Brain-Computer Interfaces

ISSN: 2326-263X (Print) 2326-2621 (Online) Journal homepage: http://www.tandfonline.com/loi/tbci20

# Improving motor imagery BCI with user response to feedback

Mahta Mousavi, Adam S. Koerner, Qiong Zhang, Eunho Noh & Virginia R. de Sa

**To cite this article:** Mahta Mousavi, Adam S. Koerner, Qiong Zhang, Eunho Noh & Virginia R. de Sa (2017): Improving motor imagery BCI with user response to feedback, Brain-Computer Interfaces, DOI: 10.1080/2326263X.2017.1303253

**To link to this article:** http://dx.doi.org/10.1080/2326263X.2017.1303253

BRAIN-COMPUTER INTERFACES, 2017
http://dx.doi.org/10.1080/2326263X.2017.1303253

# Improving motor imagery BCI with user response to feedback

Mahta Mousavi[a], Adam S. Koerner[b], Qiong Zhang[d], Eunho Noh[a] and Virginia R. de Sa[b,c]

[a]Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, USA; [b]Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA, USA; [c]Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA; [d]Carnegie Mellon University, Pittsburgh, PA, USA

**ABSTRACT**

In brain-computer interface (BCI) systems, the non-stationarity of brain signals is known to be a challenge for training robust classifiers as other brain processes produce signals that coincide with those resulting from the desired brain activity. One source of interference is the user's cognitive response to the provided BCI feedback. In the case of motor imagery paradigms, this feedback can for instance be a cursor moving on the screen. The response to such feedback has been shown in general to be a source of noise that can add to the non-stationarity of the brain signal; however, in this work, we show that the user's brain response to this feedback can be used to improve the BCI performance. We first show in a motor imagery task that the user's brain responds to the direction of cursor movement, which is different for the cursor moving towards or away from the target (i.e. BCI feedback), and this feedback-related information is present in frequency bands similar to those used in motor imagery. Next, we propose a classifier that combines the user response to feedback together with the motor imagery signal itself, and show that this combined classifier can significantly outperform a conventional motor imagery classifier. Our results show an average of 11% and up to 22% improvement in classification accuracy across 10 participants.

**ARTICLE HISTORY**
Received 3 August 2016
Accepted 1 March 2017

**KEYWORDS**
Motor imagery; visual feedback; non-stationarity of brain signals; error-related spectral perturbation; error-related potentials

## 1. Introduction

Brain-computer interface (BCI) systems collect and infer neural signals without the use of normal neuromuscular pathways [1–4]. These systems were originally developed for locked-in patients who suffer from amyotrophic lateral sclerosis (ALS) or brainstem stroke syndrome. Motor imagery (MI) – where a user imagines a movement without producing it – enables effective non-invasive BCI control when measured with EEG [5]. Imagined movements result in an event-related desynchronization (ERD) (decrease in power) in the mu band (8–13 Hz) [6, 7]; a similar ERD occurs in the beta band as well (14–30 Hz) [8, 9]. Motor imagery of different body parts results in somewhat different spatial patterns of desynchronization across the scalp and the BCI uses these features to distinguish among the imagined movement classes. For example, a user might imagine moving their right hand to move a cursor in one direction and imagine moving their left hand to move the cursor in another. The targets can be mapped to different actions to allow a user to interact with the world (e.g. turn a light on or off or move a robot arm to one object or another).

Motor imagery BCIs have an advantage over other non-invasive BCIs as they require neither external stimulation (as needed for steady-state visually evoked potential (SSVEP) [10] and P300-style BCIs [1, 11]) nor gaze control by the user. However, they also have their own challenges, such as their low reliability and information transfer rate, which can confound system use outside a controlled laboratory environment. This low reliability is due in part to the non-stationarity of the brain signals. As these BCIs are not dependent on external stimulation, they rely on signals created internally by the user and are susceptible to contamination from other brain activity of the user. Due to the low classification accuracies when classification is based on short time windows of EEG recordings containing the motor imagery signal, classification results from several time windows of motor imagery EEG are combined in order to increase the reliability of target selection. The usual method is to use several small time windows of 500–1000 ms and provide feedback in the form of an incremental cursor movement towards the decoded target (for that window) [12]. In this way, the accuracy of hitting the desired target can be boosted at the

**CONTACT** Mahta Mousavi mahta@ucsd.edu

cost of more time. Improving the classification accuracy in each time step could greatly improve the information transfer rate of the system by reducing the number of steps required to hit the target accurately.

One important factor contributing to contamination of the motor imagery signal is the feedback provided by the BCI itself. It is well understood that BCI feedback is important to help the user learn to perform motor imagery and that providing feedback affects the performance of the user, e.g. [13–17]. One type of EEG signal that can be generated in response to feedback is the error-related potential (ErrP), which can help distinguish between movements in the desired direction and those in the non-desired direction [18]. Schalk et al. [19] reported that in an EEG-based cursor control BCI through modulation of mu and beta rhythms, participants elicited error-related potentials at the end of erroneous trials. Ferrez and Millan in [20] reported detection of error-related potentials in an experiment where the participants manually controlled the cursor movement. In another study [21], they showed the application of the detected error-related potential in improving the classification rate of a motor imagery task by undoing movements associated with detected error-related potentials. Artusi et al. [22] proposed a strategy of repeating trials when an error potential is detected. The authors in [23] and [24] improved performance of speller BCIs by correcting for the detected error from user feedback. Detection of errors for adaptably calibrating a code-modulated visual evoked potential (c-VEP) classifier was proposed in a c-VEP BCI [25]. Kreilinger et al. [26] showed the classifiability of error-related potentials during continuous movement of an artificial arm provided as delayed feedback in a right-hand motor imagery task. Koerner [15, 27, 28] investigated methods to classify and use error-related potentials while the participant was performing motor imagery. Omedes et al. [29] investigated error-related potentials in the frequency domain but only in the lower frequencies (theta band, 4–8 Hz). In an electrocorticographic (ECoG) BCI study, Milekovic et al. [30] found error-related neural responses in lower-frequency bands similar to ErrP studies in EEG as well as high gamma (beyond 60 Hz) carrying partially independent information. However, high gamma information is difficult to obtain from EEG data.

The error-related potentials (ErrP) mentioned above are temporal signals usually observed in EEG signals filtered between approximately 1 and 10 (or sometimes up to 20) Hz [20, 29, 31]. In this study, we investigate the response to visual feedback of motor imagery while the user is actively performing motor imagery. Our goal is to look for error-related information in other frequency bands (not just low-frequency traditional error-related potentials) and in other spatial locations (not just on the center midline channels). We introduce the term 'error-related spectral perturbation' (ErrSP) as a certain type of event-related spectral perturbation [32] to emphasize that we look beyond error-related potentials (ErrP), i.e. in multiple frequency bands and in a data-driven manner in the spatial domain.

We show that the polarity of the feedback (whether the cursor moves in the direction the user intended or in the opposite direction) is classifiable and that some of the information used for classification is contained within the same frequency bands that are important for motor imagery. We then present a method to make use of the classifiability of the brain response to such feedback. The idea is to combine the active motor imagery that the user is generating along with the passive response [33, 34] to the cursor movements to best determine the desired cursor movement direction/target location. Our results show that the proposed approach significantly outperforms the conventional approach in motor imagery.

## 2. Methods

### 2.1. Paradigm

We collected data from 10 participants who were recruited from the UC San Diego student population after the study was approved by the University Institutional Review Board. The demographic details are specified in Table 1 as age, gender, and handedness for each participant.

Participants were naive to BCI and signed a consent form approved by the UCSD Human Research Protections Program before participating in the experiment. The experiment consisted of two parts: in phase one, the participants were trained to perform kinesthetic motor imagery of left and right hands. It has been shown previously that kinesthetic motor imagery ('imagine what it feels like to move your hand') induces a stronger EEG signal [35] than visual motor imagery ('imagine what your hand moving looks like'). This phase consisted of a total of 30 trials, divided into 3 blocks of 10 trials each. Each trial began by randomly showing an arrow pointing to the

**Table 1.** Demographics of participants recruited from the UC San Diego student population.

| Participant | Age | Gender | Handedness |
|---|---|---|---|
| One | 20 | Female | Left-handed |
| Two | 18 | Male | Right-handed |
| Three | 19 | Female | Right-handed |
| Four | 22 | Male | Right-handed |
| Five | 19 | Male | Left-handed |
| Six | 18 | Female | Right-handed |
| Seven | 15 | Female | Right-handed |
| Eight | 34 | Male | Right-handed |
| Nine | 22 | Female | Right-handed |
| Ten | 19 | Male | Right-handed |

left or right to indicate the trial being a left- or right-hand motor-imagery trial respectively. Then a cross appeared at the center of the screen, and after one second the phrase ‘motor imagery’ appeared on top of the cross. Participants were instructed to perform kinesthetic motor imagery as long as the cross and ‘motor imagery’ phrase were on the screen (i.e. for 4 s). At the end of this time period, participants were provided with true feedback in terms of two bars, whose height reflected the average power in the mu band (8–13 Hz) of small Laplacian-filtered [36] signals from channels over the left (C3, FC3) and right (C4, FC4) motor and pre-motor cortices [6]. Participants were instructed that they should aim for maximizing the difference between the height of the two bars, with the higher being on the side of the imagery target. For instance, on a right imagery trial, the bar on the right side should have a greater height than the one on the left and this difference should be as large as possible (though the participants were not informed of the reason, this is because the desynchronization –decrease in power – is greater on the side contra-lateral to the imagined body part.) The powers of the two bars were scaled if the larger power was greater than a threshold so that the bar heights were visualizable on the monitor. The bars were presented on the screen for 2 s and the inter-trial interval was set to about 10 s – a random duration between 0 and 1 s was added to the rest period to prevent adaptation. Phase 1 was designed to be a short training session to give participants the chance to learn how to perform discriminable motor imagery during phase 2. Phase 2, which will be described next, is the main part of the experiment; all analysis and the reported results are from data collected during this phase.

In phase 2, participants were instructed to use motor imagery to move a cursor on a horizontal line on the monitor to hit a target on the left or right. This paradigm is an extension to what was originally proposed in [15, 28]. Each trial began by showing the cursor in the center of the screen and the target at either end on the right or left side. The cursor and target were each represented as a circle having 2 cm diameter, and colored blue and white respectively. The center position (where the cursor would begin moving from) was three steps away from both right and left sides where the target would appear. After 1.5 s the target vanished to reduce visual distraction for participants and ensure that classification was not based on a visual signal. The cursor moved every 1 s based on a pre-determined sequence of movements. Each trial ended when the cursor hit the target position (success) or the other end of the screen (failure). An example of the paradigm is presented in Figure 1(left). There were 10 blocks in this phase, each composed of 20 trials. Participants’ performance within the past block and the overall performance were provided on the screen after each block. Participants were misled to believe they were controlling the cursor while in fact sham feedback was used to keep the stimuli consistent among them. Particpants were told that performances above a certain level would be rewarded monetarily (over the regular compensation). During the rest period in between the trials, the participants kept their eyes open but did not fixate on the center of the monitor. They could look around and blink normally. During the rest period in between blocks, the participants had as much time as they wanted to relax and stretch out and close their eyes and take some rest.

As mentioned earlier, the cursor followed a pre-determined pattern which was determined pseudo-randomly with a few conditions enforced: target hit rate for the blocks varied between 60 and 90%; in each trial – from the cursor beginning in the center until the end – at most three changes in direction were allowed; no more than two consecutive changes in direction were possible; and finally, the cursor started in the middle of the screen and ended

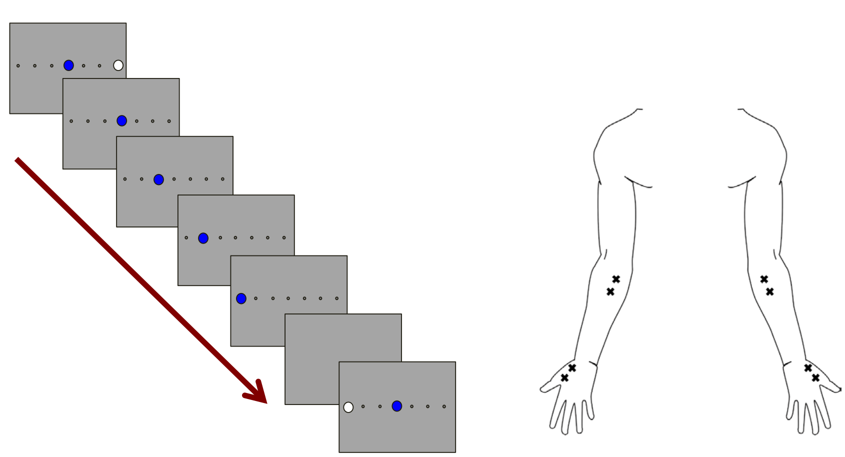

**Figure 1.** Left: one trial of the paradigm in the second (main) phase of the experiment. The participants were instructed to move the cursor to the target with motor imagery of their left or right hands. Right: bipolar electrode placements on each arm.

at the right or left side of the monitor (not in other locations). The sequence of trials and the cursor movement pattern were kept the same for all participants and was designed to have an adequate number of cursor movements towards and away from the target.

### 2.2. Data collection and processing

Data were collected with a 64-channel BrainAmp system (Brain Products GmbH), with electrodes arranged according to the International 10–20 system [37]. The impedance of the electrode connectivity was adjusted to be below 6 kΩ. Aside from EEG data, EMG data were recorded with the same system through bipolar electrodes, one on the upper forearm and another on the wrist of each hand, as shown in Figure 1(right). Both EEG and EMG data were collected at 5000 Hz sampling rate and downsampled to 500 Hz for further processing.

Pre-processing was done in MATLAB [38] and EEGLAB [39]. Data were first band-pass filtered between 1 and 200 Hz with an FIR filter with 500 taps and the Cleanline plug-in [40] was applied to remove the line noise. Data sections contaminated with large muscle artifacts were identified visually and removed. The rejected sections contained less than 5% of the data recorded during trials. Next, one to five channels with high power in the higher frequencies (above 60 Hz), indicating channels possibly contaminated by muscle or other artifacts, were removed. All these channels were from electrodes over temporal sites. Then, the EEG data were epoched into 500 ms non-overlapping intervals and automatic artifact rejection – `autorej` and `jointprob` – from EEGLAB was applied to remove at most 10% of the data. Then Infomax ICA decomposition [41] was applied and ICA components were saved. Afterwards, the raw data were once more band-pass filtered between .1 and 50 Hz, and the data sections contaminated by large muscle artifacts were visually identified and removed. No epoching or automatic artifact rejection was performed in this round and only ICA components that were saved earlier were applied to remove muscle and eye artifacts based on the instructions in [39].

### 2.3. Classification

Data were band-pass filtered with an FIR filter with 500 taps in the following frequency intervals: 1–3, 2–5, 4–7, 6–10, 7–12, 10–15, 12–19, 18–25, 19–30, 25–35, and 30–40 Hz. These intervals were selected to cover low and high theta, mu, and beta frequency bands while overlapping to compensate for individual differences [42, 43]. The pre-processed data were epoched 150 to 950 ms after each cursor movement. The notion of *step* in the rest of the paper represents this time window after each cursor movement and classification is explored at this level. This interval was selected to take into account the time that it took for the participant to perceive the cursor movement. Note that in general for reporting BCI performance, it is standard to consider all time intervals in the course of cursor sequence/trial, beginning in the middle of the screen and ending at either side (right or left), and to compute accuracy on this scale in terms of the hit-rate for hitting the correct target. However, as our goal is to look for the user response to feedback in short-time recordings of EEG while the user is performing motor imagery, we compute accuracy on a single *step* basis (only the time periods between two consecutive cursor movements).

Common spatial patterns (CSP) were applied to extract the top three filters for each class and in each frequency band [44]. After applying the trained CSP filters, features were extracted as the log of the power through each filter for each step. Other than log power in the aforementioned frequency bands, different features were also extracted from the temporal signal (single-step version of the event-related potential [ERP]). The pre-processed data were bandpass filtered from .5 to 10 Hz, with an FIR filter comprising 500 taps. Signals in each channel were averaged in non-overlapping 50 ms windows from 150 ms to 950 ms in the time domain and these values from channels Cz, Pz, CPz, and Fz were selected as ERP features.

Each step, that is each cursor movement, depending on the location of the target (either on the right or left) and the movement (towards [good] or away from the target [bad]), can be divided into four categories: good-right (GR), good-left (GL), bad-right (BR), and bad-left (BL). Classification in all cases described next, was done on balanced classes, i.e. the number of steps in GR, BR, GL, and BL classes were balanced by randomly removing steps from classes with a higher number of steps. Linear discriminant analysis (LDA) classification [45, 46] was performed over each cursor movement with two different classifiers: one being the conventional right/left (R/L) classification to classify the motor imagery signal, and the other classifying the 'goodness' of each cursor movement, i.e. to decide whether the cursor moved toward the target – good movement – or away from the target – bad movement. We call the latter a good/bad (G/B) classifier (in contrast to the standard R/L). Since we looked into two different sets of features for the G/B classifier, i.e. power and ERP features, we name the two G/B-p and G/B-erp accordingly. We present classification results for R/L, G/B-p, and G/B-erp classifiers separately, as well as for combinations of these classifiers, as discussed next.

Our first attempt to combine the two R/L and G/B-p classifiers is within each frequency band (per-frequency-band classifier). For each step, the probabilities of belonging to the right class and good class are considered as scores from the R/L and G/B-p classifiers respectively. However, to combine the two output features consistently in one classifier, the direction of the observed cursor movement must be taken into account to allow translation of the G/B outputs into the R/L output space, since the output of the combined classifier is to determine the motor imagery intention. If the cursor moved to the right, then the movement was a GR or BL. Therefore, the G/B-p classifier maps to R/L space; hence, the probability of belonging to the good class is the same as the probability of belonging to the right class. On the other hand, if the cursor moved to left, the movement was either GL or BR. In this case, the G/B-p classifier maps to L/R as opposed to R/L. Therefore, the probability of being in the good class is the same as the probability of being in the left class; hence, the G/B score is translated by one minus the probability of belonging to the good class. We call this process 'translation of features based on the observed cursor direction of movement'. After translation, R/L and G/B-p classifier scores were combined through logistic regression.

We also propose a combined R/L and a combined G/B-p classifier that both use the features across all frequency bands (across-frequency-bands classifiers). These classifiers train a logistic regression over the scores from frequency bands that show R/L or G/B-p classification rates above chance level with respect to the number of steps with significance level of .01 [42, 47]. Note that the test data (steps) remained unseen during the training session, including the logistic regression part. To see whether dividing the signal into many frequency bands is more effective for discriminating R/L motor imagery than using one classifier over one 7–30 Hz frequency band, we compare the performance of the combined across-frequency-bands R/L classifier with that of a conventional R/L classifier that trains an LDA classifier on the log power features from the top three CSP filters for each class in one 7–30 Hz frequency band.

To make use of the responses to the 'goodness' of cursor movement, we propose to augment the R/L across-frequency-bands classifier to a [R/L]+[G/B-p] across-frequency-bands classifier that uses motor imagery information as well as the state of the participant with respect to each cursor movement in relevant frequency bands. For each participant, this classifier selects LDA features in terms of probabilities of belonging to each class, based on frequency bands with significantly above chance R/L or G/B-p classification rates. The chance level was calculated based on the number of steps with significance level of .01 [42, 47]. A logistic regression classifier was trained over LDA scores from the selected R/L and G/B-p frequency bands. Translation of G/B-p scores (probabilities) into R/L was performed based on the observed cursor direction of movement as described earlier. We compare this with across-frequency-bands [R/L]+[G/B-erp] classifier that uses ERP features instead of power features in a similar way. Translation of G/B-erp scores to R/L was performed in the same way as translation of G/B-p scores to R/L, as explained earlier.

The ultimate proposed classifier is one that uses all sets of available features described earlier. Therefore, we propose a combined across-frequency-bands [R/L]+[G/B-p]+[G/B-erp] classifier that uses motor imagery information as well as the state of the participant with respect to each cursor movement in both frequency (G/B-p) and time (G/B-erp) domains. Similar to the [R/L]+[G/B-p] classifier, the frequency bands in which R/L and G/B-p perform above chance level with respect to the number of steps with significance level of .01 [42, 47] were selected. LDA scores for the G/B-erp classifier were concatenated to the selected G/B-p scores, while translated to R/L based on the observed cursor direction of movement as described earlier. A logistic regression classifier was trained over all three sets of scores: R/L as well as translated G/B-p and G/B-erp. As before, the test data (steps) remained unseen during both parts of training the classifier.

EMG data were collected to confirm that classifying right-left motor imagery is not possible from arm/wrist movements. Data from the bipolar channels on the forearm and wrist on each hand were bandpass filtered from 10 to 100 Hz with an FIR filter with 500 taps. The line noise was removed with the Cleanline plug-in [40] in EEGLAB [39]. Data were epoched 0 ms to 1000 ms after each cursor movement and the log power of the signal was used as the feature for classification in a separate control classifier using LDA.

In all of the results reported next, we made sure that train and test steps (beginning from the feature extraction phase) were absolutely separate subsets of steps and performed multiple 10-fold nested cross validations for all classification results.

## 3. Results

Table 2 compares the classification results for the combined-across-the-frequency bands R/L classifier that uses classifier scores trained on several different frequency bands and the conventional R/L classifier trained on features over one 7–30 Hz band. A paired-sample two-tailed $t$-test shows that the R/L classifier combined across multiple frequency bands performs significantly better for participants 4, 6, and 8. For the rest of the participants, the performance is not significantly different. We decided to continue using the combined-across-the-frequency-bands R/L classifier as later on we are interested in looking at both R/L and G/B classifiers in multiple frequency bands including lower than alpha frequencies. Note that the classification rates that are reported here are after each cursor movement within a time window of length 800 ms.

To make sure that the results presented here are due to motor imagery and not actual movement execution, we performed R/L classification on EMG data as well. Table 2 also shows the results for the R/L classifier on EMG data for each participant and the class conditional correlation with the combined-across-the-frequency-bands R/L classifier on EEG data. To do so, we first corrected for different means between the right and left classes in each case and then calculated the correlation coefficient and the corresponding p-value for each participant. As Table 2 presents, participants 2 and 4 show significant correlation between R/L classifiers trained on EEG and EMG data; however, the correlation coefficient for participant 4 is very small and the R/L classification rate on EMG data for participant 2 is only chance level. Therefore, we conclude that the classification rates reported for R/L classifier on EEG data are in fact from motor imagery and not actual movements. To be consistent, we also trained a G/B classifier on EMG recordings and found chance-level performance for all participants.

Figure 2 shows the results of LDA classification for R/L and G/B-p in each frequency band; i.e. the solid and dashed black lines. The magenta line shows the combined-per-frequency-band R/L and G/B-p classifiers in each frequency band separately. Each point on the plots is represented as an error bar showing the mean and standard error of results from five instances of 10-fold cross validation. The dashed green line represents the chance level .5 and the solid green line indicates the chance level calculated based on the number of steps [47] with significance level of .05.

**Table 2.** Comparison of conventional R/L classifier (7–30 Hz R/L) and the combined across-frequency-bands R/L classifier. The first number is the mean classification rate over five instances of 10-fold nested cross validation and the second number is standard error. The significantly higher rates among the two are identified in bold. The fourth column presents the R/L classifier results on EMG data with mean and standard error over five instances of 10-fold cross validation. The last two columns show the correlation coefficient and corresponding *p*-value between the combined-across-the-frequency-bands R/L classifier on EEG data and the R/L classifier on EMG data.

| Participant | R/L | 7–30 Hz R/L | EMG | Corr. coef. | *p* value |
|---|---|---|---|---|---|
| One | .87 / .005 | .86 / .005 | .63 / .008 | .008 | .719 |
| Two | .60 / .010 | .61 / .006 | .54 / .008 | .17 | <.01 |
| Three | .68 / .012 | .66 / .007 | .55 / .010 | .039 | .069 |
| Four | **.68 / .010** | .63 / .009 | .63 / .007 | .080 | <.01 |
| Five | .73 / .010 | .74 / .007 | .50 / .009 | −.032 | .277 |
| Six | **.63 / .012** | .56 / .008 | .55 / .007 | .029 | .170 |
| Seven | .78 / .009 | .77 / .007 | .61 / .009 | −.016 | .466 |
| Eight | **.79 / .010** | .74 / .007 | .56 / .008 | −.018 | .390 |
| Nine | .60 / .011 | .60 / .008 | .53 / .008 | −.026 | .379 |
| Ten | .57 / .010 | .56 / .008 | .56 / .007 | −.001 | .957 |

As can be noted from the plots, the combined classifier outperforms the R/L classifier in frequency bands where G/B-p performs above chance level.

Note that, for all participants, lower-frequency bands (below 10 Hz) show above-chance-level G/B-p classification performance. We hypothesized that this might be reflecting error-related signals that might be better classified using a conventional windowed-mean classifier on the low-pass-filtered temporal signal. In order to investigate whether G/B-p classifiers and G/B-erp classifiers are classifying different signals in both lower and higher frequency ranges, we looked at the class conditional correlation coefficients between the real-valued classifier outputs of the G/B-p and the G/B-erp classifiers. We did this separately for the good and bad classes to exclude significant correlation that may result when both classifiers perform above chance. We calculated the correlation coefficient between the LDA scores computed based on each classifier in each step. Our results show that there are no significant correlations in higher-frequency bands except for participants 6 and 10, though they are fairly low valued – *R* below .15. This implies that in fact G/B-p (when performing above chance level) is using new information which is not considered by G/B-erp in the .5–10 Hz frequency band. In fact the correlations between G/B-p and G/B-erp are small (and for some participants not significant) even in the lower frequencies. Thus we decided to keep all G/B-p features as well as the G/B-erp features in our proposed classifier [R/L]+[G/B-p]+[G/B-erp].

We introduce the term error-related spectral perturbation (ErrSP) for CSP-filtered EEG data and calculated it for each participant: see Figures 3 and 4. In each frequency band, we ran 10-fold cross validation and found the top three CSP filters in each class for G/B classification in each fold. Next LDA was trained on the log-power of the CSP-filtered train data, and the LDA-weighted log power of CSP-filtered test data in each fold in 50 ms non-overlapping time windows was calculated. A paired-sample two-tailed *t*-test was run to measure the significance of the difference in good (G) and bad (B) classes in each time window and each frequency band. *P*-values are also plotted in Figures 3 and 4. These results show that there is classifiable information beyond the low-frequency ErrP on center-line channels which could be used in better detecting the polarity of feedback, i.e. whether the cursor is moving towards or away from the target.

Table 3 shows classification results for G/B-p and G/B-erp and [G/B-p]+[G/B-erp] classifiers. For each we ran paired-sample two-tailed *t*-tests between G/B-p and [G/B-p]+[G/B-erp] and another test between G/B-erp and [G/B-p]+[G/B-erp]. Whenever [G/B-p]+[G/B-erp] results in significantly higher performance compared to both G/B-p and G/B-erp, the result is specified in bold.

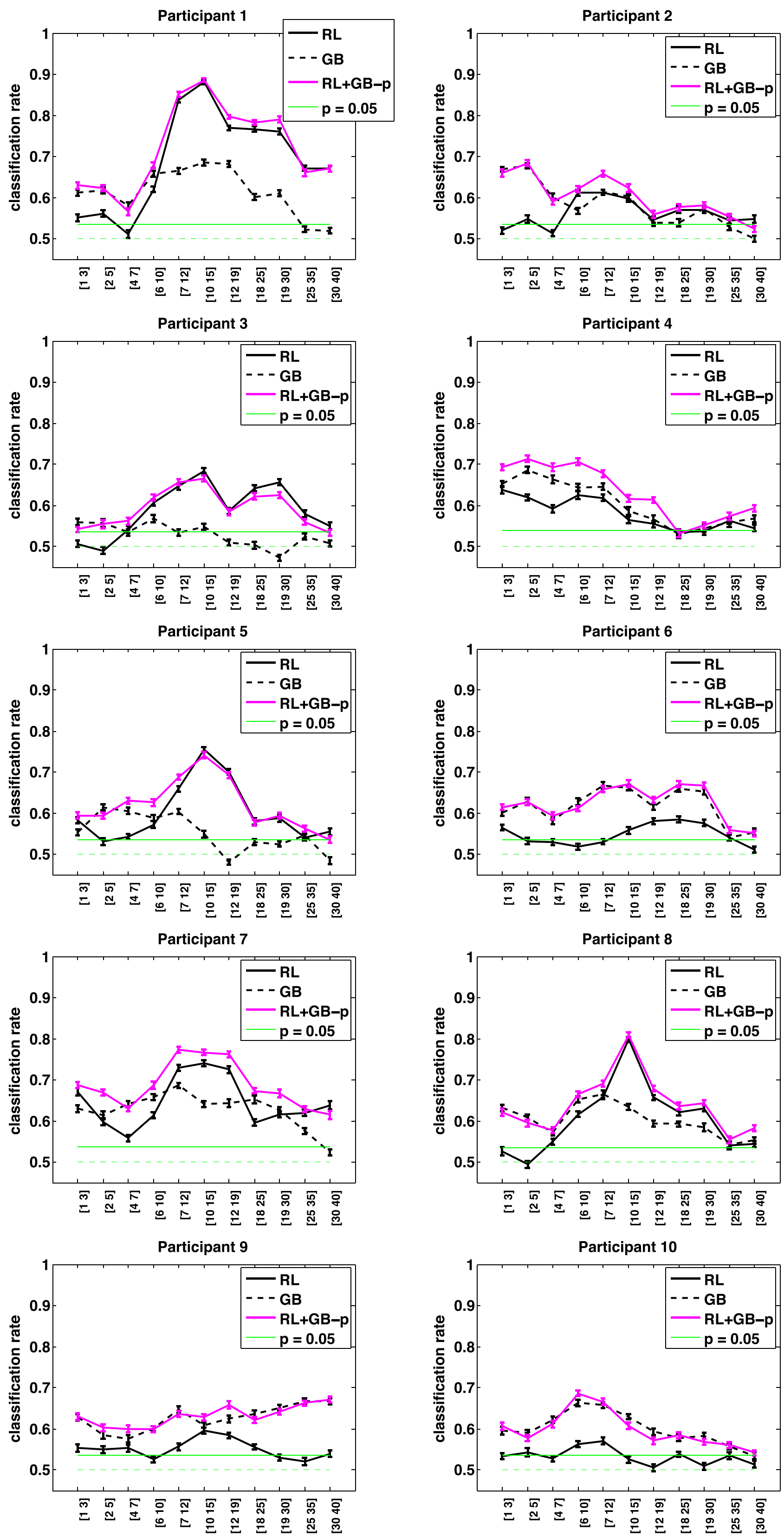


**Figure 2.** The black solid and dashed lines show the result of LDA classifiers on R/L and G/B-p respectively, trained on the individual frequency bands. The magenta line is the combined R/L and G/B-p classifier per frequency band.

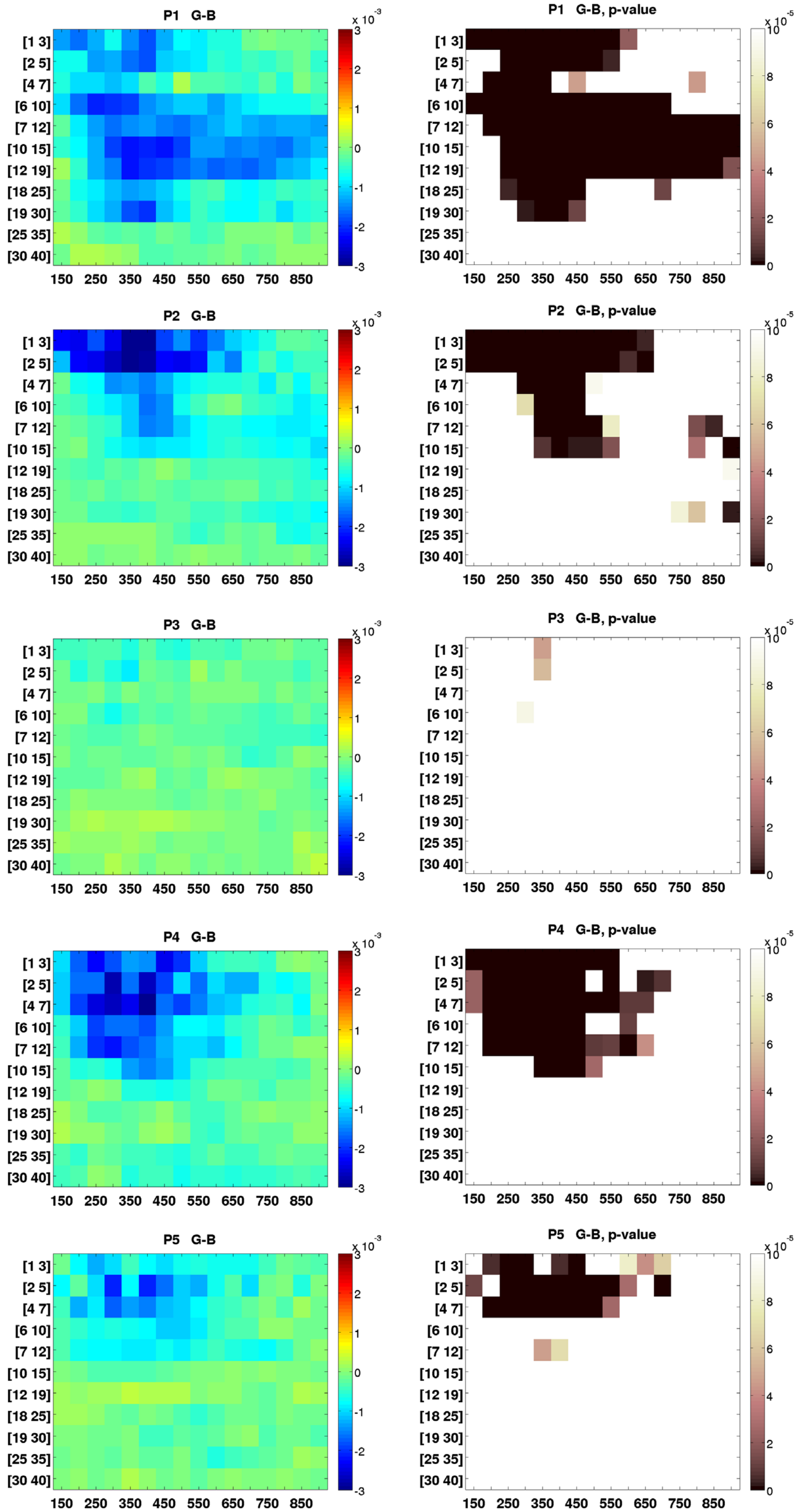


**Figure 3.** ErrSP (left column) for G-B difference and *p*-values from paired-sample two-tailed *t*-tests between good and bad classes in participants 1–5 in various frequency bands/time bins. *X*-axis shows the time in ms after cursor movement and *Y*-axis the frequency bands. *P*-values are uncorrected but are only shown for *p*-values $< 10^{-4}$.

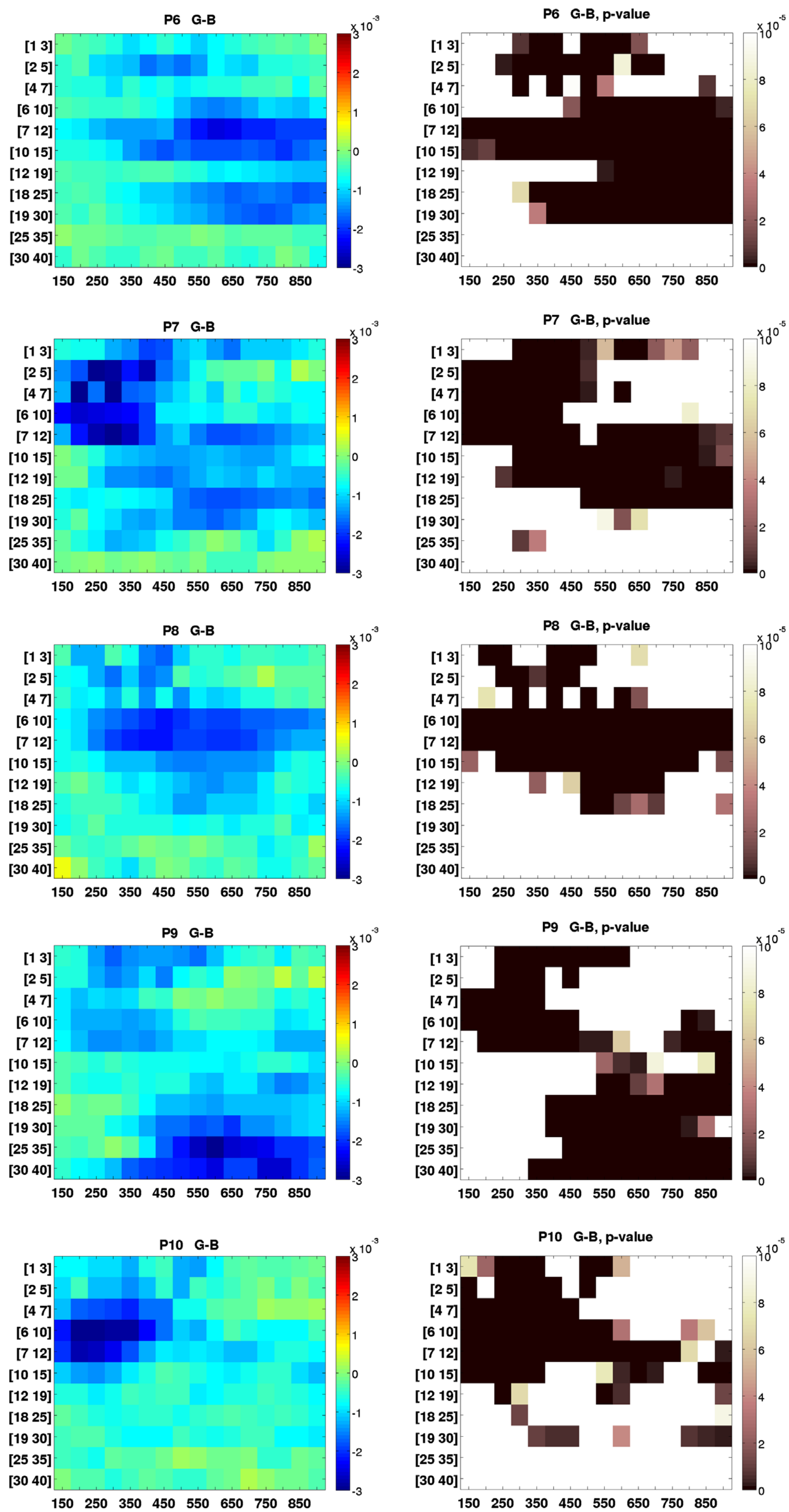


**Figure 4.** ErrSP (left column) for G-B difference and *p*-values from paired-sample two-tailed *t*-tests between good and bad classes in participants 6–10 in various frequency bands/time bins. *X*-axis shows the time in ms after cursor movement and Y-axis the frequency bands. *P*-values are uncorrected but are only shown for *p*-values < $10^{-4}$.

**Table 3.** Comparison of G/B-erp and G/B-p classification rates. The last column shows the results of the combined classifier. The first number is the mean classification rate over three instances of 10-fold nested cross validation and the second number is the standard error. Whenever the combined classification rate is significantly higher than both G/B-p and G/B-erp, the number is specified in bold.

| Participant | G/B-p | G/B-erp | G/B-p+G/B-erp |
|---|---|---|---|
| One | .76 / .010 | .73 / .007 | **.81 / .007** |
| Two | .73 / .010 | .73 / .008 | **.77 / .009** |
| Three | .54 / .009 | .60 / .009 | .59 / .010 |
| Four | .74 / .010 | .78 / .006 | **.81 / .011** |
| Five | .65 / .009 | .66 / .008 | **.71 / .007** |
| Six | .71 / .009 | .69 / .007 | **.75 / .009** |
| Seven | .75 / .009 | .72 / .010 | **.79 / .007** |
| Eight | .67 / .012 | .72 / .008 | **.75 / .010** |
| Nine | .76 / .009 | .75 / .009 | **.81 / .008** |
| Ten | .70 / .012 | .70 / .007 | **.76 / .009** |

**Table 4.** Classification results for combined classifier across frequency bands. The first number is the mean classification rate over three instances of 10-fold nested cross validation and the second number is the standard error. Whenever the results in the right-most column are significantly higher than the R/L results, numbers are specified in bold.

| Participant | R/L | R/L+G/B-p | R/L+G/B-erp | R/L+G/B-p+G/B-erp |
|---|---|---|---|---|
| One | .87 / .005 | .89 / .006 | .90 / .005 | **.91 / .006** |
| Two | .60 / .010 | .76 / .009 | .75 / .011 | **.78 / .006** |
| Three | .68 / .012 | .68 / .008 | .67 / .010 | .68 / .010 |
| Four | .68 / .010 | .79 / .008 | .80 / .009 | **.84 / .008** |
| Five | .73 / .010 | .76 / .009 | .78 / .008 | **.80 / .007** |
| Six | .63 / .012 | .71 / .011 | .73 / .007 | **.75 / .009** |
| Seven | .78 / .009 | .84 / .008 | .83 / .007 | **.87 / .007** |
| Eight | .79 / .010 | .81 / .007 | .83 / .008 | **.84 / .009** |
| Nine | .60 / .011 | .77 / .009 | .75 / .009 | **.82 / .008** |
| Ten | .57 / .010 | .69 / .009 | .70 / .009 | **.74 / .009** |

Note that for all participants except for participant 3, the combined classifier performs significantly better than either of the two pieces of information separately (with significance level of .05). Moreover, all participants except for 3, 4, and 8 show significantly improved combined classifier with significance level .01.

Table 4 presents the results for the R/L, [R/L]+[G/B-p], [R/L]+[G/B-erp], and [R/L]+[G/B-p]+[G/B-erp] classifiers when information across all frequency bands is taken into account. The first number is the mean classification rate with three instances of 10-fold nested cross validation and the second number shows the standard error. Paired-sample two-tailed *t*-tests were calculated for the proposed [R/L]+[G/B-p]+[G/B-erp] classifier, i.e. the last column in Table 4, in comparison with the R/L classifier, i.e. the second column in the same table, and if the difference is significant (with .01 significance level) the higher rate is identified in bold. We ran paired-sample two-tailed *t*-tests with .01 significance level to compare the R/L classifier with the [R/L]+[G/B-p] and [R/L]+[G/B-erp] classifiers as well. Our results show that R/L when combined with the G/B (either power, erp or power and erp) classifier outperforms the R/L classifier significantly for all participants except for participant 3. It is worth noting that participant 3 has G/B-p and G/B-erp classification rates very close to chance level (refer to Table 3) and it is not surprising that the combined classifiers do not outperform the R/L classifier. Interestingly, though, the performance of the combined classifier for this participant is not worse than our baseline R/L classifier. For easier visual comparison, the results of R/L in the second column and [R/L]+[G/B-p]+[G/B-erp] in the last column are plotted as a bar plot for each participant in Figure 5. Our results show an average of 11% improvement in classification accuracy across 10 participants.

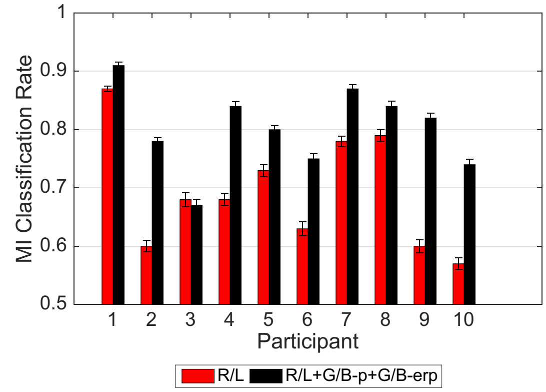


**Figure 5.** Comparison of across-the-frequency-bands [R/L] and [R/L]+[G/B-p]+[G/B-erp] classifiers, from Table 4.

## 4. Discussion and conclusion

In this study we investigated error-related spectral perturbation to parse out the effect of error-related brain processes that may occur in the same frequency bands as the motor imagery signal. There are many studies in the literature that show the effect of feedback in EEG-based BCI performance [13–17, 19, 20, 22, 26–28, 31, 48]; however, we believe that the use of learned error-related signals from multiple spectral bands and spatial locations combined with active BCI signals through learned weighted voting is unique. The learned weighted voting (for combining R/L and G/B) lets both the error-related features and the active BCI features (motor imagery in our case) have influence at the same time with the flexibility to let the classifier adjust to each participant individually. It is worth emphasizing that, at least in our 10 participants, combining R/L with G/B-p and G/B-erp is not harmful even for participants with relatively poor G/B classification.

In this study, we used a sham feedback paradigm where the participants were presented pre-determined cursor movements but were misled to believe that they were in control. Note that we do not propose that sham feedback should be used during the operation of real BCIs. The sham feedback was important here to have adequate number of good and bad cursor movements to train the G/B classifier for all participants independent of the participant's motor imagery performance. We hypothesize that during real online MI use, the user's response to cursor movements will similarly be discriminative to feedback polarity and that a combined classification system, as proposed in this work, will give improved performance for predicting desired target locations.

Although in this work we have considered a right/left-hand motor imagery paradigm, we believe that our approach is generalizable to other classes of motor imagery, other BCI modalities and perhaps other BCI paradigms. For instance, it would be interesting to look for similar G/B classifiability in other BCI modalities such as those based on functional near-infrared spectroscopy signals [50, 51]. In a recent study by Stivers et al. [52], low-frequency time-domain error-related features were used in a G/B classifier in a speller. In another study [53], Zeyl et al. demonstrated the application of error-related potentials below 20 Hz in an auditory BCI. In both these examples as well as in other BCI paradigms such as P300 or SSVEP BCIs [10, 11], it would be interesting to look for G/B signals in higher-frequency bands.

Traditional active cursor-control BCI systems use explicit control strategies, i.e. they use only R/L-type classifiers to detect motor imagery mapped to cursor movements. Zander et al. [54] in a recent study demonstrated neuroadaptive technology for implicit cursor control. Using passive BCI [33, 34], the participants controlled a cursor through passively changing future cursor movements based on an error-related potential-type signal, without any type of explicit communication or control. In our current work, we demonstrated an active R/L control combined with passive G/B classifier using a wider range of features from low to high frequency bands. As one final variation, de Sa [49, 55] proposed that active motor imagery can also be mapped to G/B-type commands (e.g. left-hand MI meaning 'change direction' and right-hand MI meaning 'stay on course'), and theoretically demonstrated increased robustness compared to direct control. For future work, we are interested in how these four strategies compare to each other in a uniform cursor-control paradigm, in terms of performance and other metrics such as user satisfaction and ease of use.

We are interested in investigating the generalizability of the G/B classifier across changes in task condition. For instance, do participants produce similar patterns of G/B across multiple sessions or when task workload varies? Omedes et al. showed this in lower-frequency bands across various tasks [29] and associated it with error-related potentials, but they have not discussed this for higher-frequency bands. A related interesting question would be to investigate the underlying brain networks that are involved in generating the good/bad signal(s).

## Funding

This work was supported by the NSF grants IIS 1219200, SMA 1041755, IIS 1528214, UCSD FISP G2171, and NIH grant R01NR013500.

## References

[1] Farwell L, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol. 1988;70(6):510–523. Available from: http://www.sciencedirect.com/science/article/pii/0013469488901496.

[2] Wolpaw JR, McFarland DJ, Neat GW, et al. An EEG-based brain-computer interface for cursor control. Electroencephalogr Clin Neurophysiol. 1991;78(3):252–259. Available from: http://www.sciencedirect.com/science/article/pii/001346949190040B.

[3] Kalcher J, Flotzinger D, Neuper C, et al. Graz brain-computer interface II: towards communication between humans and computers based on online classification of three different EEG patterns. Med Biol Eng Comput. 1996;34(5):382–388. Available from: http://dx.doi.org/10.1007/BF02520010.

[4] Millán JdR, Rupp R, Mueller-Putz G, et al. Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. Front Neurosci. 2010;4:161.

[5] Mason S, Bashashati A, Fatourechi M, et al. A comprehensive survey of brain interface technology designs. Ann Biomed Eng. 2007;35(2):137–169.

[6] Pfurtscheller G, Brunner C, Schlögl A, et al. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. Neuroimage. 2006;31:153–159.

[7] Pineda JA, Allison BZ, Vankov A. The effects of self-movement, observation, and imagination on $\mu$ rhythms and readiness potentials (RPs): towards a brain-computer interface (BCI). IEEE Trans Rehab Eng. 2000;8(2):219–222.

[8] McFarland DJ, Miner LA, Vaughan TM, et al. Mu and beta rhythm topographies during motor imagery and actual movements. Brain Topog. 2000;12(3):177–186. Available from: http://dx.doi.org/10.1023/A:1023437823106.

[9] Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. Proc IEEE. 2001;89(7):1123–1134.

[10] Allison B, Faller J, Neuper C. BCIs that use steady-state visual evoked potentials or slow cortical potentials. Brain-computer interfaces: principles and practice. Oxford University Press; 2012.

[11] Sellers EW, Arbel Y, Donchin E. 12 BCIs that use P300 event-related potentials. Brain-computer interfaces: principles and practice. Oxford University Press; 2012.
[12] Hammon PS. Adaptive online brain-computer interface for interpretation and visualization of desired reach [dissertation]. UC San Diego; 2009.
[13] Neuper C, Scherer R, Wriessnegger S, et al. Motor imagery and action observation: Modulation of sensorimotor brain rhythms during mental control of a brain-computer interface. Clin Neurophysiol. 2009;120(2):239–247. Available from: http://www.sciencedirect.com/science/article/pii/S1388245708012601.
[14] Ang KK, Guan C. Brain-computer interface in stroke rehabilitation. J Comput Sci Eng. 2013;7(2):139–146.
[15] Koerner AS, Zhang Q, de Sa VR. The effect of real-time positive and negative feedback on motor imagery performance. In: the International BCI Meeting; 2013. Available from: http://dx.doi.org/10.3217/978-3-85125-260-6-73.
[16] Darvishi S, Ridding MC, Abbott D, et al. Does feedback modality affect performance of brain computer interfaces? In: 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER); April; 2015. p. 232–235.
[17] Dolezal J, Cerny V, Stastny J. Online motor-imagery based BCI. In: Applied Electronics (AE), 2012 International Conference on; Sept; 2012. p. 65–68.
[18] Chavarriaga R, Sobolewski A, Millán JdR. Errare machinale est: the use of error-related potentials in brain-machine interfaces. Front Neurosci. 2014;8(208). Available from: http://www.frontiersin.org/neuroprosthetics/10.3389/fnins.2014.00208/abstract.
[19] Schalk G, Wolpaw JR, McFarland DJ, et al. EEG-based communication: presence of an error potential. Clin Neurophysiol. 2000;111(12):2138–2144. Available from: http://www.sciencedirect.com/science/article/pii/S1388245700004570.
[20] Ferrez PW, del R Millan J. Error-related EEG potentials generated during simulated brain computer interaction. IEEE Trans Biomed Eng. 2008 March;55(3):923–929.
[21] Ferrez PW, Millán JdR. Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy. In: Proceedings of the 4th international brain-computer interface workshop and training course; 2008. p. 197–202; CNBI-CONF-2008-004.
[22] Artusi X, Niazi IK, Lucas MF, et al. Accuracy of a BCI based on movement-related and error potentials. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug; 2011. p. 3688–3691.
[23] Spüler M, Bensch M, Kleih S, et al. Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. *Clin Neurophysiol*. 2012;123(7):1328–1337. Available from: http://www.sciencedirect.com/science/article/pii/S1388245711009059.
[24] Schmidt NM, Blankertz B, Treder MS. Online detection of error-related potentials boosts the performance of mental typewriters. BMC Neuroscience. 2012;13(1):19. Available from: http://dx.doi.org/10.1186/1471-2202-13-19.
[25] Spüler M, Rosenstiel W, Bogdan M. Online adaptation of a c-VEP brain-computer interface (BCI) based on error-related potentials and unsupervised learning. PloS One. 2012;7(12):e51077.
[26] Kreilinger A, Neuper C, Müller-Putz GR. Error potential detection during continuous movement of an artificial arm controlled by brain–computer interface. Med Biol Eng Comput. 2012;50(3):223–230. Available from: http://dx.doi.org/10.1007/s11517-011-0858-4.
[27] Koerner AS, de Sa VR. A novel method to integrate error detection into motor imagery BCI. In: Workshop on Brain-Machine body interfaces, 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; IEEE; 2012.
[28] Koerner AS. Investigating natural control signals for brain-computer interfaces [dissertation]. UC San Diego; 2013.
[29] Omedes J, Iturrate I, Montesano L, et al. Using frequency-domain features for the generalization of EEG error-related potentials among different tasks. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July; 2013. p. 5263–5266.
[30] Milekovic T, Ball T, Schulze-Bonhage A, et al. Error-related electrocorticographic activity in humans during continuous movements. J Neural Eng. 2012;9(2):026007.
[31] Tong J, Lin Q, Xiao R, et al. Combining multiple features for error detection and its application in brain–computer interface. Biomed Eng Online. 2016;15(1):17.
[32] Makeig S. Auditory event-related dynamics of the eeg spectrum and effects of exposure to tones. Electroencephalogr Clin Neurophysiol. 1993;86(4):283–293.
[33] Zander TO, Kothe C, Jatzev S, et al. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. London: Springer; 2010. p. 181–199.
[34] Zander TO, Kothe CA. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. J Neural Eng. 2011;8(2):025005.
[35] Neuper C, Scherer R, Reiner M, et al. Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. Cogn Brain Res. 2005;25(3):668–677.
[36] McFarland DJ, McCane LM, David SV, et al. Spatial filter selection for EEG-based communication. Electroencephalogr Clin Neurophysiol. 1997;103(3):386–394. Available from: http://www.sciencedirect.com/science/article/pii/S0013469497000222.
[37] Klem GH, Lüders HO, Jasper H, et al. The ten-twenty electrode system of the international federation. Electroencephalogr Clin Neurophysiol. 1999;52(3):3–6.
[38] MATLAB, 2012b STR. Natick, MA: The MathWorks Inc.; 2012.
[39] Delorme A, Makeig S. EEGlab: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004;134(1):9–21. Available from: http://www.sciencedirect.com/science/article/pii/S0165027003003479.
[40] Mullen T. Nitrc: Cleanline. 2012. Available from: http://www.nitrc.org/projects/cleanline.
[41] Makeig S, Bell AJ, Jung TP, et al. Independent component analysis of electroencephalographic data. In: Touretzky DS, Hasselmo ME, editors. Advances in neural information processing systems 8. Cambridge, MA: MIT Press; 1996. p. 145–151. Available from: http://papers.

nips.cc/paper/1091-independent-componentanalysis-of-electroencephalographic-data.pdf.
[42] Noh E, Herzmann G, Curran T, et al. Using single-trial EEG to predict and analyze subsequent memory. NeuroImage. 2014;84:712–723. Available from: http://www.sciencedirect.com/science/article/pii/S1053811913009646.
[43] Ang KK, Chin ZY, Wang C, et al. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Front Neurosci. 2012;6(39). Available from: http://www.frontiersin.org/neuroprosthetics/10.3389/fnins.2012.00039/abstract.
[44] Blankertz B, Tomioka R, Lemm S, et al. Optimizing spatial filters for robust EEG single-trial analysis. Signal Proc Mag IEEE. 2008;25(1):41–56.
[45] Blankertz B, Curio G, Müller KR. Classifying single trial eeg: towards brain computer interfacing. In: Dietterich TG, Becker S, Ghahramani Z, editors. Advances in neural information processing systems 14. Cambridge, MA: MIT Press; 2002. p. 157–164. Available from: http://papers.nips.cc/paper/2030-classifyingsingle-trial-eeg-towards-brain-computer-interfacing.pdf.
[46] Lotte F, Congedo M, Lécuyer A, et al. A review of classification algorithms for EEG-based brain computer interfaces. J Neural Eng. 2007;4(2):R1. Available from: http://stacks.iop.org/1741-2552/4/i=2/a=R01.
[47] Müller-Putz G, Scherer R, Brunner C, et al. Better than random? a closer look on BCI results. Int J Bioelectromagn. 2008;10(1):52–55.
[48] Spüler M, Niethammer C. Error-related potentials during continuous feedback: using eeg to detect errors of different type and severity. Front Hum Neurosci. 2015;9:155.
[49] de Sa VR. An interactive control strategy is more robust to non-optimal classification boundaries. In: Proceedings of the 14th ACM International Conference on Multimodal Interaction; New York, NY. ACM; 2012. p. 579–586; ICMI '12. Available from: http://doi.acm.org/10.1145/2388676.2388798.
[50] Naseer N, Hong KS. Classification of functional near-infrared spectroscopy signals corresponding to the right- and left-wrist motor imagery for development of a brain–computer interface. Neurosci Lett. 2013;553:84–89.
[51] Hong KS, Naseer N, Kim YH. Classification of prefrontal and motor cortex signals for three-class fNIRS-BCI. Neurosci Lett. 2015;587:87–92.
[52] Stivers JM, Krol LR, de Sa VR, et al. Spelling with cursor movements modified by implicit user response. In: Proceedings of the 6th International Brain-Computer Interface Meeting; 2016. p. 28.
[53] Zeyl T, Yin E, Keightley M, et al. Improving bit rate in an auditory bci: exploiting error-related potentials. Brain-Comput Interfaces. 2016;3(2):75–87.
[54] Zander TO, Krol LR, Birbaumer NP, et al. Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. Proc Natl Acad Sci USA. 2016;113(52):14898–14903.
[55] de Sa VR. Changing the commands in noisy incremental human controlled systems. In: NIPS Workshop on Human Propelled Machine Learning; 2014.