This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

MEAN$_{SP}$: How Many Channels Are Needed to Predict the Performance of a SMR-Based BCI?

Tania Jorajuría, Vadim V. Nikulin, Nikolai Kapralov, Marisol Gómez and Carmen Vidaurre

***Abstract*—Predicting whether a particular individual would reach an adequate control of a Brain-Computer Interface (BCI) has many practical advantages. On the one hand, participants with low predicted performance could be trained with specifically designed sessions and avoid frustrating experiments; on the other hand, planning time and resources would be more efficient; and finally, the variables related to an accurate prediction could be manipulated to improve the prospective BCI performance. To this end, several predictors have been proposed in the literature, most of them based on the power estimation of EEG signals at the specific frequency bands. Many of these studies evaluate their predictors in relatively small datasets and/or using a relatively high number of channels. In this manuscript, we propose a novel predictor called MEAN$_{SP}$ to predict the performance of participants using BCIs that are based on the modulation of sensorimotor rhythms. This novel predictor has been positively evaluated using only 2, 3, 4 or 5 channels. MEAN$_{SP}$ has shown to perform as well as or better than other state-of-the-art predictors. The best sets of different number of channels are also provided, which have been tested in two different settings to prove their robustness. The proposed predictor has been successfully evaluated using two large-scale datasets containing 150 and 80 participants, respectively. We also discuss predictor thresholds for users to expect good performance in feedback experiments and show the advantages in comparison to a competing algorithm.**

***Index Terms*—Brain–computer interface (BCI), Sensorimotor rhythms (SMRs), Cross-frequency coupling, Performance predictor, BCI inefficiency.**

I. Introduction

MOTOR Imagery (MI) is a mental process where an individual simulates a motor action [1], [2]. This process activates the primary sensorimotor area, similar to what occurs during motor preparation of a real movement [3], modulating the power of specific brain oscillations or sensorimotor rhythms (SMRs). These SMRs can be captured by electroencephalography (EEG) and decoded with a MI-based Brain-Computer Interface (BCI) [4]–[6], to subsequently execute some pre-programmed actions. Nevertheless, not every participant is able to control a BCI, a phenomenon known as BCI inefficiency [7]–[10].

BCI inefficiency [11] is one of the reasons why it would be desirable to know beforehand whether a specific individual would be able to control a BCI. Such information could support decisions related to the amount of research resources employed in a study, the time planned to perform experiments and the number of participants to be engaged. However, predicting BCI performance could also be useful to categorize participants beforehand. For example, research questions might advise the engaging of particular participant types, as for instance those exhibiting a low SMR peak [12].

The literature shows several ways to relate BCI performance with subject-specific metrics. In [13], a review about performance variation in MI-based BCIs was presented, focusing on aspects such as personal information of participants [14], [15], their psychological state [16]–[18], and physiological [19]–[22] or anatomical [23], [24] variables.

An example is the work of [21], where neurophysiological differences between participants were studied using recordings of non-task related states. There, a performance potential factor (PPfactor) that combines power from four different frequency bands in $[4-70]$ Hz range was proposed. PPfactor is a predictor calculated using the relative power level of channels C3 and C4 after re-referencing them with a 64-channel common average reference. This factor achieved a correlation of r = 0.48 with hand MI BCI performance obtained from 52 subjects. After including 9 more participants from a separate dataset, the authors showed that this correlation increased to r = 0.59.

The authors in [20], on the other hand, proposed a SMR-based neurophysiological predictor based on the signal-to-noise ratio (SNR) of the μ and β bands, using C3 and C4 Laplacian channels (amounting a total of 10 channels) from 2 minutes of resting-state EEG data in relax condition with eyes open. They obtained a correlation of r = 0.53 between the predictor value and BCI feedback performance applied to 80 subjects. This predictor was later tested in another dataset comprising 151 subjects [25], using 2.5 minutes of resting-state data, obtaining again a correlation of r = 0.53. Besides, the authors fitted a linear regression model with the dataset analyzed in [20] that estimates the BCI performance of new subjects from their SMR-based predictor value. This regression model

Tania Jorajuría and Marisol Gómez were with the Department of Statistics, Computer Science and Mathematics, Universidad Pública de Navarra, Pamplona, Spain (e-mail: tania.jorajuria@unavarra.es; marisol@unavarra.es).

Vadim V. Nikulin and Nikolai Kapralov were with the Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany (e-mail: nikulin@cbs.mpg.de; kapralov@cbs.mpg.de).

Nikolai Kapralov was also with the International Max Planck Research School NeuroCom, Leipzig, Germany.

Carmen Vidaurre was with Ikerbasque, Basque Foundation for Science, Bilbao, Spain; BCBL, Basque Center for Cognition Brain and Language, San Sebastián, Spain; BIFOLD, Berlin Institute for the Foundations of Learning and Data, Berlin, Germany; TECNALIA, Basque Research and Technology Alliance (BRTA), San Sebastián, Spain (e-mail: cvidaurre@bcbl.eu).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

was applied to the subjects studied in [25], obtaining a correlation between real and estimated BCI feedback accuracy of r = 0.53.

Suk and colleagues analyzed in [26] a probabilistic framework called Bayesian Spatio-Spectral Filter Optimization. They extracted subject-specific spectral characteristics to cluster subjects into groups related to their performance. This grouping was used to build a linear regression model to predict BCI performance. They used 2 minutes of resting-state EEG data from 3 Laplacian channels (C3, Cz, C4, involving 13 mounted channels) from the same dataset as in [20], obtaining a correlation coefficient of r = 0.58.

In another approach proposed by Zhang et al. [27], the spectral entropy was presented as SMR-based BCI performance predictor. The authors found that this predictor calculated from 2 minutes of resting-state EEG data of channel C3 in eyes-closed condition had a correlation of r = 0.65 with the offline performance of hand MI BCI. It was evaluated with 66 sessions, composed of 40 independent subjects from whom 26 returned in a posterior session. Besides, they also predicted inter-session performance for these 26 participants who performed 2 sessions, achieving an average classification accuracy up to 89%.

Other predictors have been defined using more complex measures such as connectivity. For example, in [28] it is described that the imaginary part of coherency, *ImCoh* [29], over the sensorimotor cortices in the μ band was positively and significantly correlated with online BCI performance. *ImCoh* is a connectivity metric robust against zero-lag interactions including those caused by the effect of volume conduction. It was computed using 61 channels from pre-stimulus data of offline MI recordings. The authors suggested the up-regulation of (undirected) functional connectivity as a possible tool to increase online BCI performance.

Also in relation to connectivity, Lee et al. [30] calculated the directed coupling strengths in resting-state between brain regions, using a dynamic causal model implemented from 56 EEG channels. They observed significant differences between low- and high-MI performance groups. They showed that the connectivity strength between the supplementary motor area and the right dorsolateral prefrontal cortex was positively correlated with MI-based BCI performance (r = 0.54 in session 1; r = 0.42 in session 2). In their paper, MI performance was also predicted with a linear regression model based on this connectivity (r-squared = 0.31) with data from 54 subjects.

Contrary to other authors who only use one resting-state condition (eyes-open or eyes-closed), Kwon et al. [31] suggested that employing power estimates from both brain states may lead to a more robust predictor of MI-BCI performance. In their analyses with 15 subjects, they obtained a correlation of r = 0.71 between this modified predictor and MI-based BCI online performance, using only two channels. However, these two channels were selected among C3, C4 and Cz depending on the pair of classes chosen for the online MI task. Thus, unless the pair of tasks were fixed for every participant, task-related data would be necessary to estimate the predictor in new users.

In summary, a number of different predictors described in BCI literature are based on studies that use a relatively high number of channels or a relatively low number of participants, and even task information. More channels translates into longer preparation for fixing the electrodes, tiredness of the subjects and also, more expensive equipment. On the other hand, using task information to obtain a predictor value implies the need of performing BCI experiments beforehand, which is a strong requirement because predictors are usually studied to anticipate the performance. Finally, correlation frameworks as the ones presented in predictor studies need a high number of participants to deliver robust and generalizable results. Depending on the expected effect size, the number of participants needed to obtain a significant result can largely exceed the rule of thumb of "10 times the number of independent variables", suggested by Roscoe in [32].

As previously discussed, resting-state power of different electrodes in different frequency bands is a very common choice to estimate predictors. The hypothesis behind this choice is that people who exhibit high power peaks computed over sensorimotor electrodes are more likely to modulate them according to the task. However, power computed at sensor level might originate from different regions, and thus, it might not have exclusive sensorimotor origin. In our paper we exploit that SMRs are non-sinusoidal due to the synchrony between μ- and β-rhythms [33], [34]. Other dominant EEG rhythms such as those of occipital origin also exhibit α-β synchronization (note that the α frequency range coincides with the μ band, but it is only referred to as μ when its origin is sensorimotor). However, although α power might be captured by sensors over the sensorimotor cortices, the power of the occipital β oscillations is too low to reach these channels. Thus, extracted cross-frequency synchronized sources are likely to have sensorimotor origin [35]. Therefore, our hypothesis is that by finding μ-β synchronized sources from sensorimotor electrodes, we can also reduce the influence of other unrelated rhythms in the computation of EEG signatures, and thus obtain a robust predictor with a low number of channels.

Grounded on the aforementioned, we propose a novel predictor named MEAN_{SP} for SMR-based BCI performance. Our predictor is based on the Nonlinear Interaction Decomposition (NID) [36] method to find spatial filters that decompose multichannel EEG into non-linearly coupled pairs of sources. Here, we used NID to obtain pairs of sources that are 1:2 phase-coupled between μ and β frequencies; that is, we aimed to extract pairs of sources of sensorimotor origin, as their μ- and β-rhythms are sought to be synchronized. After this, the SMR predictor from [20] is performed and combined with the Phase-Locking Value (PLV) [34], which measures the strength of this synchronization.

In this work we show that the proposed predictor achieved good results with very few channels. We also

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

provide sets of standard electrode locations for 2, 3, 4 and 5 channels, obtained from two different initial montages to demonstrate the robustness of the selection. We evaluated $\mathrm{MEAN_{SP}}$ using 2 minutes of resting-state EEG data. The correlation between $\mathrm{MEAN_{SP}}$ and the SMR-based BCI performance in a MI paradigm was used for its validation. A total of 230 participants belonging to two large datasets, with 150 and 80 subjects, respectively, were employed to test its robustness. Besides, we also suggest threshold values for $\mathrm{MEAN_{SP}}$ that would allow researchers assessing the BCI performance potential of a new subject.

We benchmarked our predictor against the SMR predictor by Blankertz et al. [20], the PPfactor by Ahn et al. [21] and the spectral entropy predictor presented by Zhang et al. [27]. These predictors were selected for benchmarking because they used a relatively low number of channels and were based on power-related features. The results obtained show that in all the evaluated settings, $\mathrm{MEAN_{SP}}$ predictor achieved similar or significantly better results than the benchmarked methods with a low number of channels, showing that a setting composed of only C3 and C4 is sufficient to obtain state-of-the-art results. In conclusion, we present a novel predictor to estimate SMR-based BCI performance of new subjects that can be quickly and robustly computed because it requires very few spatially unfiltered channels.

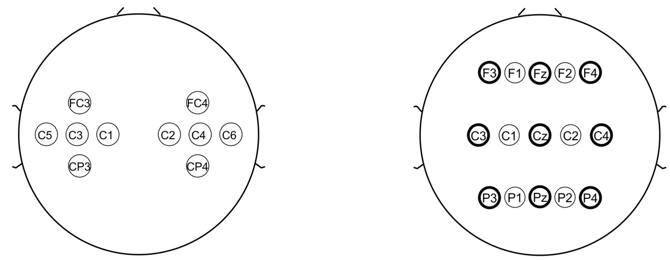
This paper is organized as follows: in section II we detail the three state-of-the-art predictors used as benchmark. Here, we also introduce our proposed predictor, $\mathrm{MEAN_{SP}}$, together with the details about the method we used to select channels from an initial montage setting, and how NID was applied. In section III, the experimental data are described, together with the evaluation settings. Section IV details the scores we used to validate and compare our predictor with baselines, and in section V we explain the statistical analysis done. In section VI, obtained results are shown, which are discussed in section VII. Finally, a brief conclusion of the study is presented in section VIII.

II. Methods

Our proposed predictor for SMR-based BCI performance is inspired by the SMR predictor by Blankertz et al. [20]. We chose this predictor as baseline because it is easy to compute and its robustness was demonstrated with two independent and large datasets. In addition, we also compared $\mathrm{MEAN_{SP}}$ with two other different predictors, since they can be computed using very few channels, are also based on band power estimates and they showed to outperform Blankertz et al. predictor in their respective datasets. In this section we explain the details to calculate all analyzed predictors. The BBCI Toolbox [37] of MATLAB® was used to process EEG data and calculate predictor values. We also used the EEGLAB [38] Toolbox of MATLAB® to generate scalp plots.

A. Benchmarking

We compared our proposed predictor, $\mathrm{MEAN_{SP}}$, with the following state-of-the-art estimators.



(a) (b)

Fig. 1: Scalp plot of the electrode configurations used in: a) SMR predictor from Blankertz to calculate C3 and C4 Laplacian channels, b) each initial montage setting for our $\mathrm{MEAN_{SP}}$ predictor ($M = 15$ channels: all plotted electrodes; $M = 9$ channels: bold-circled electrodes).

1) SMR predictor (SMR): Blankertz et al. defined for the first time a SMR predictor in [20]. It is based on the SNR estimation of the sensorimotor rhythm, using two small Laplacian channels located over C3 and C4 (amounting a total of 10 electrodes, see Fig.1a). For more information, please refer to section I-A of Supp. Material.

2) Performance potential factor (PPfactor): Ahn et al. proposed a predictor named performance potential factor (PPfactor) [21] that combines power from different bands. Specifically, it computes spectral power of channels C3 and C4, and then band power is calculated in θ (4-8 Hz), α (8-13 Hz), β (13-30 Hz) and γ (30-70 Hz) frequency bands. In this last frequency range we used the interval $[30-50]$ Hz because our data had been downsampled to 100 Hz, and thus previously filtered to remove frequencies over 50 Hz. For more information, see section I-B of Supp. Material.

3) Spectral entropy predictor (SH): Zhang et al. also proposed a predictor for SMR-based BCI performance based on the spectral entropy [27]. We followed the procedure explained in [27] to implement this predictor. For more details, refer to section I-C of Supp. Material.

B. *Our proposed approach:* $\mathrm{MEAN_{SP}}$

In this section we explain the details of our proposed predictor (see Fig.2), named $\mathrm{MEAN_{SP}}$, since it is calculated by averaging the SMR predictor from Blankertz and PLV values.

a) EEG channels: The number of required electrodes is directly related to the required EEG montage time. Hence, one of the main goals of this study was to find the smallest possible set that would still lead to at least as good performance as other published predictors. To achieve that, we studied different settings of number and locations of channels, C (see section III-B). A Least Absolute Shrinkage and Selection Operator (LASSO) regression [39] was employed to determine the position of a given number of channels C that would maximize the correlation between BCI performance and the predictor in each analyzed setting. It is known that LASSO regression does not cope well with correlated predictors [40], so its

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

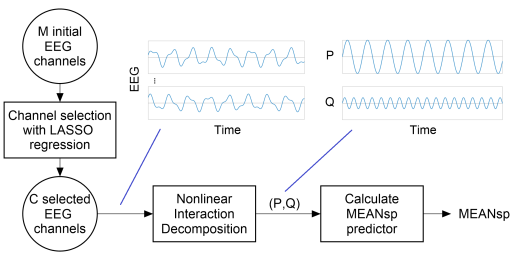


Fig. 2: Pipeline of MEAN_{SP} predictor. M and C depend on the evaluated setting. Note that P and Q components illustrated here are narrow band, but the actual input to calculate MEAN_{SP} predictor is broadband.

results might depend on the initially available channels. Therefore, to avoid overfitting and obtain robust and stable results, we analyzed two different initial EEG montages M with different amount and location of channels (see section III-B), to which we applied LASSO regression.

LASSO performs a regression analysis including variable selection to improve the correctness and interpretability of the statistical model. The LASSO regression was computed between the SMR predictor described in II-A1 of each monopolar (not spatially filtered) channel and the BCI performance, thereby also reducing the risk of overfitting because none of the channels were pre-processed as in MEAN_{SP}. The code corresponding to LASSO regression was part of the SpaSM Toolbox [41] of MATLAB®, which is a variant of the LARS algorithm [42] with elements from [43].

In particular, the procedure was as follows: we first calculated the SMR predictor of each of the M channels in the initial montage setting, for each of the N subjects in the selected dataset (see section IV). Then, for each analyzed setting, we performed a leave-one-subject-out (LOSO) procedure to find a robust set of channels; in each iteration, the previously calculated matrix of SMR predictors, with dimensions $[(N-1)\times M]$ was column-wise normalized and the performance vector $[(N-1)\times 1]$ was centered. The used toolbox allows directly selecting the number of features to be kept. Thus, we specified the desired number of non-zero variables that should be returned by LASSO as the number of channels, C, that should be retained in each evaluated setting. Then LASSO regression was applied. Finally, a $[N\times C]$ matrix with the channels selected in each iteration of the LOSO procedure was obtained. The electrodes that were more often selected were chosen to form the final channel set.

b) Spatial filter: In this paper, we propose the use of the Nonlinear Interaction Decomposition (NID) [36] method to spatially filter a set C of selected channels. We used the implementation of the NID Toolbox [36] in MATLAB® to obtain the necessary spatial filters.

NID is a method for extracting non-linearly coupled neural sources oscillating in two different frequency bands from multichannel recordings of brain activity (i.e. non-linear cross-frequency interaction). The main idea behind NID is that the linear mixture of two narrow band oscillations, centered at f_n and f_m, respectively, will follow a non-Gaussian distribution if they are non-linearly coupled (alternatively, they will follow an approximately normal distribution in case they are independent). Therefore, by maximizing non-Gaussianity, NID extracts cross-frequency coupled sources and returns spatial filters corresponding to the frequency bands of interest.

In the case of sensorimotor rhythm, μ and β components are phase-to-phase synchronized [33], [34]. This non-linear interaction can be extracted with NID as pairs of coupled sources (P, Q). To achieve this goal, first Spatio-Spectral Decomposition (SSD) [44] is applied to the multichannel data. SSD is a method that calculates the spatial filters that maximize the SNR of extracted oscillatory sources at the frequency band of interest. It reduces to a generalized eigenvalue decomposition and thus it is very fast to compute. For NID in particular, SSD is applied separately at two narrow bands respectively centered at f_n and f_m to extract neuronal oscillations in each band, resulting in $\mathbf{W_{SSD\_P}}$ and $\mathbf{W_{SSD\_Q}}$ matrices of SSD spatial filters. Then, the two matrices of SSD components at f_n and f_m are put together to form an augmented matrix, on which a non-Gaussianity maximization decomposition (NGMD) is applied. This method finds a subspace, given by $\mathbf{W_{NGMD\_P}}$ and $\mathbf{W_{NGMD\_Q}}$ spatial-filter matrices, that maximizes the non-Gaussianity of the linear mixtures of the estimated SSD oscillatory sources. As a result of this process, cross-frequency coupled oscillations are separated. For more information about the use of the NID method, we refer the reader to [36].

In this paper, NID was applied separately to each subject, calculating their specific SSD and NGMD spatial filters, and filtering the continuous broadband EEG data with them to obtain pairs of coupled sources P and Q. We only kept the first pair of sources to compute our novel predictor MEAN_{SP}, which indicated the pair of sources with the greatest synchronization index.

c) The MEAN_{SP} *predictor:* In order to obtain MEAN_{SP}, we computed the phase-phase coupling between μ and β components in SMR. We used synchronization index between the coupled pair of sources (P, Q) returned by NID. This index consists in measuring the Phase-Locking Value (PLV) [34] between the selected NID source pairs. PLV is defined as $|< e^{j\psi_{n,m}(t)} >|$, where j is the imaginary unit number and $< \cdot >$ represents the average over time samples. The term $\psi_{n,m}(t) = m\phi_n(t) - n\phi_m(t)$ stands for the difference of the instantaneous phases ϕ_n and ϕ_m of two oscillations with frequencies f_n and $f_m = (m/n)f_n$, $n, m \in \mathbb{N}$. These oscillations are then said to be $n : m$ phase-coupled if $|\ \psi_{n,m}(t)\ | < const$. In this work we were interested in $1 : 2$ phase-coupling between μ and β, so we selected $f_n = 11$ Hz and $f_m = 22$ Hz values as the respective frequency centers of these bands. Note that NID does not directly maximize PLV in phase-phase coupled sources, but the non-Gaussianity of their linear mixture; nonetheless, as a consequence, the PLV is also indirectly maximized as shown with simulations and analytically

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

[36].

Apart from the aforementioned PLV obtained with NID, the final predictor value also included the SNR estimation given by [20] (see section II-A1), which was shown to be robust in two independent and large datasets [20], [25].

Thus, we used the SMR predictor computed on the first pair of sources obtained with NID, together with the PLV value of that pair. The final predictor value $\mathrm{MEAN_{SP}}$ was then the average between SMR and PLV. Note here that by definition, PLV values range between 0 and 1, but the SMR predictor is not bounded. Therefore, SMR values need to be normalized before averaging (we provide this information in the results section, so that researchers can estimate $\mathrm{MEAN_{SP}}$ in new participants). Hence, the final $\mathrm{MEAN_{SP}}$ estimate carries information about the possible facility for power modulation of the participant in the SMR predictor and also an indication of the sensorimotor origin of the SMR (through the amount of synchrony that exists between μ and β rhythms).

III. Experimental Data

A. Data Description

We used two large datasets, which are described below.

a) Dataset 1: This dataset was presented in [25], where the SMR predictor proposed in [20] was re-tested in a large number of subjects. Specifically, 168 naive BCI subjects participated in the study; among them, 17 participants were excluded for different reasons, leading to a total of 151 analyzed subjects. We further had to discard one of these subjects because their corresponding resting-state data was not available, amounting a total of 150 subjects. As explained in [25], brain activity was recorded from 64 electrodes with 1 kHz sampling frequency, referenced to the left mastoid and grounded to the forehead. The data were later filtered under 50 Hz and downsampled to 100 Hz. During the BCI session, resting-state EEG data was recorded in eyes closed condition for 15 seconds and eyes open for another 15 seconds, repeating this cycle 10 times in total. Afterwards, a co-adaptive MI-BCI [11] was employed to provide feedback while the subjects performed imaginary movements of right hand, left hand or feet during four runs. The provided online feedback was continuous. We refer reader to [25] for more details about their experimental setup.

b) Dataset 2: This dataset was described and used in the study by Blankertz et al. [20] to present their SMR predictor, where 80 healthy novel BCI users took part. EEG data were recorded from 119 electrodes in an extended 10-20 system, with reference at nasion, 1 kHz sampling rate and a band-pass filter from 0.05 Hz to 200 Hz. The data were later filtered under 50 Hz and downsampled to 100 Hz. During the session, EEG artifacts were first recorded (eye movements, blinking, ...). Then, ten periods of 15 seconds were also recorded, using “relax with eyes open” and “relax with eyes closed” alternating tasks. Afterwards, subjects performed some MI tasks (left hand, right hand, and right foot or feet movement, this last one according to the participant’s preference) that were used to calibrate a MI-based BCI. Later a BCI feedback session took place. It consisted of three runs of 100 trials each, where the MI continuous classification result of left hand, right hand or foot classes were presented to the participants. For more details about the experimental setup regarding this dataset, please refer to [20].

The data employed to estimate $\mathrm{MEAN_{SP}}$ comprised EEG signals recorded in resting-state and eyes open condition from each subject in both datasets. Since in these two large datasets less than half of the participants improved their performance in the last run with respect to the first one, the BCI performance of all feedback runs was averaged to obtain the final BCI accuracy for each participant. Note that had the majority of participants exhibited performance improvement from beginning to end of the session, one could also have considered using the last run or last two runs to compute the final BCI accuracy. Note that in these two datasets, the feedback performance of each participant was obtained with the pair of classes out of three possible that achieved the best cross-validation accuracy in a calibration (offline) recording performed previously to the feedback runs. We used 120 seconds of resting-state eyes-open data to obtain the predictors because for some participants the recording was shorter than 150 s. $\mathrm{MEAN_{SP}}$ and also the other three benchmarked state-of-the-art predictors were then computed. In the case of the spectral entropy (SH) predictor [27], the authors suggested its computation from resting-state EEG data in the eyes closed condition. However, to calculate it we used data in the eyes open condition since the correlations obtained were higher.

B. Evaluation Settings

As explained in section II-B, we analyzed different settings with different number of channels, C, whose location was selected by LASSO regression from an initial EEG montage containing a total of M electrodes.

a) Initial montage: It is known that LASSO regression is sensitive to correlated predictors [40]. In order to at least partially reduce correlations between channels, we selected initial montages that do not contain a large number of closely adjacent electrodes. This correlation reduction has the potential to enhance the effective information of the channel set so that NID can deliver better decomposition results. In particular, we studied two different initial montages (see Fig.1b), with a total number of electrodes, M, equal to 9 and 15, respectively. Having two initial sets also allowed us testing the stability of the channel selection performed by LASSO.

b) Number of selected channels: For each of the two initial montages, C channels were selected by LASSO among the M total number of electrodes. Our goal was to obtain number of channels C lower than 10, which is the number of electrodes used in [20]. Also, C should be as small as possible but should be larger than 1 to be able to find NID sources. Thus, we investigated results obtained

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

with 2, 3, 4 and 5 channels. Nevertheless, results with 10 channels selected by LASSO using the 15-channel montage setting are also provided for further analysis, as well as the result of the SMR predictor with two raw channels.

IV. Validation

In order to validate and compare our proposed predictor with the benchmarked ones, we calculated the Pearson correlation between predictors and BCI feedback accuracy, as in [20], [25]. In the case of $\mathrm{MEAN_{SP}}$, this correlation was computed with the predictor values obtained in each evaluated setting.

The larger dataset with 150 participants (Dataset 1) was selected for performing channel selection with LASSO in each evaluated setting. Also, the range limits for SMR normalization were taken from this dataset (see section II-B). Then, correlations were computed with Dataset 2, which contains unseen data of 80 participants. We chose to also show correlation results obtained with Dataset 1, for the sake of comparison with those from Dataset 2.

Besides, in order to assess the ability of the predictors to classify a new subject in terms of their online BCI performance potential, we analyzed the precision-recall curves. Precision in this context is the percentage of participants correctly classified from those assigned by the classifier to the good performers group. Recall is the correctly classified subjects from those who actually are good performers. There is a trade-off between them, since when the precision increases, the recall decreases. In the case of $\mathrm{MEAN_{SP}}$, this was done for the best set with the minimal number of channels. We compared these results to the best benchmarked predictor, using both datasets. The BCI accuracy threshold set to distinguish between good and poor BCI performance was 70%. This probability was experimentally established to distinguish between random and voluntary control of a two class BCI [45]. For specific precision values (80%, 85% and 90%), we found the corresponding predictor thresholds and studied the recall values obtained.

V. Statistical Analysis

The Meng's z-test for dependent samples with overlapping pairs of variables was used to compare correlation coefficients by pairs [46]. The total number of compared pairs was m=3: $\mathrm{MEAN_{SP}}$ vs. SMR, $\mathrm{MEAN_{SP}}$ vs. PPfactor and $\mathrm{MEAN_{SP}}$ vs. SH. The correlation values of $\mathrm{MEAN_{SP}}$ were obtained for each montage and number of channels. Significance tests were one-tailed, being the alternative hypothesis that the highest correlation was significantly higher than the other one.

Besides, for Dataset 2 we also compared correlations of $\mathrm{MEAN_{SP}}$ and BCI performance between the different evaluated settings. In particular, for each number of channels, we compared correlations achieved with the two analyzed montage settings (m=2), and vice-versa; for each initial montage setting, we compared correlations obtained with each number of channels setting (m=6), by pairs. In this case significance tests were two-tailed, to check for differences in correlation values.

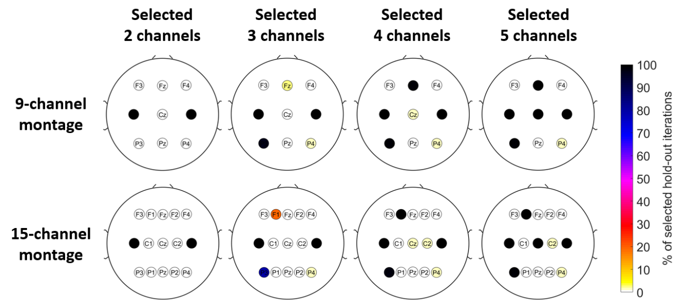


Fig. 3: Scalp plots showing the number of iterations (%) in which each channel was selected during the LOSO procedure for LASSO regression in each analyzed setting (i.e. number of channels and initial montage).

Finally, in order to account for multiple comparisons ($m \geq 3$), obtained p-values were corrected with the Holm-Bonferroni method. This correction procedure has more power than Bonferroni correction and is also less conservative [47].

VI. Results

In this section, we show the results achieved with our proposed $\mathrm{MEAN_{SP}}$ predictor in each of the studied settings (i.e. number of channels and initial montage settings), as well as with all three benchmarked predictors.

A. Best channels selection

A LOSO procedure with LASSO regression was performed to select the best channel combination for each number of channels, in each initial montage setting. In Fig.3, the number of iterations (in percentage) in which each electrode was individually selected during cross-validation is depicted.

In Tables I and II found in the Supp. Material, we also show the percentage of iterations in which each electrode was chosen by LASSO regression during LOSO procedure for each number of channels and both the 9-ch and the 15-ch montage settings, respectively. The electrodes that were more often chosen by LASSO regression were then selected to compose the best channel combination for each number of channels. They are marked in bold in Tables I and II of the Supp. Material.

It can be seen that for 2 and 3 channels, the selected electrodes are the same for both initial montage settings. Regarding the selected channels for groups of 4 and 5 electrodes, they differ only in one electrode.

B. Normalization limits and $\mathrm{MEAN_{SP}}$ computation steps

As aforementioned, SMR values need to be normalized before computing the $\mathrm{MEAN_{SP}}$ predictor. In Table I we provide the employed range limits for each analyzed setting, taken from Dataset 1.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

TABLE I: Range limits taken from Dataset 1 to normalize SMR values before computing the MEAN_{SP} predictor, for each analyzed setting.

SMR ranges	MIN		MAX	
	9-ch	15-ch	9-ch	15-ch
2-ch	1.09		18.71	
3-ch	1.44		19.12	
4-ch	0.84	0.95	20.23	19.58
5-ch	0.85	0.98	20.76	20.65

The information presented in Table I can be used by researchers to compute MEAN_{SP} in new data. Here we enumerate the steps necessary to obtain the predictor.

1) Select one channel combination from Fig.3.
2) Run NID and obtain as outputs the spatial filters of P and Q and also PLV.
3) Take the resting-state EEG broadband data and spatially filter them with the NID filters.
4) Calculate the SMR predictor of the two resulting NID sources and normalize it using the limits presented in Table I.
5) Average the normalized SMR predictor and the PLV to obtain the MEAN_{SP} value.

The NID code can be found in [48] and the SMR code is described in [20].

C. Validation with correlations between predictors and BCI performances

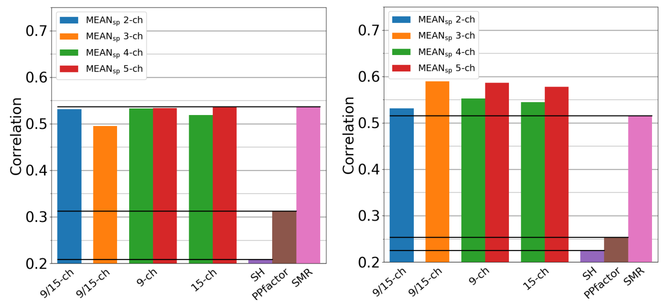
Here, we present the correlation results obtained with our MEAN_{SP} predictor in each analyzed setting and also with the three benchmarked predictors.

In Fig.4 we show the obtained Pearson correlations between predictors and BCI feedback performances. For each channel combination (see section VI-A) and initial montage (9 or 15 channels), we calculated MEAN_{SP} for each subject. Then, we obtained a correlation value of the predictor with BCI performance. This was done for each of the two studied datasets, and correlation values were compared against the ones obtained with the benchmarked predictors.

As shown in Fig.4, the specific correlations obtained with MEAN_{SP} in the 9-ch montage setting were 0.53, 0.50, 0.53 and 0.53, calculated with Dataset 1, and 0.53, 0.59, 0.55 and 0.59, in Dataset 2, for 2-, 3-, 4- and 5-channels, respectively.

On the other hand, correlation results obtained in the 15-ch montage setting for MEAN_{SP} were 0.53, 0.50, 0.52 and 0.54 with Dataset 1, and 0.53, 0.59, 0.54 and 0.58 with Dataset 2, using 2-, 3-, 4- and 5-channels, respectively. Note that the correlation values obtained with the 9- and 15-ch montages in the 2- and 3-channel settings were identical when calculated over the same dataset, since the selected channels in these two cases were the same.

The previous outcomes were obtained using fixed frequency centers of the μ and β bands. However, when the recorded data of each participant is sufficiently clean, it is also possible to define subject-specific frequency



(a) Dataset 1. (b) Dataset 2.

Fig. 4: Correlations of predictors and BCI performances in the two analyzed datasets: a) Dataset 1 (150 subjects), b) Dataset 2 (80 subjects). Predictors are MEAN_{SP} in each number of channels and montage setting, and SMR, PPfactor and SH.

centers for both bands. These results are presented in Table VI of Supp. Material. Our analyses showed slight differences between both procedures, which however were non-significant.

Regarding the results obtained with benchmarked predictors, SMR achieved a correlation value of 0.54 and 0.52, PPfactor 0.31 and 0.25, and SH a correlation coefficient of 0.21 and 0.23, for Dataset 1 and Dataset 2, respectively.

As aforementioned, we compared the correlation coefficients calculated with MEAN_{SP} in each analyzed setting to those obtained with each benchmarked predictor. The results are presented in Tables III and IV in the Supp. Material, which were obtained with the Meng's z-test by pairs and later corrected with the Holm-Bonferroni method. Here we want to remark that the correction did not affect the final results.

The correlation values achieved by SMR with Dataset 1 were quantitatively higher than with MEAN_{SP}. On the contrary, MEAN_{SP} obtained greater correlation coefficients than SMR with Dataset 2. However, in none of these two cases one predictor was significantly better than the other.

Regarding PPfactor and SH predictors, they both achieved significantly worse results than MEAN_{SP} in every evaluated setting, and for both datasets.

As aforementioned, we also compared correlations calculated in Dataset 2 with MEAN_{SP} between different settings. On the one hand, for each number of channels, we compared correlation values obtained with the two analyzed montage settings. Since the channels selected and hence the calculated correlations in the 2- and 3-channels settings were the same for the two montages, these two cases were not compared. Regarding the 4- and 5-channels settings, correlation coefficients were slightly higher for the initial montage of 9-ch than for 15-ch. However, no significant differences were found in any of these two cases ($p = 0.7675$ and $p = 0.7285$ for 4- and 5-channel settings, respectively).

On the other hand, for each montage setting we compared correlations obtained with different number of chan-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

nels, using the Meng's z-test by pairs. The results are shown in Table V of the Supp. Material. They were corrected for multicomparisons with the Holm-Bonferroni method. As seen, none of the differences were found significant. Again, the correction did not affect final results.

Besides, we also investigated the correlation obtained after selecting 10 channels with LASSO, starting from the initial 15-ch montage, since 10 channels were also necessary in the SMR predictor. The obtained correlation coefficients with $\mathrm{MEAN_{SP}}$ were 0.53 for Dataset 1, and 0.55 for Dataset 2 (respectively, SMR achieved 0.54 and 0.52). In order to find out whether results of $\mathrm{MEAN_{SP}}$ and SMR were significantly different, we performed two two-tailed Meng's z-test, one for each dataset. The result was that there were not significant differences between the correlations ($p = 0.9506$ in Dataset 1 and $p = 0.6783$ in Dataset 2).

Finally, and for the completion of the results, we also calculated the correlation between SMR predictor and BCI performance, with C3 and C4 raw channels (i.e. without applying any spatial filtering). The resulting correlation values were 0.41 with Dataset 1 and 0.22 with Dataset 2, which were found to be significantly worse than $\mathrm{MEAN_{SP}}$ with 2 channels ($p = 0.0361$ in Dataset 1 and $p = 0.0011$ in Dataset 2) after performing a two-tailed Meng's z-test for each dataset.

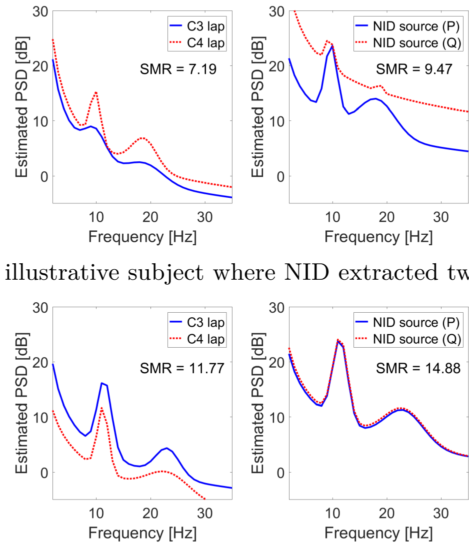
D. Analysis of spatial filters

Here we present results exploring the differences between Laplacian and NID filters as well as the results of the NID optimization.

a) Laplacian vs. NID spatial filters: We explored how NID helps in the enhancement of the SNR in μ and β frequency bands, as well as the μ-β synchronization of the sources found by the algorithm. In Fig.5 we show the estimated PSDs of two illustrative subjects. On the one hand, for C3 and C4 Laplacian channels. On the other hand, for the coupled pair of sources (P, Q) extracted with NID in the 2-channel setting. When computing SMR predictor from C3 and C4 Laplacian channels, the result is 7.19 and 11.77 for these two subjects, whereas it becomes 9.47 and 14.88, respectively, when calculating it from the sources returned by NID.

The SMR predictor values of Laplacian C3 and C4 channels were investigated for the 230 participants composing Datasets 1 and 2. The mean and standard error of the SMR using Laplacian channels was 7.4 ± 0.2, and the same for NID sources extracted from two electrodes was 7.8 ± 0.2. A non-parametric Wilcoxon signed rank test revealed a significant difference between them ($p = 0.0068$).

As expected, the PLV computed between P and Q sources extracted with NID was also significantly higher (using the same test as above) than that computed using C3 and C4 Laplacian channels, with a p-value $<< 0.0001$. Mean and standard error were 0.046 ± 0.001 for Laplacian channels and 0.091 ± 0.006 for NID sources.



(a) An illustrative subject where NID extracted two sources.

(b) An exemplary subject where NID extracted one source.

Fig. 5: PSDs estimated from C3 and C4 Laplacian channels of two exemplary subjects, and also from the coupled pair of sources (broadband) returned by NID in the 2-channel setting.

b) Number of NID sources: We analyzed spatial filters produced by NID in the 2-channel setting, aggregating all participants from both datasets. The goal of this analysis was to understand whether the algorithm extracted one source of highly pronounced non-sinusoidality (see Fig.5b) or, rather, two different μ-β synchronized sources (see Fig.5a). To this end, we computed the dot product between the spatial filters returned by NID corresponding to P and Q components. When the absolute value of this product was greater than 0.9, we considered both components to be the same. As a result, the spatial filters of 77% of subjects were almost identical, thus, for them NID returned a single source of non-sinusoidal SMR rhythm. On the contrary, for the remaining subjects NID extracted a pair of distinct non-linearly coupled sources.

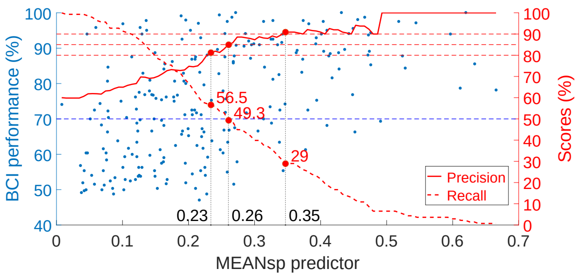
E. Participant selection based on predictor values

As stated in section IV, we analyzed threshold values for predictors to select participants who can obtain voluntary control of a BCI with high probability. Since no significant differences were found between settings, we only investigated $\mathrm{MEAN_{SP}}$ with 2 channels. Regarding the benchmarked predictors, no significant differences were found between SMR and $\mathrm{MEAN_{SP}}$, so SMR was also selected for this analysis.

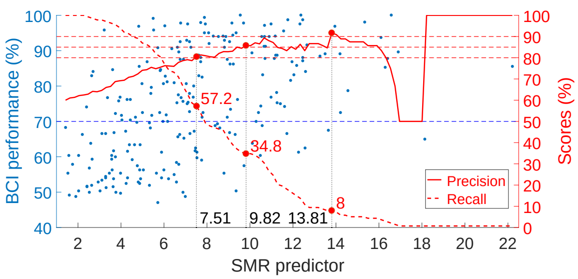
In Fig.6 scatter plots of predictor values vs. BCI online performances, together with precision-recall curves for $\mathrm{MEAN_{SP}}$ in the 2-channel setting and SMR are shown. Both studied datasets were employed for this analysis. BCI accuracies under 70% were labeled as poor performance. This figure shows the obtained predictor thresholds and recall values for specific precisions. In the case of $\mathrm{MEAN_{SP}}$,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612



(a) $\mathrm{MEAN_{SP}}$ with 2 channels.



(b) SMR.

Fig. 6: Scatter plots with predictor values and BCI online performances (left y-axis), together with precision-recall curves (right y-axis) for subjects in the two analyzed datasets. Poor performance was labeled under 70% of accuracy. Predictor threshold values and corresponding recalls are labeled for each specified precision. Predictors: a) $\mathrm{MEAN_{SP}}$ in 2-channel setting, b) SMR.

the obtained predictor thresholds were 0.23, 0.26 and 0.35, and the recalls were 56.5%, 49.3% and 29%, for precision values of 80%, 85% and 90% respectively. For the SMR predictor, the thresholds were 7.51, 9.82 and 13.81, and the recalls 57.2%, 34.8% and 8%, for the same precisions respectively.

Here we see that for a precision of 80% both predictors obtained similar recall values (a bit above 55%), whereas as the precision increased, $\mathrm{MEAN_{SP}}$ outperformed SMR, in terms of recall. In particular, for a precision of 90%, the recall of SMR fell below 10%, which means that from actual good performers, less than 10% were identified as such. In contrast, $\mathrm{MEAN_{SP}}$ still captured around 30% of those participants.

In Fig.6b, it is also visible that the precision suddenly drops for high SMR predictor thresholds. This means that there can be participants having a very high SMR predictor indicator whose obtained performance is actually below 70%. This is not the case of $\mathrm{MEAN_{SP}}$, where the precision steadily increases.

VII. Discussion

In this section, different aspects of the results are discussed. Although correlation values of $\mathrm{MEAN_{SP}}$ were quantitatively higher for Dataset 2 than for Dataset 1, all statistical outcomes comparing $\mathrm{MEAN_{SP}}$ and the benchmarked predictors are the same in both datasets.

A. $\mathrm{MEAN_{SP}}$ *versus benchmarked predictors*

$\mathrm{MEAN_{SP}}$ is a predictor based on power estimates and the synchronization between μ and β rhythms, which is one of the characteristics of oscillations with sensorimotor origin. The other selected predictors are also based on different power measures. The results presented in section VI-C show that PPfactor and SH are significantly worse than $\mathrm{MEAN_{SP}}$ in every evaluated setting (different number of channels and initial montage), including $\mathrm{MEAN_{SP}}$ computed with only 2 channels.

A reason why PPfactor might have performed worse in this paper than in the original one [21] might be related to the actual number of channels employed. In the original paper, the authors mention the need of only two channels for calculating the predictor (C3 and C4). However, they also mentioned a common average reference (CAR) filter that required 64 channels. CAR also removes spatial artifacts from data and might have helped to obtain better results. However, our goal in this paper is to test predictors using a low number of channels, thus we computed PPfactor in C3 and C4 raw electrodes. Another reason might be related to the fact that we could not use the γ band up to 70 Hz as in [21], because our data was filtered below 50 Hz.

In relation to SH, the reported correlation in [27] is related to BCI offline performance instead of online performance. However, online performance is the accuracy that actually informs about the ability of a participant to control a BCI system. Hence, we selected this online assessement as the target of all predictors. In order to account for the possible differences in visual input between calibration and feedback settings, we tested SH in both eyes-closed and eyes-open conditions and selected the most favorable for SH (eyes-open). Nevertheless, the differences between $\mathrm{MEAN_{SP}}$ and SH are still significant.

Besides, both PPfactor and SH were only correlated with performance obtained classifying right vs. left hand MI tasks in their respective papers, whereas in the datasets selected in this work, any 2-class combination out of the three MI tasks performed by the participants in a calibration recording (left hand, right hand, feet/right foot MI) were possible. Thus, PPfactor and SH might not cope well with feet/right foot MI. Indeed, in the specific case of SH, which is computed in just one channel, adding foot-related tasks might affect its performance.

Regarding SMR, correlations of $\mathrm{MEAN_{SP}}$ and SMR were not found significantly different in any case. However, the number of channels needed by SMR is more than twice the number of electrodes used in $\mathrm{MEAN_{SP}}$. In particular, $\mathrm{MEAN_{SP}}$ evaluated in just two channels is not significantly different from SMR either. On the contrary, SMR computed in two raw channels (C3 and C4) is significantly worse than $\mathrm{MEAN_{SP}}$.

One difference between SMR and $\mathrm{MEAN_{SP}}$ is related to the spatial filter. The Laplacian derivations employed in SMR are spatial filters with fixed weights [49] that aim to reduce the volume conduction artifact of a channel by

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

subtracting the activity of neighboring electrodes. Hence, they need a relatively high number of sensors to obtain one virtual channel. On the other hand, NID is a data-driven method that can be computed even with two channels. It finds spatial filters of synchronized oscillations between different frequency bands. Thus, it can be used to assess the phase-phase coupling observed between oscillations in μ and β frequency bands of sensorimotor origin [33], [34]. Furthermore, as shown in Fig.5 from section VI-D, it also increases the SNR of the electrodes used in the montage, facilitating the estimation of the SNR of sensorimotor rhythms which are known to include both μ and β oscillations as well as interactions between them.

As seen in section VI-D, NID mostly extracts one μ-β coupled single source. This might be seen as a limitation of NID, because it cannot differentiate harmonics of the same source from a cross-frequency interaction from two different sources. However, in our work, it is not very relevant whether the coupling comes from distinct sources or not, as even the single source interaction is an evidence of highly pronounced non-sinusoidal activity, again pointing to the sensorimotor origin of the extracted sources.

Finally, our analyses were performed with fixed frequency centers for μ and β bands, where the β oscillation was placed at twice the frequency of the μ oscillation. However, some participants might exhibit a β peak that cannot be defined as a harmonic of the μ rhythm. Furthermore, some individuals might not display a clear β peak in sensor space. Regarding the first aspect, in our additional analyses with subject-specific centers (see Table VI of Supp. Material), the β peak frequency was individually selected and NID modified accordingly. However, no significant differences were found between fixed and subject-specific frequency centers. Concerning the second point, even though a β peak might not be visible in sensor space, the oscillating source might still exist [44]. Thus, approximating this oscillation using twice the μ frequency is a robust approach, as shown by our results. Both outcomes are related to the bandwidth used to decompose μ and β oscillation within NID, which were 10 to 12 and 20 to 24 Hz, both covering a large proportion of the usual μ (8 to 12) and β (16 to 24) frequency ranges [50].

B. *Channel selection for* $\mathrm{MEAN_{SP}}$

The selection of channels for $\mathrm{MEAN_{SP}}$ was performed with LASSO regression on the SMR predictor of single raw channels. In order to test the robustness of the selection method, two initial montages were employed from which to select a predefined number of electrodes (from 2 to 5). The specific electrodes selected in the settings of 2 and 3 channels were the same for both initial montages. In regard to settings of 4 and 5 channels, they only differed in one electrode comparing the selection done from 9- and 15-channel initial montages. However, both differing electrodes were immediately adjacent to each other (see Fig.3), suggesting that results are stable across the initial montages. In fact, no significant differences were found between correlation results obtained with each of the two initial montages for a specific number of channels (see section VI-C). Hence, as correlations were in general quantitatively higher with electrodes selected using the 9-channel montage, we recommend the channel combinations shown in the top row of Fig.3.

Furthermore, the channels selected in each iteration of the LOSO procedure were almost always the same. For instance, C3 and C4 were selected every time in all settings (see Tables I and II in Supp. Material), which is in line with previous evidence [4]. In summary, the low variability in the channel selection outcomes suggests that the results are robust across subjects.

C. $\mathrm{MEAN_{SP}}$ *with different number of channels*

$\mathrm{MEAN_{SP}}$ was evaluated with different settings regarding the number of channels used, specifically with 2, 3, 4 and 5 electrodes. No significant differences were found across settings. This means that using 2 or 5 electrodes do not provide a significantly different correlation result.

Besides, $\mathrm{MEAN_{SP}}$ was also tested with 10 channels, selected by LASSO from the 15-ch montage. No significant differences were found between this setting and the SMR predictor (also computed on 10 electrodes). These results show that NID on a low number of channels (2 to 5), as studied in this work, is sufficient to capture the information provided by the SMR predictor on C3 and C4 Laplacian derivations (needing a total of 10 channels). This result suggests that 2 channels, in particular C3 and C4, are enough to compute $\mathrm{MEAN_{SP}}$, saving setup time.

D. *Threshold to predict good performance*

The scatter plots in Fig.6 show that for both $\mathrm{MEAN_{SP}}$ and SMR predictors, poor performance classification is unreliable, i.e. low predictor values are not representative of poor performance. The reasons of low predictor values for good accuracy are discussed in [20]. They suggest that some participants can control BCI by means of the peri-imagery Event-Related Synchronization phenomenon instead of the Desynchronization effect, that is far more common. This would imply that a low SNR of the sensorimotor rhythms could still allow obtaining good performances. In our case, another reason might be related to a low SNR of oscillations in either the μ or the β frequency bands, underestimating the level of cross-frequency synchrony in those participants (see Fig.5 of [36]).

Regarding the classification of subjects achieving more than 70% of accuracy, in the same figure it is shown that as the precision increases, $\mathrm{MEAN_{SP}}$ can retain better recall percentages than the SMR predictor. This means that good performers are better identified by $\mathrm{MEAN_{SP}}$ rather than by the SMR indicator. Furthermore, very high SMR values might also be obtained by subjects who performed worse than 70% (see the precision drop in Fig.6b caused by an outlier). It is apparent that adding cross-frequency related information to the predictor helps identifying good performers by ensuring the sensorimotor

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

origin of the oscillations [35]. MEAN_{SP} might discard artefactual signals (e.g. occipital oscillations), reducing the occurrence of outliers for high predictor values.

Thus, although none of the two predictors can actually predict whether a person will perform worse than 70% of accuracy (low predictor values are not representative of low accuracy), MEAN_{SP} is a better option than SMR to select participants based on good predicted performance.

In general, the stratification of participants on the basis of BCI predictors is important not only in the context of BCI, but also when considering its clinical applications (e.g. stroke [51]–[53]). In this case, motor rehabilitation with BCI may require many sessions. Therefore, one can primarily focus on patients who have a potential to use and benefit from sensorimotor BCI. In particular, when a system delivers complex neurofeedback by means of robotic arms or exoskeletons, high predictor values would indicate that the patient is likely to perform well. In the case of low predictor values, the clinician may carry out additional tests or specific training with simpler paradigms before deciding whether the patient will benefit from a complex BCI rehabilitation system. Furthermore, additional rehabilitation techniques such as non-invasive brain stimulation can also be considered for these patients.

VIII. Conclusion

MEAN_{SP} is a novel predictor to assess SMR-based BCI performance. It finds mixtures of μ-β synchronized channels, optimizing the sensorimotor origin of the estimated SNR. It is easily calculated with just 2 minutes of data, requiring only the provided SMR normalization limits, together with the NID algorithm and the SMR predictor.

We evaluated MEAN_{SP} with two large-scale datasets in different settings regarding the number of channels used. For each of them, we provided the optimal location of EEG electrodes. In all cases, MEAN_{SP} performed similar or significantly better than other benchmarked predictors. Moreover, MEAN_{SP} proved to be robust even when only two channels were available, showing that C3 and C4 are sufficient to obtain reliable results. Lastly, we showed that MEAN_{SP} can robustly detect good BCI performers. The provided precision-recall curves of MEAN_{SP} allow the selection of thresholds that can be used in future studies.

Acknowledgments

C.V. was supported by the Spanish Ministry of Research and Innovation PID2020-118829RB-100, Diputacion Foral de Gipuzkoa (DFG) Brain2Move project, DFG Neurocog Project, and Ikerbasque. This research was also supported by the Basque Government through the BERC 2022-2025 program and Funded by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010/AEI/10.13039/501100011033.

References

[1] M. Lotze and U. Halsband, "Motor imagery," *Journal of Physiology-paris*, vol. 99, no. 4-6, pp. 386–395, 2006.

[2] J. Decety, "The neurophysiological basis of motor imagery," *Behavioural brain research*, vol. 77, no. 1-2, pp. 45–52, 1996.

[3] G. Pfurtscheller and C. Neuper, "Motor imagery activates primary sensorimotor area in humans," *Neuroscience letters*, vol. 239, no. 2-3, pp. 65–68, 1997.

[4] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. Da Silva, "Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

[5] C. Park, D. Looney, N. ur Rehman, A. Ahrabian, and D. P. Mandic, "Classification of motor imagery BCI using multivariate empirical mode decomposition," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 21, no. 1, pp. 10–22, 2012.

[6] H.-J. Hwang, K. Kwon, and C.-H. Im, "Neurofeedback-based motor imagery training for brain–computer interface (BCI)," *Journal of neuroscience methods*, vol. 179, no. 1, pp. 150–156, 2009.

[7] C. Vidaurre and B. Blankertz, "Towards a cure for BCI illiteracy," *Brain topography*, vol. 23, no. 2, pp. 194–198, 2010.

[8] B. Z. Allison and C. Neuper, "Could anyone use a BCI?" in *Brain-computer interfaces*. Springer, 2010, pp. 35–54.

[9] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an EEG-based brain-computer interface (BCI)?" *IEEE transactions on neural systems and rehabilitation engineering*, vol. 11, no. 2, pp. 145–147, 2003.

[10] M. C. Thompson, "Critiquing the concept of BCI illiteracy," *Science and engineering ethics*, vol. 25, no. 4, pp. 1217–1233, 2019.

[11] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Co-adaptive calibration to improve BCI efficiency," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025009, 2011.

[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*, vol. 14, no. 1, pp. 1–37, 01 2019.

[13] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain–computer interface: a brief review," *Journal of neuroscience methods*, vol. 243, pp. 103–110, 2015.

[14] A. B. Randolph, "Not all created equal: individual-technology fit of brain-computer interfaces," in *2012 45th Hawaii International Conference on System Sciences*. IEEE, 2012, pp. 572–578.

[15] A. B. Randolph, M. M. Jackson, and S. Karmakar, "Individual characteristics and their effect on predicting mu rhythm modulation," *Intl. Journal of Human–Computer Interaction*, vol. 27, no. 1, pp. 24–37, 2010.

[16] 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," *Clinical Neurophysiology*, vol. 124, no. 8, pp. 1586–1595, 2013.

[17] W. Burde and B. Blankertz, *Is the locus of control of reinforcement a predictor of brain-computer interface performance?*, 2006, vol. 2006.

[18] E. M. Hammer, S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, K.-R. Müller, and A. Kübler, "Psychological predictors of SMR-BCI performance," *Biological psychology*, vol. 89, no. 1, pp. 80–86, 2012.

[19] C. Sannelli, M. Braun, M. Tangermann, and K.-R. Müller, "Estimating noise and dimensionality in BCI data sets: Towards illiteracy comprehension," 2008.

[20] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance," *Neuroimage*, vol. 51, no. 4, pp. 1303–1309, 2010.

[21] 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.

[22] M. Ahn, S. Ahn, J. H. Hong, H. Cho, K. Kim, B. S. Kim, J. W. Chang, and S. C. Jun, "Gamma band activity associated with BCI performance: simultaneous MEG/EEG study," *Frontiers in human neuroscience*, vol. 7, p. 848, 2013.

[23] S. Halder, D. Agorastos, R. Veit, E. M. Hammer, S. Lee, B. Varkuti, M. Bogdan, W. Rosenstiel, N. Birbaumer, and A. Kübler, "Neural mechanisms of brain–computer interface control," *Neuroimage*, vol. 55, no. 4, pp. 1779–1790, 2011.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Transactions on Neural Systems and Rehabilitation Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2023.3339612

[24] S. Halder, B. Varkuti, M. Bogdan, A. Kübler, W. Rosenstiel, R. Sitaram, and N. Birbaumer, "Prediction of brain-computer interface aptitude from individual brain structure," *Frontiers in human neuroscience*, vol. 7, p. 105, 2013.

[25] L. Acqualagna, L. Botrel, C. Vidaurre, A. Kübler, and B. Blankertz, "Large-scale assessment of a fully automatic co-adaptive motor imagery-based brain computer interface," *PloS one*, vol. 11, no. 2, p. e0148886, 2016.

[26] H.-I. Suk, S. Fazli, J. Mehnert, K.-R. Müller, and S.-W. Lee, "Predicting BCI subject performance using probabilistic spatio-temporal filters," *PloS one*, vol. 9, no. 2, p. e87056, 2014.

[27] R. Zhang, P. Xu, R. Chen, F. Li, L. Guo, P. Li, T. Zhang, and D. Yao, "Predicting inter-session performance of SMR-based brain–computer interface using the spectral entropy of resting-state EEG," *Brain topography*, vol. 28, no. 5, pp. 680–690, 2015.

[28] C. Vidaurre, S. Haufe, T. Jorajuría, K.-R. Müller, and V. V. Nikulin, "Sensorimotor functional connectivity: A neurophysiological factor related to BCI performance," *Frontiers in Neuroscience*, vol. 14, 2020.

[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," *Clinical neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.

[30] M. Lee, J.-G. Yoon, and S.-W. Lee, "Predicting motor imagery performance from resting-state EEG using dynamic causal modeling," *Frontiers in human neuroscience*, vol. 14, p. 321, 2020.

[31] M. Kwon, H. Cho, K. Won, M. Ahn, and S. C. Jun, "Use of both eyes-open and eyes-closed resting states may yield a more robust predictor of motor imagery BCI performance," *Electronics*, vol. 9, no. 4, p. 690, 2020.

[32] J. T. Roscoe, *Fundamental research statistics for the behavioral sciences [by] John T. Roscoe*, 1975.

[33] H. Carlqvist, V. V. Nikulin, J.-O. Strömberg, and T. Brismar, "Amplitude and phase relationship between alpha and beta oscillations in the human electroencephalogram," *Medical and Biological Engineering and Computing*, vol. 43, no. 5, pp. 599–607, 2005.

[34] V. V. Nikulin and T. Brismar, "Phase synchronization between alpha and beta oscillations in the human electroencephalogram," *Neuroscience*, vol. 137, no. 2, pp. 647–657, 2006.

[35] N. Schaworonkow and V. V. Nikulin, "Spatial neuronal synchronization and the waveform of oscillations: Implications for EEG and MEG," *PLoS Computational Biology*, vol. 15, no. 5, p. e1007055, 2019.

[36] M. J. Idaji, K.-R. Müller, G. Nolte, B. Maess, A. Villringer, and V. V. Nikulin, "Nonlinear interaction decomposition (NID): A method for separation of cross-frequency coupled sources in human brain," *NeuroImage*, vol. 211, p. 116599, 2020.

[37] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio *et al.*, "The berlin brain–computer interface: non-medical uses of BCI technology," *Frontiers in neuroscience*, p. 198, 2010.

[38] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[39] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[40] M. Hebiri and J. Lederer, "How correlations influence lasso prediction," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1846–1854, 2012.

[41] K. Sjöstrand, L. H. Clemmensen, R. Larsen, G. Einarsson, and B. K. Ersbøll, "SpaSM: A MATLAB toolbox for sparse statistical modeling," *Journal of Statistical Software*, vol. 84, no. 10, 2018.

[42] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[43] S. Rosset and J. Zhu, "Piecewise linear regularized solution paths," *The Annals of Statistics*, pp. 1012–1030, 2007.

[44] V. V. Nikulin, G. Nolte, and G. Curio, "A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition," *NeuroImage*, vol. 55, no. 4, pp. 1528–1535, 2011.

[45] A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. P. Birbaumer, "Brain-computer communication: self-regulation of slow cortical potentials for verbal communication," *Archives of physical medicine and rehabilitation*, vol. 82, no. 11, pp. 1533–1539, 2001.

[46] X.-L. Meng, R. Rosenthal, and D. B. Rubin, "Comparing correlated correlation coefficients." *Psychological bulletin*, vol. 111, no. 1, p. 172, 1992.

[47] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, pp. 65–70, 1979.

[48] M. J. Idaji, "NID," 2021. [Online]. Available: https://github.com/minajamshidi/NID/

[49] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalography and clinical Neurophysiology*, vol. 103, no. 3, pp. 386–394, 1997.

[50] C. Vidaurre, J. Pascual, A. Ramos-Murguialday, R. Lorenz, B. Blankertz, N. Birbaumer, and K.-R. Müller, "Neuromuscular electrical stimulation induced brain patterns to decode motor imagery," *Clinical Neurophysiology*, vol. 124, no. 9, pp. 1824–1834, 2013.

[51] M. Sebastián-Romagosa, W. Cho, R. Ortner, N. Murovec, T. Von Oertzen, K. Kamada, B. Z. Allison, and C. Guger, "Brain computer interface treatment for motor rehabilitation of upper extremity of stroke patients—a feasibility study," *Frontiers in Neuroscience*, vol. 14, p. 591435, 2020.

[52] M. A. Cervera, S. R. Soekadar, J. Ushiba, J. d. R. Millán, M. Liu, N. Birbaumer, and G. Garipelli, "Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis," *Annals of clinical and translational neurology*, vol. 5, no. 5, pp. 651–663, 2018.

[53] E. Buch, C. Weber, L. G. Cohen, C. Braun, M. A. Dimyan, T. Ard, J. Mellinger, A. Caria, S. Soekadar, A. Fourkas *et al.*, "Think to move: a neuromagnetic brain-computer interface (BCI) system for chronic stroke," *Stroke*, vol. 39, no. 3, pp. 910–917, 2008.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/