] G. Rousselet and C. Pernet, "Improving standards in brain-behavior correlation analyses," *Frontiers in Human Neuroscience*, 2012.

[18] M. Acciarresi, J. Bogousslavsky, and M. Paciaroni, "Post-stroke fatigue: Epidemiology, clinical characteristics and treatment," *European Neurology*, 2014.

[19] C. Benaroch, "Contribution to the understanding of mental task BCI performances using predictive computational models," These de doctorat, Bordeaux, 2021.

Published by
Verlag der Technischen Universität Graz