Journal of Neural Engineering

PAPER • OPEN ACCESS

Hybrid brain-computer interface with motor imagery and error-related brain activity

To cite this article: Mahta Mousavi *et al* 2020 *J. Neural Eng.* **17** 056041

View the article online for updates and enhancements.

This content was downloaded from IP address 130.149.173.119 on 14/01/2021 at 09:18

Journal of Neural Engineering



OPEN ACCESS

RECEIVED
13 February 2020

REVISED
20 July 2020

ACCEPTED FOR PUBLICATION
29 July 2020

PUBLISHED
21 October 2020

Original Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



PAPER

Hybrid brain-computer interface with motor imagery and error-related brain activity

Mahta Mousavi[1,5], **Laurens R Krol**[2,3] and **Virginia R de Sa**[1,4]

[1] Department of Cognitive Science, UC San Diego, La Jolla, CA, United States of America
[2] Biological Psychology and Neuroergonomics, Technische Universität Berlin, Berlin, Germany
[3] Neuroadaptive Human-Computer Interaction, Brandenburg University of Technology, Cottbus-Senftenberg, Germany
[4] Halıcıoğlu Data Science Institute, UC San Diego, La Jolla, CA, United States of America

E-mail: mahta@ucsd.edu, lrkrol@gmail.com and desa@ucsd.edu

Keywords: brain-computer interface, error-related potentials, motor imagery, EEG, common spatial patterns, hybrid BCI, BCI feedback

Abstract

Objective. Brain-computer interface (BCI) systems read and interpret brain activity directly from the brain. They can provide a means of communication or locomotion for patients suffering from neurodegenerative diseases or stroke. However, non-stationarity of brain activity limits the reliable transfer of the algorithms that were trained during a calibration session to real-time BCI control. One source of non-stationarity is the user's brain response to the BCI output (feedback), for instance, whether the BCI feedback is perceived as an error by the user or not. By taking such sources of non-stationarity into account, the reliability of the BCI can be improved. *Approach*. In this work, we demonstrate a real-time implementation of a hybrid motor imagery BCI combining the information from the motor imagery signal and the error-related brain activity simultaneously so as to gain benefit from both sources. *Main results*. We show significantly improved performance in real-time BCI control across 12 participants, compared to a conventional motor imagery BCI. The significant improvement is in terms of classification accuracy, target hit rate, subjective perception of control and information-transfer rate. Moreover, our offline analyses of the recorded EEG data show that the error-related brain activity provides a more reliable source of information than the motor imagery signal. *Significance*. This work shows, for the first time, that the error-related brain activity classifier compared to the motor imagery classifier is more consistent when trained on calibration data and tested during online control. This likely explains why the proposed hybrid BCI allows for a more reliable means of communication or rehabilitation for patients in need.

1. Introduction

Since the introduction of a brain-computer interface (BCI) by Vidal [1], there have been many implementations as potential communication or rehabilitation interventions for patients (e.g. [2–6]). In BCI systems, brain activity is read directly from the brain and a control command is generated based on the user's intentions while bypassing the common neuromuscular pathways. Electroencephalography (EEG) is one technique to read brain activity non-invasively and is thus widely used in BCI applications. Motor imagery (MI) BCI is a popular category of BCI systems that relies on the user-initiated movement imagination of different body parts, which results in distinguishable patterns of brain activity [3, 7]. By detecting and classifying these patterns in real time, imagining movement of different body parts can then be mapped to, for instance, a cursor moving on a screen or a switch controlling a wheelchair.

The MI signal is user-generated; therefore, the user's ability to generate distinguishable motor imagery signals plays an essential role in reliable BCI control. Since the generated patterns are relatively subtle to detect, pattern recognition methods are necessary to train relevant classifiers [8-10]. BCIs usually use supervised machine learning in which the training data come from a calibration session [11]. Due to the inherent non-stationarity of brain activity, re-calibration is often necessary after some time. Without such re-calibration, the BCI system suffers a

[5] Author to whom any correspondence should be addressed.

© 2020 The Author(s). Published by IOP Publishing Ltd

decrease in accuracy over time, limiting the transfer of a classifier trained during calibration to subsequent real-time control [12].

To improve reliability in BCI control while avoiding time-consuming re-calibration sessions, one approach is to use other available sources of information in order to support, adjust, or correct the information from the primary detected signal. One potential such source is the brain activity that occurs in response to the BCI output (feedback). User brain activity was shown to be different when observing a successful execution of an intended task by the BCI versus an unsuccessful execution [13–20]. This so-called 'error-related brain activity' is classifiable and can be used to alleviate the reliability limitation [17, 21–23]. Of course, during online use of a BCI system, classifiers focusing on these secondary sources of information could potentially suffer the same degradation over time. Aside from the issue of how to optimally combine multiple sources of information, another important question is how the accuracy changes when a classifier trained on the calibration data is used in online control for both the primary (e.g. motor imagery) and the secondary (e.g. error-related brain activity) signals.

There are multiple approaches towards using error-related brain activity to improve BCI performance. One approach is to discard a BCI output and repeat the trial or to execute an action in the reverse direction upon detection of an error [14, 24, 25]. A second approach, based on error-driven learning, attempts to limit the occurrence of a future error by updating the classifiers upon the detection of error-related brain activity or to discard the unsupervised adaptation temporarily if an error is detected [26, 27]. A third approach based on error integration proposed a hybrid BCI for a 1-D cursor control by combining the motor imagery signal with the user brain activity in response to the cursor's changes in the direction of movement [22, 23].

Recent studies in invasive BCIs also show evidence for the possible detection of error-related brain activity and propose to use it to improve the BCI performance. For instance, the authors in [28] showed that error signals can be detected from human electrocorticography (ECoG) in a continuous task comprising a video game. In other work, the authors in [29] showed the detection of error-related brain activity in a motor imagery task in human ECoG. Moreover, the authors in [30] proposed to augment an intracortical BCI with error detection. They showed in an experiment with non-human primates that a classifiable error signal can be detected from electrodes located in the premotor and primary motor cortices and proposed a system to automatically undo or prevent mistakes.

In earlier work [17, 31], we showed that the motor imagery signal and feedback/error related brain activity occur in overlapping frequency bands and may therefore interfere with each other if not handled appropriately. We also proposed to combine these two signals through a more sophisticated error integration approach that simultaneously combined a right/left hand motor imagery classifier with a classifier detecting whether the user perceived the last BCI output as an error or not. The proposed hybrid BCI system translated the classification score from the domain of the error-related brain activity to that of the motor imagery classifier and learned a logistic regression classifier to best combine the two sources of information for each user. This allows for a system that relies more/less on either the motor imagery or error-related brain activity signals depending on how classifiable each source of information is for a specific user. We showed the efficacy of such a system in an offline BCI with sham feedback across 10 participants [17].

Since real-time feedback can itself modify brain activity, it is vital to evaluate the efficacy of our earlier proposed hybrid BCI in real-time control. This is the goal of the current study where we answer two questions: 1) how the performance of the proposed hybrid BCI in real-time control compares to a conventional motor imagery BCI, and 2) how the underlying two aspects of the proposed hybrid BCI (i.e. error-related brain activity and motor imagery signal) compare with each other in their contributions to robust and reliable control.

2. Methods

The study was approved by the University Institutional Review Board at UC San Diego and all participants signed an informed consent form prior to their participation. Data were recorded from 12 participants who were naive to BCI experiments (7 females, 1 left-handed, average age = 20.4± 1.0). Each participant took part in one session of roughly 2.5 hours in length while they comfortably sat in front of a screen (Dell 19" CRT monitor), centered with respect to the screen and about one meter away.

EEG data were recorded using a 64-channel BrainAmp system (Brain Products GmbH). The active electrodes were located according to the international 10–20 system [32]. The impedance of the electrode connectivity was adjusted to be below 20 kΩ. Electromyography (EMG) data were also recorded with bipolar electrodes using BrainAmp ExG from the wrists and upper forearms; however, the recorded EMG data were not used for the analysis in this paper. Data were recorded at a sampling rate of 5000 Hz with online reference and ground electrodes located at FCz and AFz, respectively. To ensure accurate recording and inference of the brain activity, a small photo sensor was placed at the bottom right corner of the screen and was connected to the ExG box. A white circle (with the same diameter as the the photo sensor) was turned on and off at the same time that

the cursor moved on the screen. This allowed for an accurate recording of the stimuli presentation time and removing any potential jitter in the measurement system.

Python was used to design and present the stimuli as well as for the real-time processing of the EEG signal. Simulation and Neuroscience Application Platform (SNAP) [33] was used for stimuli presentation, Lab Streaming Layer (LSL) [34] to interface the EEG system with the computer, and Numpy [35], SciPy [36] and scikit-learn [37] for data processing. Python was used for the offline analysis of the data as well. MATLAB [38] and EEGLAB [39] were used for offline epoching and plotting of the results. Temporal filters were implemented through the `filtfilt` function from SciPy [36] that applies a filter twice, once forward and once backward, to ensure zero-phase filtering. Also, wherever not otherwise indicated, all 64 EEG channels were used for feature extraction and classification.

2.1. Experiment

Each participant took part in one experiment session that was comprised of two phases. The first phase was primarily designed for the participants to gain experience with the motor imagery of their right/left hand. Data from the first phase were not analyzed in this work. Phase 2, which is the focus of this work, involved using motor imagery to control a cursor on the screen. This phase had two parts: calibration and online control. Details of the stimuli presentation are described next.

2.1.1. Phase 1

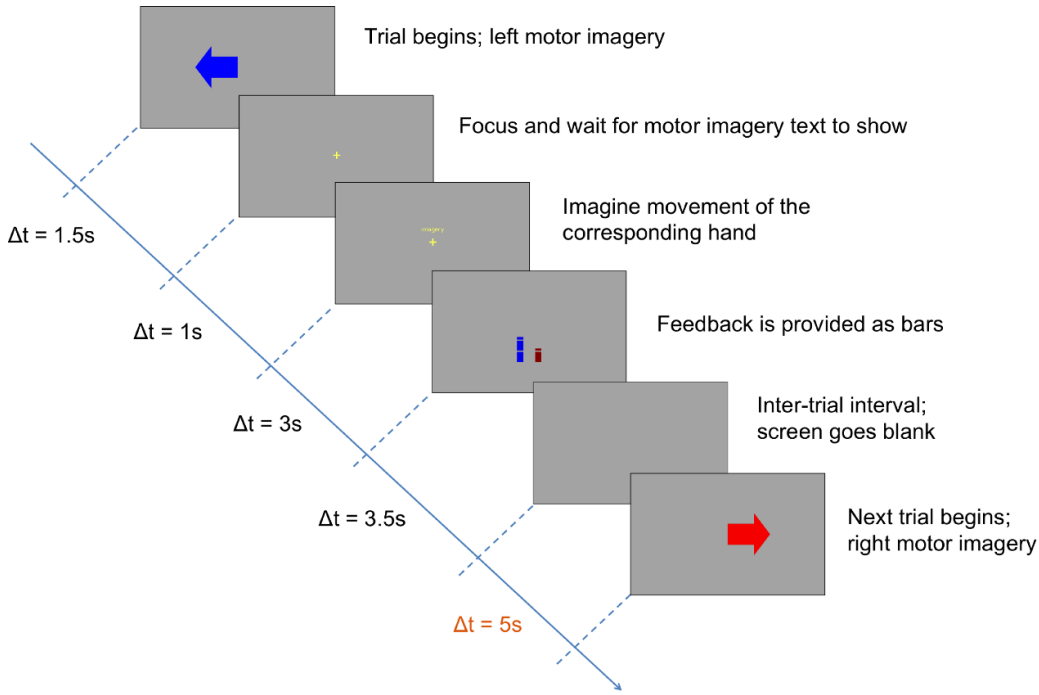
Each trial began by showing a right/left arrow representing the side of imagery. Next, participants fixated on a fixation cross for 1 second. Then the text 'imagery' appeared above the fixation cross instructing participants to begin motor imagery of the corresponding hand. The movement imagination time was set to 3 seconds and afterwards participants were provided with feedback in the form of two bars whose height represented the power in the 7–30 Hz frequency band averaged on the electrodes over the right (EEG channels FC4, C4, CP4) and left (EEG channels FC3, C3, CP3) motor cortices, separately. Since motor imagery of the right/left hand results in contralateral event-related desynchronization (i.e. decreased power over right/left motor cortex for left/right motor imagery), participants were instructed to maximize the height of the bar on the side of imagery and minimize the height of the one on the other side [40]. If the bar height was larger than a set threshold, the scale was adjusted for easy observation and interpretation of the bar heights. An example of a trial is depicted in figure 1(a) and an explanation of the provided feedback is shown in figure 1(b).

Phase 1 had a total of 30 trials (15 right hand and 15 left hand motor imagery) divided into three blocks and there was a 5-second break in between two consecutive trials. The order of the trials was randomly selected once and the order was kept the same across participants. Participants were given instructions and suggestions on what to imagine for their hand movements; however, they had a chance to explore different imagined movements and decide on what worked best for them. After each block, participants could take as much rest as needed. Upon completion of this phase, participants filled in a short questionnaire in which they answered what movement they imagined for their right and left hands. They were instructed not to change their selected movement imagination throughout the rest of the experiment.

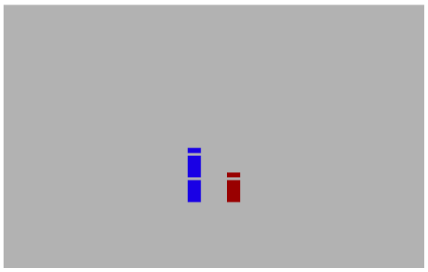
2.1.2. Phase 2

In phase 2, participants were instructed to use their selected right/left hand movement imagination to move a cursor to the right/left towards a target. At the beginning of every trial, the cursor (blue circle, 2 cm in diameter) appeared at its starting position, i.e. the center of the screen. The target (white circle, 2 cm in diameter) also appeared at either side of the screen, exactly three horizontal steps away from the cursor. After 1.2 seconds, the target disappeared and the cursor moved at the rate of one movement or 'step' per 1.2 seconds to either the left or the right. The trial ended when the cursor hit the target location or the corresponding location on the other side or after 12 cursor steps, whichever came first. Therefore, trial duration varied across different trials, depending on the cursor movements, approximately between 4 to 15 seconds. Participants had 5 seconds to rest before the next trial began. The choice of a maximum of 12 steps in each trial was selected to ensure that the trial ended with a maximum duration of about 15 seconds. An example of a trial is depicted in figure 2. In this example, the target appears at the right side of the screen indicating a right hand motor imagery trial where the participant should imagine movement of the right hand throughout the trial. Since the goal in each trial is for the cursor to hit the target, a movement towards the target is called a 'good' movement and a movement away from the target is called a 'bad' movement. Therefore, in this example, the first cursor movement towards the right is a good movement, the second one is a bad movement as the cursor moves away from the target. The third movement is a good movement and so on. This trial is a success (or hit) as the cursor reached the target location.

Phase 2 had a total of nine blocks and each block was comprised of 20 trials. In each trial, the target appeared randomly at the right or the left side of the screen while maintaining that each block had 10 right and 10 left trials. In the first three blocks, participants received sham feedback while they were led to believe that they were in control of the cursor movements.



(a) An example of a trial in phase 1. Note that the background of the frames was dark gray in the experiment but is depicted lighter here for easier visualization of the details.



(b) An example of the MI training feedback in phase 1. The power on the left side (blue) is two times the maximum height, plus a remainder value. The power on the right side (red) is one times this maximum height plus a remainder. This was a successful left-hand motor imagery trial.

Figure 1. In phase 1, participants gained some experience with motor imagery: (a) depicts an example of a trial in this phase, and (b) shows the provided feedback.

This was to have the necessary labels to train the classifiers. In these first three blocks, the cursor movements were predefined and randomly generated with the following criteria: the cursor had a fixed probability of going towards the target ($p = 0.60$) following a Bernoulli distribution until the target or the corresponding location on the other side was reached, or a maximum of 12 steps occurred; however, any generated trials with more than three consecutive changes in direction were not used. The cursor movements for the first three blocks were generated ahead of time and kept the same across participants. The hit rates were 0.75, 0.6 and 0.75 in blocks 1, 2 and 3, respectively. The recorded EEG data in the first three blocks were used to calibrate the classifiers as will be explained in detail later.

In the latter six blocks of phase 2, participants received real online feedback in which three blocks used the conventional right/left hand motor imagery (R/L) control and the remaining three blocks used our proposed BCI control that combined the right/left hand motor imagery with the error-related brain activity signal. The latter classifier detects whether the user perceived the last cursor movement as 'good' or 'bad', i.e. going towards or away from the target, respectively. This is called a good/bad classifier (G/B) and the proposed control is therefore called R/L+G/B control. The blocks with R/L and R/L+G/B control were alternated and the order was counterbalanced across participants. After each block in this phase, participants could take as much rest as needed.

Participants were not aware of the sham feedback in the first three blocks and the different controls in the online blocks. After each block in phase 2 (including the calibration and online control blocks), participants answered the following question: *from 1 to 10 where 1 represents the least and 10 the most amount of control, how much in control of the cursor did you feel?*

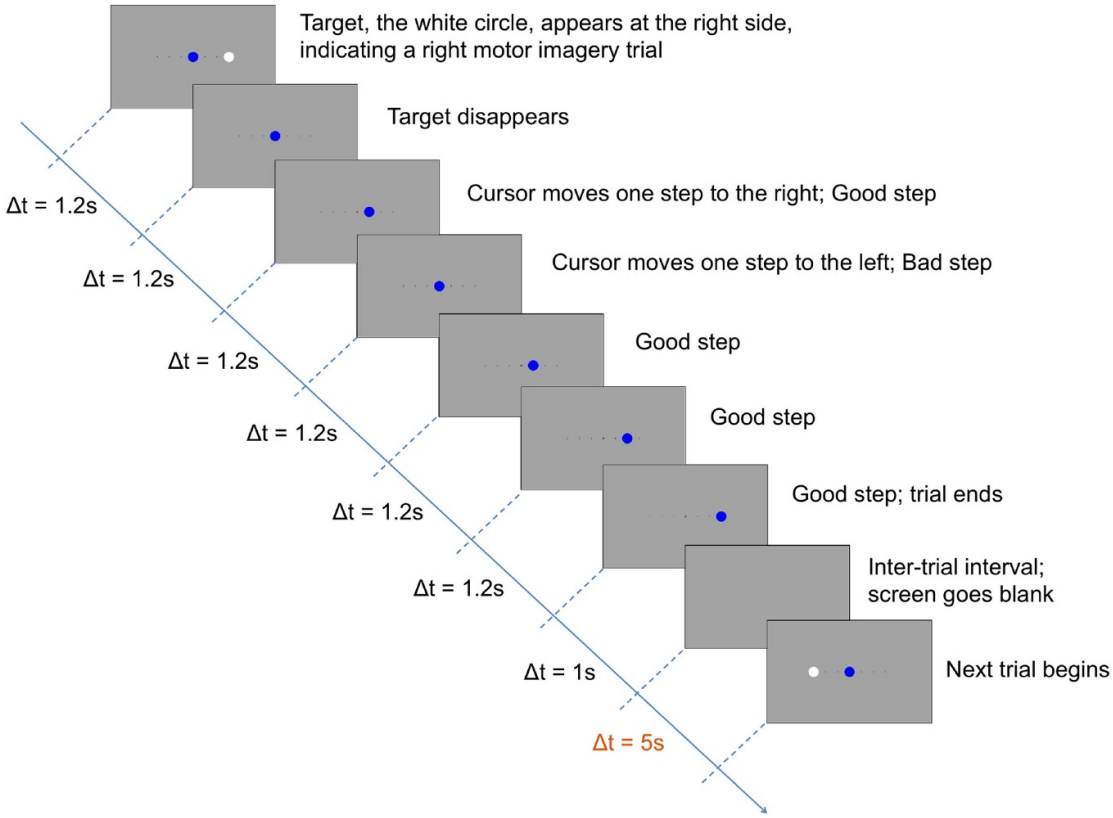


Figure 2. An example of a trial in phase 2. Each trial began with the cursor (blue circle) at the center and the target (white circle) at either side of the screen, exactly three steps from the cursor. Trial ended when the cursor hit the target location or the corresponding location on the other side or after 12 cursor steps, whichever came first. Participants had 5 seconds to rest before the next trial began. Note that the background of the frames was dark gray in the experiment but is depicted lighter here for easier visualization of the details.

Participants also filled in questionnaires aimed at quantifying their handedness and various aspects of their personality [41–44]. However, the data were not used for the analysis of this work.

2.2. Calibration and online control

We performed extensive analysis on our previously recorded data [17] comprising data from 10 separate participants each participating in ten blocks with 20 trials of an experiment with sham feedback (similar to the calibration phase in this work). We compared different classifiers in their performance when applied to three blocks of their data, i.e. the planned duration of the calibration phase, to decide the specifications of the feature extraction and classification methodologies such as the type of classifier, number of spatial filters, specifications of the temporal filters, etc as explained next. Note that the data from our previous work were not directly used for calibration in this work.

2.2.1. Calibration

The recorded EEG data from the first three blocks of phase 2, called the 'calibration data' during which the cursor moved according to a pre-determined set of movements, were used to train the classifiers to be applied later for online control.

To train classifiers, the recorded calibration data were downsampled to 100 Hz, re-referenced to common average, and epoched 0–1 seconds time-locked to each step (cursor movement) excluding the last cursor movement. The last cursor movement was excluded as it indicated the end of a trial and no motor imagery signal was generated by the participants when a trial ended. In the trial example depicted in figure 2, epoching the EEG data resulted in 4 epochs. Each epoch was labeled as good-right (GR), bad-right (BR), good-left (GL) or bad-left (BL) depending on whether it belonged to a right/left motor imagery trial and whether the cursor moved towards/away from the target resulting in either a good or a bad movement. In the trial example in figure 2, the first epoch was labeled as GR, the second as BR, and the third and fourth as GR. Similarly, a left motor imagery trial (with the target at the left side of the screen), resulted in GL and BL epochs for cursor movements towards and away from the target, respectively.

Three classifiers were trained on the selected epochs: two to classify the error-related brain activity (called the G/B-csp and G/B-wm classifiers) and one to classify the motor imagery signal (called the R/L classifier). Classifiers were trained for each participant separately on her/his calibration data. Since the number of GR, BR, GL and BL steps in calibration

data was not balanced (mainly because the average hit rate for the calibration blocks were by design higher than 0.5 to maintain participants' motivation), prior to training the classifiers, the population of the GR, BR, GL and BL epochs were balanced by subsampling the larger groups. This was to have an unbiased classifier. The subsampling occurred at random and no steps from the calibration data were removed except for balancing the populations of the four aforementioned groups of steps.

(a) **R/L classifier**: For the right/left hand motor imagery classification (R/L), the GR and BR epochs were concatenated and labeled as the right class. The GL and and BL epochs were also concatenated and labeled as the left class. The R/L classifier was trained to distinguish between the right and left classes. To do so, first the epoched data were filtered to 7–30 Hz with an IIR filter (6th order Butterworth). Then the method of common spatial patterns (CSP) [9, 45] was applied and the top 3 CSP filters for each of the right and left classes were selected (i.e. a total of 6 filters). Temporally filtered epochs were passed through the selected CSP filters and the logarithm of the variance of the filtered data across time were selected as features. A linear discriminant analysis (LDA) with automatic shrinkage using the Ledoit-Wolf lemma [46] was trained on the selected features [37]. The choice of shrinkage was to encourage better generalization.

(b) **G/B-csp classifier**: The GR and GL epochs were concatenated and labeled as the good class. The BR and and BL epochs were also concatenated and labeled as the bad class. The G/B-csp classifier was trained to distinguish between the good and bad classes [17]. To train the G/B-csp classifier, we filtered the data to 1–30 Hz with an IIR filter (6th order Butterworth). A smaller time window of 50–950 ms was selected from each filtered epoch. Next, the CSP technique [9, 45] was applied and the top 3 CSP filters for each of the good and bad classes were selected (i.e. a total of 6 filters). Temporally filtered epochs were passed through the selected CSP filters and the logarithm of the variance of the filtered data across time were selected as features.To encourage better generalization, LDA with automatic shrinkage using the Ledoit-Wolf lemma [46] was trained on the selected features [37].

(c) **G/B-wm classifier**: we trained another G/B classifier called G/B-wm following the windowed-means approach for single trial classification of an event-related potential (ERP) [14, 47]. We considered EEG activity on channels Fz, Cz, CPz and Pz as the error-related brain activity is considered to be a fronto-central signal that is best picked up by the mid-line channels [14]. The EEG signal on these 4 channels was filtered to 1–10 Hz with an IIR filter (6th order Butterworth). A smaller time window of 50–950 ms was selected from each filtered epoch and baselined to the first 50 ms. Then the average of the signal in 100 ms non-overlapping windows in 9 windows (50–950 ms) were selected as features. An LDA classifier with automatic shrinkage using the Ledoit-Wolf lemma [46] was trained on the selected features [37, 47]. The choice of shrinkage was to encourage better generalization.

The proposed R/L+G/B classifier: Our proposed hybrid MI-BCI combined the scores from the aforementioned classifiers as follows. Each R/L, G/B-wm and G/B-csp is a binary LDA classifier. Let the event spaces for the R/L and G/B (-csp and -wm) classifiers be r, l and g, b where r, l, g and b represent right, left, good and bad cursor movements. Then, $P_{RL}(r)$ or $P_{RL}(l)$ indicate the score/probability of the R/L classifier for outputting r or l, respectively. On the other hand, $P_{GB-csp}(g)$ or $P_{GB-csp}(b)$ indicate the score/-probability of the G/B-csp classifier for outputting g or b and similarly, $P_{GB-wm}(g)$ or $P_{GB-wm}(b)$ indicate the score/probability of the G/B-wm for outputting g or b. Then the following holds:

$$P_{RL}(r) + P_{RL}(l) = 1;$$
$$P_{GB-csp}(g) + P_{GB-csp}(b) = 1;$$
$$P_{GB-wm}(g) + P_{GB-wm}(b) = 1.$$

To combine the three classifiers, we trained a logistic regression to combine the scores from the aforementioned classifiers and output r/l as this is the goal of a motor imagery BCI. However, this would not be possible directly since the domain of the G/B (including G/B-csp and G/B-wm) and R/L classifiers are not aligned in general. If the cursor last moved to the right, then $P_{GB-csp}(g)$ and $P_{GB-wm}(g)$ would be directly mapped to $P_{RL}(r)$. However, if the cursor last moved to the left, then $P_{GB-csp}(b)$ and $P_{GB-wm}(b)$ would be mapped to $P_{RL}(r)$ instead. So the selected features were the following depending on the last cursor movement: if the cursor last moved to the right $[P_{RL}(r), P_{GB-csp}(g), P_{GB-wm}(g)]$, and if the cursor last moved to the left $[P_{RL}(r), 1 - P_{GB-csp}(g), 1 - P_{GB-wm}(g)]$ were selected as features and a logistic regression (with three weights and one bias term) was trained on the selected features [17].

2.2.2. Online control

As mentioned earlier, blocks 4-9 comprised online control of the cursor where half of them used the R/L control and the rest used the proposed R/L+G/B control. In each trial, after the target disappeared, the classifier (selected with respect to the block's control

type) was applied to every 1.2 seconds of EEG data at a time and the cursor moved to the right or left according to the output of the classifier.

R/L control: In the R/L control blocks, only the R/L classifier was used to control the cursor the same way as in a conventional motor imagery BCI. This is depicted in figure 3.

R/L+G/B control: The R/L+G/B blocks used the proposed R/L+G/B classifier. As explained earlier, the scores from the R/L, G/B-csp and G/B-wm classifiers were combined through a logistic regression based on the direction of the last cursor movement (CD). This is depicted in figure 4. Note that in the R/L+G/B blocks, in each trial, the first cursor movement was based on the R/L classifier only as no feedback was provided to the participant yet. From the second movement onward, the R/L+G/B classifier was used.

2.3. Metrics

2.3.1. Hit rate and subjective rate

We compared R/L and R/L+G/B controls (blocks 4-9) in various aspects. A successful trial was when the cursor hit the target. Therefore, 'Hit rate (HR)' is defined as the rate/percentage of hit targets in each block (of 20 trials). We define another online score based on the participants' ratings of how much in control they felt in each block. We call this 'subjective rate (SR)'. We report the participants' online scores including hit rates and their subjective ratings.

2.3.2. Information transfer rate

Since there is a trade-off between accuracy and time, we also calculated the information transfer rate (ITR) for each type of control. Let x_i and y_j represent the intended and decoded classes, respectively. Since the target is located at either the right or left, then $i, j \in r, l$. We used the following equation to calculate the ITR [48, 49]:

$$ITR = \Bigg(\sum_{j \in \{r,l\}} -p(y_j) log_2(p(y_j)) + \sum_{i \in \{r,l\}} \sum_{j \in \{r,l\}} p(x_i) p(y_j|x_i) log_2(p(y_j|x_i)) \Bigg) / T,$$

where T is the trial duration. We estimated T using the average number of steps (AST) per trial across the R/L or R/L+G/B blocks for each participant. Note that the timed-out trials, taking the maximum allowed number of steps (i.e. 12), were also included in the calculation of the AST. Since each step took about 1.2 seconds to complete, we estimated $T = 1.2 \times AST + 5$ as the average trial duration including the inter-trial interval of 5 seconds. Since the target location was balanced by design, then $p(x_r) = p(x_l) = 0.5$. Also, $p(y_j|x_i)$ for $i, j \in r, l$, was estimated as the rate of trials that hit the target at the j side when the target was in fact located at the i side. ITR was calculated for each participant, separately for the three R/L and the three R/L+G/B control blocks.

2.3.3. Classification accuracy of steps

As another metric, we looked at the performance of the R/L, G/B-csp, G/B-wm and R/L+G/B classifiers for classifying each cursor movement or step through offline analysis of the recorded data. Our goal was to investigate how the trained classifiers compared with each other and how they contributed to our proposed R/L+G/B classifier. To do so, the calibration (blocks 1-3 of phase 2) and the online recorded data (blocks 4-9 of phase 2) for each participant were epoched 0-1 seconds time-locked to each cursor movement while excluding the last cursor movements. All trials were used for the analyses except for three trials from participant B5 and two trials from participant B6 (all from the recorded online data of these participants) that were removed due to technical issues during recording.

As mentioned earlier, each cursor step may have any of the following 4 different labels: GR, BR, GL or BL. For training and testing the classifiers, these four groups of steps were balanced by subsampling the groups with the larger population. After balancing the groups, there were 136 steps in the calibration blocks, and on average 272.7± 38.4 steps in the online blocks across participants. Note that the steps in the calibration blocks were the same across participants by design, but the steps differed between participants in the online blocks due to the difference in participants' performance. Subsampling (separately within the calibration and online data steps) was done 10 times where each is called an instance of the data with balanced GR, BR, GL and BL groups leading to balanced right, left, good and bad classes. Balanced classes allow for easier interpretation of the results.

We compared the classification accuracy of the R/L, G/B-csp, G/B-wm and our proposed R/L+G/B classifiers in three conditions: 1) trained and tested on the calibration data using cross-validation (TRcal-TEcal), 2) trained and tested on the online data using cross-validation (TRon-TEon) and 3) trained on the calibration data and tested on the online data (TRcal-TEon). The last condition is what took place during the online control. We did these analyses two ways: with and without separating the online steps that belonged to the R/L and R/L+G/B blocks.

3. Results

3.1. Hit rate and subjective rate

Figure 5 reports the online hit rates (HR) and subjective rates (SR) across participants. For each participant, the online part of our experiment comprised six blocks (three R/L and three R/L+G/B blocks). Scores were averaged across the R/L and R/L+G/B blocks separately and reported as HR and SR for

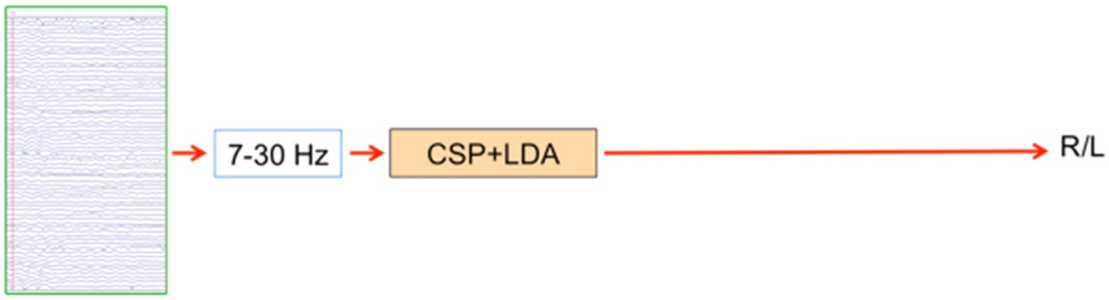


Figure 3. LR/ control using the R/L classifier.

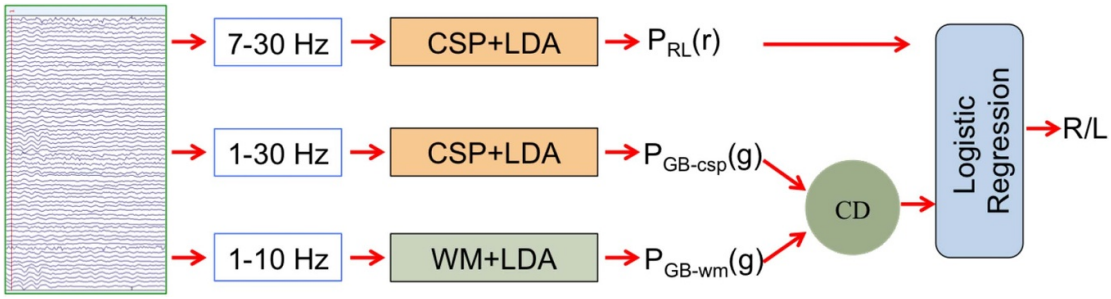


Figure 4. R/L+G/B control using the R/L+G/B classifier. CD indicates the direction of the last cursor movement. If the cursor last moved to the right, $P_{RL}(r)$, $P_{GB-csp}(g)$ and $P_{GB-wm}(g)$ were combined through the logistic regression. However, if the cursor last moved to the left, $P_{RL}(r)$, $1-P_{GB-csp}(g)$ and $1-P_{GB-wm}(g)$ were combined.

Table 1. The sensitivity of the right (R) and left (L) target hit rates for the R/L and R/L+G/B blocks. Note that columns 4 and 7 report the online hit rates (HR) of the R/L and R/L+G/B blocks and are plotted in figure 5(a) as bar plots. The first number in each entry is the average hit rate across the three blocks with the R/L or R/L+G/B control for the corresponding participant and the second number indicates the standard error of the mean.

ID	R/L (R)	R/L (L)	R/L (HR)	R/L+G/B (R)	R/L+G/B (L)	R/L+G/B (HR)
B1	0.43/0.30	0.67/0.19	**0.55/0.08**	0.67/0.03	0.90/0.10	**0.78/0.04**
B2	0.57/0.28	0.60/0.23	**0.58/0.08**	0.63/0.12	0.77/0.12	**0.70/0.03**
B3	0.23/0.12	1.00/0.00	**0.62/0.06**	0.23/0.09	1.00/0.00	**0.62/0.04**
B4	0.97/0.03	0.73/0.07	**0.85/0.03**	0.90/0.10	0.83/0.12	**0.87/0.04**
B5	0.90/0.10	0.66/0.09	**0.78/0.03**	0.93/0.07	1.00/0.00	**0.96/0.04**
B6	0.27/0.12	0.52/0.23	**0.40/0.06**	0.57/0.15	0.70/0.15	**0.63/0.03**
B7	0.33/0.19	0.83/0.17	**0.58/0.02**	0.63/0.12	0.93/0.03	**0.78/0.07**
B8	0.77/0.13	0.53/0.09	**0.65/0.03**	0.90/0.06	0.87/0.03	**0.88/0.03**
B9	1.00/0.00	0.57/0.12	**0.78/0.06**	0.90/0.06	0.90/0.00	**0.90/0.03**
B10	0.40/0.12	0.50/0.17	**0.45/0.13**	0.53/0.13	0.77/0.09	**0.65/0.09**
B11	0.33/0.28	1.00/0.00	**0.67/0.14**	0.40/0.31	0.90/0.10	**0.65/0.10**
B12	0.33/0.20	0.67/0.23	**0.50/0.05**	0.47/0.03	0.77/0.09	**0.62/0.03**

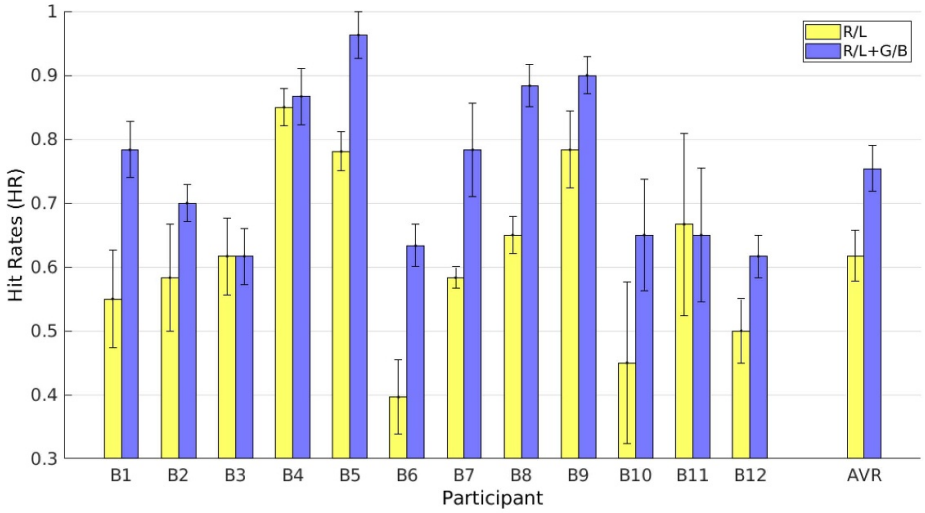
the corresponding control types. Error bars show the standard error of the mean. Wilcoxon signed rank tests show that the HR and SR are significantly better in the R/L+G/B blocks than in the R/L blocks ($p < 0.003$). This indicates a more accurate control in the proposed BCI both objectively and subjectively.

Table 1 presents the sensitivity of the right (R) and left (L) target hit rates in the R/L and R/L+G/B blocks. Even though for some participants (e.g. B4, B8 and B9) there is either negligible bias or bias towards the right, on average across participants there is a slight bias towards the left in both R/L and R/L+G/B blocks. We will show later in section 3.3.1 that this bias is most probably induced by the bias in the R/L classifier at the level of every cursor movement.

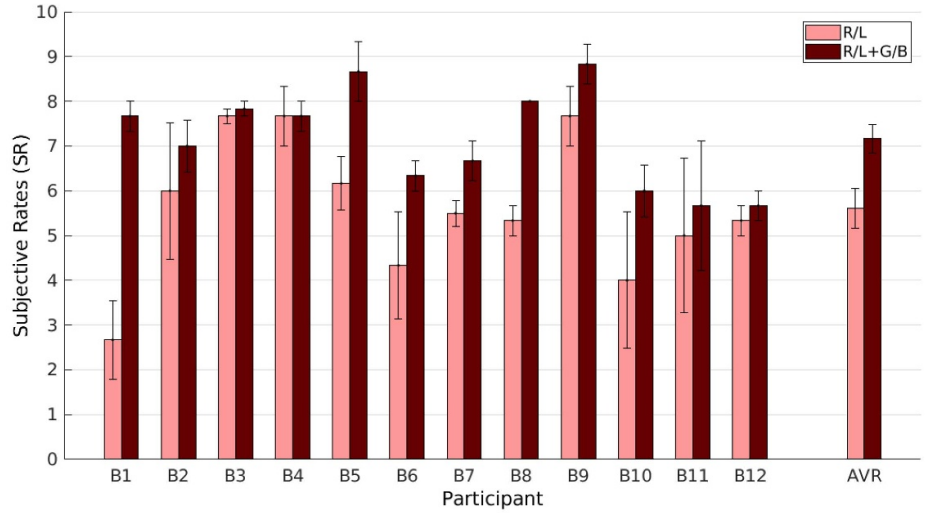
As mentioned earlier, trials could end when the number of cursor movements reached its maximum (i.e. 12 movements). Figure 6 shows the average number of timed-out trials for R/L and R/L+G/B blocks before hitting the target or the corresponding location on the other side. Wilcoxon signed rank test shows no significant difference across participants in the number of timed-out trials between the R/L and R/L+G/B blocks ($p = 0.86$).

3.2. Information transfer rate

AST and ITR for the R/L and R/L+G/B blocks are reported in table 2. Note that 1 bit of information was conveyed when the cursor hit the target. Since the target was 3 steps away from the cursor, the



(a) Online hit rates (HR). The upper limit of the chance interval with significance of p=0.05 is 0.62 [49].



(b) Subjective rates (SR).

Figure 5. The bar heights represent the average online scores of the R/L and R/L+G/B blocks for each participant and AVR indicates the averages across participants. Error bars show the standard error of the mean. Wilcoxon signed rank tests indicate that across participants, the average online hit rates (HR) and subjective rates (SR) in the R/L+G/B blocks are significantly better than in the the R/L blocks ($p < 0.003$).

best achievable ITR was 1/T when AST = 3, i.e. 0.11 628 bit/sec as $T = 1.2 \times AST + 5$ also included the 5 seconds inter-trial interval. This explains the low ITR values reported in table 2. Nevertheless, the ITR in the R/L+G/B blocks is significantly higher across participants than the ITR in the R/L blocks (Wilcoxon signed rank test, $p = 0.007$).

3.3. Classification accuracy on steps

The classification accuracy of the classifiers on steps are reported in figures 7, 8 and 9. The red bars represent the 'cross-validated calibration accuracy' (TRcal-TEcal), that is the accuracy of a classifier trained and tested on the calibration steps. The blue bars on the other hand, represent the 'cross-validated online accuracy' (TRon-TEon), that is the accuracy of a classifier trained and tested on the online steps. The blue and red bar heights indicate the average accuracy of 10 instances of a 5-fold cross-validation over balanced steps. Bar heights represent the average and the error bars indicate the standard deviation for individual participants and standard error of the mean for the average bar (AVR) which is across participants.

The green bars, on the other hand, represent the 'transferred online accuracy' (TRcal-TEon), that is the classification accuracy of a classifier trained on the calibration steps and tested on the online steps, again over balanced classes. Bar heights represent the average and the error bars indicate the standard deviation for individual participants and standard error of the mean for the average bar (AVR) which is across participants. Note that we did not separate

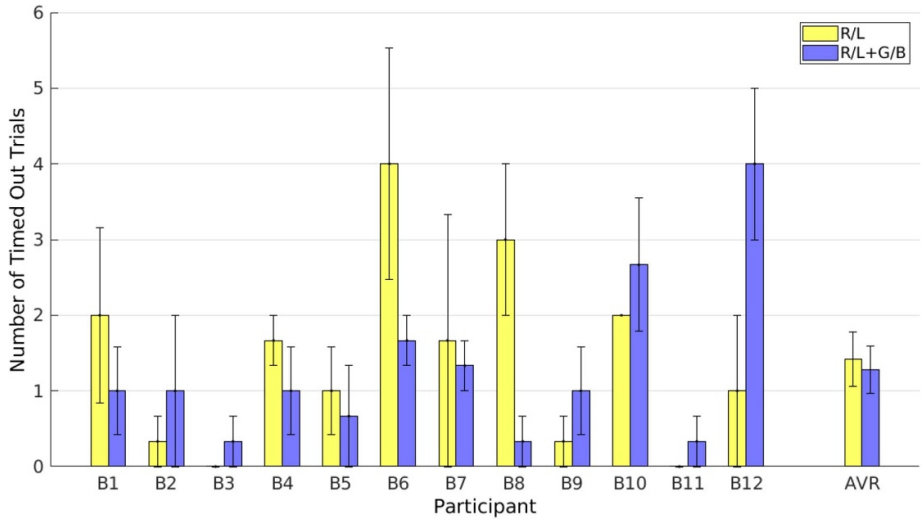


Figure 6. The bar heights represent the average number of timed-out trials per block in R/L and R/L+G/B blocks for each participant and AVR indicates the averages across participants. Error bars show the standard error of the mean. Wilcoxon signed rank test shows no significant difference across participants in the number of timed-out trials between the R/L and R/L+G/B blocks ($p = 0.86$). Note that every block had 20 trials.

Table 2. AST and ITR for each participant across the R/L and R/L+G/B blocks. AST (which includes the timed-out trials) is not different across participants in the R/L and R/L+G/B blocks (Wilcoxon signed rank test, $p = 0.97$). However, the ITR in the R/L+G/B blocks is significantly higher across participants than the ITR in the R/L blocks (Wilcoxon signed rank test, $p = 0.007$).

ID	AST-R/L	AST-R/L+G/B	ITR-R/L (bit/sec)	ITR-R/L+G/B (bit/sec)
B1	5.23	4.98	0.00 292	0.03 045
B2	4.92	4.88	0.00 234	0.01 537
B3	4.00	3.88	0.01 304	0.01 373
B4	5.78	4.62	0.05 638	0.05 430
B5	4.43	3.67	0.02 901	0.08 672
B6	6.38	5.92	0.00 017	0.00 938
B7	5.02	5.67	0.00 830	0.03 993
B8	5.75	4.35	0.01 735	0.05 213
B9	4.02	4.95	0.03 905	0.06 931
B10	5.43	6.17	0.00 001	0.015 6
B11	3.43	4.62	0.02 093	0.01 131
B12	5.18	6.47	0.00 016	0.01 672
AVR	4.96	5.01	0.01 580	0.03 458

Table 3. The upper limit of the chance level (at $p = 0.05$) for each participant for the cross-validated online and the transferred online accuracy reported in figures 7, 8 and 9—i.e. the blue and green bars.

B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
0.57	0.55	0.61	0.56	0.58	0.54	0.56	0.56	0.57	0.54	0.58	0.55

the steps from the R/L and R/L+G/B blocks in these analyses.

Since calibration steps were the same across participants, the upper limit of the chance level (at $p = 0.05$) [50] indicates 0.58 for the red bars. However, the number of available steps in the online blocks varied across participants and the upper limit of the chance level (at $p = 0.05$) for each participant is reported in table 3.

As can be noted from figures 7, 8 and 9, even though on average the cross-validated calibration accuracy (TRcal-TEon) is lower for both G/B-csp and G/B-wm compared to the R/L classifier, the loss from transferring the classifier to the online data is much larger for the R/L classifier. We used Wilcoxon signed rank tests to compare the classification accuracy of TRcal-TEcal (the red bars) and TRcal-TEon (the green bars) for each classifier, across participants. For G/B-csp and G/B-wm, the difference is not statistically significant across participants (Wilcoxon signed rank test, $p > 0.6$). However, for the R/L classifier, TRcal-TEon is worse than TRcal-TEcal and this difference is statistically significant across participants (Wilcoxon signed rank test, $p < 0.001$).

We also compared the cross-validated calibration accuracy (TRcal-TEcal) and the cross-validated online accuracy (TRon-TEon), for each classifier.

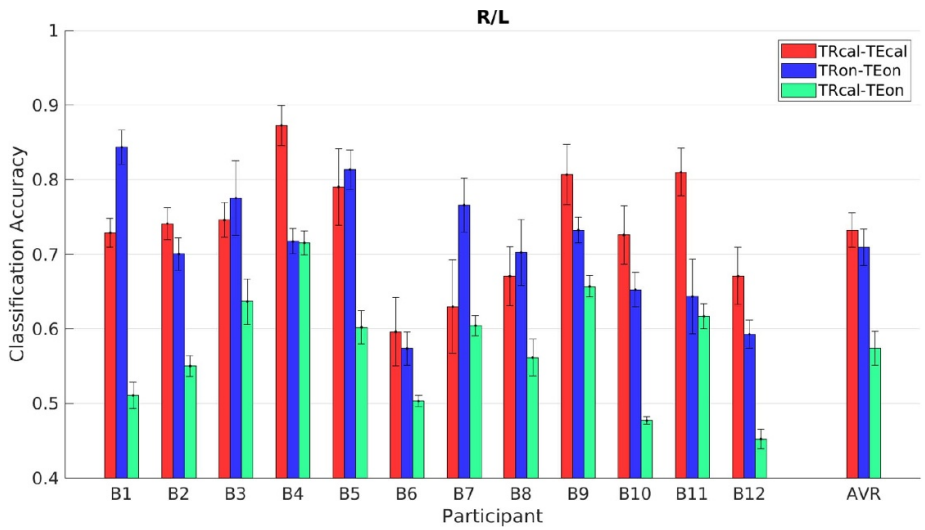


Figure 7. The transferability of the R/L classifier from calibration to online data. The red and blue bars in each plot indicate the cross-validated calibration accuracy (TRcal-TEcal, i.e. trained and tested on the calibration data) and the cross-validated online accuracy (TRon-TEon, i.e. trained and tested on the online data), respectively. The green bars on the other hand, represent the transferred online accuracy (TRcal-TEon, i.e. trained on the calibration data and tested on the online data). AVR indicates the average across participants. Bar heights represent the average and the error bars indicate the standard deviation for individual participants and the standard error of the mean for the average across participants (AVR). Across participants, TRcal-TEon (green bars) is significantly worse than TRcal-TEcal (red bars) – Wilcoxon signed rank test, $p < 0.001$.

This is because one can argue that the different performance as represented by the lower R/L transferred online accuracy is mainly because the data quality is different and less classifiable in the online data. However, a Wilcoxon signed rank test shows that the difference between TRcal-TEcal (the red bars) and TRon-TEon (the blue bars) for the R/L classifier across participants is not statistically significant ($p = 0.42$), suggesting that the R/L motor imagery data quality is not the main cause of the drop in the transferred online accuracy.

Across participants, the transferred online accuracy (TRcal-TEon) of the R/L+G/B classifier as depicted by the green bars in figure 9 is significantly better than the transferred online accuracy (TRcal-TEon) for the R/L classifier as depicted by green bars in figure 7 (Wilcoxon signed rank test, $p < 0.001$). This indicates a higher accuracy in each cursor movement that resulted in the higher hit rates in the RL/+G/B blocks during the online control.

The sensitivity of TRcal-TEcal and TRcal-TEon for the R/L, G/B-csp, G/B-wm, and R/L+G/B classifiers are presented in tables 4, 5, 6 and 7. TRcal-TEon for the R/L classifier seems to have a bias towards the left side for more than half of the participants which we believe could be the cause for the hit rate bias in table 1. However, since this is not the case for all 12 participants, it is most probably not due to the experiment design. Data from more participants are needed for a more detailed sensitivity investigation. It is also important to note that this bias is slightly alleviated in the R/L+G/B classifier again pointing towards the superiority of the R/L+G/B control.

3.3.1. Comparison of the transferred online performance (TRcal-TEon)

We compared the performance of TRcal-TEon for the R/L, G/B-csp, G/B-wm and R/L+G/B classifiers in table 8. Note that these values are also presented as green bars in figures 7, 8 and 9. Wilcoxon signed rank test shows that the difference between TRcal-TEon in the R/L and R/L+G/B classifiers is significant ($p < 0.001$) indicating that the proposed R/L+G/B control provides a more accurate and reliable control for every cursor movement.

The difference in TRcal-TEon across participants for the G/B-csp and the R/L+G/B classifiers is also significant (Wilcoxon signed rank test, $p < 0.001$). However, the difference between the G/B-wm and R/L+G/B is not significant across participants (Wilcoxon signed rank test, $p = 0.2$). Wilcoxon signed rank test for each participant reveals that the R/L+G/B classifier performs significantly different from the G/B-wm classifier for B1, B2, B4, B5, B8, B9, B10 and B11 ($p = 0.02$, Bonferroni corrected for the number of comparisons, i.e. 12). In fact, for B10 and B11, the G/B-wm classifier provides a significantly higher accuracy than the R/L+G/B classifier for the classification of every step.

3.3.2. Comparison of the classification accuracy in the R/L and R/L+G/B blocks

We separated the steps in the R/L and R/L+G/B blocks and redid the analysis in figures 7, 8 and 9 to investigate whether there is a difference for the results of the offline analysis with respect to which type of block the steps were randomly selected

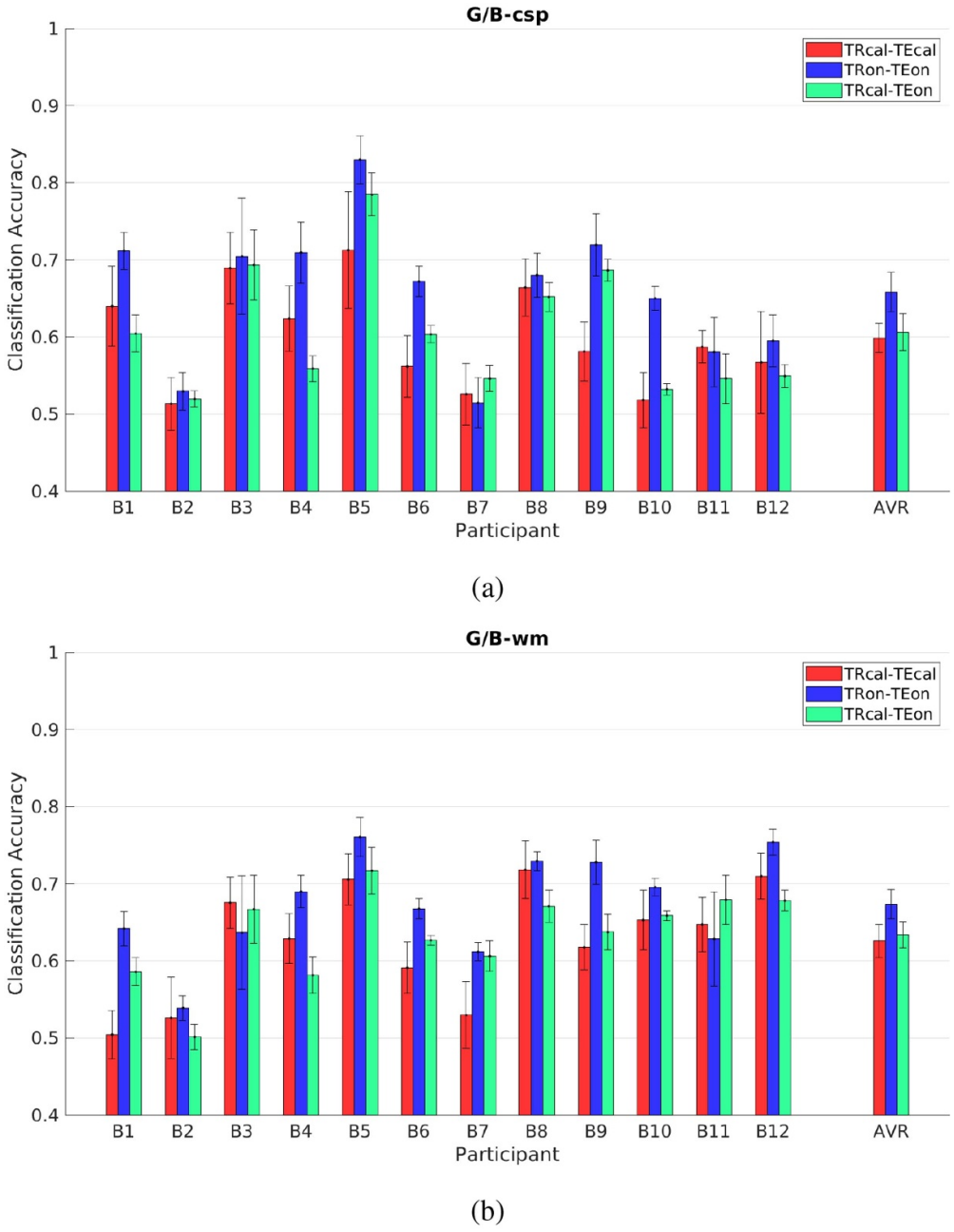


Figure 8. The transferability of the G/B-csp and G/B-wm classifiers from calibration to online data. The red and blue bars in each plot indicate the cross-validated calibration accuracy (TRcal-TEcal, i.e. trained and tested on the calibration data) and the cross-validated online accuracy (TRon-TEon, i.e. trained and tested on the online data), respectively. The green bars on the other hand, represent the transferred online accuracy (TRcal-TEon, i.e. trained on the calibration data and tested on the online data). AVR indicates the average across participants. Bar heights represent the average and the error bars indicate the standard deviation for individual participants and the standard error of the mean for the average across participants (AVR). Across participants, the difference between TRcal-TEcal (red bars) and TRcal-TEon (green bars) is not statistically significant for G/B-csp and G/B-wm classifiers—Wilcoxon signed rank test, $p > 0.60$.

from in the train-test sets. However, we found that across participants, the transferred online accuracy (TRcal-TEon) in the R/L, G/B-csp, G/B-wm and R/L+G/B classifiers were not different between the R/L and R/L+G/B blocks (Wilcoxon signed rank test, $p > 0.11$, Bonferroni corrected with the number of performed tests, i.e. 4). Moreover, across participants, the cross-validated online accuracy (TRon-TEon) in the R/L, G/B-csp, G/B-wm and R/L+G/B classifiers were also not different between the R/L and R/L+G/B blocks (Wilcoxon signed rank test, $p > 0.61$, Bonferroni corrected with the number of performed tests, i.e. 4).

4. Discussion

In hybrid BCI systems, multiple sources of information can be processed *simultaneously* or *sequentially*

Table 4. The sensitivity of the cross-validated calibration classifier (TRcal-TEcal) in predicting the right (R) and left (L) for the R/L and R/L+G/B classifiers. The first number in each entry is the average accuracy across the 10 instances of the data and the second number indicates the standard deviation.

ID	R/L (R)	R/L (L)	R/L	R/L+G/B (R)	R/L+G/B (L)	R/L+G/B
B1	0.74/0.03	0.72/0.04	0.73/0.02	0.68/0.05	0.74/0.07	0.72/0.04
B2	0.76/0.02	0.73/0.05	0.74/0.02	0.68/0.05	0.61/0.06	0.65/0.04
B3	0.76/0.04	0.73/0.03	0.75/0.02	0.81/0.04	0.80/0.03	0.81/0.03
B4	0.87/0.04	0.88/0.04	0.87/0.03	0.88/0.04	0.85/0.04	0.86/0.03
B5	0.85/0.03	0.73/0.09	0.79/0.05	0.85/0.05	0.84/0.07	0.84/0.05
B6	0.54/0.06	0.65/0.05	0.60/0.05	0.55/0.04	0.69/0.05	0.61/0.03
B7	0.63/0.08	0.63/0.06	0.63/0.06	0.62/0.10	0.64/0.05	0.63/0.06
B8	0.69/0.06	0.65/0.04	0.67/0.04	0.76/0.05	0.75/0.05	0.75/0.03
B9	0.82/0.05	0.79/0.05	0.81/0.04	0.76/0.04	0.74/0.03	0.75/0.04
B10	0.78/0.03	0.68/0.06	0.73/0.04	0.77/0.05	0.74/0.04	0.74/0.02
B11	0.80/0.03	0.82/0.06	0.81/0.03	0.82/0.03	0.81/0.06	0.82/0.03
B12	0.65/0.05	0.69/0.04	0.67/0.04	0.72/0.04	0.77/0.05	0.75/0.04

Table 5. The sensitivity of the cross-validated calibration classifier (TRcal-TEcal) in predicting the good (G) and bad (B) for the G/B-csp and G/B-wm classifiers. The first number in each entry is the average accuracy across the 10 instances of the data and the second number indicates the standard deviation.

ID	G/B-csp(G)	G/B-csp(B)	G/B-csp	G/B-wm(G)	G/B-wm(B)	G/B-wm
B1	0.62/0.05	0.66/0.06	0.64/0.05	0.50/0.03	0.53/0.05	0.50/0.03
B2	0.47/0.03	0.55/0.05	0.51/0.03	0.53/0.05	0.51/0.06	0.53/0.05
B3	0.68/0.06	0.70/0.04	0.69/0.05	0.67/0.05	0.69/0.02	0.68/0.03
B4	0.61/0.05	0.63/0.05	0.62/0.04	0.64/0.05	0.62/0.04	0.63/0.03
B5	0.73/0.08	0.70/0.08	0.71/0.08	0.74/0.05	0.70/0.04	0.71/0.03
B6	0.58/0.06	0.55/0.07	0.56/0.04	0.62/0.04	0.59/0.05	0.59/0.03
B7	0.56/0.07	0.49/0.04	0.53/0.04	0.55/0.08	0.54/0.05	0.53/0.04
B8	0.66/0.03	0.67/0.06	0.66/0.04	0.76/0.04	0.71/0.03	0.72/0.04
B9	0.57/0.05	0.59/0.05	0.58/0.04	0.64/0.06	0.63/0.04	0.62/0.03
B10	0.51/0.08	0.53/0.04	0.52/0.04	0.68/0.03	0.65/0.04	0.65/0.04
B11	0.67/0.05	0.50/0.04	0.59/0.02	0.68/0.04	0.63/0.03	0.65/0.03
B12	0.62/0.09	0.51/0.06	0.57/0.07	0.76/0.04	0.69/0.04	0.71/0.03

Table 6. The sensitivity of the transferred online classifier (TRcal-TEon) in predicting the right (R) and left (L) for the R/L and R/L+G/B classifiers. The first number in each entry is the average accuracy across the 10 instances of the data and the second number indicates the standard deviation.

ID	R/L (R)	R/L (L)	R/L	R/L+G/B (R)	R/L+G/B (L)	R/L+G/B
B1	0.44/0.04	0.59/0.02	0.51/0.02	0.58/0.04	0.67/0.03	0.62/0.02
B2	0.53/0.02	0.57/0.01	0.55/0.01	0.61/0.02	0.53/0.01	0.57/0.01
B3	0.37/0.05	0.91/0.03	0.64/0.03	0.46/0.06	0.93/0.02	0.70/0.03
B4	0.77/0.01	0.66/0.03	0.72/0.02	0.77/0.02	0.69/0.02	0.73/0.02
B5	0.61/0.02	0.59/0.04	0.60/0.02	0.82/0.01	0.79/0.04	0.80/0.02
B6	0.43/0.01	0.58/0.01	0.50/0.01	0.59/0.02	0.64/0.01	0.62/0.01
B7	0.52/0.02	0.69/0.02	0.60/0.01	0.53/0.03	0.74/0.02	0.64/0.01
B8	0.59/0.02	0.53/0.03	0.56/0.02	0.73/0.02	0.71/0.02	0.72/0.01
B9	0.65/0.03	0.66/0.02	0.66/0.01	0.81/0.03	0.71/0.03	0.76/0.01
B10	0.39/0.01	0.56/0.01	0.48/0.01	0.54/0.01	0.68/0.01	0.61/0.01
B11	0.42/0.03	0.81/0.02	0.62/0.02	0.49/0.04	0.73/0.03	0.61/0.02
B12	0.40/0.02	0.51/0.02	0.45/0.01	0.60/0.04	0.68/0.01	0.64/0.02

for improved performance [51]. Earlier work in the literature proposed to use the error-related brain activity *sequentially* to either correct an executed action [14, 24, 25] or to adapt a trained classifier [26, 27]. A different approach in [22, 23], proposed a motor imagery BCI for 1-D cursor control that *simultaneously* integrated the user brain activity in response to the cursor change in its direction of movement. In this study, we went beyond cursor direction changes and proposed a hybrid BCI to *simultaneously* combine the feedback-related brain activity to every cursor movement with the motor imagery signal. We showed that across 12 participants, they were able to control the BCI using the proposed method significantly better than using the conventional motor imagery BCI. Our results showed that this improvement is significant in terms of the classification accuracy of single cursor movements,

Table 7. The sensitivity of the transferred online classifier (TRcal-TEon) in predicting the good (G) and bad (B) for the G/B-csp and G/B-wm classifiers. The first number in each entry is the average accuracy across the 10 instances of the data and the second number indicates the standard deviation.

ID	G/B-csp(G)	G/B-csp(B)	G/B-csp	G/B-wm(G)	G/B-wm(B)	G/B-wm
B1	0.71/0.03	0.50/0.02	0.60/0.02	0.62/0.04	0.56/0.02	0.59/0.02
B2	0.72/0.01	0.32/0.01	0.52/0.01	0.59/0.02	0.48/0.01	0.50/0.02
B3	0.74/0.07	0.65/0.05	0.69/0.05	0.72/0.06	0.65/0.06	0.67/0.04
B4	0.61/0.02	0.51/0.01	0.56/0.02	0.73/0.04	0.43/0.02	0.58/0.02
B5	0.87/0.04	0.70/0.02	0.78/0.03	0.84/0.06	0.58/0.02	0.72/0.03
B6	0.58/0.01	0.62/0.01	0.60/0.01	0.61/0.02	0.65/0.01	0.63/0.01
B7	0.43/0.03	0.66/0.01	0.55/0.02	0.55/0.03	0.67/0.03	0.61/0.02
B8	0.76/0.03	0.55/0.01	0.65/0.02	0.80/0.03	0.65/0.02	0.67/0.02
B9	0.74/0.03	0.63/0.01	0.69/0.01	0.63/0.03	0.73/0.02	0.64/0.02
B10	0.27/0.01	0.80/0.01	0.53/0.01	0.71/0.01	0.64/0.01	0.66/0.01
B11	0.47/0.06	0.62/0.03	0.55/0.03	0.71/0.05	0.68/0.03	0.68/0.03
B12	0.48/0.01	0.62/0.02	0.55/0.02	0.81/0.02	0.59/0.01	0.68/0.01

Table 8. The transferred online accuracy (TRcal-TEon) for the R/L, G/B-csp, G/B-wm and R/L+G/B classifiers. For B1-B12, the first number in each entry is the average accuracy across the 10 instances of the data and the second number indicates the standard deviation. In the last row, each entry indicates the average across participants for the corresponding column and the second number indicates the standard error of the mean.

ID	R/L	G/B-csp	G/B-wm	R/L+G/B
B1	0.51/0.02	0.60/0.02	0.59/0.02	0.62/0.02
B2	0.55/0.01	0.52/0.01	0.50/0.02	0.57/0.01
B3	0.64/0.03	0.69/0.05	0.67/0.04	0.70/0.03
B4	0.72/0.02	0.56/0.02	0.58/0.02	0.73/0.02
B5	0.60/0.02	0.78/0.03	0.72/0.03	0.80/0.02
B6	0.50/0.01	0.60/0.01	0.63/0.01	0.62/0.01
B7	0.60/0.01	0.55/0.02	0.61/0.02	0.64/0.01
B8	0.56/0.02	0.65/0.02	0.67/0.02	0.72/0.01
B9	0.66/0.01	0.69/0.02	0.64/0.02	0.76/0.01
B10	0.48/0.01	0.53/0.01	0.66/0.01	0.61/0.01
B11	0.62/0.02	0.55/0.03	0.68/0.03	0.61/0.02
B12	0.45/0.01	0.55/0.02	0.68/0.01	0.64/0.02
AVR	0.57/0.02	0.61/0.02	0.63/0.02	0.67/0.02

average hit rate, the perceived subjective rate and information-transfer rate.

We further showed that the performance of the motor imagery classifier (R/L) was negatively affected when transferred from calibration to the online control. We believe that this is in part due to the drift in the EEG data, but also in part due to the fact that some participants performed worse than the sham calibration feedback and this may be affecting the transferability of the R/L classifier – as mentioned earlier, during calibration, participants were provided with sham feedback but were not aware of it. At the same time, the error-related brain activity classifiers (G/B-csp and G/B-wm) were both better transferred from calibration to the online control than the R/L classifier. In other words, G/B-csp and G/B-wm classifiers were able to provide a more reliable (consistent from calibration to online) classifier than the motor imagery R/L classifier. This difference may also be influenced by the fact that the error-related brain activity was time-locked to the stimulus onset as opposed to the user-generated motor imagery signal. Moreover, we hypothesize that the R/L signals could be impacted by the G/B signals which occur in similar frequency bands [17, 23]. However, for G/B, the signals of 'goodness' and 'badness' were either not affected or actually improved [49, 52] resulting in a more consistent classifier from calibration to online use.

Other work in the literature also mention that the error-related brain activity can result in a reliable classifier over time. For instance, the authors in [53] and [54] showed that a trained error-related brain activity classifier could be reliably used in a test session over long periods of time. However, to the best of our knowledge, our work is the first that provides evidence comparing the reliability of the error-related brain activity with that of the motor imagery signal.

As mentioned earlier, error-related brain activity was used in hybrid BCIs before, e.g. to correct mistakes of the primary motor imagery classifier, or to improve it by providing labels for additional training data [14, 23, 24, 26, 27]. Error-related brain activity has also been used as the sole input for 2D cursor control using a passive BCI [55–58]. This was implemented using cognitive probing, i.e. the active elicitation of automatic (in this case, error- or prediction-related) brain response by the system without the explicit involvement of the participant [59]. This was implemented using a windowed-means approach similar to the G/B-wm classifier used here. The current data (see table 8) show that in the transferred online accuracy while R/L+G/B outperforms both R/L and G/B-csp classifiers, the performance of the R/L+G/B classifier is not significantly different than that of the G/B-wm classifier across participants. Given this, one may ask whether a passive control based on the G/B-wm classifier—i.e. a control based solely on the error-related brain activity, as in [56] – is capable of providing a comparable or even better control compared to a hybrid BCI. Further comparisons for each participant revealed that the R/L+G/B classifier provided significantly better classification

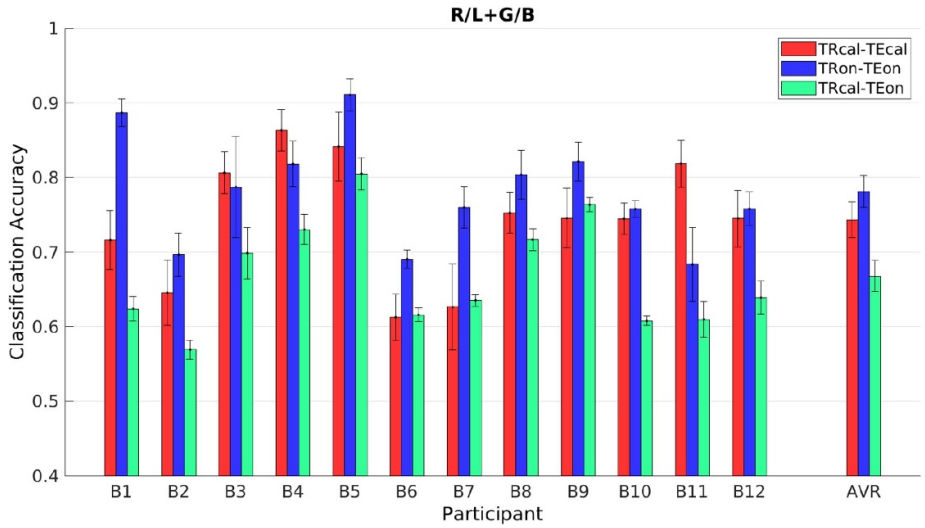


Figure 9. The transferability of the R/L+G/B classifier from calibration to online data. The red and blue bars in each plot indicate the cross-validated calibration accuracy (TRcal-TEcal, i.e. trained and tested on the calibration data) and the cross-validated online accuracy (TRon-TEon, i.e. trained and tested on the online data), respectively. The green bars on the other hand, represent the transferred online accuracy (TRcal-TEon, i.e. trained on the calibration data and tested on the online data). AVR indicates the averages across participants. Bar heights represent the average and the error bars indicate the standard deviation for individual participants and the standard error of the mean for the average across participants (AVR).

accuracy for 6 participants (B1, B2, B4, B5, B8 and B9). On the other hand, the G/B-wm classifier was significantly better than the R/L+G/B for 2 participants (B10 and B11). While the answer to this question could very well be user-dependent, a possible factor that is not answered in this study is whether the error-related brain activity is different (potentially more pronounced) when users are actively controlling a BCI [60] than in a passive BCI.

5. Conclusion and future work

In this work, we showed the efficacy of our proposed hybrid BCI that combines the motor imagery signal with the error-related brain activity in response to the BCI error. Our proposed BCI significantly outperforms the conventional motor imagery BCI in terms of accuracy, information transfer rate and the perceived subjective rate. We further analyzed the two components of our proposed hybrid BCI, namely the motor imagery and the error-related brain activity classifiers. We showed for the first time that compared to the motor imagery classifier, the classifier based on the error-related brain activity is more consistent with respect to transferring the trained classifier to the online control. This finding helps explain why the proposed hybrid BCI outperforms a conventional motor imagery BCI, and may help improve other forms of BCI as well.

Future work should compare the transferability of the classifiers from calibration to online using other approaches beyond common spatial patterns and windowed-means (e.g. [11, 61–65]). Further study is also required to conclude if a passive BCI based on either the G/B-csp or G/B-wm classifiers or a combination of the two, may be a better choice for some participants as opposed to the proposed hybrid BCI. Future work can also shed light on why the performance of the motor imagery classifier is less consistent than the error-related brain activity classifier. This can allow us to better understand the interaction between the two sources of information (i.e. the error-related brain activity and the motor imagery signal) and direct us to design an even more reliable BCI system.

Acknowledgment

This work was supported by NSF IIS 1219200, IIS 1817226, SMA 1041755, and IIS 1528214, FISP G2171, G3155, NIH 5T32MH020002-18, the Deutsche Forschungsgemeinschaft (ZA 821/3-1), and UCSD Mary Anne Fox dissertation year fellowship.

ORCID iDs

Mahta Mousavi
https://orcid.org/0000-0002-0139-7295
Laurens R Krol
https://orcid.org/0000-0002-6585-2795
Virginia R de Sa
https://orcid.org/0000-0002-0989-3576

References

[1] Vidal J J 1973 Toward direct brain-computer communication *Ann. Rev. Biophys. Bioeng.* **2** 157–80
[2] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23
[3] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain-computer communication *Proc. IEEE* **89** 1123–34

[4] Wolpaw J R, Birbaumer N, McFarland D J, G Pfurtscheller and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
[5] Nicolas-Alonso L F and Gomez-Gil J 2012 Brain computer interfaces, a review *Sensors* **12** 1211–79
[6] Cervera M A, Soekadar S R, Ushiba J, Millán J del R, Liu M, Birbaumer N and Garipelli G 2018 Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis *Ann. Clin. Translational Neurol.* **5** 651–63
[7] Pfurtscheller G, Brunner C, Schlögl A and Da Silva F H L 2006 Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks *NeuroImage* **31** 153–9
[8] Müller-Gerking J, Pfurtscheller G and Flyvbjerg H 1999 Designing optimal spatial filters for single-trial EEG classification in a movement task *Clin. Neurophysiol.* **110** 787–98
[9] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Muller K-R 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Process. Mag.* **25** 41–56
[10] Hammon P S and de Sa V R 2007 Pre-processing and meta-classification for brain-computer interfaces *IEEE Trans. Biomed. Eng.* **54** 518–25
[11] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A and Yger F 2018 A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update *J. Neural Eng.* **15** 031005
[12] Shenoy P, Krauledat M, Blankertz B, Rao R P N and Klaus-Robert M 2006 Towards adaptive classification for BCI *J. Neural Eng.* **3** R13
[13] Schalk G, Wolpaw J R W, McFarlandand D J and Pfurtscheller G 2000 EEG-based communication: presence of an error potential *Clin. Neurophysiol.* **111** 2138–44
[14] Ferrez P W and Millán J del R 2008 Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy *Proc. of the 4th Int. Brain-Computer Interface Workshop and Training Course* pp 197–202
[15] Koerner A S, Zhang Q and de Sa V R 2013 The effect of real-time positive and negative feedback on motor imagery performance *Int. BCI Meeting*
[16] Chavarriaga R, Sobolewski A and Millán J del R 2014 Errare machinale est: the use of error-related potentials in brain-machine interfaces *Front. Neuroscie.* **8** 208
[17] Mousavi M, Koerner A S, Zhang Q, Noh E and de Sa V R 2017 Improving motor imagery BCI with user response to feedback *Brain-Computer Interfaces* **4** 74–86
[18] Lopes-Dias C, Sburlea A I and Müller-Putz G R 2019 Online asynchronous decoding of error-related potentials during the continuous control of a robot *Sci. Rep.* **9** 1–9
[19] Yousefi R, Sereshkeh A R and Chau T 2019 Development of a robust asynchronous brain-switch using ErrP-based error correction *J. Neural Eng.* **16** 066042
[20] Kumar A, Gao L, Pirogova E and Fang Q 2019 A review of error-related potential-based brain–computer interfaces for motor impaired people *IEEE Access* **7** 142451–66
[21] Parra L C, Spence C D, Gerson A D and Sajda P 2003 Response error correction-a demonstration of improved human-machine performance using real-time EEG monitoring *IEEE Trans. Neural Syst. Rehabil. Eng.* **11** 173–7
[22] Koerner A S and de Sa V R 2012 A novel method to integrate error detection into motor imagery BCI *Workshop on Brain-Machine Body Interfaces, 34th Annual Int. Conf. of the IEEE Eng. Med. Biol. Soc.*
[23] Koerner A S 2013 Investigating natural control signals for brain-computer interfaces *PhD Thesis*, UC San Diego
[24] Spüler M, Bensch M, Kleih S, Rosenstiel W, Bogdan M and Andrea K 2012 Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI *Clin. Neurophysiol.* **123** 1328–37
[25] Bhattacharyya S, Konar A and Tibarewala D N 2017 Motor imagery and error related potential induced position control of a robotic arm *IEEE/CAA J. Automatica Sinica* **4** 639–50
[26] Artusi X, Niazi I K, Lucas M-Fçoise and Farina D 2011 Performance of a simulated adaptive BCI based on experimental classification of movement-related and error potentials *IEEE J. Emerging Sel. Top. Circuits Syst.* **1** 480–8
[27] Spüler M, Rosenstiel W and Bogdan M 2012 Online adaptation of a c-VEP brain-computer interface (BCI) based on error-related potentials and unsupervised learning *PloS one* **7** e51077
[28] Milekovic T, Ball T, Schulze-Bonhage A, Aertsen A and Mehring C 2013 Detection of error related neuronal responses recorded by electrocorticography in humans during continuous movements *PloS one* **8** e55235
[29] Wilson N R, Sarma D, Wander J D, Weaver K E, Ojemann J G and Rao R P N 2019 Cortical topography of error-related high-frequency potentials during erroneous control in a continuous control brain–computer interface *Front. Neurosci.* **13** 502
[30] Even-Chen N, Stavisky S D, Kao J C, Ryu S I and Shenoy K V 2017 Augmenting intracortical brain-machine interface with neurally driven error detectors *J. Neural Eng.* **14** 066007
[31] Mousavi M 2019 Novel machine learning and design methods to improve EEG-based motor imagery brain-computer interfaces *PhD Thesis* UC San Diego
[32] Klem G H, Lüders H O, Jasper H H and Elger C 1999 The ten-twenty electrode system of the Int. Federation *Electroencephalogr. Clin. Neurophysiol.* **52** 3–6
[33] San Diego U C Swartz Center for Computational Neuroscience Simulation and Neuroscience Application Platform (SNAP) (https://github.com/sccn/SNAP)
[34] San Diego U C Swartz Center for Computational Neuroscience Lab Streaming Layer (LSL) (https://github.com/sccn/labstreaminglayer)
[35] Oliphant T E 2006 *A Guide to NumPy* vol 1 (USA: Trelgol Publishing)
[36] Jones E, Oliphant T and Peterson P 2001 SCiPy: Open source scientific tools for Python (www.scienceopen.com/document?vid\protect$\relax=$ab12905a-8a5b-43d8-a2bb-defc771410b9)
[37] Pedregosa F *et al* 2011 Scikit-learn: Machine learning in Python *J. Machine Learning Res.* **12** 2825–30
[38] MATLAB and Statistics Toolbox Release 2018b 2018 The MathWorks Inc., Natick, Massachusetts, United States
[39] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–21
[40] Mousavi M and de Sa V R 2017 Towards elaborated feedback for training motor imagery brain computer interfaces *Proc. of the 7th Graz Brain-Computer Conf. 2017* pp 332–7
[41] Oldfield R C 1971 The assessment and analysis of handedness: the Edinburgh inventory *Neuropsychologia* **9** 97–113
[42] John O P and Srivastava S *et al* 1999 The Big Five trait taxonomy: History, measurement and theoretical perspectives *Handbook Personality: Theory Res.* **2** 102–38
[43] Rotter J B 1966 Generalized expectancies for internal versus external control of reinforcement *Psychol. Monographs: General Appl.* **80** 1
[44] Scheier M F, Carver C S and Bridges M W 1994 Distinguishing optimism from neuroticism (and trait anxiety, self-mastery and self-esteem): a reevaluation of the Life Orientation Test *J. Personality Social Psychol.* **67** 1063
[45] Ramoser H, Muller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441–6
[46] Ledoit O and Wolf M 2004 Honey, I shrunk the sample covariance matrix *J. Portfolio Manage.* **30** 110–19

[47] Blankertz B, Lemm S, Treder M, Haufe S and Klaus-Robert M 2011 Single-trial analysis and classification of ERP components – a tutorial *NeuroImage* **56** 814–25
[48] Fatourechi M, Mason S G, Birch G E and Ward R K 2006 Is information transfer rate a suitable performance measure for self-paced brain interface systems? *2006 IEEE Int. Symp. on Signal Process. Information Technol.* pp 212–16
[49] de Sa V R 2012 An interactive control strategy is more robust to non-optimal classification boundaries *Proc. of the 14th ACM Int. Conf. on Multimodal Interaction*, ICMI '12 New York, NY ACM USA pp 579–86
[50] Müller-Putz G, Scherer R, Brunner C, Leeb R and Pfurtscheller G 2008 Better than random: a closer look on BCI results *Int. J. Bioelectromagnet.* **10** 52–5
[51] Pfurtscheller G *et al* 2010 The hybrid BCI *Front. Neurosci.* **4** 3
[52] de Sa V R 2017 Improving information transfer rate in active BCIs *Proc. of the 7th Graz Brain-Computer Conf. 2017*
[53] Ferrez P W and Millán J del R 2008 Error-related EEG potentials generated during simulated brain–computer interaction *IEEE Trans. Biomed. Eng.* **55** 923–9
[54] Chavarriaga R and Millán J del R 2010 Learning from EEG error-related potentials in noninvasive brain-computer interfaces *IEEE Trans. Neural Syst. Rehabil. Eng.* **18** 381–8
[55] Zander T O and Kothe C 2011 Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general *J. Neural Eng.* **8** 025005
[56] Zander T O, Krol L R, Birbaumer N P and Gramann K 2016 Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity *Proc. of the National Academy of Sciences* **113** 14898–903
[57] Stivers J M, Krol L R, de Sa V R and Zander T O 2016 Spelling with cursor movements modified by implicit user response *Proc. of the 6th Int. Brain-Computer Interface Meeting* p 28
[58] Krol L R, Andreessen L M and Zander T O 2018 Passive brain-computer interfaces: A perspective on increased interactivity *Brain-Computer Interfaces Handbook: Technological and Theoretical Advances* pp 69–86
[59] Krol L R, Haselager P and Zander T O 2020 Cognitive and affective probing: a tutorial and review of active learning for neuroadaptive technology *J. Neural Eng.* **17** 012001
[60] van Schie H T, Mars R B, Coles M G H and Bekkering H 2004 Modulation of activity in medial frontal and motor cortices during error observation *Nat. Neurosci.* **7** 549–54
[61] Noh E and De Sa V R 2013 Canonical correlation approach to common spatial patterns *2013 6th Int. IEEE/ Conf. on Neural Engineering (NER)* pp 669–72
[62] Lotte F and Guan C 2010 Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms *IEEE Trans. Biomed. Eng.* **58** 355–62
[63] Mousavi M and de Sa V R 2019 Spatio-temporal analysis of error-related brain activity in active and passive brain–computer interfaces *Brain-Computer Interfaces* **6** 118–27
[64] Congedo M, Barachant A and Bhatia R 2017 Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review *Brain-Computer Interfaces* **4** 155–74
[65] Mousavi M and de Sa V R 2019 Temporally adaptive common spatial patterns with deep convolutional neural networks *2019 41st Annual Int. Conf. of the IEEE Eng. Med. Biol. Soc.* pp 4533–6