When should MI-BCI feature optimization include prior knowledge, and which one?

Camille Benaroch, Maria Sayu Yamamoto, Aline Roc, Pauline Dreyer, Camille Jeunet, Fabien Lotte

When should MI-BCI feature optimization include prior knowledge, and which one?

Camille Benaroch[1,2*], Maria Sayu Yamamoto[1], Aline Roc [1,2], Pauline Dreyer[1,2], Camille Jeunet[3], Fabien Lotte[1,2],

1 Inria Bordeaux Sud-Ouest, Talence, France
2 LaBRI (CNRS, Univ. Bordeaux, INP), Talence, France
3 CLLE Lab, CNRS, Univ. Toulouse Jean Jaurès, Toulouse, France
3 Univ. Bordeaux, CNRS, EPHE, INCIA, UMR5287 F-33000 Bordeaux, France

* camille.benaroch@inria.fr

Abstract

Motor imagery-based brain-computer interfaces (MI-BCIs) rely on interactions between humans and machines. Therefore, the (learning) characteristics of both components are key to understand and improve performances. Data-driven methods are often used to select/extract features with very little neurophysiological prior. Should such approach include prior knowledge and, if so, which one? This paper studies the relationship between BCI performances and characteristics of the subject-specific Most Discriminant Frequency Band (MDFB) selected by a popular heuristic algorithm. First, our results showed a correlation between the selected MDFB characteristics (mean and width) and performances. Then, to investigate a possible causality link, we compared, online, performances obtained with a constrained (enforcing characteristics associated to high performances) and an unconstrained algorithm. Although we could not conclude on causality, average performances using the constrained algorithm were the highest. Finally, to understand the relationship between MDFB characteristics and performances better, we used machine learning to 1) predict MI-BCI performances using MDFB characteristics and 2) select automatically the optimal algorithm (constrained or unconstrained) for each subject. Our results revealed that the constrained algorithm could improve performances for subjects with either clearly distinct or no distinct EEG patterns.

Introduction

Brain-Computer Interfaces (BCIs) identify users' intent by analyzing their brain activity, and thereby enable users to interact with the external world without any movement [8]. This activity is most often measured by ElectroEncephaloGraphy (EEG). There are several BCI applications that exploit different types of electrophysiological activities. In particular, motor imagery BCIs (MI-BCIs) are promising as a rehabilitation tool and an assistive technology for motor-impaired users [19, 23, 25] or as a new interaction modality for healthy users [16]. Motor imagery generally leads to changes in brain rhythms originating from the sensory motor cortices, i.e., sensorimotor rhythms (SMRs) [4]. SMRs, in the broad sense, refer to oscillations recorded over the sensorimotor cortex in the 8-30Hz frequency-band. An SMR desynchronisation, i.e., a decrease of signal power, is typically observed in the contro-lateral sensorimotor cortex during the execution, or imagination [5], of a hand movement [1].

Therefore, ideally, when learning how to control an MI-BCI, the user also learns how to generate stable and distinct brain signals for each class (i.e. task) [31]. To be able to do so, extracted features from the recorded EEG signals should be consistent. In other words, they should ideally correspond to the neurophysiological patterns identified in the literature. During the calibration phase of an MI-BCI system, discriminative data-driven learning methods are commonly used to perform EEG feature extraction. One popular signal processing technique for EEG-based MI-BCIs is the Common Spatial Pattern (CSP) algorithm, which learn spatial filters that best discriminate between two MI classes [6, 11]. CSP filters maximize the variance of the spatially filtered signal under one condition while minimizing it for the other class. Before using CSP filters, several parameters have to be selected: the band pass filter, the time interval and the number of filters to use. When calibrating the system, it is common to use general settings (i.e, the same settings) for all subjects, e.g., the standard 8-30 Hz pass-band for MI-BCIs [6]. Yet, individually choosing the hyperparameters for CSP-based classification could improve online performances [10]. Some data-driven methods have been developed to select subject-specific hyperparameters [11]. Once the discriminative features identified, linear classifiers (e.g., LDA...) are most often used to distinguish classes [13].

Although these methods are commonly used and have proven to be effective, they are almost exclusively data-driven. They include very little neurophysiological prior, and rather trust the (potentially noisy) EEG data recorded during the calibration phase. In this paper, we hypothesise that, to be more effective, these methods should take into consideration some constraints that may need to be identified. Indeed, are all properties of the features or classifiers learned from BCI data equally likely to be associated with good performances in practice? If not, what properties are more often associated with superior decoding performances? Would enforcing such properties in machine learning BCI algorithms lead to better decoding performances in practice? In this paper, we aim at answering these questions for a specific machine learning BCI algorithm: the frequency band selection algorithm of Blankertz et. al [11]. This algorithm has been shown to improve user performances [11]. This heuristic algorithm selects the frequency band the power of which, in the sensorimotor channels, maximally correlates with the class labels. While often effective in practice, the lack of constraints in this algorithm may lead to the selection of frequency bands that are suboptimal. For instance, for illustrative purpose, this algorithm could select a most discriminant frequency band that is only 0.5 Hz wide whereas one may wonder whether such a narrow - and thus probably too specific - band is likely to lead to high decoding performances. Note that we selected this specific data-driven method as it is commonly used in order to improve BCI performances but, to our knowledge, no introspective work has been done on this algorithm. In the future, similar introspective work should be performed on other popularly used machine learning algorithms, such as LDA and CSP. Thus, in this paper, we study the impact of the properties of the most discriminant frequency band (MDFB) selected by machine learning algorithms on online performances, with three different analyses. First, we investigated possible relationships between the characteristics of this selected MDFB and online BCI performances. Then, in order to study a possible causality link between these characteristics and performances, we designed an experiment using a new version of the algorithm with added constraints to enforce characteristics associated with superior performances. We compared online BCI performances obtained using the constrained algorithm with those of subjects using the unconstrained (i.e., original) algorithm. Finally, to better understand the relationships between the characteristics of the MDFB and online BCI performances, and to find out whether and if so when to use constraints, we used a LASSO (least absolute shrinkage and selection operator)[3] regression to determine explanatory and predictive models of MI-BCI performances based on MDFB characteristics. We then used a classification algorithm (here a decision tree) to automatically select the best algorithm (i.e. constrained or unconstrained) for each subject, based on the factors identified by these models.

This paper is organized as follows. In Section 1, we study the relationship between characteristics of the MDFB and online performances. Section 2 is dedicated to the analysis of possible causal relationships between such characteristics and performances while in Section 3 we go a step beyond the correlation and designed models to (1) better understand the relationships between characteristics of the MDFB and online performances and (2) select automatically the best user-specific algorithm (with or without constraints). The last two sections are a general discussion and a general conclusion of the paper.

1 Studying the relationships between MI-BCI performances and subject-specific frequency band characteristics

1.1 Objective

The frequency band selection algorithm is a data-driven method that selects the frequency band the power of which, in the sensorimotor channels, maximally correlates with the class labels, with little consideration for the resulting human performances and with little constraints based on prior knowledge. In this section, in order to determine if such an approach could benefit from using prior knowledge, we first investigated the relationships between the subject-specific frequency bands selected by the algorithm and users' online BCI performances. Our first objective was to determine if there was a correlation between characteristics of the selected band and online performances. Then, based on those first results, we aimed at identifying and proposing constraints to add to the original frequency band selection algorithm, in order to exploit this new knowledge as prior, and thus possibly improve the approach.

1.2 Materials and Methods

1.2.1 Dataset

We used the data of fifty-five subjects (N=55) out of the sixty subjects from the experiment in [36] (29 women; age 19-59; $\bar{X}$ =29; SD=9.318). This dataset is named **Dataset A** in the rest of the paper. We discarded five subjects due to technical issues during the BCI recording session (e.g., a missing calibration run, or a mislabelling of EEG channels that resulted in an erroneous calibration). The main goal of this previous experiment was to assess the impact of the interaction between experimenters' and participants' gender on MI-BCI performances. Each subject participated in one MI-BCI session of 2 hours, comprising two three-minute recordings in resting state condition (one with eyes open and one with eyes closed) followed by six 7-minute blocks (a.k.a runs) of trials in which the subject had to perform MI-tasks, i.e., imagine right or left-hand movements. Each subject performed 40 trials in a block, 20 per MI-task, that appeared in random order. Each trial lasted for 8s. At t=0s a cross was displayed on screen. After two seconds the subject was provided with an acoustic signal which indicated the appearance of a red arrow, that appeared one second later. The red arrow direction indicated the type of MI-task to perform, e.g., a left pointing arrow to indicate the imagination of a movement of the left hand. From t=4.250s a blue bar was displayed as feedback, representing the classifier output: its length changed according to the classifier output and it appeared only if the instruction matched the recognized task (positive feedback only). The feedback lasted 3.75s and was updated at 16Hz, using a 1s sliding window. Afterwards the screen turned black and the user had a few seconds to rest before the next trial started.

1.2.2 EEG Recordings and Signal Processing

To record the EEG signals, 27 active scalp electrodes, referenced to the left ear, were used (Fz, FCz, Cz, CPz, Pz, C1, C3, C5, C2, C4, C6, F4, FC2, FC4, FC6, CP2, CP4, CP6, P4, F3, FC1, FC3, FC5, CP1, CP3, CP5, P3, 10-20 system). EEG signals were measured using a g.USBAmp (g.tec, Austria), sampled at 256 Hz and processed online using OpenviBE 2.1.0 [20].

In order to discriminate the MI tasks from EEG data, participant-specific spectral and spatial filters have been optimized on the data of the first two runs (calibration runs). First, the participants' MDFB was selected using the heuristic algorithm proposed in [11] (see Algorithm 1). It selects the frequency band the power of which, in the sensorimotor channels, maximally correlates with the class labels. More precisely, the EEG trials were first spatially filtered using a Laplacian filter around C3 $[4C3 - FC3 - CP3 - C5 - C1]$ and C4 $[4C4 - FC4 - CP4 - C6 - C2]$. Then, for each Laplacian channel and each frequency, the correlation coefficient between the log-transformed band power of the Laplacian channel signal and the class labels was computed across trials. The frequency for which the sum of the two Laplacian channel correlation coefficients was maximal was used as a reference. In the case where the correlation coefficient for this frequency of reference was negative in a Laplacian channel, all the signs of the correlation coefficient were reversed. The frequency for which the sum of correlation coefficients across the two Laplacian channels was maximal was selected and named $fmax$. Finally, the frequency band was enlarged around $fmax$ until the correlation signal (named $fscore$) exceeded a threshold of $fscore(fmax) \times 0.05$.

Algorithm 1 Unconstrained Algorithm

Let $X_{(c,i)}$ denote trial i at channel c with label y_i and let C denote the set of channels.
$dB_c(f, i) \leftarrow$ log band-power of$X_{(c,i)}$at frequency f (f from 5 to 20 Hz)
$score_c(f) \leftarrow \text{corrcoef}(dB_c(f, i), y_i)_i$
$f_{max} \leftarrow argmax_f \sum_{c \in C} score_c(f)$
$score_c^*(f) \leftarrow \begin{cases} score_c(f) & \text{if } score_c(f_{max}) > 0 \\ -score_c(f) & \text{otherwise} \end{cases}$
fscore(f) $\leftarrow \sum_{c \in C} score_c^*(f)$
$f_{max}^* \leftarrow argmax_f fscore(f)$
$f_0 \leftarrow f_{max}^* f_1 \leftarrow f_{max}^*$
while $fscore(f_0 - 1) \geq fscore(f_{max}^*) * 0.05$ **do**
    $f_0 \leftarrow f_0 - 1$
while $fscore(f_1 + 1) \geq fscore(f_{max}^*) * 0.05$ **do**
    $f_1 \leftarrow f_1 + 1$
Return frequency band$[f_0; f_1]$

Then EEG was filtered in that selected band with a Butterworth filter of order 5. To build the classifier, the Common Spatial Pattern (CSP) algorithm [6, 11] was first used to optimize 3 pairs of spatial filters, still using the data from the two calibration runs. Such spatially filtered EEG signals should thus have a band power, which is maximally different between the two MI conditions. We then computed the band power of these spatially filtered signals by squaring the EEG signals, averaging them over a 1 second sliding window (with 1/16th second between consecutive windows). Finally, the log-transformed band power of these signals was used as input of a linear discriminant analysis (LDA) classifier [13].

1.2.3 Variables and factors

The metric used for online performance was the the online trial-wise accuracy (TAcc), i.e. the default accuracy measure provided online in the MI-BCI scenarios of OpenViBE, which is also the performance metric that the experimenters were seeing online.This metric is computed by first summing the (signed) LDA classifier outputs (distance to the separating hyperplane) over all epochs (1s long epochs, with 15/16 s overlap between consecutive windows) during the trial feedback period. If this sum sign matched the required trial label, i.e., negative for left hand MI and positive for right hand MI, then the trial was considered as correctly classified, otherwise it was not. The TAcc for each run was estimated as the percentage of trials considered as correctly classified using this approach. We explored whether this classification accuracy value (dependant variable) correlated with two characteristics of the MDFB (independent variables), namely its mean frequency and width, the computation of which is explained below.

The optimization algorithm presented previously provided us with the MDFB boundaries f_{min} and f_{max}. From these two values, for each subject, we computed the mean value (MDFB mean) and the width (MDFB width) of the selected MDFB:

- MDFB mean = $\frac{f_{min}+f_{max}}{2}$
- MDFB width = f_{max} - f_{min}

1.2.4 Analyses

None of the three variables (namely the MDFB mean, width and the MI-BCI performance) followed a normal distribution. Thus we used non-parametric tests:

- A Spearman's rank correlation test to measure the relationship between the MDFB mean or width and MI-BCI performances.
- A Wilcoxon rank sum test to compare the MI-BCI performances of two groups of subjects with different MDFB mean or width (i.e., one group below a threshold and the other above) in order to identify possible constraints to add to a new improved version of the algorithm. The thresholds were fixed visually to allow a separation between the cluster of high performers and the rest of the participants. We based this separation on the bivariate distributions of the mean online classification accuracy and the MDBF width and mean as visual support. In addition, we made sure that the two groups were of approximatively similar size (32 vs 23 for the MDBF mean and 30 vs 25 for the MDFB width).

1.3 Results

First results showed correlations between performances and the two characteristics of the selected frequency band. A significant negative correlation was observed between MI-BCI performances and the *MDFB mean* frequency (r=-0.28, $p \leq 0.01$) and a significant positive one between MI-BCI performances and the *MDFB width* (r=0.70, $p \leq 0.01$).

Figure 1 represents the bivariate distribution of the *MDFB width* and the *MDFB mean* with the MI-BCI performances, i.e., the mean classification accuracy.

Other analyses revealed that:

- Subjects with an *MDFB width* lower than 3.5Hz have lower MI-BCI performances compared to subjects with a larger *MDFB width* (U=-4.9, p<10-5) (see Figure 2 A).
- Subjects with an *MDFB mean* value above 16Hz seem to have lower performances than subjects with an *MDFB mean* value under 16Hz (U=2.2, p<0.05) (see Figure 2 B).

Thresholds (i.e., 3.5Hz and 16Hz) were fixed visually using Figure 1 results.

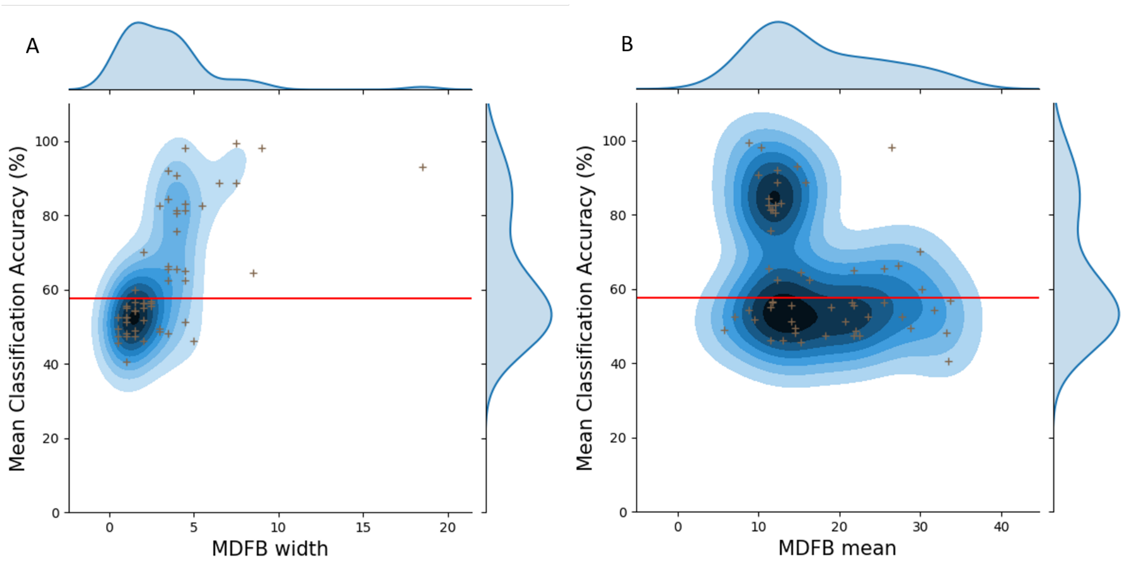

Figure 1: Bivariate distribution of the the mean online classification accuracy and (a) the frequency band width and (b) the mean frequency of the band. In red, the statistical chance level for 2-classes, 80 trials per class and $\alpha = 5\%$ [17]

1.4 Intermediate Discussion

These results suggest that there is a correlation between the selected user-specific frequency band characteristics (i.e., the mean value and the width of the MDFB selected) and the classification accuracy. The user-specific frequency bands were selected within the α-β range $(5 - 35\text{Hz})$ by using the heuristic proposed by Blankertz et al. in [12]. Results also suggest that (1) subjects with MDFB Length < 3.5Hz seem to experience difficulties controlling a BCI compared to subject with higher *MDFB width*; (2) subjects with an *MDFB mean* value above 16Hz (in the β band) seem to have lower performances than subjects with an *MDFB mean* value under 16Hz (in the α band). However, this correlation does not imply any causal relationship between those factors. A selected user-specific frequency band in α-lowβ could cause higher performance, or users with higher performance could modulate their user-specific frequency bands in α-lowβ range, or a third factor, not assessed here, could have impacted both the user-specific frequency band and performances. Therefore, we cannot conclude on causal effects from this study alone.

The next step in our work would thus be to determine if there is a causal link between both variables and, if any, the direction of this link. In the following section, we thus investigated how manipulating the MDFB by restricting the selection of frequency bands affected BCI performance, to assess possible causal effects.

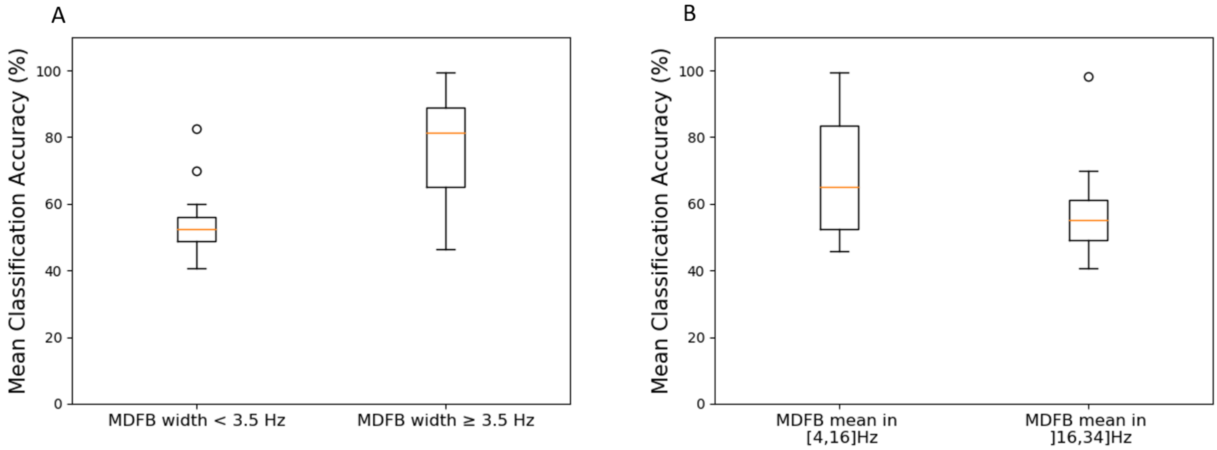

Figure 2: Boxplot that show (A) a significant difference of the median of the two groups (*MDFB mean* value in [4,16] or in]16,35] (U =2.2, p<0.05) and (B) a significant difference of the median of the two groups (*MDFB width* <3.5 or *MDFB width* ≥3.5 (U =-4.9, p<e-5)

2 Investigating the causal relationship between MI-BCI performances and subject-specific frequency band characteristics

2.1 Objective

In order to investigate a possible causation between BCI performances and characteristics of the MDFB, we designed a constrained version of the MDFB selection algorithm that enforced i) an *MDFB width* larger than 3.5Hz and ii) an *MDFB mean* value under 16Hz. We used it in an online experiment specifically designed for this study (**Dataset B**). We then compared the online performances obtained for subjects in this experimental group, **Dataset B** (using the constrained algorithm) with online performances of matched subjects in a control group, named Dataset A*. These subjects were selected from **Dataset A** and were then provided with the unconstrained, i.e., original algorithm.

2.2 Materials and Methods

2.2.1 Constrained version of the frequency band selection algorithm

In order to enforce (1) an *MDFB width* larger than 3.5Hz and (2) an *MDFB mean* value under 16Hz in the MDFB selection algorithm, we did the following:

(1) We smoothed, using a Savitzky-Golay filter [21], for each set of Laplacian channels, the $scorec(f)$ signal (Algorithm 2, line 6), which represents the correlation, at each frequency, between the band power of each trial and its class label. Smoothing the signal avoids having high isolated peaks in the scorec(f) signal which had the consequence of leading to narrow *MDFB width*. Then, we proceeded as in the original algorithm. If the *MDFB width* was still below 3.5Hz, we increased the band on the side where the sum of the the two channel correlation coefficients was the highest until reaching the

desired width, i.e., at least 3.5Hz (Algorithm 2, lines 13-20).

(2) We constrained the search of the frequency (Algorithm 2, lines 22-26) for which the sum of the two Laplacian channel correlation coefficients was maximal (i.e., f^*_{max}) to frequencies between 8 and 16Hz (i.e., in α-lowβ range) and we used the algorithm for frequencies restricted to be between 5 and 20Hz, which added to the constraints, avoids having an MDFB mean value above 16Hz.

Algorithm 2 Constrained Algorithm

1: Let $X_{(c,i)}$ denote trial i at channel c with label y_i and let C denote the set of channels.
2: $dB_c(f,i) \leftarrow$ log band-power of$X_{(c,i)}$at frequency f (f from 5 to 20 Hz)
3: $score_c(f) \leftarrow$ corrcoef$(dB_c(f,i), y_i)_i$
4: $f_{max} \leftarrow argmax_f \sum_{c \in C} score_c(f)$
5: $score^*_c(f) \leftarrow \begin{cases} score_c(f) & \text{if } score_c(f_{max}) > 0 \\ -score_c(f) & \text{otherwise} \end{cases}$
6: fscore(f) $\leftarrow$ savgolFilter($\sum_{c \in C} score^*_c(f)$)
7: $f^*_{max} \leftarrow argmax_f fscore(f)$
8: $f_0 \leftarrow f^*_{max} f_1 \leftarrow f^*_{max}$
9: **while** $fscore(f_0 - 1) \geq fscore(f^*_{max}) * 0.05$ **do**
10: $\quad f_0 \leftarrow f_0 - 1$
11: **while** $fscore(f_1 + 1) \geq fscore(f^*_{max}) * 0.05$ **do**
12: $\quad f_1 \leftarrow f_1 + 1$
13: **while** $fscore(f_1) - fscore(f_0) < 3.5$ **do**
14: $\quad$ **if** $fscore(f_1) < fscore(f_0)$ **then**
15: $\quad\quad f_0 \leftarrow f_0 - 1$
16: $\quad$ **else if** $fscore(f_1) > fscore(f_0)$ **then**
17: $\quad\quad f_1 \leftarrow f_1 + 1$
18: $\quad$ **else if** $fscore(f_1) = fscore(f_0)$ **then**
19: $\quad\quad f_0 \leftarrow f_0 - 1$
20: $\quad\quad f_1 \leftarrow f_1 + 1$
21:
22: **if** $\frac{f_0+f_1}{2} > 16$ or $\frac{f_0+f_1}{2} < 8$ **then**
23: $\quad$ **while** $\frac{f_0+f_1}{2} > 16$ **do**
24: $\quad\quad f_0 \leftarrow f_0 - 1$
25: $\quad$ **while** $\frac{f_0+f_1}{2} < 8$ **do**
26: $\quad\quad f_1 \leftarrow f_1 + 1$
27: Return frequency band$[f_0; f_1]$

2.2.2 Datasets

Dataset B: The experimental group

We included twenty-one (21) participants (8 women; age: 19-37 year-old; $\bar{X}$ =29; SD=9.318) in this experimental group with the same inclusion criteria and experimental protocol (instructions, tasks, feedback and training environment) as **Dataset A** in order to match this dataset as closely as possible to **Dataset A**. The only difference was the algorithm that we used to select the MDFB (see 2.2.1).

Dataset A*: The control group

As the objective was to study the effect of the new constrained algorithm on online performances, we paired subjects from **Dataset A and B** as follows. We first obtained, offline, the MDFB mean and width of subjects from **Dataset B** using the unconstrained (i.e., original) algorithm. Then, each subject was paired with a subject from **Dataset A** with similar MDFB mean and width. This enabled us to compare online the effects, if any, of the constrained algorithm (from **Dataset B**) for pairs of subjects with similar MDFB with the original algorithm. In the end, in **Dataset A*** we included 21 subjects (11 women; age: 20-57 year-old; $\bar{X}$ =29; SD=10). Table. 1 shows selected MDFB characteristics and the mean classification accuracy for subjects in **Dataset B** and **Dataset A***.

Experimental Group (Dataset B)						Control Group (Dataset A*)			
ID	Constrained Algorithm			Unconstrained Algorithm		ID	Unconstrained Algorithm		
	MDFB Mean (Hz)	MDFB Width (Hz)	Classification Accuracy (%)	MDFB Mean (Hz)	MDFB Width (Hz)		MDFB Mean (Hz)	MDFB Width (Hz)	Classification Accuracy (%)
B1	10	8	80.6	11.75	4.5	A43	11.75	4.5	81.3
B2	12	5	86.3	25.5	15	A44	25.5	4	65,6
B3	12	4	44.4	12	2	A06	11.5	2	46.3
B4	10.25	10.5	53.8	18.25	0.5	A10	18.25	0.5	47.5
B5	11.75	3.5	55	11.5	1	A52	11.5	1	55
B6	12.25	8.5	97.5	12.5	8	A23	12.25	6.5	88.8
B7	12.25	3.5	82.5	12.25	1.5	A48	11.75	1.5	56.6
B8	12.25	3.5	60	11.25	1.5	A24	8.75	1.5	54.4
B9	13.25	6.5	55.6	20.25	4.5	A40	20.75	4.5	51.3
B10	13.25	13.5	55	26.75	3.5	A43	27.25	3.5	66.3
B11	8.75	3.5	58.8	33	3	A02	33.25	3.5	48.1
B12	10.75	9.5	84.4	13.25	4.5	A01	12.25	4.5	62.5
B13	10.5	6	90.6	11	4	A13	11.5	4	75.6
B14	15	8	47.5	29	1	A22	28.75	0.5	49.4
B15	9.75	3.5	54.4	21.75	1.5	A25	21.75	1.5	47.5
B16	12	15	55	32	6	A29	26.5	9	98
B17	13.5	11	49.4	17	2	A03	19	2	55
B18	14.75	3.5	60.6	20.5	3	A11	22	3	48.8
B19	13.75	7.5	50	29	2	A30	30	2	70
B20	9.5	8	68.1	11.5	5	A19	13	5	46.5
B21	12	10	71.9	12.25	5.5	A21	11.25	5.5	82.5

Table 1: Data for experimental and control groups

2.2.3 Analyses

Our objective was to compare the MI-BCI performances obtained with a constrained vs. unconstrained MDFB algorithm. Therefore, we selected in **Dataset A*-B** only the participants whose MDFB characteristics differed in both conditions. The following heuristics were used to select the participants:

$$\left| \frac{m_{constrained}}{m_{unconstrained}} - 1 \right| \leq 0.10 \tag{1}$$

where $m_{constrained}$ is the MDFB mean selected by the constrained algorithm and $m_{unconstrained}$ is the MDFB mean selected by the unconstrained algorithm.

$$\frac{d}{(l_{constrained} + l_{unconstrained}) - d} \geq 0.60 \tag{2}$$

where $l_{constrained}$ is the *MDFB width* selected by the constrained algorithm, $l_{unconstrained}$ is the *MDFB width* selected by the unconstrained algorithm, and d is the width of the range of frequencies that are common to the MDFB obtained by both algorithm (e.g., if the unconstrained algorithm selected 11-19 Hz while the constrained algorithm selected 8-13Hz, their common frequency range was 11-13Hz and thus $d = 2$ Hz). Equation 1 consists of selecting pair of subjects whose MDFB mean rate of increase or decrease (between the two algorithms) is less than 10%. Equation 2 selects subjects whose MDFB width overlap (between the two algorithms) is over 60%. Such thresholds were selected manually to ensure pairs of similar subjects in terms of MDFB. As a result, the first analyses are based on the results of 18 participants out of 21 for each group (B06, B08 and B09 were removed as their MDFB characteristics were similar with both algorithms as well as A23, A24 and A40 as they were associated to B06, B08 and B09).

As the classification accuracy of all the different groups did not follow a normal distribution, we used a Wilcoxon signed-rank test to compare the MI-BCI classification accuracy between the two algorithms.

2.3 Results

While the average online performances for the constrained algorithm are higher (M=64.7 ± 16.5) than for the unconstrained algorithm (M=62.8 ± 16.6), our analyses showed that this difference is not significant (W=63.5, p=0.17). Figure 3 illustrates our results.

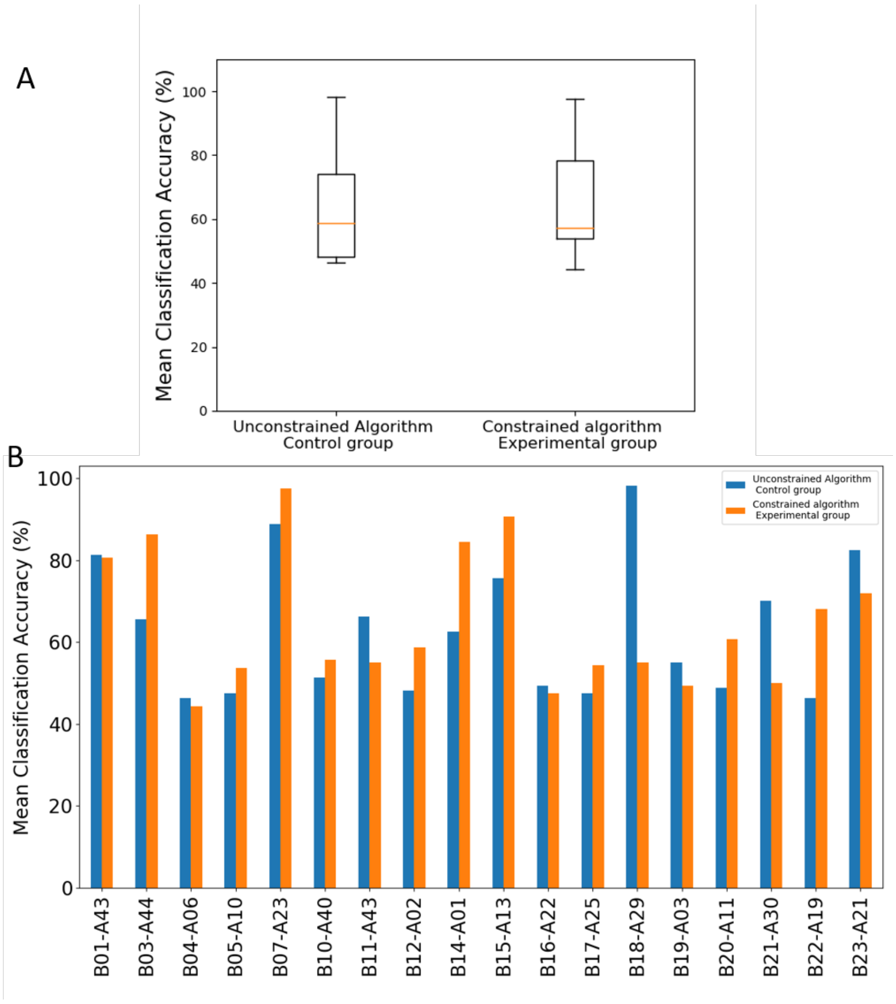

Figure 3: (A) Boxplot representing the distribution of the online classification accuracy (%) for the two groups (i.e., the experimental group using the constrained algorithm and the control group using the unconstrained algorithm, before using the CSP algorithm.) (B) Graph representing the online classification accuracy (%) for each subject.

2.4 Intermediate Discussion

These results suggest that there is no statistical difference in the classification accuracy when using the constrained or the unconstrained algorithm. The unconstrained algorithm was associated with similar or better performances than the constrained algorithm in most subjects (i.e., between -4% and 45% of improvement in 78% of the subjects), but also detrimental for 22% of the subjects (a drop between 10% and 43% in online performances). It is interesting to note that for one pair of subjects (B18-A29), adding constraints was strongly detrimental and that in this case, it seems that subject A29 was able to produce discriminant signals in the high β band (22–31 Hz). Three other pairs of subject had better performances with the unconstrained algorithm, which could be due to the fact that the *MDFB width* selected with the constrained algorithm was not specific enough, i.e., too broad (larger than 10 Hz). Indeed, in the constrained algorithm, we decided to smooth, for each Laplacian channel, the signal representing the correlation, at each frequency, between the band power of each trial and its class label. Although this prevented us from having narrow MDFBs, it resulted in increasing the *MDFB width* for several subjects by including most of the pre-determined maximal band (i.e. 5-20Hz) in the MDFB.

Finally, despite these four cases, modifying the algorithm by putting constraints led to better or similar online performances in most cases (N=14) . This observation may suggest that adding constraints could be a solution to help most subjects learn how to control a BCI but not all of them. Is it possible to find other characteristics based on the MDFB to predict online performances? Could we use these characteristics to automatically select the optimization algorithm to use during the calibration for each subject? These questions are investigated in the next section.

3 Gaining a better understanding of the relationship between MI-BCI performances and subject-specific frequency band characteristics

3.1 Objective

In the two previous sections, we showed that even though there was a correlation between the selected user-specific frequency band characteristics (i.e., the mean value and the width of the MDFB selected) and the classification accuracy, adding constraints to the frequency band selection algorithm did not statistically improve user performances, even though this led to similar or better performance for 78% of the subjects. The reasons for this variability being unclear, we led additional analyses in order to:

- identify new predictors of MI-BCI performances originating from both constrained and unconstrained algorithms.
- create a statistical model that can predict MI-BCI performances based on characteristics of the MDFB.
- create a statistical model to select the most relevant algorithm for each subject.

3.2 Materials and Methods

3.2.1 Datasets

For this study we will use Datasets A and B where the frequency band selection algorithm was used for online experiments, during the calibration before using the CSP algorithm. In Dataset A the unconstrained algorithm was used whereas in Dataset B, we used the constrained algorithm.

3.2.2 Variable and Factors

To identify new predictors, we used the signal representing the sum of the two $fscore$ coefficients (of the two Laplacian channels) at each frequency, between the band power of each trial and its class label used to select the MDFB.

Then, as the MDFB can only be chosen in one specific band, we modeled each $fscore$ curve as a Gaussian function g of the frequency f, by using non-linear least squares to fit the function g to $fscore$:

$$g(f) = ae^{\frac{-(f-f0)^2}{b}} \tag{3}$$

where the coefficient a represents the maximum value of the g curve, $f0$ the mean value of the Gaussian g and b is a representation of its variance.

In the case in which the optimization algorithm did not converge (18% of the subjects), we enforced default values for a (we used the maximum value of $fscore$ in the MDFB) and $f0$ (we used the MDFB mean) and used the optimization algorithm to obtain b.

Coefficient a allowed us to assess the $fscore$ variability within the MDFB. Indeed, the more stable and high the $fscore$ signal within the MDFB, the higher a. Coefficient $f0$ gave information about the most discriminant frequency and b about the $fscore$ signal distribution.

Coefficients a, $f0$ and b were used as predictors, together with the MDFB width and mean obtained using the algorithm used online (cf Fig. 4).

3.2.3 Analyses

statistical model to predict/explain MI-BCI performances: We used a LASSO [3] regression to obtain models that could predict the performances of MI-BCI users from the characteristics of the MDFB. The LASSO regression uses an L1 norm regularization with a penalty parameter λ (see Eq. 4) that promotes sparse solutions, i.e., that selects only a small number of variables (many coefficients will be zero using this regularization). It is particularly adapted to reduce the number of relevant features when those features are more numerous than the training data [30] and enables the creation of interpretable models. As for any linear regression set up, we have a continuous output vector $Y \in \mathbb{R}^n$ (here the MI-BCI performance to be explained/predicted), a matrix $X \in \mathbb{R}^{n \times p}$ of p normalized features using a z-score normalization (here the users' MDFB characteristics) for n examples (the subjects) and a coefficient vector $\beta \in \mathbb{R}^p$ (the regression weight). The LASSO estimator is defined as:

$$\beta_{lasso} = \text{argmin}_{\beta \in \mathbb{R}^P} ||Y - X\beta||_2^2 + \lambda||\beta||_1 \tag{4}$$

where, $||u||_2^2 = \sum_{i=1}^{n} u^i$ for $u \in \mathbb{R}^n$ and $||\beta||_1 = \sum_{j=1}^{p} |\beta_j|$. For some values of the penalty

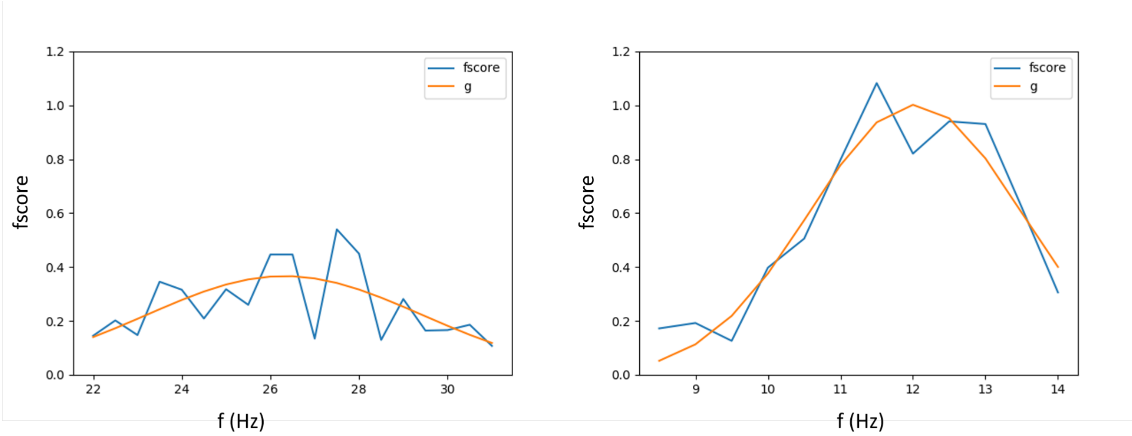

Figure 4: Two illustrations of the computation of the new performance predictors. The plots depict the $fscore$ of two participants (solid blue line) and the fitted values $g(f; a, f0, b)$ (solid orange line).

parameter λ, some components of β_{lasso} will be set exactly to 0. Once β_{lasso} obtained, the MI-BCI performances of the i^{th} user, $Perf^{i}_{pred}$, can be predicted from this user's MDFB characteristics x_i, as $Perf_{pred} = \beta^T_{lasso} x_i$ (T denotes transpose). In order to evaluate the stability and reliability of the different models, we used a leave-one-subject-out cross validation process. We also used an inner cross-validation (total number of training subjects: N-2) to automatically find the optimal λ among 100 possible $\lambda \in [0.001; 10]$, i.e., the one that minimizes the mean absolute inner-cross validation error and provides us with sparse features. We used this optimal λ to build a regression model and then, the outer cross-validation (N-1 subjects for training, 1 for testing) is used to evaluate this model.

For each cross-validation, different features were selected with different weights. Studying these features and weights first allowed us to assess the stability of the results. In a second step, we determined the reliability of the models by testing each of them on the participant not included in the training set during the cross-validation process. We then computed the mean absolute error of all the models:

$$\sum_{i=1}^{N} \frac{|Perf_{pred(i)} - Perf_{real(i)}|}{N} \tag{5}$$

n being the total number of models generated for the group.

In order to determine whether the models could make better-than-chance predictions, we estimated the empirical chance level in terms of mean absolute error, given our data. First, we randomly permuted the mean MI-BCI performances of the training set, while keeping features identical, thus breaking the relationship between predictors and performance. Then, we used the LASSO as explained above to predict the MI-BCI performance of the left-out subject. We repeated this process 1000 times and stored each mean absolute error to obtain the distribution of the chance prediction performance. Then, we sorted those values in descending order and the 99^{th} and 95^{th} percentiles were used to identify the chance level for the mean absolute error for $p = .01$ and $p = .05$ respectively.

statistical model to select the most relevant algorithm for each subject For this analyse, we used M=20 pairs of subjects from Datasets A* and B (see Table. 1). Subject A52 paired with subject B06 were removed as their online performance were exactly the same in the two conditions (i.e., with constrained and unconstrained algorithm). This subject pair was thus not informative to predict which algorithm to use. In order to build a classification model to automatically select the most suitable algorithm for each subject, we used decision trees [2] to obtain models that could predict the algorithm (constrained or unconstrained) with which a user reaches the highest MI-BCI performances. We only used a as predictive variable among the ones available before the use of the MDFB selection algorithm (i.e., a, $f0$, b) as according to our results below it was selected in 100% of the generated models with a strong positive weight. In order to evaluate the reliability of the different models, we used a leave-one-subject-out cross validation process. We also used an inner cross-validation (total number of training subjects: M-2) to automatically select hyper-parameters (i.e., the maximum depth of the tree and the minimum number of samples required to be at a leaf node). For each cross-validation, we compared the online performances of subjects from Dataset A* (all using the unconstrained algorithm) with performances obtained using the algorithm (either constrained or unconstrained) automatically selected by the decision tree model. We used the Wilcoxon signed-rank test to compare the two groups of performances. We also built a classification tree trained on all 20 subjects from Datasets A*, for interpretation purposes, in order to determine the critical values of a that could give information about which algorithm to use.

3.2.4 Results

Statistical model to predict/explain MI-BCI performances: Our results revealed a model that predicted performances better than chance (p<0.01) with a mean absolute error of 10.56%. The mean absolute error of the random model (generated through the permutation process) at $p = 0.01$ is 17.3%.

The *MDFB width* and the parameter a of the function g were selected in 100% of the generated models with a strong positive weight in comparison to the other factors' weights (see Fig. 5). As a reminder, a represents the maximimun $fscore$ value of the fitted function g. Subjects with a higher parameter a and a higher *MDFB width* seem to obtain better performances than those with lower ones. In addition, the parameter $f0$ was selected in 93.5% of the generated models as well with a negative weight enhancing the hypothesis that subject with an MDFB mean value in the α-lowβ range tend to perform best. Indeed, as $f0$ represents the fitted Gaussian g mean value, its value should be close to the MDFB mean value.

statistical model to select the most relevant algorithm for each subject: The average online performances using the algorithm selected by the classification tree are higher (M=65.8 $\pm$ 17.4) than the average online performances using the unconstrained algorithm (M=62.1 $\pm$ 15.9), or the constrained algorithm (M=64.82 $\pm$ 15.9). Our analyses showed that the difference between performances obtained using the unconstrained algorithm and those obtained using the classification tree tend towards significance (W=22.5, p=0.054) but is not statistically significant at $\alpha = 0.05$. However, there is no statistical significance between online performances using the algorithm selected by the classification tree and performances using the constrained algorithm.

Our final classification model (see Figure 6) suggests that subjects with an intermediate a value (between 0.3 and 0.4) have higher online performances using the unconstrained algorithm whereas subjects with a low a value (lower than 0.3) or a high a value (higher than 0.4) perform better with the constrained algorithm.

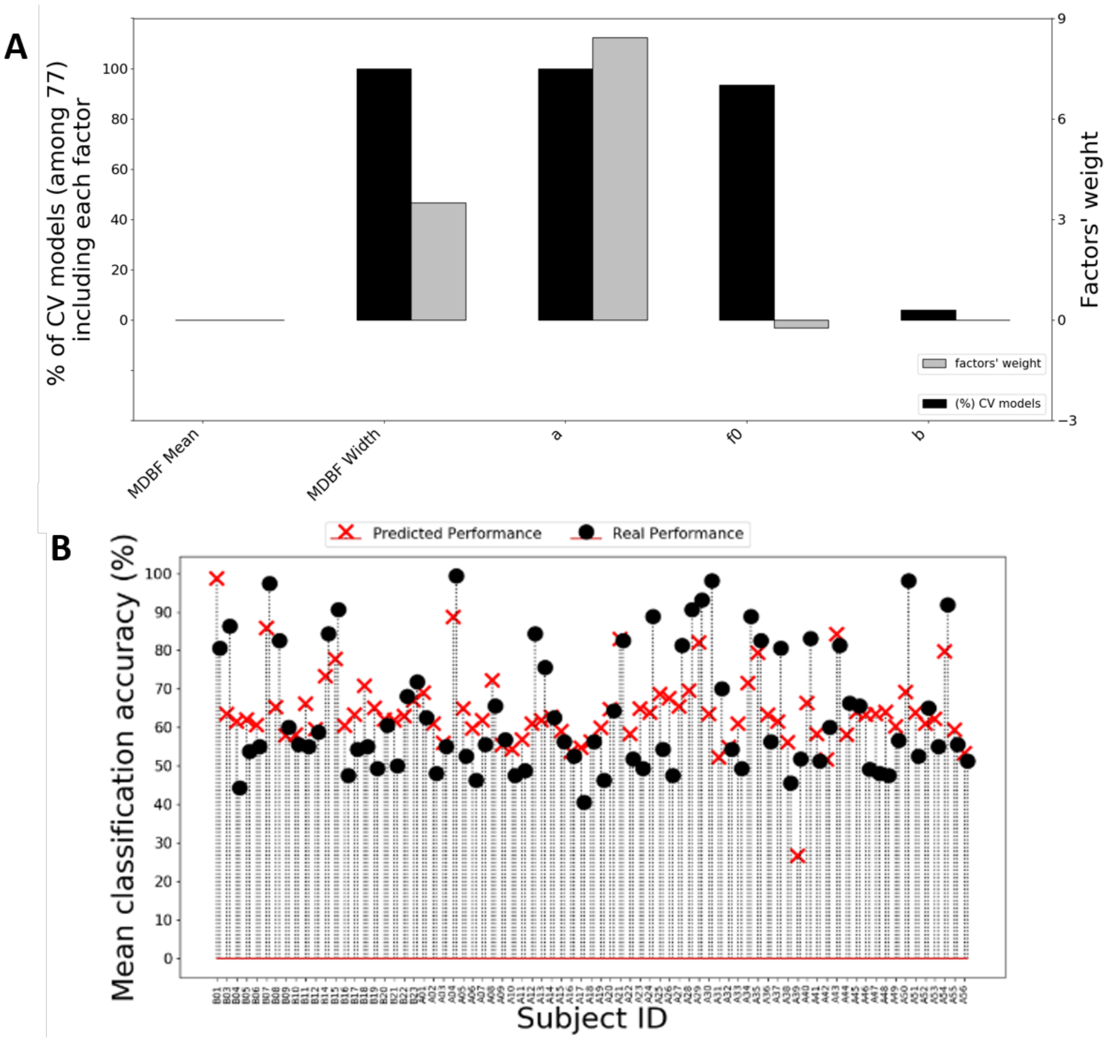

Figure 5: Results of the different models generated. (a) represents the percentage of Cross-Validation models including each factor (black) and their average weight (grey). (b) illustrates in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one).

3.2.5 Intermediate discussion

Our analyses suggested that it was possible to predict the MI-BCI classification accuracy of users using parameters extracted from the optimization algorithms: the frequency band selection algorithms. The prediction models included two main factors: The *MDFB width* and the parameter a of the function g and one other factor with a negative smaller weight $f0$. Results from Section 1 already suggested that subjects with a higher *MDFB width* tend to perform better than subjects with a lower one. It also suggested that subjects with an MDFB mean value above 16 Hz (in the β band) seem to have lower performances, which is confirmed by the negative weight of $f0$ in all our models. Therefore, both those results indicate that the *MDFB width* and *mean* (or $f0$) could be a good predictor of MI-BCI performances. In addition, all models included the parameter a of the function g with a strong positive

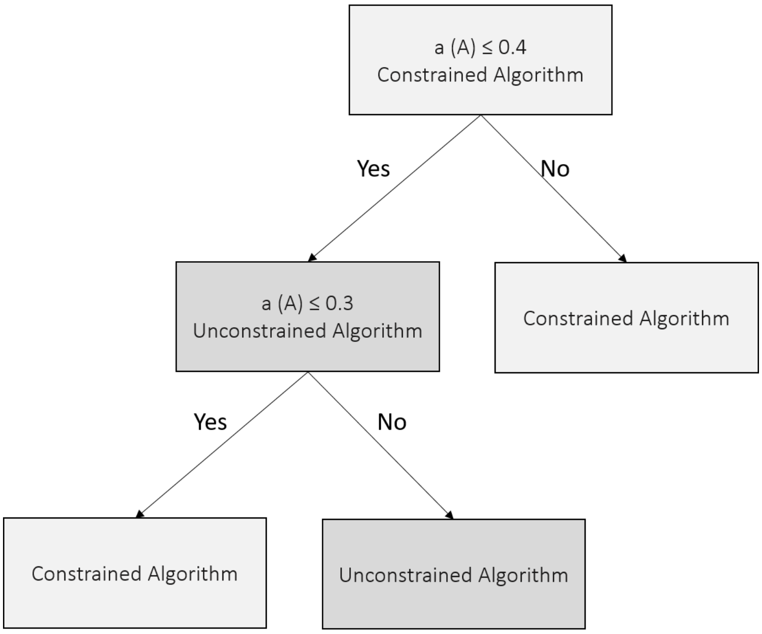

Figure 6: Representation of the decision tree.

weight. As a represents the ability to produce strong and distinct EEG oscillations in the sensorimotor cortices, we believed that it could be interesting to explore the value of this parameter during the calibration of the system in order to refine the algorithm selection. We used parameter a to study the possibility of having a classification model that could select the most optimal algorithm (constrained or unconstrained) for each subject. Our results suggest that subjects with a low parameter a (i.e., subjects that have difficulties producing strong and distinct signals) could benefit a little from the added constraints (an average an improvement of 6%). This may suggest that for those subjects with very weak SMR modulations, using default values (here our constraints), might be better than letting the classifier learn from data, as there is nothing clear in the data to learn from. Such a small improvement might also suggests that in order to help these subjects to control an MI-BCI, another approach is to be considered (e.g. a longer and/or more appropriate user training [34]). Subjects with a high parameter a who can produce strong and distinct signals seem to have higher online performances using the constrained algorithm with an increase of 15% in online performances. For those subjects, adding constraints either selected an MDFB in the α low-β band while it was not the case with no constraints, or it widened the MDFB which might lead to a better learning of BCI control (at it may allow for more flexibility about which frequency to modulate, as compared to a narrow band). For subjects with an intermediate a value, we observed that adding constraints lowered their online performances by 15%. This final result suggests that when parameter a is not at an extreme, constraints might have a negative effect on performances. When looking at the MDFB characteristics of those subjects, we observed that subject A29 (paired with B18) who was able to produce discriminant signals in the high β band (22–31 Hz) was included in that group. When removing this pair, adding constraints lowered the online performances of this group by 2% only (versus 15% when including this subject). As we only have one subject in this situation (i.e., producing discriminant signals in the high β band), it is difficult to conclude for this group.

Finally, we find it important to note that our classification model was build on a small number of subjects (i.e., 20) and that our results do not allow us to conclude about a statistical improvement of online performances using the algorithm chosen by the model. Although our results do not allow us to conclude, we believe that such an investigation into data-driven BCI methods could be applied to other BCI algorithms, e.g., spatial filters, in the future, to better understand and possibly improve them.

4 Limitations

We hope this work represents a significant step towards the study of data-driven methods to better understand performances of MI-BCI users' performances. Nevertheless, some general limitations have to be mentioned. First, for Section 2 the different values of the thresholds for the constraints were fixed visually, i.e., manually in order to best delineate the performance clusters. We used this method as we had a small number of subjects and the clusters were visually identifiable. Therefore, in the future, such thresholds could be more precisely defined by using a larger number of subjects. In addition, we could also explore using unsupervised machine-learning algorithms such as clustering methods to obtain statistically more precise and relevant thresholds. In addition, concerning the prediction of BCI performances, we believe it is relevant to note that a large majority of our subjects (52%) had performances under chance level and would thus be considered users with BCI illiteracy/deficiency issues [33]. This can be explained by the fact that in those experiments there was only one BCI session. Therefore we do not know if those subjects could improve with more sessions (and thus would not suffer from BCI illiteracy/deficiency anymore), if their low performances were due to their current mental state which could impact performances [28] or if it were due to the training protocol, i.e., the standard Graz training protocol [7] which is known to be suboptimal, both theoretically [24, 27] and practically [29]. Finally, aiming at predicting performances that are not better than chance could be discussed. However we believe that, as there is a margin of error and statistical variability concerning the estimated online performances, we believed that we could still obtain interesting information about performances using all users, including even those with performances under chance level. Alternatively, rather than predicting users' performances using regression methods, future work could rather aim at predicting users with BCI illiteracy/deficiency versus other users, using classification methods.

5 Conclusion

In this study, we used the data from two different experiments to study the impact of the MDFB selected by an optimization algorithm during the calibration of an MI-BCI system. The objective was to determine if there was a relationship between characteristics of the chosen MDFB and subjects' online performances (i.e., classification accuracy), the causality of their relationship, if any, and if we could find a parameter that could be used to refine the algorithm to optimize it further. Our analyses suggested that there is a correlation between the selected user-specific frequency band characteristics (i.e., the mean value and the width of the MDFB selected) and the classification accuracy. To study a possible causality link between them, we added constraints to the algorithm, enforcing characteristics associated to higher performances (based on our first results), and used it in an experiment specifically designed for this study (**Dataset B**). We then compared the online performances obtained for subjects in this experimental group, **Dataset B** (using the constrained algorithm) with online performances of matched subjects in a control group Dataset A* from **Dataset A** (using the unconstrained, i.e., original algorithm). This study did not allow us to conclude on a causality link. Indeed, while average online performances obtained with the constrained algorithm were slightly higher than those obtained with the unconstrained algorithm, this difference did not reach significance.

Finally, in a last part we were able to predict the MI-BCI classification accuracy of users using parameters extracted from the MDFB optimization algorithms. To build our models, we added three parameters in addition to the *MDFB width* and *mean*. We modeled each correlation signal as a Gaussian function of the frequency and used the parameters of the Gaussian model as predictors (i.e., its mean and variance and the maximum value of the curve).

The prediction models included three factors: The *MDFB width* and the maximum value of the Gaussian model with strong weights compared to other factors, and Gaussian mean value with a smaller negative weight. This result reinforced our previous results, i.e., subjects with a higher *MDFB width* tend to perform better than subjects with a lower one and subjects with an *MDFB mean* value above 16 Hz (in the β band) seem to have lower performances. In addition, we were able to find a new predictor of MI-BCI performances: the maximum value of the optimized Gaussian function. In a last part, we decided to use this coefficient to chose a subject-specific algorithm (constrained or unconstrained) in order to possibly increase online performances. Our results revealed that using a constrained algorithm could help improve online performances of subjects with either strong and distinct EEG signals or no distinct EEG signals. All together, our studies revealed that studying features extracted from data-driven methods could be interesting to better understand why some subjects have difficulties controlling a BCI. It could also give an indication about the path to follow to adapt and improve these methods by adding relevant constraints. Finally, by combining features extracted from data-driven methods with user-related features, we believe that we could have a better understanding of MI-BCI performances. Indeed, in the literature, other studies used performance prediction models in order to better understand the users' performances and to identify user-related factors influencing it. Factors were related to users' stable characteristics (i.e., traits) [26, 32] or users' neurophysiological characteristics [22, 35]. In studies using users' traits, models reach, an error of less than 5% in specific conditions (high number of BCI sessions) and did not succeed in predicting performances better than chance in other studies. In studies focusing on users' neurophysiological characteristics for instance, the model of Tzdaka et. Al [35] reached an error of 12.4%. The model of Ahn et .al [22] reached a root mean square error (RMSE) of 0.101. In comparison with those studies, characteristics extracted from data-driven methods could be informative enough to better understand BCI performances. Finally, to our knowledge, before ours, there was no study exploring the relationship between data-driven algorithm results and BCI performances. Therefore, future works could consist in exploring similar approaches for other machine learning algorithms used in BCIs. For instance, many other algorithms are available to extract subject-specific discriminative frequency components to improve discrimination between two MI classes. For instance, in the literature, Common Sparse Spectral Spatial Pattern [9], Sub-Band CSP [14], Filter Bank CSP [15] and Discriminative FBCSP [18] have been proposed for selecting the optimal frequency band automatically. In addition other machine-learning methods such as spatial filters or classifiers could be studied to identify various characteristics of successful spatial filters and classifiers, and possibly improve such algorithms by enforcing these characteristics, if they are causally linked to higher performances.

6 References

Bibliography

[1] G. Pfurtscheller and A Aranibar. “Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement”. In: *Electroencephalography and clinical neurophysiology* 46.2 (1979), pp. 138–146.

[2] J. R. Quinlan. “Learning efficient classification procedures and their application to chess end games”. In: *Machine learning*. Springer, 1983, pp. 463–482.

[3] R Tibshirani. “Regression shrinkage and selection via the lasso”. In: *J R Stat Soc Series B Stat Methodol* (1996), pp. 267–288.

[4] R. Hari and R. Salmelin. “Human cortical oscillations: a neuromagnetic view through the skull”. In: *Trends in neurosciences* 20.1 (1997), pp. 44–49.

[5] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw. “Mu and beta rhythm topographies during motor imagery and actual movements”. In: *Brain topography* 12.3 (2000), pp. 177–186.

[6] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. “Optimal spatial filtering of single trial EEG during imagined hand movement”. In: *IEEE transactions on rehabilitation engineering* 8.4 (2000), pp. 441–446.

[7] G. Pfurtscheller and C. Neuper. “Motor imagery and direct brain-computer communication”. In: *Proceedings of the IEEE* 89.7 (2001), pp. 1123–1134.

[8] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. “Brain–computer interfaces for communication and control”. In: *Clinical neurophysiology* 113.6 (2002), pp. 767–791.

[9] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller. “Combined optimization of spatial and temporal filters for improving brain-computer interfacing”. In: *IEEE transactions on biomedical engineering* 53.11 (2006), pp. 2274–2281.

[10] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio. “The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects”. In: *NeuroImage* 37.2 (2007), pp. 539–550.

[11] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller. “Optimizing spatial filters for robust EEG single-trial analysis”. In: *IEEE Signal processing magazine* 25.1 (2007), pp. 41–56.

[12] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller. “Optimizing spatial filters for robust EEG single-trial analysis”. In: *IEEE Signal processing magazine* 25.1 (2007), pp. 41–56.

[13] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. “A review of classification algorithms for EEG-based brain–computer interfaces”. In: *Journal of neural engineering* 4.2 (2007), R1.

[14] Q. Novi, C. Guan, T. H. Dat, and P. Xue. “Sub-band common spatial pattern (SBCSP) for brain-computer interface”. In: (2007), pp. 204–207.

[15] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. “Filter bank common spatial pattern (FBCSP) in brain-computer interface”. In: (2008), pp. 2390–2397.

[16] A Lécuyer, F Lotte, R. Reilly, R Leeb, M Hirose, and M Slater. “Brain-computer interfaces, virtual reality, and videogames”. In: *Computer* 41.10 (2008).

[17] G Müller-Putz, R Scherer, C Brunner, R Leeb, and G Pfurtscheller. “Better than random: a closer look on BCI results”. In: *Int J Bioelectromagn* 10 (2008), pp. 52–55.

[18] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang. “A new discriminative common spatial pattern method for motor imagery brain–computer interfaces”. In: *IEEE Transactions on Biomedical Engineering* 56.11 (2009), pp. 2730–2733.

[19] J. d. R. Millán, R. Rupp, G. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, et al. “Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges”. In: *Frontiers in neuroscience* 4 (2010), p. 161.

[20] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, et al. “OpenViBE: An Open-Source Software Platform to Design, Test and Use Brain-Computer Interfaces in Real and Virtual Environments”. In: *Presence: Teleoperators and Virtual Environments* 19.1 (2010), pp. 35–53.

[21] R. W. Schafer. “What is a Savitzky-Golay filter?[lecture notes]”. In: *IEEE Signal processing magazine* 28.4 (2011), pp. 111–117.

[22] 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”. In: *PloS one* 8.11 (2013), e80886.

[23] K. K. Ang and C. Guan. “Brain-computer interface in stroke rehabilitation”. In: (2013).

[24] F. Lotte, F. Larrue, and C. Mühl. "Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design". In: *Frontiers in human neuroscience* 7 (2013), p. 568.

[25] H. Yuan and B. He. "Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives". In: *IEEE Transactions on Biomedical Engineering* 61.5 (2014), pp. 1425–1435.

[26] C. Jeunet, B. N'Kaoua, S. Subramanian, M. Hachet, and F. Lotte. "Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns". In: *PloS one* 10.12 (2015).

[27] F. Lotte and C. Jeunet. "Towards improved BCI based on human learning principles". In: (2015), pp. 1–4.

[28] A. Myrden and T. Chau. "Effects of user mental state on EEG-BCI performance". In: *Frontiers in human neuroscience* 9 (2015), p. 308.

[29] C. Jeunet, E. Jahanpour, and F. Lotte. "Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study". In: *Journal of neural engineering* 13.3 (2016), p. 036024.

[30] V Fonti and E Belitser. "Feature selection using lasso". In: *VU Amsterdam Research Paper in Business Analytics* (2017).

[31] F. Lotte and C. Jeunet. "Defining and quantifying users' mental imagery-based BCI skills: a first step". In: *Journal of neural engineering* 15.4 (2018), p. 046030.

[32] C. Benaroch, C. Jeunet, and F. Lotte. "Are users' traits informative enough to predict/explain their mental-imagery based BCI performances?" In: 2019.

[33] M. C. Thompson. "Critiquing the concept of BCI illiteracy". In: *Science and engineering ethics* 25.4 (2019), pp. 1217–1233.

[34] A. Roc, L. Pillette, J. Mladenovic, C. Benaroch, B. N'Kaoua, C. Jeunet, and F. Lotte. "A review of user training methods in brain computer interfaces based on mental tasks". In: *Journal of Neural Engineering* (2020).

[35] 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), pp. 2490–2495.

[36] L. Pillette, A. Roc, B. N'kaoua, and F. Lotte. "Experimenters' Influence on Mental-Imagery based Brain-Computer Interface User Training". In: *International Journal of Human-Computer Studies* 149 (2021), p. 102603.