



HAL
open science

Why Standard Brain-Computer Interface (BCI) Training Protocols Should be Changed: An Experimental Study

Camille Jeunet, Emilie Jahanpour, Fabien Lotte

► To cite this version:

Camille Jeunet, Emilie Jahanpour, Fabien Lotte. Why Standard Brain-Computer Interface (BCI) Training Protocols Should be Changed: An Experimental Study. *Journal of Neural Engineering*, 2016. hal-01302154

HAL Id: hal-01302154

<https://inria.hal.science/hal-01302154v1>

Submitted on 13 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Why Standard Brain-Computer Interface (BCI) Training Protocols Should be Changed: *An Experimental Study.*

CAMILLE JEUNET^{1,2,*}, EMILIE JAHANPOUR^{1,2}, FABIEN LOTTE²

1 University of Bordeaux - Bordeaux, France - - Laboratoire Handicap & Système Nerveux

2 Inria Bordeaux Sud-Ouest - Talence, France - - Project-Team Potioc

E-mail: camille.jeunet@inria.fr

Abstract

OBJECTIVE While promising, ElectroEncephaloGraphy based Brain-Computer Interfaces (BCIs) remain barely used due to their lack of reliability: 15% to 30% of users are unable to control a BCI. Standard training protocols may be partly responsible as they do not satisfy recommendations from psychology. Our main objective was to determine in practice to what extent standard training protocols impact users' Motor-Imagery based BCI (MI-BCI) control performance.

APPROACH We performed two experiments. The first consisted in evaluating a standard BCI training protocol efficiency for the acquisition of non-BCI related skills in a BCI-free context, which enabled to rule out the possible impact of BCIs on the training outcome. Thus, participants (N=54) were asked to perform simple motor-tasks. The second experiment aimed at measuring the correlations between motor-task and MI-BCI performances. The 10 best and 10 worst performers of the first study were recruited for an MI-BCI experiment during which they had to learn to perform 2 MI-tasks. We also assessed users' spatial abilities and pre-training mu rhythm amplitude, as both have been related to MI-BCI performance in the literature.

MAIN RESULTS Around 17% of the participants were unable to learn to perform the motor-tasks, which is close to the BCI-illiteracy rate. It suggests that standard training protocols are suboptimal for skill-teaching. No correlation was found between motor-task and MI-BCI performance. However, spatial abilities played an important role in MI-BCI performance. Besides, once the "spatial ability" covariable had been controlled for, using an ANCOVA, it appeared that participants who faced difficulty during the first experiment improved during the second while the others did not.

SIGNIFICANCE These studies suggest that 1) standard MI-BCI training protocols are suboptimal for skill-teaching, 2) spatial abilities are confirmed to impact MI-BCI performances and 3) when faced with difficult pre-training, subjects seem to explore more strategies and therefore learn better.

1. INTRODUCTION

Brain-computer interfaces (BCIs) are communication and control systems which allow users to interact with the environment using their brain activity alone [47], which is often measured using Electroencephalography (EEG). A prominent type of BCI, called Motor-Imagery based BCIs (MI-BCIs), makes use of control signals sent via the execution of *motor-imagery tasks*, such as imagining hand movements. They are indeed very promising, in particular for the rehabilitation of stroke patients [3], for controlling assistive technologies such as neuroprosthetics or smart wheelchairs [31], or even for gaming for healthy users [23]. However, MI-BCIs remain barely used outside laboratories due to their lack of reliability [47]. Two main factors responsible for this low reliability have been identified. The first, extensively investigated, concerns brain signal processing with current classification algorithms being still imperfect [2]. The second concerns the users themselves: between 15% and 30% cannot control a BCI at all (so-called "BCI illiteracy" or

"BCI deficiency"), while most of the remaining 80% have relatively modest performances [2]. It is known that controlling an MI-BCI requires the acquisition of specific skills, and particularly the ability to generate stable and distinct brain activity patterns while performing the different MI-tasks [36, 46]. Just as with any skill, appropriate training is required to acquire BCI control [36]. Yet, current standard training protocols, which do not take into account the recommendations from psychology and instructional design (such as proposing adaptive and progressive tasks or explanatory, supportive and multimodal feedback), do not seem to be theoretically appropriate, and thus might be partly responsible for BCI illiteracy and modest user performance [27, 28].

However, while being instructive, insights such as those presented in [27, 28] only provide theoretical considerations about the flaws associated with training approaches used in MI-BCI that could be responsible for modest user performance and BCI illiteracy. It is therefore necessary to *concretely* assess whether standard training protocols used in MI-BCI paradigms are appropriate to train a skill, and how much they impact BCI performance and BCI skill acquisition. Moreover, it is necessary to perform this evaluation independently of MI-BCIs, to rule out possible biases due to BCI complexity, EEG non-stationarity and poor signal-to-noise ratio. Indeed, if a BCI training results in poor performances (i.e., the subject fails to obtain BCI control), this might not be due to the training protocol itself but simply to poor EEG signal processing, noisy or non-stationary signals, or to the fact that the relevant neural signals cannot be found in the EEG signals of the user due to the orientation of the user's cortex, for instance. Therefore, to study the impact and usefulness of a given training protocol, it is necessary to study it without the possible confounding factors originating from the BCI design.

Thus, the objective of this paper is to evaluate the efficiency of a standard training protocol [40] for the acquisition of MI-BCI related skills. In particular, we focused here on the MI-BCI training protocol proposed by the Graz group [40] which is a widely used BCI training protocol [36]. Moreover, most other existing MI-BCI training protocols can be seen as variants of the Graz training protocol as they use similar timings, feedback and training tasks, see, e.g., [4, 32].

In order to achieve this objective, two experiments were conducted. The first consisted in studying the efficiency of a standard MI-BCI training protocol [40] for skill acquisition in an MI-BCI free context: participants (N=54) were asked to learn specific and simple motor tasks, i.e., drawing circles and triangles on a graphic tablet, using this standard training approach¹ [40]. The second experiment aimed at studying the correlations between motor-task performance and MI-BCI performance. The 10 best and 10 worst performers of the first study were selected to participate in an MI-BCI experiment during which they had to learn to perform 2 MI-tasks: left- and right-hand movement imagination. We hypothesised that poor performers in the first experiment would also experience difficulties in the MI-BCI experiment, while the best performers of the first experiment would also perform well in the second.

In the following sections, we first present the details of the standard training protocol initially proposed by the Graz group [40], on which we based our study. Then we present a quick review of the literature that has been published on human training in BCI, and more particularly on the impact of user profile on performance and on the improvement of training protocols and feedback. Subsequently, both experiments are introduced and their results presented and discussed.

2. DESCRIPTION OF A STANDARD TRAINING APPROACH: THE GRAZ PROTOCOL [40]

This protocol was first proposed by the Graz BCI group as an alternative to the Operand Conditioning (OC) approach, enabling to provide the participants with a shorter training. Indeed, because in the OC approach the user has to adapt to the system, training can take several weeks.

¹Preliminary results (N=20 participants) of this first study have been published in a short conference paper [18].

In the Graz approach however, which is based on Machine Learning, the system adapts to the user, thus enabling training time to be reduced from weeks to few days [41]. The Graz protocol also has the specificity of being externally paced, since it is based on stimuli, and of being specific, since EEG is recorded on specific areas, i.e., most commonly over the sensori-motor cortex (while for the OC approach, undefined mental processes can be used for control). Indeed, the most used tasks in the context of the Graz protocol are motor-imagery tasks (such as the imagination of hand movements) which are known to be associated with an activation of the motor cortex. The Graz protocol is divided into two steps: (1) training of the system and (2) training of the user. During the first step, the user is instructed to perform several successive motor imagery tasks such as the imagination of left- and right-hand movements. From the recorded EEG signals collected during the different MI tasks, the system extracts characteristic EEG patterns which are specific to each MI task. These extracted patterns are then used to train a classifier the goal of which is to determine the class to which the EEG signals belong (i.e., imagination of left- or right-hand movements). For MI-BCI training protocols that last over several sessions (i.e., days), it is common to regularly retrain the classifier on newly acquired data in order to take into account cap variations and the condition/state in which the user is (which can change from one session to another). Step 2 consists in training the user. To do so, the user is instructed to perform the same MI tasks, but this time feedback (provided by the classifier, which was optimised in Step 1) is provided to inform the user which MI-task the system has recognised and how confident the system is that the task it has recognised is the one being performed by the user. Thus, the goal of the user will be to find strategies so that the system recognises the mental task he/she is performing. This training protocol is most often performed over different sessions divided into runs of approximatively 7 minutes each. One session typically includes 4 to 6 runs, in order to avoid the fatigue which is usually felt after more runs. Runs are themselves divided into trials, usually between 10 to 20 per class (i.e., per MI-task). One trial typically lasts 8s. At the beginning of each trial, a fixation cross is displayed to announce the start of the trial and to avoid eye movements during the following 2-second long rest period (which is usually used as a reference period for event-related synchronisation and desynchronisation calculation). Then, after 2s, a beep is used to trigger the attention of the user and prepare him/her for the oncoming instruction. One second later, at $t = 3s$, the instruction appears as an arrow the direction of which indicates the MI task to be performed, i.e., an arrow pointing left indicates a left hand MI and an arrow pointing right a right hand MI. From $t = 3.250s$, a feedback is provided for 4s in the shape of a bar the direction of which indicates the mental task that has been recognised and the length of which represents the confidence of the system in the recognition of the MI-task. This sequence of events is depicted in Figure 1.

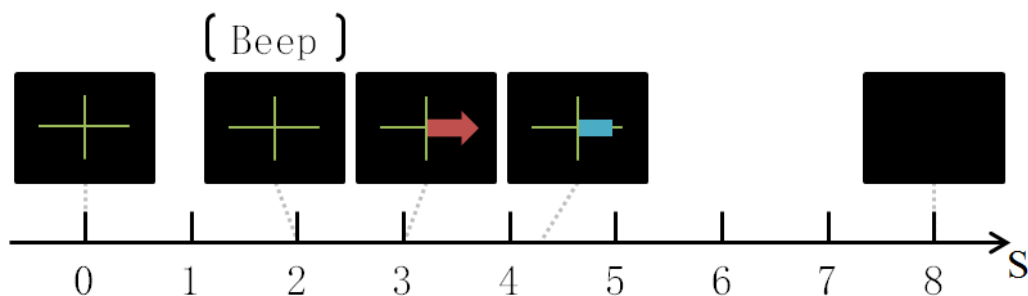


Figure 1: *Timing of one trial in the Graz Protocol.*

3. RELATED WORK - HUMAN TRAINING IN MI-BCI

Research about human training in MI-BCI mainly focuses on two aspects: (1) the influence of the user's profile on his MI-BCI control performance and (2) the enhancement of the *communication/comprehension* between the user and the system by improving the training protocols and feedback. A brief state of the art of these two points of interest is presented in this section.

3.1 Impact of the User's Profile on BCI performance

The important inter-subject and inter-session variability in terms of MI-BCI performance led the community to look for predictors of performance. Two main categories of predictors have been studied: neurophysiological and psychological predictors. A review of neurophysiological predictors is presented in [1]. Among them, a prominent one, whose usefulness has been reproduced and confirmed across multiple experiments [14, 15], is the predictor proposed by Blankertz et al. [5]. They showed that the amplitude of sensorimotor-rhythms (SMRs) at rest was a good predictor of subsequent BCI-performance in motor-imagery paradigms: a correlation [$r=0.53$] was found between this neurophysiological predictor based on the μ rhythm (about 9-14 Hz) over sensorimotor areas and MI-BCI performance ($N = 80$). Furthermore, concerning psychological factors, mood and motivation [39], as well as the locus of control score related to dealing with technology [8], have been shown to be correlated with MI-BCI performance. Fear of the BCI system has also been shown to affect performance [8, 38]. The study of Hammer et al. [14] revealed that attention span, personality and motivation played a moderate role for one-session MI-BCI performance, but a significant predictive model of performance, including visuo-motor coordination and the degree of concentration, was proposed. They tested this model in a 4 session experiment within a neurofeedback paradigm [15]. Results revealed that these parameters explained almost 20% of the BCI performance within a linear regression, even if visuo-motor coordination failed significance. Finally, we recently showed a strong and significant correlation between users' spatial abilities (assessed using the "mental-rotation test" [44]) and mental-imagery based BCI performance [$r=0.696$] [17, 19]. We also defined a predictive model of MI-BCI performance [$R_{adj}^2=0.809$] which included four parameters: tension, abstractness abilities, self-reliance and the "active/reflective" dimension of the learning style [19].

3.2 MI-BCI Training Protocols

Different training protocols have been proposed in the literature, most of them being based on the Graz protocol described here-above or being similar to it. They focus on improving either the instructions provided to the user at the beginning of the experiments, the training tasks proposed to the participant to control the MI-BCI, the feedback provided concerning the system's decision (i.e., about which MI task was recognised) or the training environment. Only two studies considered the instructions and showed that it is beneficial to incite the users to perform kinesthetic motor-imagery [37] and to not give them over-specific strategies so that their cognitive resources are not overtaxed [22]. Concerning the training tasks, several studies have proposed to use either progressive or adaptive tasks, instead of fixed and repetitive tasks, to increase performance [11, 35, 45]. The feedback certainly is the aspect of training protocols for which the most alternatives have been tested. Indeed, the bar representing the classifier output has been replaced by smileys [25] to increase motivation, or by auditory [13, 16, 30, 39] or tactile [9, 10, 20, 21, 24] feedback in order to reduce cognitive workload related to the overtaxed visual channel. While auditory feedback seems to be the best solution for Locked-In patients, tactile feedback appeared to be at least as efficient as visual feedback, and more efficient in interactive

contexts [20]. Finally, concerning the training environment, some authors have proposed to gamify the training process to increase motivation and improve user experience [30] and some have even integrated virtual reality (for a review, see [23, 26]), which appeared to efficiently improve performance.

It should be noted that a vast majority of the research about training protocols aims at improving the Graz training protocol or similar protocols. Similarly, a large proportion of the work performed to determine predictors of mental-imagery based performance are based on experiments in which either the Graz protocol, or protocols that can be seen as variants of it, were used. Yet, the efficiency of this type of training approach has not been questioned nor extensively evaluated. Thus, before improving these protocols, it would be worth testing their efficiency in terms of skill acquisition. This is what we aimed to do with the two experiments we present in the next sections.

4. EXPERIMENT # 1: USING A STANDARD MI-BCI TRAINING PROTOCOL TO LEARN TO PERFORM SIMPLE MOTOR TASKS

The objective of this first study was to evaluate the impact of a standard training approach (the Graz protocol, introduced here-above [40]) on participants' ability to acquire a skill in an MI-BCI free context.

4.1 Material & Methods

Participants were asked to learn to perform two motor tasks: drawing triangles and circles with a pen on a graphic tablet (see Figure 2), using the Graz protocol [40] (i.e., same instructions, tasks and feedback). As would have been the case in an MI-BCI training context, in which users have to learn a suitable movement imagination strategy, the participants here had to learn the strategy which allowed the system to correctly recognise their drawing, e.g. they had to identify the suitable shape size, angles and speed of drawing. The participants were divided into two groups: one used a "Standard" training approach [40] while the other one used a "Partially Self-Paced" BCI training approach, which provides the user with more autonomy. Indeed, with the standard approach, no autonomy is given to the user, who always has to perform the tasks required by the protocol. Yet, autonomy is known to increase motivation and learning efficiency in general [28]. Interestingly enough, the study described in [35] obtained promising results when providing more autonomy to a single BCI user. These two approaches are described here after.

4.1.1 Participants

54 BCI-naive and healthy participants (20 females; aged 25.1 ± 4.6 year-old) took part in this study, which was conducted in accordance with the relevant guidelines for ethical research according to the Declaration of Helsinki. All the participants signed an informed consent form at the beginning of the experiment.

4.1.2 Experimental protocol

Each participant (N=54) had to learn to do 2 motor tasks, namely to draw circles and triangles on a graphic tablet so that they were recognised by the system. The training session was divided into runs which were either standard (s) or self-paced (sp). S-runs were composed of 20 trials per task. At the beginning of each trial a green cross was displayed. After 2s, an auditory cue (a beep) triggered the attention of the participant towards the red arrow, which was displayed at

3s for 1s, and indicated which task the participant had to perform, i.e., draw triangles or circles continuously upon appearance of a right or left arrow, respectively. The mapping between the task (drawing circles or triangles) and the instructions/feedback (arrow/bar extending to the left or right) being incongruent, we helped the participants to remember it by providing them with a picture representing the cross of the Graz protocol with a circle on its left and a triangle on its right side. This picture was visible at all times to ensure subjects could refer to it whenever needed. At 4.25s, a blue feedback bar appeared and was updated at 16Hz for 3.75s. Its direction indicated the shape recognised by the classifier (left: circle, right: triangle) and its length was proportional to the classifier output. During sp-runs, no instructions were given: the participants were asked to do the motor tasks in an autonomous way, i.e., they could do the task they wanted to, whenever they wanted to.

All participants were provided with the following instruction: “Your goal is to find the right strategy so that the system recognises as well as possible the shape you are drawing, which will concretely correspond to having the feedback bar as long as possible in the correct direction: left for circles and right for triangles”.

Half the participants (N=27) were asked to learn using a Standard (S) training approach: they completed 4 seven-minute-long s-runs. The other half learned using the Partially Self-Paced (PSP) training approach: the 1st and 4th runs were s-runs, while the 2nd run was replaced by a 3.5 minute long sp-run followed by a shortened s-run (10 trials per task, 3.5 minutes), and the 3rd run was replaced by a shortened s-run followed by a 3.5 minute long sp-run. Total training duration was the same in both conditions. We studied the impact of the condition, S vs. PSP, on the recognition accuracy of triangles and circles by the system and on subjective experience (measured by a usability questionnaire -UQ-).

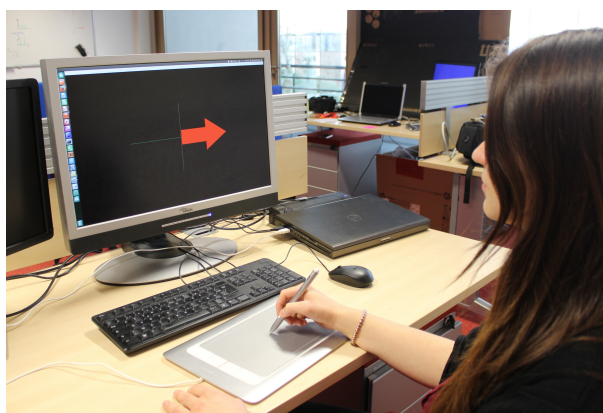


Figure 2: *Picture of a participant during the experiment. The instruction (red arrow pointing right) indicates that the participant has to draw triangles on the graphic tablet.*

4.1.3 Signal Processing

In order to discriminate triangular from circular pen movements on the graphic tablet, we used a pattern recognition approach as in BCIs. To this end, the 2D position of the pen on the tablet was acquired at a sampling frequency of 16 Hz. From the past 1s-long time window (in a sliding window scheme, with 1/16s step between consecutive time windows, with overlap) of the 2D pen position, a histogram of angles was computed. More precisely, the angles between each consecutive segment of the time window were first computed. Then the number of angles falling

in the ranges 0-30°, 30-75°, 75-105°, 105-150° and 150-180° were counted, and these 5 count values were used as input features for a Linear Discriminant Analysis (LDA) classifier. The (subject-independent) LDA classifier was trained on 60 trials from each movement, from 2 persons (1 left-handed, 1 right-handed). The resulting classifier could discriminate triangles from circles with 73.8% classification accuracy (10-fold cross-validation on the training set), which is an accuracy equivalent to the average accuracy of an MI-BCI [5]. The output of the LDA was mapped to the direction and length of the feedback bar, as in a typical MI-BCI.

Classically, subject-specific classifiers are used in BCI experiments. Nonetheless, here, the task being extremely simple, such a classifier would most likely have been perfect, i.e., with 100% classification accuracy, which is not the case in BCI experiments. We thus used a subject-independent classifier which enabled us to have a classification accuracy similar to that obtained for BCI. Furthermore, a subject-specific classifier would have added another bias to the training protocol evaluation as the obtained accuracy would also have depended on how well the two gestures were performed during the calibration run, and not only on the training protocol (instructions, tasks and feedback). Again, here we wished to isolate the training protocol in order to study it, hence the use of a subject-independent classifier (i.e., the same classifier for all), in order to obtain results that were independent from the classifier.

4.1.4 Analyses

To study how well subjects could learn the motor tasks, we measure their performance as the average classification accuracy obtained to discriminate triangular from circular pen movements, averaged over the whole feedback period, i.e., from $t=4.25s$ to $t=8s$ after the start of the trial. In order to analyse the interaction between the "Condition" (2 modalities: *S* and *PSP*; independent measures) and the performance obtained at each "Run" (4 modalities: *run1*, *run2*, *run3* and *run4*; repeated measures), we performed a 2-way ANOVA for repeated measures. Moreover, we asked the participants to complete a UQ which measured 4 dimensions: learnability/memorability (LM), efficiency/effectiveness (EE), safety and satisfaction. Thus, we did four one-way ANOVAs, each of them aiming at analysing the impact of the "Condition" on one "Evaluated Dimension" (4 modalities: *LM*, *EE*, *safety* and *satisfaction*; repeated measures).

4.2 Results

4.2.1 Performance analyses

Results (depicted in Figure 3) showed that 45 out of 54 participants managed to learn the task, i.e. obtained more than 70% average performance² -classification accuracy- [34] ($\bar{X} = 89.09\%$; $SD = 6.35$; $range = [72.84, 98.26]$) while 9 did not manage ($\bar{X} = 55.68\%$; $SD = 6.35$; $range = [50.23, 65.64]$). This rate of 16.67% of people who did not manage to learn is of the same order of magnitude as the BCI-illiteracy rate (between 15% and 30% [2]). Thus, one can hypothesise that BCI illiteracy could not only be due to the user, but also partly to the training protocol. Indeed, it has been hypothesized that BCI illiteracy/deficiency could be due to the user, who may generate noisy or non-stationary signals, who may have a cortex whose orientation prevents the relevant neural signals from reaching the scalp and thus EEG sensors, or who may fail to produce the desired EEG patterns [2]. Our experiment suggests that some subjects may fail to reach BCI control because the training protocol is not suited to everyone.

²This 70% accuracy is a threshold often used in the BCI community to distinguish subjects that achieved BCI control from those who did not achieve such a control [2]

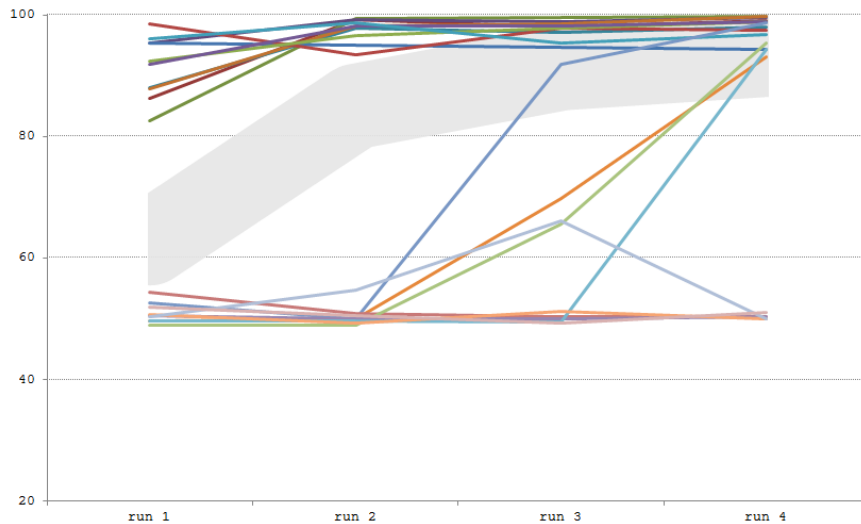


Figure 3: Graphic representing the performance of the participants (mean classification accuracy) as a function of the run. We chose to represent the 10 best and 10 worst performers -who took part in the next experiment-. The average performance of the 34 other participants is represented by the large grey line.

Furthermore, we performed a 2-way ANOVA for repeated measures to evaluate the impact of the *Condition* on motor performance according to the *Run* number. Checking the assumptions revealed that the normality [Skewness test - $s_{Run1}=-0.203$; $s_{Run2}=-1.295$; $s_{Run3}=-1.709$; $s_{Run4}=-1.961$] and equality of variance [Levene test - $p_{Run1}=0.044$; $p_{Run2}=0.024$; $p_{Run3}=0.160$; $p_{satisfaction}=0.128$] were not totally respected. Nonetheless, given that the results were close to the threshold and the ANOVA being a robust analysis [43], we decided to use this analysis. The 2-way ANOVA revealed neither a main effect of the *Condition* [$F(1,52)=1.997$; $p=0.164$] nor a *Condition*Run* interaction [$F(3,212)=1.301$; $p=0.259$]. However, it revealed a main effect of the *Run* [$F(3,50)=46.178$; $p\leq 0.001$]. Post-hoc analyses -Student t-test for paired samples- showed a significant increase in performance between the Runs #1 and #2 [$perf_{run1}=72.88\%$, $perf_{run2}=84.48\%$; $p\leq 0.001$] and between Runs #2 and #3 [$perf_{run2}=84.48\%$, $perf_{run3}=87.62\%$; $p\leq 0.005$] but not between Runs #3 and #4 [$perf_{run3}=87.62\%$, $perf_{run4}=89.11\%$; $p=0.277$].

4.2.2 Usability questionnaires

Each participant was asked to complete a usability questionnaire at the end of the experiment. This questionnaire measured four dimensions: learnability/memorability (LM), efficiency/effectiveness (EE), safety and satisfaction. Four one-way ANOVAs were performed to evaluate the impact of the *Condition* (S vs. PSP) on these dimensions. The prerequisites of the ANOVA were satisfied: all the dimensions had a normal distribution [Skewness test - $s_{LM}=-0.072$; $s_{EE}=0.046$; $s_{safety}=0.098$; $s_{satisfaction}=0.232$] and the variances were equal [Levene test - $p_{LM}=0.938$; $p_{EE}=0.415$; $p_{safety}=0.861$; $p_{satisfaction}=0.143$]. However, results revealed no effect of the *Condition*: LM [$F(1,53)=2.257$; $p=0.139$], EE [$F(1,53)=0.089$; $p=0.766$], safety [$F(1,53)=0.166$; $p=0.686$] and satisfaction [$F(1,53)=0.895$; $p=0.349$].

4.3 Discussion

The aim of this first study was to concretely assess whether training approaches used in BCI are appropriate to train a skill in general. Half the participants were asked to learn to perform simple motor tasks using a "Standard" (S) training approach while the other half used a "Partially Self-Paced" (PSP) one, in order to increase the feeling of autonomy. Results showed no differences between the conditions (S vs. PSP) in terms of performance or in terms of usability. This might be explained by the fact that most participants of the PSP group had found the right strategy, and thus had good performance, before the first sp-run. It might be that sp-runs could be useful for participants who still needed to explore strategies to find the right one. But once the right strategy found, sp-runs might not bring any further help to the participants. In future experiments, it could be worth modifying the protocol so that the sp-runs come earlier in the training.

A very relevant result is the fact that while a learning effect was noted for the whole group on average over the 4 runs, around 17% of the participants (9 out of 54) seemed unable to learn to perform the motor tasks (their performances were below 70% on average over the 4 runs). It is noteworthy that this rate is of the same order of magnitude as the BCI-illiteracy rate (between 15% and 30% - [2]). Thus, it seems most likely that a substantial proportion of illiterates are illiterate partly due to the training protocols given the fact that all subjects were cognitively able to understand the instructions and had the motor abilities to perform the tasks. This result emphasises the fact that such protocols should be improved to enable an efficient BCI training. In particular, numerous subjects reported verbally that the feedback was too poor as it did not indicate what they should do or change in order to succeed. It has to be noted that the poor performances of the participants might also be due to the difficulty of processing the mapping between the tasks and the protocol, i.e., drawing circles and triangles upon the appearance of a left- or right-facing arrow respectively. Indeed, the incongruence of this mapping could have led to a high workload and a low feeling of agency. In order to avoid such an effect, participants were provided with a picture representing this mapping which was available during the entirety of the experiment. Moreover, none of the participants reported difficulties in processing the mapping.

These results lead to two questions needing further investigations: (1) is the ability to learn using this kind of protocol correlated to some aspects of the user's personality, neurophysiological or cognitive profiles? and (2) are the performances obtained at these simple motor tasks predictive of MI-BCI performance?

Some aspects of these questions are investigated in the second study introduced here after.

5. EXPERIMENT # 2: INVESTIGATING THE RELATIONSHIP BETWEEN MOTOR PERFORMANCE AND MI-BCI PERFORMANCE

This second experiment aimed at investigating the relationship between the ability to learn to perform simple motor tasks (as done in the first experiment) and the ability to learn to control an MI-BCI using a standard training approach: the Graz protocol [40]. Indeed, we hypothesised that there could be a positive correlation between the performance obtained at the motor tasks (introduced in the Experiment #1) and MI-BCI performance. Indeed, we hypothesised that subjects who could learn motor tasks using the Graz training protocol would be likely to learn MI task using the same protocol as they managed to learn a skill using this approach. We also hypothesised that some aspects of the participants' profile would impact their MI-BCI performance. We focused on the two predictors which seemed to be the most reliable and adapted to our experiment context according to the literature, namely the spatial abilities and the Blankertz SMR-predictor. We thus selected the ten best and the ten worst performers from the first experiment, based on the averaged

classification accuracy they obtained, and asked them to take part in an MI-BCI experiment during which they had to learn to perform motor-imagery tasks, i.e. imagination of left- and right-hand movements.

5.1 Material & Methods

5.1.1 Participants

20 BCI-naive participants (10 females; aged 24.7 ± 4.0 year-old) took part in this second study, which was also conducted in accordance with the relevant guidelines for ethical research according to the Declaration of Helsinki. Participants were selected from the first experiment and divided into two groups, the *good* and the *bad* performers. The 10 best performers of the first experiment [$\bar{X} = 96.00\%$ of performance - classification accuracy; $SD = 1.13$] were in the *good* group while the 10 worst performers of the first experiment [$\bar{X} = 63.12\%$ of performance - classification accuracy; $SD = 11.54$] were in the *bad* group. These two groups happened to be composed of 5 women and 5 men each. Moreover, in each group, 7 participants were in the Standard (S) and 3 were in the Partially Self-Paced (PSP) Conditions during the first study. Considering the results of the first experiment as well as the distribution of the Conditions into the groups, we decided not to consider this variable (S vs. PSP) in this second experiment. In other words, the MI-BCI training only comprised standard runs.

5.1.2 Experimental protocol

Each participant (N=20) had to learn to do 2 MI-tasks, namely imagining left- and right-hand movements, so that they were recognised by the system. Participants first had to complete a "calibration" run which aimed at providing the system with examples of EEG patterns associated with each of the MI-tasks. This run and the whole classifier training process are explained here after (see Section 5.1.5). Then, as in the first experiment, user training lasted 4 runs, each of them being composed of 20 trials per task. As shown in Figure 1, at the beginning of each trial a green cross was displayed. After 2s, an auditory cue (a beep) triggered the attention of the participant towards the red arrow, which was displayed at $t = 3s$ for 1s, and indicated which task the participant had to perform (imagining right- or left-hand movements upon appearance of a right or left arrow, respectively). At 4.25s, a blue feedback bar appeared and was updated at 16Hz for 3.75s. Its direction indicated the imagined movement recognised by the classifier and its length was proportional to the classifier output. This was thus exactly the same training protocol as used in the first experiment.

Here as well, all participants were provided with the following instruction: "Your goal is to find the right strategy so that the system recognises as well as possible the motor task you are doing, which will concretely correspond to having the feedback bar as long as possible in the correct direction: left for left-hand and right for right-hand movements".

Added to these MI-tasks, participants were asked to complete a mental rotation questionnaire which is depicted in the next section.

5.1.3 Spatial Ability Assessment using the Mental Rotation Test

Participants were asked to complete the Mental Rotation test [44] - which is a validated "paper and pen" psychometric questionnaire assessing spatial abilities - at the beginning of the experiment. Spatial abilities have been related to mental-imagery based BCI performance [17, 19], but not yet to purely motor imagery-based BCI. This test is composed of two sets of 10 items. Each set has

to be completed in 3 minutes maximum. An item consists in a 3D shape on the left and four 3D shapes on the right of the page. Among the four 3D shapes, two are the same as the one presented on the left with a rotation of 60° , 120° or 180° around the vertical axis (see Figure 4). The other two are mirror-reversed and rotated images of the 3D shape on the left. For each item, the participant has to find the two 3D shapes that are the same than the one on the left (i.e., only rotated). Since there is a strong gender effect associated with this test (men usually perform better than women), we had to take the participants' gender into account in the analyses in order to study the impact of spatial abilities on performance.

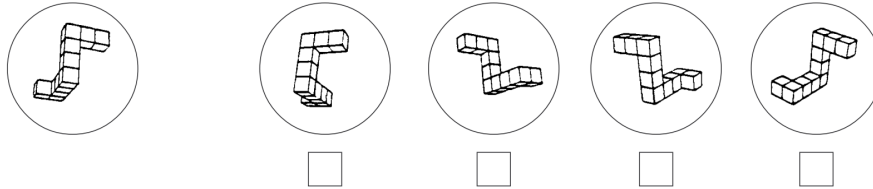


Figure 4: First item of the Vandenberg & Kuse Mental Rotation Test [44]

5.1.4 Blankertz's SMR-predictor of MI-BCI performance

The Blankertz SMR-predictor [6] is currently one of the most replicated and reliable neurophysiological predictors of MI-BCI performance (correlation of $r = 0.53$ with SMR performance over a large dataset, $N = 80$). It is computed from a 2 min baseline in a "rest with eyes open" state using two Laplacians over the motor cortex, i.e., C3 and C4. This predictor allows to quantify the potential for desynchronisation of the SMRs at rest, which can be used as an indicator of SMR strength during the performance of motor-imagery tasks. As no 2 minute-long baseline had been recorded with our protocol, we used the forty 3 second-long pre-trial time windows (3000 ms before the instruction) of each run, which gave a 2 minute-long time-window and enabled to compute the predictor on this sequence. More precisely, we computed the power spectrum of each 3 second-long time window, averaged these spectrums (i.e., over time windows), and computed the predictor on this averaged spectrum.

5.1.5 EEG Recordings & Signal Processing

The EEG signals were recorded from two g.USBamp amplifiers (g.tec, Graz, Austria), using 30 scalp electrodes (F3, Fz, F4, FT7, FC5, FC3, FCz, FC4, FC6, FT8, C5, C3, C1, Cz, C2, C4, C6, CP3, CPz, CP4, P5, P3, P1, Pz, P2, P4, P6, PO7, PO8, 10-20 system) [12], referenced to the left ear and grounded to AFz. Such electrodes cover the sensori-motor cortex, where EEG variations due to MI can be measured. EEG signals were sampled at 256 Hz. First, EEG signals were band-pass filtered in 8-30Hz (containing the sensori-motor rhythms) [40]. The first run (20 trials per MI task) was used to train the classifier. While 20 trials per class is not much, it has been shown to be sufficient to setup a motor imagery classifier [7, 29]. For instance, in [12], a successful mental imagery BCI classifier was setup with only 10 trials per class. Moreover, here we do not focus on the impact of the classifier but on the impact of the training protocol. Thus, at the end of the first run, which served for training the classifier, a Common Spatial Pattern algorithm [33] was used for each user on the collected data, to find 6 spatial filters whose resulting EEG power was maximally different between the two MI tasks. The spatially filtered EEG signal power (computed on a 1s time window, with 250ms overlap between consecutive windows) was used to train a

Linear Discriminant Analysis (LDA) classifier [33]. The LDA was then used online to differentiate between left- and right-hand MI during the 5 user-training runs.

5.1.6 Analyses

In this study, we analysed the effect of the "Group" of the first experiment (2 modalities: *good* vs. *bad*; independent measures), of the "Run" (4 modalities: *run1*, *run2*, *run3* and *run4*; repeated measures), of the "Mental Rotation Score" (continuous covariable) and of the "Gender" (2 modalities: *men* vs. *women*; independent measures) on participants' MI-BCI performance, that is to say their classification accuracy. We considered their "Gender" because of the important gender effect associated with the Mental Rotation test. Thus, we performed an ANCOVA with the "Mental Rotation scores" as the covariable and the "Group", the "Run" and the "Gender" as independent variables. We also studied the correlations between participants' MI-BCI performance and the Blankertz SMR-predictor.

5.2 Results

5.2.1 MI-BCI Performance

In our analysis aiming at evaluating the effect of the group (*bad* vs. *good* performers in the first experiment), gender (*men* vs. *women*) and run (*run1*, *run2*, *run3*, *run4*) on users' MI-BCI performance once the effect of the mental rotation had been controlled for, we considered two different measures of MI-BCI performance: (1) the **peak** classification accuracy (measured at the time window of the feedback period for which the classification accuracy over all trials is maximal), which is the typical performance measure used with the Graz protocol, see, e.g., [42] and (2) the **mean** classification accuracy over the whole feedback period of all trials. We thus performed two ANCOVAs. Note that the mean accuracy being the averaged accuracy over the whole feedback period, it is bound to be substantially lower than the usually reported peak accuracy, identified for the best time window. The mean accuracy is therefore a rather pessimistic performance estimate. We nonetheless believe it is useful as it reflects participant's ability to produce a long and stable BCI control signal.

PEAK PERFORMANCE

The average peak performance of the 20 participants was 66.95% (SD = 6.24; range = [57.09 ; 82.69]). Assumptions checking is depicted in Figure 5. It shows that the criteria for a normal distribution was satisfied for the mental rotation scores, for the peak performance of run1 and run4 but not for run2 and run3 (which was anecdotal, especially given the low number of subjects per group, and thus should not impact the analysis reliability [43]). The homogeneity of the regression slopes and the equality of variance criteria were satisfied. However, it has to be noted that the linearity criteria was not: which could also be explained by the important inter-run variability due to the small sample size. Indeed, when considering the mean performance over the four runs, a linear relation with mental rotation scores is revealed. The ANCOVA with the **peak** MI-BCI performance as the dependent variable revealed a main effect of *Mental Rotation Scores* [$F(1,15) = 6.991$; $p \leq 0.05$; $\eta^2 = 0.318$] as well as a strong tendency towards a main effect of the *Run* [$F(1,15) = 3.638$; $p = 0.076$; $\eta^2 = 0.195$]. However, neither a main effect of the *Group* [$F(1,15) = 0.388$; $p = 0.789$; $\eta^2 = 0.050$] nor a main effect of the *Gender* [$F(1,15) = 0.719$; $p = 0.410$; $\eta^2 = 0.046$] were revealed. Post-hoc analyses (Student t-tests for paired measures) about the tendency towards a Run main effect revealed a significant increase in peak performance on average between the first and the last runs [$t = -2.360$; $p \leq 0.05$] thus suggesting a learning effect. The ANCOVA also revealed significant

interactions. First, a *Run * Mental Rotation Scores* interaction [$F(1,15) = 6.269$; $p \leq 0.05$; $\eta^2 = 0.295$] suggesting an impact of mental rotation on the ability to improve in terms of performance across the runs. Second, a *Run * Gender* interaction [$F(1,15) = 7.936$; $p \leq 0.05$; $\eta^2 = 0.346$] (see Figure 6) which suggests that, if we consider performance independently from participants' spatial abilities, while men's MI-BCI performances were stable across the 4 runs, women's increased significantly. Furthermore, the *Run * Group* interaction [$F(1,15) = 4.907$; $p \leq 0.05$; $\eta^2 = 0.246$] revealed that, again if we consider performance independently from participants' spatial abilities, participants from the *good* group performed better than those of the *bad* group in the first run, but then they did not improve while participants from the *bad* group improved in terms of performance (see Figure 7). Finally, this ANCOVA revealed a strong tendency towards a *Run * Gender * Group* interaction [$F(1,15) = 4.221$; $p = 0.058$; $\eta^2 = 0.220$] (see Figure 8) but no *Gender * Group* interaction [$F(1,15) = 2.982$; $p = 0.105$; $\eta^2 = 0.166$].

	NORMALITY	LINEARITY	HOMOGENEITY OF REGRESSION SLOTS	EQUALITY OF VARIANCES
MENTAL ROTATION SCORES	s = 0.049		P = 0.155	
RUN 1	s = 0.676	P = 0.090		P = 0.261
RUN 2	s = 1.259	P = 0.183		P = 0.410
RUN 3	s = 1.138	P = 0.508		P = 0.069
RUN 4	s = 0.421	P = 0.174		P = 0.203

Figure 5: Table representing the assumptions checking for the ANCOVA on peak performance.

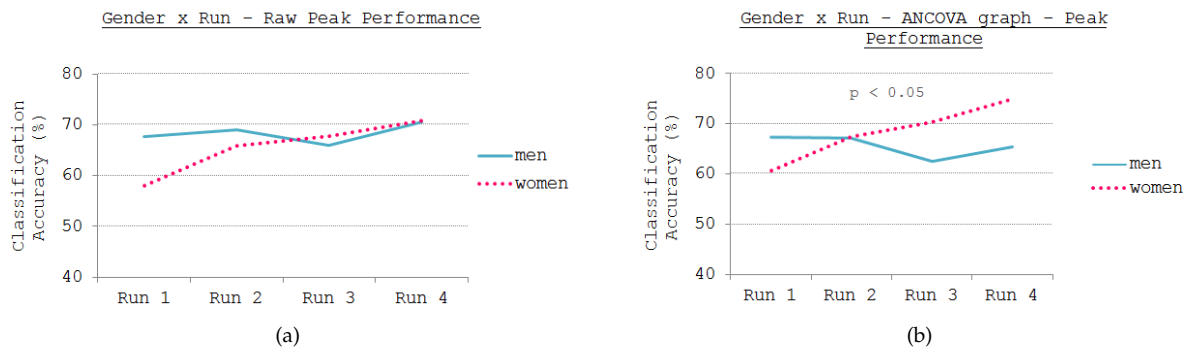


Figure 6: (a) Graph representing participants' raw MI-BCI peak performance (i.e., without the ANCOVA correction) as a function of their gender and of the run; (b) Graph representing the ANCOVA results for the Gender*Run interaction ($p < 0.05$), considering the Mental Rotation Scores as a covariable. When considering the performance independently from the mental rotation scores, women increase in performance across the 4 runs while men do not.

MEAN PERFORMANCE

The 20 participants obtained an average mean classification accuracy of 54.89% (SD = 6.56; range

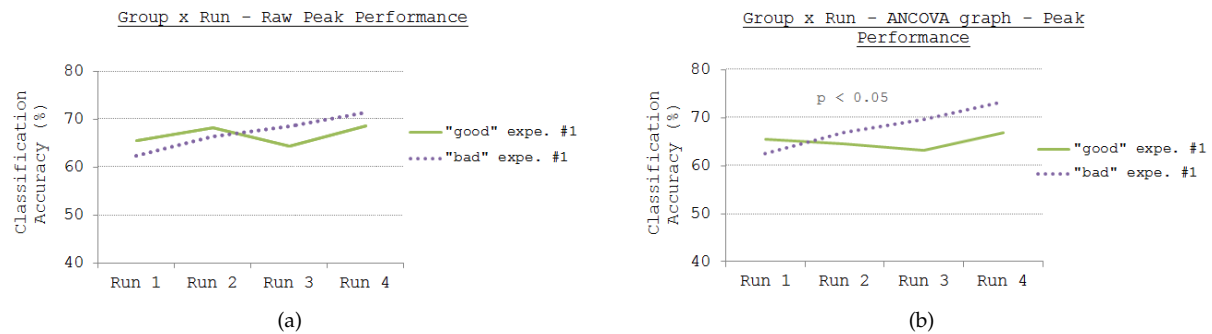


Figure 7: (a) Graph representing participants' raw MI-BCI peak performance (i.e., without the ANCOVA correction) as a function of their group from experiment 1 and of the run; (b) Graph representing the ANCOVA results for the Group*Run interaction ($p < 0.05$), considering the Mental Rotation Scores as a covariable. When considering the performance independently from the mental rotation scores, participants from the "good" group of the first experiment obtain stable performance across the four runs while participants from the "bad" group of the first experiment begin with lower performance but then improve and outperform the other group in the third and fourth runs.

= [46.41 ; 68.12]). As expected, this measure leads to much lower and pessimistic performance estimates. The analysis of the assumptions satisfaction for the ANCOVA are represented in Figure 9. Mental rotation scores as well as mean performance of run1, run2 and run4 satisfied the criteria for a normal distribution, but run 3 did not. As stated in the previous paragraph, this can be explained by the low number of participants per group and should not impact the analysis reliability [43]. Moreover, the homogeneity of the regression slopes as well as the equality of variance criteria were both satisfied. However, as was the case for the peak performance analysis, the criteria of linearity was violated which can be explained by the small sample size. The ANCOVA with the **mean** MI-BCI performance as the dependent variable was associated with quite similar results as for the peak performance. Indeed, it revealed a main effect of *Mental Rotation Scores* [$F(1,15) = 5.817$; $p \leq 0.05$; $\eta^2 = 0.279$] as well as a strong tendency towards a main effect of the *Run* [$F(1,15) = 4.100$; $p = 0.061$; $\eta^2 = 0.215$]. However, no main effect of the *Group* [$F(1,15) = 0.403$; $p = 0.535$; $\eta^2 = 0.026$] or of the *Gender* [$F(1,15) = 2.965$; $p = 0.106$; $\eta^2 = 0.165$] was revealed. Concerning the strong tendency towards a Run main effect, post-hoc analyses (Student t-tests for paired measures) revealed a significant increase of the mean performance between the first and the last runs [$t = -2.542$; $p \leq 0.05$] thus suggesting a learning effect, as it was the case in the peak performance analyses. This ANCOVA also revealed several significant interactions. First, there was a *Run * Mental Rotation Scores* interaction [$F(1,15) = 7.545$; $p \leq 0.05$; $\eta^2 = 0.335$]. Second, the *Run * Gender* interaction [$F(1,15) = 7.381$; $p \leq 0.05$; $\eta^2 = 0.330$] suggests that while men's MI-BCI performances (corrected so that they are independent from spatial ability scores) were stable across the 4 runs, women's increased significantly (see Figure 10). Furthermore, the *Run * Group* interaction [$F(1,15) = 6.376$; $p \leq 0.05$; $\eta^2 = 0.298$] revealed that, considering performance independently from participants' spatial abilities, participants from the *good* group obtained a better performance than those of the *bad* group at the first run, but then they did not improve while participants from the *bad* group improved in terms of performance (see Figure 11), as was the case with the previous ANCOVA. Finally, contrary to what we observed with peak MI-BCI performance, it revealed a strong tendency towards a *Gender * Group* interaction [$F(1,15) = 3.833$; $p = 0.069$; $\eta^2 = 0.204$] (see Figure 12) but no *Run * Gender * Group* interaction [$F(1,15) = 2.319$; $p =$

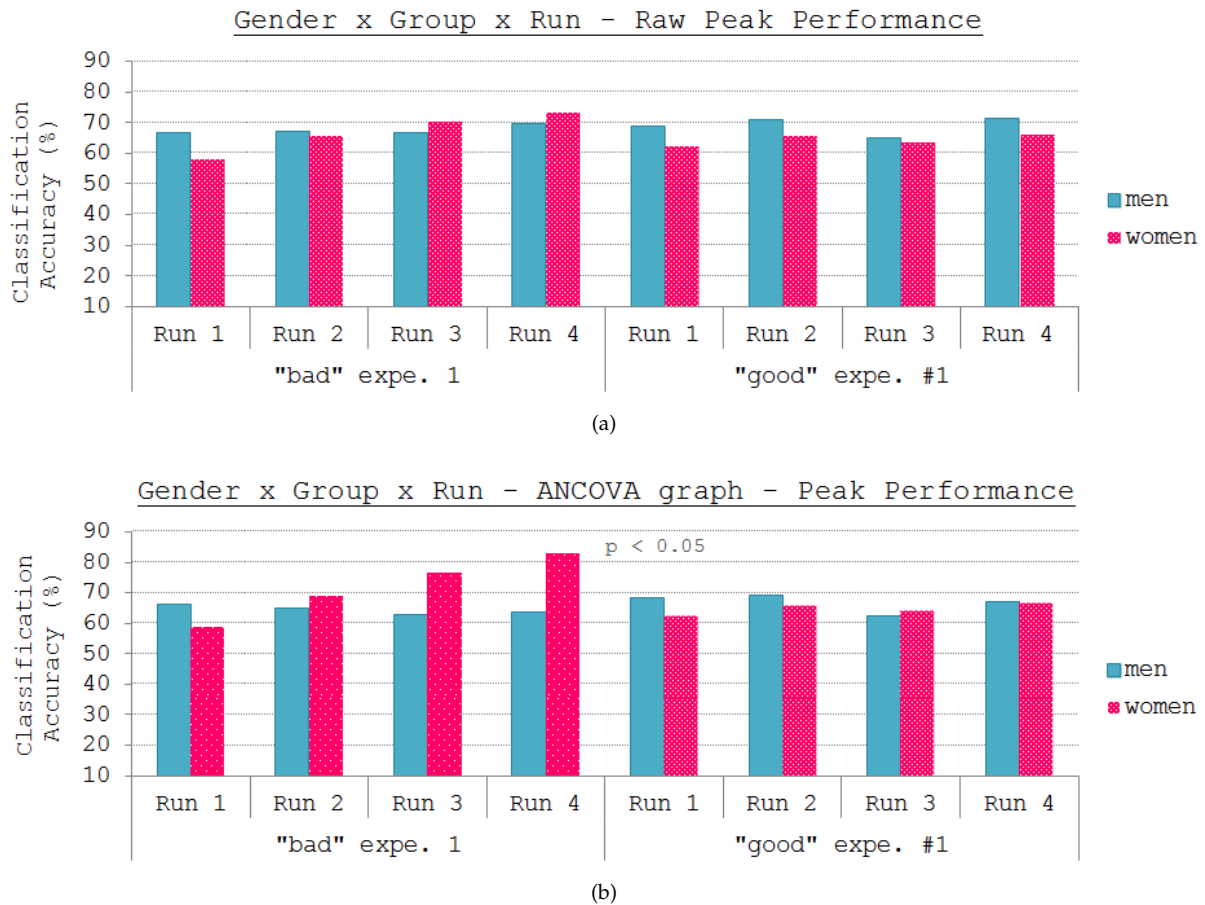


Figure 8: (a) Graph representing participants' raw MI-BCI peak performance (i.e., without the ANCOVA correction) as a function of their gender, of their group and of the run; (b) Graph representing the ANCOVA results for the Gender*Group*Run interaction ($p < 0.05$), considering the Mental Rotation Scores as a covariable. When considering the performance independently from the mental rotation scores, it can be noticed that women from the group "bad" [dark red on the left] improve in terms of performance across the runs while all the other participants do not.

0.149; $\eta^2 = 0.134$].

5.2.2 MI-BCI Performance and Spatial Abilities

First, there was a clear gender effect on the mental rotation score, consistent with the literature: $mean_{men} = 30.5 \pm 7.12 - mean_{women} = 20.7 \pm 7.21$ [t-test - $t = 3.058$; $p \leq 0.01$]. Then, both the ANCOVAs (on mean and peak performance) revealed the important impact of Spatial Abilities (SA) with a main effect of Mental Rotation scores on performance. Moreover, while Mental Rotation scores were not correlated with mean MI-BCI performance [$r = 0.266$; $p = 0.257$], they were correlated with the peak MI-BCI performance [$r = 0.464$; $p = 0.039$]. These results confirm the important impact of SA on MI-BCI performance which was demonstrated in [17, 19]. More specifically, the positive correlation indicates that people with better spatial abilities (i.e., higher mental rotation scores in this instance) obtain higher MI-BCI control performance.

	NORMALITY	LINEARITY	HOMOGENEITY OF REGRESSION SLOTS	EQUALITY OF VARIANCES
MENTAL ROTATION SCORES	$s = 0.049$		$P = 0.450$	
RUN 1	$s = 0.533$	$P = 0.159$		$P = 0.326$
RUN 2	$s = 0.344$	$P = 0.600$		$P = 0.700$
RUN 3	$s = 1.287$	$P = 0.684$		$P = 0.465$
RUN 4	$s = 0.577$	$P = 0.280$		$P = 0.767$

Figure 9: Table representing the assumptions checking for the ANCOVA on mean performance.

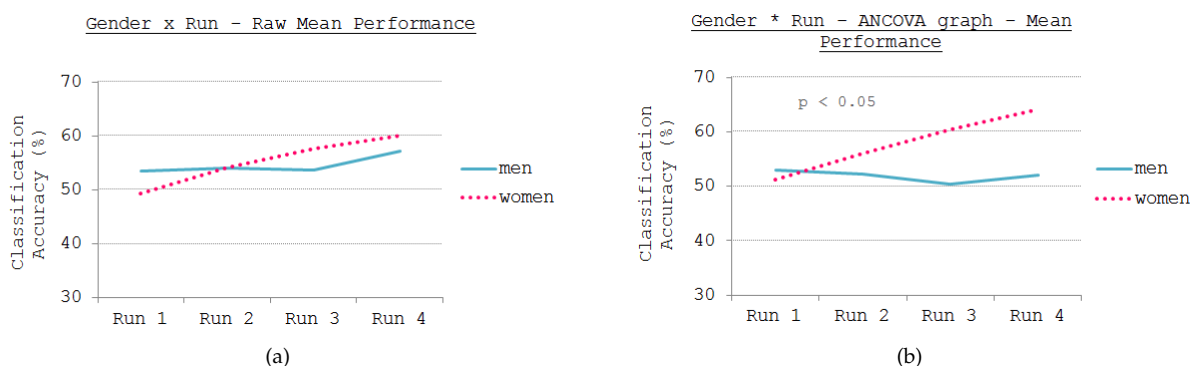


Figure 10: (a) Graph representing participants' raw MI-BCI mean performance (i.e., without the ANCOVA correction) as a function of their gender and of the run; (b) Graph representing the ANCOVA results for the Gender*Run interaction ($p < 0.05$), considering the Mental Rotation Scores as a covariable. When considering the performance independently from the mental rotation scores, women performances increase while men's do not.

5.2.3 MI-BCI Performance and the Blankertz SMR Predictor

We performed bivariate Pearson correlation analyses to assess the relation between users' **mean** and **peak** MI-BCI performance and the mean Blankertz's SMR predictor (averaged over the 4 runs). Results revealed no significant correlation between the predictor and the **mean** MI-BCI performance [$r = 0.151$; $p = 0.525$] nor with the **peak** MI-BCI performance [$r = 0.078$; $p = 0.743$].

5.2.4 Usability Questionnaires

We also evaluated the score associated with the four dimensions of the usability questionnaire [*learnability/memorability, efficiency/effectiveness, safety and satisfaction*] as a function of the participant's "Group" (*good vs. bad*), "Gender" (*men vs. women*) and of their "Mental Rotation Score". We thus performed four ANCOVAs. The prerequisite checking is depicted in Figure 13. The data satisfied the criteria for a normal distribution, homogeneity of the regression slopes and equality of variances.

However, it has to be noticed that the linearity criteria was not satisfied. No effect of the Group,

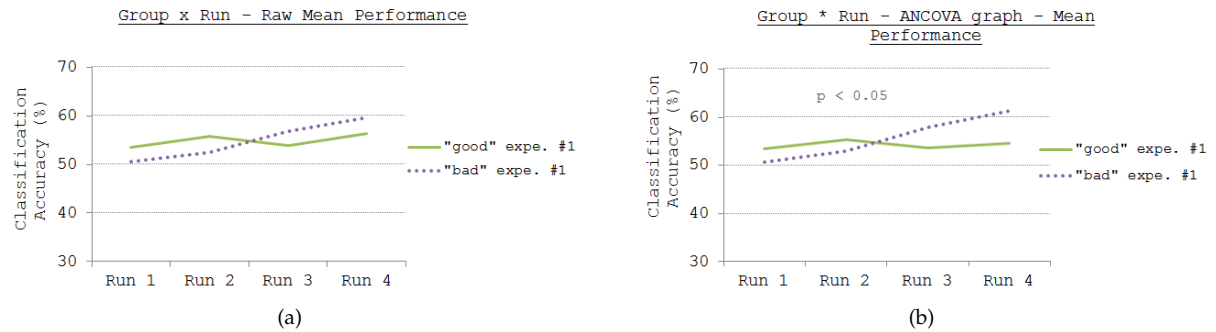


Figure 11: (a) Graph representing participants' raw MI-BCI mean performance (i.e., without the ANCOVA correction) as a function of their group from experiment 1 and of the run; (b) Graph representing the ANCOVA results for the Group*Run interaction ($p < 0.05$), considering the Mental Rotation Scores as a covariable. When considering the performance independently from the mental rotation scores, it can be noticed that participants from the group "good" of the first experiment obtain stable performance across the four runs while participants from the group "bad" of the first experiment begin with lower performance but then improve and outperform the other group from the third run.

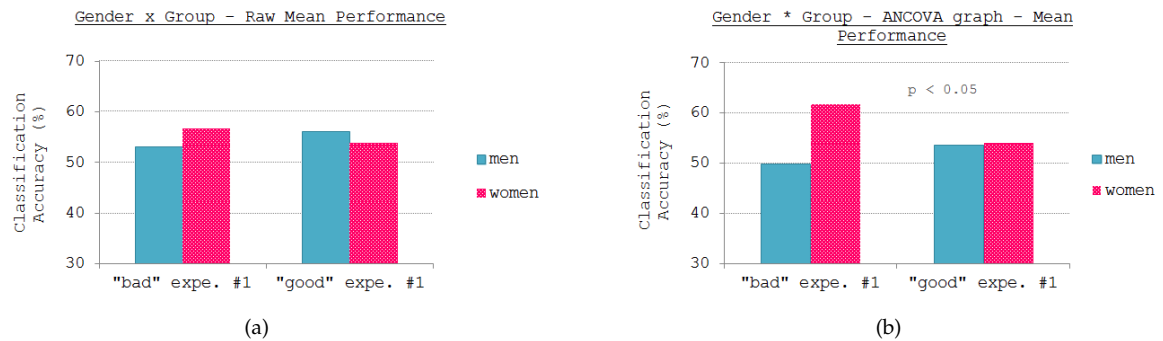


Figure 12: (a) Graph representing participants' raw MI-BCI mean performance (i.e., without the ANCOVA correction) as a function of their gender and of their group; (b) Graph representing the ANCOVA results for the Gender*Group interaction ($p < 0.05$), considering the Mental Rotation Scores as a covariable. When considering the performance independently from the mental rotation scores, it can be noticed that men from both groups ("good" and "bad") keep the same ratio at the second experiment: participants from the "good" group outperform the ones from the "bad" group. It is not the case for women. Indeed, while women from the "good" group obtain similar performance to men of their group, women from the group "bad" outperform all the other participants.

of the Gender nor an interaction of both was revealed for the *Learnability/Memorability*, the *Safety* and the *Satisfaction* dimensions. For the *Efficiency/Effectiveness* dimension however, two strong tendencies were revealed: a tendency towards a main effect of the group [$F(1,19) = 3.508$; $p = 0.081$; $\eta^2 = 0.190$] and towards a group*gender interaction [$F(1,19) = 3.439$; $p = 0.083$; $\eta^2 = 0.187$]. These interactions suggest that men evaluated the *Efficiency/Effectiveness* of the MI-BCI protocol the same whatever their performance at the first experiment, while women evaluated this dimension with lower scores when they had difficulties at the first experiment, and with higher scores when they managed at the first experiment. The fact we only have tendencies could be due to the relatively

low number of participants (N=20, i.e., only 5 per group*gender).

	NORMALITY	LINEARITY	HOMOGENEITY OF REGRESSION SLOTS	EQUALITY OF VARIANCES
MENTAL ROTATION SCORES	s = 0.049			
LEARNABILITY / MEMORABILITY	s = -0.542	P = 0.543	P = 0.335	P = 0.103
EFFICIENCY / EFFECTIVENESS	s = 0.191	P = 0.582	P = 0.230	P = 0.217
SAFETY	s = 0.344	P = 0.874	P = 0.564	P = 0.891
SATISFACTION	s = -0.044	P = 0.366	P = 0.741	P = 0.213

Figure 13: Table representing the assumptions checking for the ANCOVA. It has to be noticed that all the assumptions but the linearity were respected.

5.3 Discussion

This second experiment aimed at assessing the relationships existing between MI-BCI performance, motor-task performance (obtained in the first experiment) and spatial abilities (measured by the Mental Rotation test). Participants globally obtained modest performances, probably due to the fact they only took part in one session, while several sessions are necessary to acquire MI-BCI skills and thus improve in terms of performance. Nonetheless, the ANCOVA results showed that whatever performance measure was used (peak or mean classification accuracy), there is a main effect of the Mental Rotation scores as well as significant Run*Mental-Rotation, Run*Gender and Run*Group interactions. First, the main effect of Mental Rotation scores confirms the important impact of spatial abilities on BCI performance that was suggested in our previous papers for mental imagery based BCI (not purely motor ones) [17, 19]. The important role of spatial abilities was also strengthened by the significant correlation between MI-BCI performance (peak classification accuracy) and Mental Rotation scores. Second, the interactions suggest that when the effect of the spatial abilities is controlled for, (1) *women* improved across the runs while *men* did not and (2) participants who were *bad* performers in the first experiment began with lower MI-BCI performance than *good* performers. However the former improved across the runs whereas the latter did not. There is in fact a strong tendency [$p=0.058$] towards a Run*Gender*Group interaction when assessing performance using peak classification accuracy. This last interaction indicates that men kept the same ratio between the first and the second experiment: men who were good at the first experiment remained better at the second experiment than the others (i.e., the bad performers of the first experiment) but none of them improved during the second experiment. Women who were good at the first experiment remained good at the second (at the same performance level as the men of their group), but they did not progress. However, women from the *bad* group began with low performances in the second experiment but their performances quickly improved and eventually surpassed the others. Thus, it would seem that participants who faced difficulty during the first experiment, especially women, improved more easily in terms of performance during the second experiment. This could be explained by the fact that facing difficulty in the

context of a complex task (such as MI tasks, for which we are not trained and for which we do not have any proprioceptive feedback) requires substantial cognitive resources. Thus, these resources are not available to understand how to use the information provided by the training protocol or by the feedback. By opposition, when users face difficulty to find the right strategy in a less complex context (such as performing motor tasks which they know they can do and for which they have proprioceptive feedback) their available resources allow them to pay attention to the training protocol and feedback and to understand how the latter could be used to improve their performance. Once the process has already been executed, a re-exposition to this protocol would not require as many resources and so could be used efficiently in a more complex context.

In our case, understanding the feedback would mean understanding what it means, when the user should maintain his strategy/when he should modify it. People who faced difficulty during the motor task learnt how and when to test and modify their strategies, which helped them when confronted with the MI-BCI experiment which was more complex. Of course, this hypothesis needs to be tested in a future experiment.

It is not entirely clear why gender plays a role in BCI performance and observed training effects. A possible interpretation could be that since women have lower spatial abilities than men on average, and that spatial abilities are correlated to BCI performances, they have more room for improvement, which could explain why they improved over the runs while men did not. Another interpretation could be that women may rely on different cognitive mechanisms and strategies when faced with a difficult learning problem, although we are not aware of any literature on this topic. This should therefore be investigated further.

6. GENERAL DISCUSSION

The two experiments we conducted provide a number of relevant insights regarding MI-BCI training with standard training protocols. Our first experiment used a standard BCI training protocol, the widely used Graz group protocol, to teach non-BCI related skills, simple motor tasks in this case. It showed that with such training tasks and feedback, a substantial proportion of subjects - here 16.67%, i.e., 9 subjects out of 54, failed to acquire the targeted skills despite their simplicity (drawing triangles and circles). This suggests that such a training protocol is suboptimal for skill teaching, and therefore, that BCI illiteracy/deficiency is most likely due, at least in part, to the limitations of the training protocol. In particular, many participants reported that the feedback they received, i.e., the bar feedback, provided too little information to help them to improve. Thus, future research aiming at reducing BCI illiteracy/deficiency and improving BCI training should consider providing richer and more explanatory feedback which helps users identify what they should change in their EEG patterns and mental strategies to achieve successful BCI control. Interestingly enough, this is also what is theoretically recommended for successful training in human learning and education psychology literature [28].

Our second experiment aimed at measuring possible relationships between the performance obtained during the motor-task training (first experiment) and MI-BCI training performance (second experiment). To do so, we trained the 10 best and 10 worst participants from the first experiment, to perform left-hand and right-hand MI. Contrary to our hypothesis, this second experiment did not reveal any significant linear correlation between motor-task performance (first experiment) and MI-BCI performance. It also did not reveal significant performance differences between the 10 best and 10 worst participants from the first experiment. This may be due to the fact that different tasks were performed in the first and second studies, which could represent a limitation. However, the same protocol (in terms of instructions and feedback) was used in both studies. Moreover, in retrospect, this is not so surprising since, as mentioned before, when dealing

with BCIs, a number of factors other than the training protocol and users' learning abilities appear to affect BCI performances. In particular, the EEG signal-to-noise ratio, the participants' ability or experience at performing motor imagery or the orientation of their sensori-motor cortex with respect to the scalp, among others, can all impact the quality of the EEG patterns which in turn can impact BCI performance, independently from the training protocol and the users' learning abilities. Furthermore, our results revealed no relationship between MI-BCI performance and the Blankertz SMR predictor. The fact that this predictor is not significantly correlated with MI-BCI performance could be partly due to our experimental protocol. Indeed, as no 2 minute-long baseline was recorded the predictor was computed based on the concatenation of all the 3 second-long pre-trials of the runs, which could impact its performance. Nonetheless, this second experiment did reveal some other interesting insights. First it confirmed that spatial abilities are related to mental imagery-based BCI performances. We have shown such a relationship before for a mental imagery based BCI that was not based purely on motor imagery, but on left-hand motor imagery, mental geometric figure rotation and mental subtraction. The study introduced in this paper suggests that spatial abilities also play a role in purely motor imagery based BCI performances, in which no mental rotation tasks are involved. This thus confirms the importance of spatial abilities for successful BCI control, and reinforces the idea that spatial ability training should be explored for BCI training. Second it showed that when subjects are faced with a pre-training session they perceive as difficult (here motor learning tasks for subjects with poor performances), they seem to explore more strategies and therefore learn better in a subsequent training task, here MI-BCI training. This is an interesting result as it suggests that in the future, it would be worth exposing naive BCI subjects to a pre-training that forces them to explore multiple strategies which might in turn help them to find a better MI strategy in a following classical BCI training. Such an approach should be explored in the future. Finally, this second experiment suggests that, when spatial abilities are controlled for, women (especially the ones who faced difficulty in the first experiment) seem to improve over time whereas men do not with such a short term training. However, this might be explained by women having more room for improvement than men and/or different cognitive approaches to learning. Further studies (both experimental and theoretical) should be performed to elucidate this observation.

In summary, our studies have shown that current standard BCI training protocols such as the Graz training protocol are suboptimal and most likely responsible for a substantial part of the observed BCI illiteracy/deficiency. They should therefore be changed, in particular by providing more explanatory feedback. Our two experiments also revealed the importance of spatial abilities and of a pre-training session on subsequent MI-BCI performances. This therefore suggests that exploring spatial ability training and specific pre-training tasks for MI-BCI training are promising future prospects. Altogether, our studies have opened the door to several new research questions aimed at improving MI-BCI training and thus at further increasing the reliability and potential of MI-BCIs.

7. ACKNOWLEDGEMENTS

This research was supported by fundings from IdEx Bordeaux. The authors also thank Benjamin Blankertz and Laura Acqualagna for providing their code to compute the Blankertz's SMR predictor, Alison Cellard for coding the OpenViBE scenario and her help during the experiments, John Kelway for proof-reading the manuscript and Martin Hachet, Bernard N'Kaoua and Sriram Subramanian for their advices.

REFERENCES

- [1] M. Ahn and Sung C. Jun. Performance variation in motor imagery brain–computer interface: A brief review. *Journal of neuroscience methods*, 2015.
- [2] B. Allison and C. Neuper. *Could Anyone Use a BCI?* Springer London, 2010.
- [3] K. K. Ang and C. Guan. Brain–computer interface for neurorehabilitation of upper limb after stroke. *Proceedings of the IEEE*, 103:944–53, 2015.
- [4] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, V. Kunzmann, and F. Losch. The Berlin Brain-Computer Interface: EEG-based communication without subject training. *IEEE Trans. Neural Sys. Rehab. Eng.*, 14(2):147–152, 2006.
- [5] 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*, 51(4):1303–1309, 2010.
- [6] 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*, 51:1303–09, 2010.
- [7] Benjamin Blankertz, Florian Losch, Matthias Krauledat, Guido Dornhege, Gabriel Curio, and Klaus-Robert Muller. The berlin brain–computer interface: accurate performance from first-session in bci-naive subjects. *Biomedical Engineering, IEEE Transactions on*, 55(10):2452–62, 2008.
- [8] W. Burde and B. Blankertz. Is the locus of control of reinforcement a predictor of brain-computer interface performance? *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*, 2006:108–09, 2006.
- [9] A. Chatterjee, V. Aggarwal, A. Ramos, S. Acharya, and N. V. Thakor. A brain-computer interface with vibrotactile biofeedback for haptic information. *Journal of Neuroengineering and Rehabilitation*, 4:40, 2007.
- [10] F. Cincotti, L. Kauhanen, F. Aloise, T. Palomäki, N. Caporusso, P. Jylänki, D. Mattia, F. Babiloni, G. Vanacker, M. Nuttin, M. G. Marciani, and J. D. R. Millán. Vibrotactile feedback for brain-computer interface operation. *Computational Intelligence and Neuroscience*, 2007:12, 2007.
- [11] P. Eskandari and A. Erfanian. Improving the performance of brain-computer interface through meditation practicing. In *Engineering in medicine and biology society, 2008. EMBS 2008. 30th Annual international conference of the IEEE*, pages 662–665. IEEE, 2008.
- [12] E. V. Friedrich, C. Neuper, and R. Scherer. Whatever works: A systematic user-centered training protocol to optimize brain-computer interfacing individually. *PloS one*, 8(9):e76214, 2013.
- [13] G. D. Gargiulo, A. Mohamed, A. L. McEwan, P. Bifulco, M. Cesarelli, C. T. Jin, M. Ruffo, J. Tapson, and A. van Shaik. Investigating the role of combined acoustic-visual feedback in one-dimensional synchronous brain computer interfaces, a preliminary study. *Medical Devices: Evidence and Research*, 5:81–88, 2012.

- [14] 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*, 89:80–86, 2012.
- [15] E.-M. Hammer, T. Kaufmann, S. C. Kleih, B. Blankertz, and A. Kübler. Visuo-motor coordination ability predicts performance with brain-computer interfaces controlled by modulation of sensorimotor rhythms (SMR). *Frontiers in Human Neuroscience*, 8, 2014.
- [16] T. Hinterberger, N. Neumann, M. Pham, A. Kübler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor, and N. Birbaumer. A multimodal brain-based feedback and communication system. *Experimental Brain Research*, 154:521–26, 2004.
- [17] C. Jeunet. Training users’ spatial abilities to improve brain-computer interface performance: A theoretical approach. *CJCSC*, 2015.
- [18] C. Jeunet, A. Cellard, S. Subramanian, M. Hachet, B. N’Kaoua, and F. Lotte. How well can we learn with standard BCI training approaches? a pilot study. *6th International Brain-Computer Interface Conference*, pages 332–35, 2014.
- [19] C. Jeunet, B. N’Kaoua, M. Hachet, and F. Lotte. Predicting mental-imagery based brain-computer interface performance from personality, cognitive profil and neurophysiological patterns. *PLOS ONE*, 10(12):e0143962, 2015.
- [20] C. Jeunet, C. Vi, D. Spelmezan, B. N’Kaoua, S. Subramanian, and F. Lotte. Continuous tactile feedback for motor-imagery based brain-computer interaction in a multitasking context. *Interact2015*, 2015.
- [21] L. Kauhanen, T. Palomäki, P. Jylänki, F. Aloise, M. Nuttin, and J. del R. Millán. Haptic feedback compared with visual feedback for BCI. *Proceedings of the 3rd International Brain-Computer Interface Workshop & Training Course*, 2006.
- [22] S. E Kober, M. Witte, M. Ninaus, C. Neuper, and G. Wood. Learning to modulate one’s own brain activity: the effect of spontaneous mental strategies. *Frontiers in human neuroscience*, 7, 2013.
- [23] A. Lécuyer, F. Lotte, R. B. Reilly, R. Leeb, M. Hirose, and M. Slater. Brain-computer interfaces, virtual reality, and videogames. *IEEE Computer*, 41(10):66–72, 2008.
- [24] R. Leeb, K. Gwak, D-S. Kim, and J. d. R. Millan. Freeing the visual channel by exploiting vibrotactile brain-computer interface feedback. *IEEE Engeneering in Medicine and Biology Society*, 2013:3093–6, 2013.
- [25] R. Leeb, F. Lee, C. Keinrath, R. Scherer, and G. Bischof, H.and Pfurtscheller. Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engeneering*, 15(4):473–82, 2007.
- [26] F. Lotte, J. Faller, C. Guger, Y. Renard, G. Pfurtscheller, A. Lécuyer, and R. Leeb. Combining BCI with Virtual Reality: Towards new applications and improved BCI. In Brendan Z. Allison, Stephen Dunne, Robert Leeb, José Del R. Millán, and Anton Nijholt, editors, *Towards Practical Brain-Computer Interfaces*, Biological and Medical Physics, Biomedical Engineering, pages 197–220. Springer Berlin Heidelberg, 2013.
- [27] F. Lotte and C. Jeunet. Towards improved BCI based on human learning principles. *3rd Int. Winter Conf. on Brain-Computer Interfaces*, 2015.

- [28] F. Lotte, F. Larrue, and C. Mühl. Flaws in current human training protocols for spontaneous BCI: lessons learned from instructional design. *Frontiers in Human Neurosciences*, 7:568, 2013.
- [29] Fabien Lotte. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces. *Proceedings of the IEEE*, 103(6):871–90, 2015.
- [30] K. A. McCraedie, D. H. Coyle, and G.P. Prasad. Is sensorimotor BCI performance influenced differently by mono, stereo, or 3-D auditory feedback? *IEEE Transactions on Neural S*, 22(3):431–40, 2014.
- [31] J. d R. Millán, R. Rupp, G. R. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, R. Leeb, C. Neuper, K.-R. Müller, and D. Mattia. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in neuroscience*, 4, 2010.
- [32] J.R. Millán, J. Mouri no, M. Franzé, F. Cincotti, M. Varsta, J. Heikkonen, and F. Babiloni. A local neural classifier for the recognition of eeg patterns associated to mental tasks. *IEEE transactions on neural networks*, 13(3):678–686, 2002.
- [33] K.-R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz. Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *Journal of neuroscience methods*, 167:82–90, 2008.
- [34] G. R. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller. Better than random? a closer look on BCI results. *International Journal of Bioelectromagnetism*, 10:52–55, 2008.
- [35] C. Neuper, G. Müller, A. Kübler, N. Birbaumer, and G. Pfurtscheller. Clinical application of an EEG-based BCI. *Clinical Neurophysiology*, 114:399–409, 2003.
- [36] C. Neuper and G. Pfurtscheller. *Brain-Computer Interfaces*, chapter Neurofeedback training for BCI control, pages 65–78. Springer, 2010.
- [37] C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller. Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial eeg. *Cognitive Brain Research*, 25:668–77, 2005.
- [38] F. Nijboer, N. Birbaumer, and A. Kübler. The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis—a longitudinal study. *Frontiers in neuroscience*, 4, 2010.
- [39] F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N. Birbaumer, and A. Kübler. An auditory brain-computer interface (BCI). *Journal of Neuroscience Methods*, 167:43–50, 2008.
- [40] G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *proceedings of the IEEE*, 89(7):1123–1134, 2001.
- [41] Gert Pfurtscheller, Doris Flotzinger, and Joachim Kalcher. Brain-computer interface—A new communication device for handicapped persons. *Journal of Microcomputer Applications*, 16(3):293–99, 1993.
- [42] R. Scherer, J. Faller, D. Balderas, E.V.C. Friedrich, M. Pröll, B. Allison, and G. Müller-Putz. Brain-computer interfacing: more than the sum of its parts. *Soft Computing*, 17(2):317–331, 2013.

- [43] A. J. Schwab. *The analysis of covariance*, 2007.
- [44] S. G. Vandenberg and A. R. Kuse. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills*, 47:599–604, 1978.
- [45] C. Vidaurre and B. Blankertz. Towards a cure for BCI illiteracy. *Brain Topography*, 23:194–198, 2010.
- [46] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113:767–779, 2002.
- [47] J. R. Wolpaw and E. W. Wolpaw. *BCI: Principles and Practice*. Oxford University Press, 2012.