

# Towards Explanatory Feedback for User Training in Brain–Computer Interfaces

Julia Schumacher, Camille Jeunet, Fabien Lotte

► **To cite this version:**

Julia Schumacher, Camille Jeunet, Fabien Lotte. Towards Explanatory Feedback for User Training in Brain–Computer Interfaces. IEEE International Conference on Systems Man & Cybernetics (IEEE SMC), Oct 2015, Hong-Kong, China. hal-01179329

**HAL Id: hal-01179329**

**<https://hal.inria.fr/hal-01179329>**

Submitted on 22 Jul 2015

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

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

# Towards Explanatory Feedback for User Training in Brain–Computer Interfaces

Julia SCHUMACHER  
Inria Bordeaux Sud-Ouest / BCCN  
Talence, France / Berlin, Germany  
julia.schumacher1@gmail.com

Camille JEUNET  
University of Bordeaux/Inria  
Bordeaux, France  
camille.jeunet@inria.fr

Fabien LOTTE  
Inria Bordeaux Sud-Ouest/LaBRI  
Talence, France  
fabien.lotte@inria.fr

**Abstract**—Despite their potential for many applications, Brain–Computer Interfaces (BCI) are still rarely used due to their low reliability and long training. These limitations are partly due to inappropriate training protocols, which includes the feedback provided to the user. While feedback should theoretically be explanatory, motivating and meaningful, current BCI feedback is usually boring, corrective only and difficult to understand. In this study, different features of the electroencephalogram signals were explored to be used as a richer, explanatory BCI feedback. First, based on offline mental imagery BCI data, muscular relaxation was notably found to be negatively correlated to BCI performance. Second, this study reports on an online BCI evaluation using muscular relaxation as additional feedback. While this additional feedback did not lead to significant change in BCI performance, this study showed that multiple feedbacks can be used without deteriorating performance and provided interesting insights for explanatory BCI feedback design.

**Index Terms**—Brain-Computer Interfaces, Training, Feedback

## I. INTRODUCTION

Brain-Computer Interfaces (BCI) are systems that enable their users to control an external device such as a computer without the need for any muscular movement [1]. Instead they only rely on a measure of brain signals, e.g., electroencephalography (EEG). BCI systems can be used as assistive technology to restore communication with patients who have severe motor disabilities. Despite their potential in this and many other areas, most BCIs are still not used outside laboratory settings due to their low reliability and long training times. Furthermore, roughly 15–30% of users fail to gain any control over a BCI [1]. Recently, [2] have identified potential reasons for these limitations in the usability of BCI systems. They argue that since BCIs are co-adaptive systems, the two parts of the system might be sources of bad performance and hence are possible targets for improvement: the user and the machine. The user has to learn to produce specific brain patterns by performing mental tasks while at the same time the machine has to learn to recognize and classify these brain patterns by undergoing machine learning.

While there has been a lot of research exploring new signal processing approaches to improve the machine learning component of BCIs, this study will focus on the user’s side. BCI use can be seen as a skill and requires training [1]. Training is

particularly important in the case of spontaneous BCIs which rely on the voluntary modulation of certain brain patterns by the user. An example of such a paradigm is a mental imagery–based BCI (MI-BCI) where users try to modulate their brain activity by performing different mental imagery tasks, e.g. the imagination of movements. BCI user training can be divided into three parts: the instructions, the task and the feedback. While all these aspects are potential targets for improvement [2] this study will focus on the feedback that is provided to the user. Feedback is essential for learning to operate a BCI since it is generally not clear to the user from the beginning what exactly they are required to do in order for the computer to be able to pick up useful signals. In a classical MI-BCI the user is asked to perform different mental imagery tasks such as the imagination of a left or a right hand movement. On the basis of a calibration period a classifier is trained to distinguish between the classes by learning the differences in the recorded brain patterns which underlie the execution of the tasks. During subsequent runs feedback is given to the users to inform about their current performance. The classical feedback that is used in MI-BCIs is shown in form of a moving bar which corresponds to the strength and direction of the output of the previously trained classifier [2]. The feedback thus indicates whether the classifier was able to identify the correct class and the certainty of the classifier in its decision.

Several aspects of this feedback are not in line with current opinions on good feedback from educational research [2]. Generally speaking, while feedback should be explanatory (i.e., explain what was good or bad and why), motivating, supportive, meaningful, specific, and multimodal [3], currently used BCI feedback is usually boring, corrective only (i.e., only indicates whether it was good or bad), not meaningful to people who are not familiar with the concept of a classifier, and limited to the visual modality. Thus, there are a lot of possibilities for improvement some of which have already been explored in previous studies. While most BCI systems use visual feedback, several studies explored the auditory and haptic modalities. Giving haptic feedback either leads to comparable results as visual feedback [4] or leads to higher performances [5]. Using the auditory modality for feedback does not seem to be as promising, the performances are either comparable to visual feedback [6] or lower [7]. Regarding the

motivational aspect of feedback several studies have shown that virtual reality feedback was more motivating than the classical bar feedback, and could lead to a higher performance [8].

The present study is concerned with another aspect of feedback which is its actual content, i.e. which information is presented to the user during training. Thereby we aim to address the problem that currently used BCI feedback is corrective only, i.e. it only indicates whether the outcome of a trial was successful or not. Instead, it has been shown that feedback which provides information on which aspects of the task were done wrong/correctly and information on how to improve in the case of failure is more valuable [3]. In other words, an ideal BCI feedback should be explanatory. To improve the BCI feedback, [9] presented a surface topography of the EEG signal to train people to use a MI-BCI, i.e. instead of giving abstract feedback about the classifier output they provided more direct feedback on the subjects' brain activity. They found that the group that received the neurofeedback training outperformed a second group which did not receive any training. This suggests that providing more direct feedback about brain activity can be efficient. However, since this approach was not compared to the standard BCI feedback no claims can be made on whether neurofeedback training is better than the classical moving bar feedback. In general, there are large performance variations in BCI experiments – between as well as within subjects. Understanding the underlying neurophysiological differences might help to understand the reason for these performance variations and to find approaches to deal with this problem [10], [11]. This can also be transferred to our search for new feedback strategies: if we know what makes an EEG signal a “good” signal (i.e. one that is correctly classified by the machine learning part and thus leads to high BCI performance) we could exploit this information to develop new feedback strategies by telling the user how they have to change their brain signals to improve their performance. Since it is in general not possible to change one's brain signals directly it would be desirable to have an intuitive strategy that leads to the change in the signals. Several studies have tried to find a link between neurophysiological characteristics and BCI performance [12]. It was shown that the sensory–motor rhythm (SMR) at rest is positively correlated with subsequent performance in a MI-BCI with a correlation coefficient of  $\rho=0.53$  across subjects [12]. To study the underlying causes of BCI illiteracy, [12] investigated differences between a group of BCI illiterates and a group of BCI literates. They found a negative correlation between  $\theta$  bandpower and BCI performance and a positive correlation between  $\alpha$  bandpower and BCI performance. In a second study using combined EEG and MEG recordings the same group found a high positive correlation between  $\gamma$  bandpower in prefrontal areas (as measured by MEG) and BCI performance (using the EEG signals) [12]. They attribute high prefrontal  $\gamma$  activity to high levels of concentration indicating that the more concentrated users are, the better they perform at the BCI task. Another predictor of BCI performance was proposed by [13]. They used a coefficient of frontal  $\theta$  divided by posterior  $\alpha$  and central  $\beta$  bandpower that showed a positive

correlation of 0.53 with BCI performance.

However, all the aforementioned studies investigated the correlation between the respective feature and BCI performance on the subject–level, i.e. they tried to explain performance variations across subjects. These results might be useful for a prescreening of subjects to find those which will most likely perform well in a subsequent BCI experiment. However, in the context of exploring new feedback strategies it is also important to investigate performance variations within subjects, preferably on the basis of single trials. [10] demonstrated that SMR modulation is positively correlated with frontal and occipital  $\gamma$  power while it is negatively correlated with centroparietal  $\gamma$  power on the basis of single trials. In addition they could show that  $\gamma$  rhythms have a causal influence on SMR modulation. [11] tried to predict SMR modulation by pre–trial  $\gamma$  bandpower. Using linear regression they learned a model to predict SMR modulation by pre–trial  $\gamma$  and showed that the predicted SMR modulation was positively correlated with actual SMR modulation with a correlation coefficient of 0.1. These results can be exploited for the development of a novel feedback approach by informing BCI users about the level of a certain feature before the actual start of a trial. Thus, the user would have the chance to up– or downregulate the feature in question before starting the BCI task, thereby creating a state of mind which is beneficial for subsequent BCI performance.

In order to determine which additional information could be presented as feedback we analyzed data from a previous experiment to find features that are indicative of performance and more easily understandable for the users. Providing additional feedback raises the question whether it is possible for the users to effectively deal with multidimensional feedback during a BCI experiment. Comparing a condition that uses only the classical bar feedback with a condition in which information about three different aspects of the classifier was presented, [14] showed that the higher complexity of the feedback did not lead to difficulties for the subjects, i.e. performance was comparably high in both conditions. The fact that this multidimensional feedback did not improve performance might be due to the fact that during the condition with additional feedback, still only information about the classifier output was presented, i.e. the problem of the feedback being not explanatory was not addressed. However, the study showed that it is possible to present multidimensional feedback without deteriorating performance. Thus, we hypothesize that using multidimensional feedback that is more meaningful for the participants will help to improve performance.

## II. FEEDBACK FEATURES

Following the suggestions in [2] we explored different EEG features that might be presented as feedback to support BCI user training. These features were assessed according to their ability to predict performance by using data from a previously conducted BCI experiment. This section describes the data used and the results of the evaluation of two such features.

### A. Data Used for Evaluation of Features

The EEG data used for the evaluation of possible feedback features had been recorded previously [15] in the context of a three-class MI-BCI experiment using the classical moving bar feedback. The data of 15 BCI-naive subjects were analyzed, using data from subjects with different levels of performance. These data consisted of four feedback runs. Every run comprised 45 trials, 15 trials for each of the three mental imagery tasks. The sequence of trials from the three classes was randomized for every run. The three mental imagery tasks representing the three classes were left hand motor imagery, mental subtraction and mental rotation of objects. These tasks were chosen because they yielded the best performances in a study exploring different tasks for BCI use [16]. The calculation task consisted of repeatedly subtracting a two-digit from a three-digit number that were shown on the screen. For the mental rotation task a three-dimensional object was displayed on the screen and subjects were asked to imagine this object rotating along one of the axes. In addition to the four feedback runs, a calibration run was performed in the beginning of the experiment to train a classifier to distinguish between the three classes. The structure of this run was equivalent to the feedback runs (see below), the only difference being that no real feedback was provided due to the lack of a classifier. However, to avoid large visual differences between the calibration and the feedback runs which could lead to biases, sham feedback was provided in form of a moving blue bar. Subjects were instructed that this feedback did not reflect their actual performance during the calibration run.

Figure 1, left shows the structure of a trial. A trial began with the presentation of a white fixation cross. After two seconds an acoustical “beep” marked the actual trial start and after three seconds an arrow pointing right, left or upward indicated which task to perform. The arrow pointing right, left or upward signaled that mental rotation, left hand motor imagery or mental subtraction should be performed, respectively. Subjects were instructed to start the mental imagery as soon as the arrow appeared. To help the users remember which task to perform, little pictures were shown on screen (not shown in Figure 1). A blue bar was shown for the duration

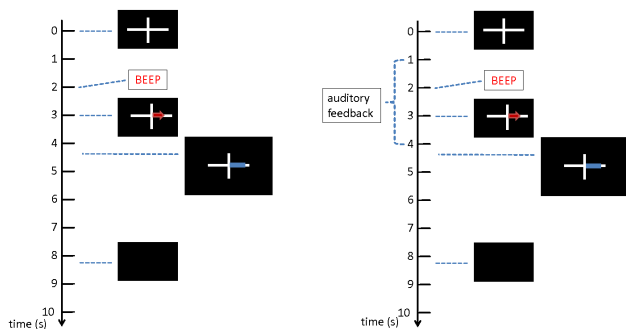


Fig. 1. Structure of a single trial, Left: data used for evaluation of features, Right: BCI experiment with additional auditory feedback

of the feedback period. This bar extended into the direction corresponding to the respective mental imagery task, but only when the correct task was identified by the classifier. In the case of a wrong classification no bar was shown (positive feedback only). The bar length corresponded to the distance of the classifier output to the separating hyperplane and thus reflected the confidence of the classifier in its decision.

EEG signals were recorded from 30 electrodes (C6, CP4, CPZ, CP3, P5, P3, P1, PZ, P2, P4, P6, PO7, PO8, OZ, F3, FZ, F4, FT8, FC6, FC4, FCZ, FC3, FC5, FT7, C5, C3, C1, CZ, C2, C4) using two 16-channels g.tec g.USBamp amplifiers and the OpenViBE software (<http://openvibe.inria.fr>). Data from the calibration run were used to determine common spatial pattern (CSP) filters to extract class-specific features which were then used to build a linear discriminant analysis (LDA) classifier [17]. This classifier was subsequently applied during the feedback runs to assess performance and present the bar feedback. First, data were bandpass filtered in the  $\mu$  and  $\beta$  frequency bands (8–30 Hz) using a fourth-order butterworth filter. Three sets of CSP filters were determined corresponding to the discrimination of each of the three classes from the other two classes. We used two pairs of CSP filters for each class. After the determination of the CSP filters, the data were filtered correspondingly and the logarithmic variance of the filtered signals was calculated. A One-Vs-the-Rest (OVR) multiclass LDA classifier was then trained on these data. During the online sessions data were classified continuously by using the last 1s of EEG signal with a stepsize of 62.5ms. The logarithmic variance of the bandpass and CSP filtered signals was used as input to the previously trained LDA classifier. Thus, the feedback was updated every 62.5ms and corresponded to the mean performance over the past 1s.

### B. Tested Features

The additional feedback information should follow three main objectives. Firstly, it should be available online during the experiment and it should be possible to update it regularly. Secondly, the feature should have a clear connection to BCI performance so that information about it might help people to improve performance. Also, the direction of the connection should be clear, i.e. the user should know if they have to up- or down-regulate the respective feature to improve performance. Thirdly, the feature should not involve any complex computations that users might not understand and/or have an intuitive explanation. With these three objectives in mind, we notably determined the following two features:

1) *Muscle bandpower*: One important aspect for the EEG signal recording in general and for the use of a BCI in particular is the reduction of muscular artifacts, as it can produce electromyogram (EMG) activity that interferes with the EEG signals. To measure muscular tension we used three electrodes from the front (F3, Fz and F4) and three electrodes from the back of the head (PO7, Oz and PO8) and determined the mean bandpower in a frequency band from 40 to 70 Hz which is associated with muscular activity. Note that the correlation between each of these channels bandpower and BCI

performance was in the same direction, which is why we used them together. The hypothesis for this feature was that it should be negatively correlated to BCI performance, i.e. the lower the bandpower in the muscle frequency band the better the performance in the BCI task.

2) *Gamma bandpower*: Another important aspect influencing performance during BCI experiments is the level of attention the user is paying to the task. The ability to concentrate accounts for 19% of the variance in BCI performance [18]. As mentioned before, [10], [11] identified a neurophysiological predictor of BCI performance with a similar interpretation. They showed that the modulation of the SMR is correlated with bandpower in the  $\gamma$  frequency range. Since  $\gamma$  oscillations are associated with shifts in attention the authors conclude that BCI performance is influenced by attentional processes [10], [11]. As in [19] we used a linearly-constrained-minimum-variance (LCMV) beamformer to determine the relevant brain region in which  $\gamma$  power is associated with BCI performance. A beamformer can be seen as a spatial filter which in this case was aimed at the superior parietal cortex, a region that was found to best predict BCI performance in [11]. In the present study a 5s long baseline period from the beginning of the calibration run was used to compute the beamformer for every subject individually. The resulting spatial filter was then applied to the signals from the four feedback runs before the bandpower in the  $\gamma$  frequency range (55-85 Hz) was calculated.

To evaluate these features regarding their relation to BCI performance two selection criteria were tested. First, we looked for correlations between the respective feature and BCI performance (mean classification accuracy over the feedback period) across subjects. However, in order for the feedback to be useful online during the experiment, we also looked for correlations with performance within-subjects, on smaller time windows. The time windows here corresponded to the moving windows of 1s length and with a stepsize of 62.5ms that were described earlier in the context of the feedback update.

To take into account not only the class label predicted by the classifier, but also the certainty of the classifier in its prediction, we followed the approach in [10]. They proposed a score - which we denote here as LDA-score - which is calculated by taking the absolute classifier output (output of the LDA for the winning class in the OVR scheme) in case of a correct time window, and by multiplying the absolute classifier output by  $-1$  in the case of a wrongly classified time window. This results in high positive values for a very certain correct classification, small absolute values in the case of uncertain classification and high negative values if the classifier was very certain in predicting the wrong class. By this means a continuous-valued performance measure on the basis of single time windows is obtained which can then be correlated with the values of the respective feedback feature.

### C. Results

Both features were assessed according to their relation to BCI performance across subjects as well as on the basis of small time windows. To check whether the features can explain

variability in performance over different runs and subjects we considered correlations between the respective feature and mean BCI performance across 15 subjects with four runs each. In both cases we can observe a significant negative correlation between the feature and BCI performance, with  $\rho = -0.3391$  ( $p = 0.008$ ) for the muscle bandpower and  $\rho = -0.4243$  ( $p = 0.0007$ ) for the gamma bandpower.

Correlations between the two features and the LDA-score for each subject are displayed in Figure 2. A star indicates that the correlation is significant ( $p < 0.05$ , t-test). For the  $\gamma$  bandpower analyses the direction of the correlation is almost consistent across subjects, i.e., the feature is almost always negatively correlated to the performance score. For the muscle bandpower the direction of the correlation is more subject-specific. While it is negative for most subjects, there are three subjects who show a significant positive correlation. In all cases the strength of the correlation varies across subjects.

### D. Discussion

Both features explored here were significantly correlated with BCI performance across subjects and runs. Thus, they can - to a certain extent - explain the variability in BCI performance between different runs and subjects. In the case of the muscle bandpower the correlation was negative which means that it went in the expected direction, i.e. the more tensed a subject the lower the BCI performance. The analyses of  $\gamma$  bandpower also yielded a negative correlation which is not what we expected. If we adopt the proposed interpretation of high  $\gamma$  bandpower being related to attentional processes, this would mean that the less concentrated a subject is the higher the BCI performance. This also seems to contradict findings by [18] who found a positive correlation between the ability to concentrate and BCI performance. However, there might be other explanations for the negative correlation. First, in [11], a multivariate model was used whereas we used here a univariate model. According to [20], a positive regression weight in a multivariate model does not necessarily imply a

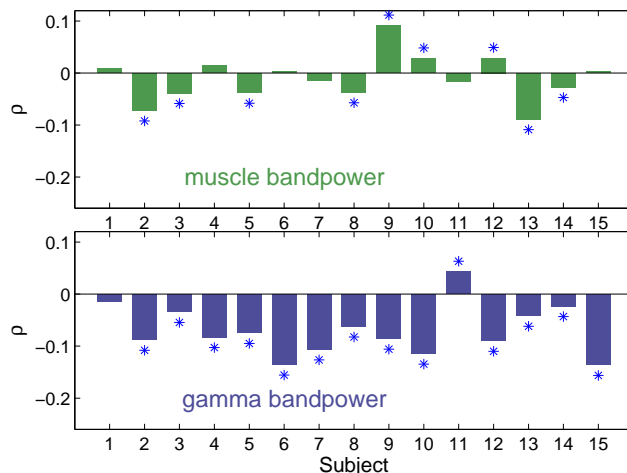


Fig. 2. Correlation between the LDA-score and the tested feedback features per subject. Stars indicate significant correlation coefficients ( $p < 0.05$ ).

positive correlation of that feature with the target variable, in a univariate model. Second, this multivariate model was built on a BCI purely based on motor imagery whereas ours used other, non-motor mental tasks. Finally, when dealing with bandpower features in higher frequency bands (which includes the 55–85 Hz band that was used here) the contamination by muscular artifacts might be high [21]. Because we want to use the feedback features in an online fashion it was not possible to perform rigorous artifact removal procedures as in [11]. Actually, the correlation between the muscle bandpower and the gamma bandpower was  $\rho = 0.57$  ( $p < 0.00001$ ), which suggests that our gamma bandpower feature was probably measuring a mix of muscle activity, and, hopefully, attention.

While we found a clear correlation between the tested features and BCI performance across runs the picture is less clear when we consider single subjects. High inter-subject variability has also been reported by [11] which is consistent with our finding that the value and direction of the correlation between BCI performance and the tested features is very subject-dependent. Also the strength of the correlations decreases when single subjects are used as compared to the correlation across subjects. It seems to be easier to explain variability in BCI performance between than within subjects.

Based on the results from the feature evaluation we decided to test the muscle bandpower to use it as feedback in an actual BCI experiment. Indeed, the muscle bandpower being linked to performance, informing subjects about their level of muscular tension might help them to improve BCI performance. Moreover, users can easily interpret the muscle bandpower as a measure of muscular tension and muscular artifacts, and thus regulate it. The next section reports on the evaluation of an online BCI using the muscle bandpower as additional feedback feature to help users acquire BCI control.

### III. BCI EXPERIMENT WITH ADDITIONAL FEEDBACK

#### A. Methods

The experiment structure was similar to the one described in Section II-A. It also consisted of four feedback runs preceded by a calibration run with sham feedback, with the same trial timing. The only difference lies in the administration of a new feedback that was provided before the start of the visual bar feedback. It was presented from 1s after the appearance of the fixation cross for a duration of 3s, i.e. it ended right before the blue bar feedback began (see Figure 1 right). The new feedback was given in auditory form as a sound that was played from loudspeakers. In a pilot experiment different sounds were tested. Most subjects preferred a relaxing sound which was therefore chosen. It was presented during the 3s long period whenever the muscle bandpower exceeded a certain threshold. Participants were asked to relax their muscles and to avoid any movement in order to make the sound stop during the 3s period. They were informed that the sound stopped right before the visual feedback started in any case.

The auditory feedback was only given in the three last feedback runs. The first feedback run was exactly the same as described in Section II-A. Data from this run was used

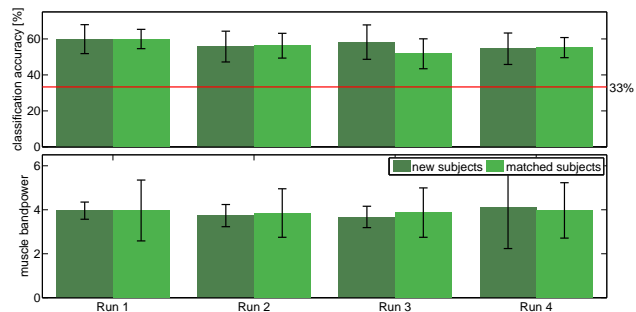


Fig. 3. Top: mean classification accuracies for subjects from the experiment and control group. Subjects from our experiment received an auditory feedback about muscular tension in addition to the visual moving bar feedback. The control group only received the visual moving bar feedback. The red line indicates chance performance for this three-class experiment. Bottom: corresponding bandpower in the 40–70 Hz frequency band. In this case control group is chosen to match the muscle bandpower during Run 1.

to determine the threshold which was computed as the 60<sup>th</sup> percentile of the distribution of the muscle bandpower during this first feedback run. Thresholds were subject-specific. The pilot experiment showed that the distributions of the muscle bandpower for a separate baseline period as well as for the calibration run were too different from the bandpower distribution during the feedback runs. Therefore, the first feedback run was used to determine a realistic threshold. 10 BCI-naïve participants took part in this experiment (6 males, 4 females, mean age 29.1, range 23–42). The data from ten subjects was selected from the previous experiment (see Section II-A) so that it matched in terms of performance during the first feedback run. They constituted the control condition.

#### B. Results

Figure 3 shows the mean performance over subjects for the three runs and compares it to the mean performance over 10 control subjects from the previous experiment. Note that the first run was the same for both conditions while run two to four included the additional auditory feedback in our experiment. Control subjects were chosen such that the average performance in the first feedback run (59.89%), was as similar as possible to the performance in the first run of our experiment (59.95%,  $p=0.9830$ ). Although the performance is slightly higher with additional auditory feedback in run 3, a two-way analysis of variance (ANOVA) for repeated measures with the factors *Condition* ( $C_2$ : auditory vs no auditory feedback) and *Runs* ( $R_3$ : run 2, 3 and 4) did not reveal any significant main effect of the *Condition* [ $F(1,18)=0.369$ ,  $p=0.551$ ], the *Run* [ $F(1,18)=0.480$ ,  $p=0.497$ ] nor a *Condition\*Run* interaction [ $F(1,18)=0.003$ ,  $p=0.960$ ]. To evaluate the effect of the additional feedback on muscular relaxation, 10 subjects were chosen from the previous experiment such that they matched in terms of muscle bandpower during the first run ( $p=0.9855$ ). In line with the slightly better performance the muscle bandpower is a little lower for subjects with additional auditory feedback in run 3 (see Figure 3). Again a two-way ANOVA for repeated measures showed no main effect of the *Condition*

[F(1,18)=0.017,  $p=0.897$ ], the *Run* [F(1,18)=0.845,  $p=0.370$ ] nor a *Condition\*Run* interaction [F(1,18)=0.220,  $p=0.644$ ].

### C. Discussion and Conclusion

We did not find any significant effects of the additional auditory feedback, neither in terms of performance nor in terms of muscle bandpower. It may be due to its duration (3s), which might have been too short for the user to understand the feedback and react to it by relaxing. This duration was chosen to fit the trial timing from the previous experiment, used as a control condition. It would thus be interesting (1) to study the effect of a longer auditory feedback period and (2) to postpone the start of the BCI task until the feature has reached a desired value [11]. In our case, this would mean to wait with the presentation of the BCI task until the user is relaxed enough, i.e. the muscle bandpower is below a predefined threshold.

Besides, BCIs being co-adaptive systems both the user and the machine might be responsible for low performance. When evaluating the different feedback features according to their ability to predict performance we implicitly assumed that the user was responsible for incorrectly classified time windows, e.g. he was not relaxed enough or was not paying enough attention to the task. Yet, it could also be due to the computer, e.g., to an imperfect signal processing or classification algorithm. In the future, distinguishing between human and computer errors should help to identify more specific feedback features.

Altogether, this study aimed at exploring various EEG features that could be used as explanatory feedback for BCI training. The challenge was to find a feature that could be used online. As such we explored a measure of muscle tension from the frontal and neck areas and the  $\gamma$  bandpower as a measure of the attentional level. Both appeared to be correlated to BCI performance. In the case of the  $\gamma$  bandpower it might be problematic to distinguish  $\gamma$  power associated with attention from muscular artifacts. However, if a contamination by EMG activity can be excluded it might be a useful feature for BCI feedback. In the future, it would also be worth exploring channel selection or spatial filtering to refine and improve our muscle bandpower feature which only used fixed channels so far. We also showed that the correlation between a feature and performance is very subject-specific. Understanding these differences might help to find out which information is useful for which group of subjects. By this approach different features could be used as feedback for different groups of subjects.

Furthermore, we tested some of these features in a BCI experiment to investigate whether they are useful to improve user performance. Although no improvement was found in the experiment using feedback about muscular relaxation, adjustments to the experimental design might help to make this feature useful to improve BCI performance, as discussed above. Moreover, we could show that it is possible to use an additional feedback during a BCI experiment without deteriorating performances. Using information about muscular activity might not be very useful for paralyzed patients. However, BCI systems are also used in various other contexts, such as stroke rehabilitation or gaming [1], involving users for which

contamination of the EEG signals by muscular activity might be a problem. We hope this study could be a first step towards designing explanatory feedback for BCI with the objective of improving BCI training and thus BCI reliability.

### ACKNOWLEDGMENT

The authors thank M. Grosse-Wentrup, A. Cellard and J. Frey for their help with this study.

### REFERENCES

- [1] J. R. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, 2012.
- [2] F. Lotte, F. Larrue, and C. Mühl, "Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design," *Front. hum. neurosci.*, vol. 7, 2013.
- [3] J. Hattie and H. Timperley, "The power of feedback," *Rev. educ. res.*, vol. 77, no. 1, pp. 81–112, 2007.
- [4] F. Cincotti, L. Kauhainen, F. Aloise, T. Palomäki, N. Caporusso, P. Jylänki, D. Mattia, F. Babiloni, G. Vanacker, M. Nuttin *et al.*, "Vibrotactile feedback for brain-computer interface operation," *Comp. intel. neurosci.*, vol. 2007, 2007.
- [5] M. Gomez-Rodriguez, J. Peters, J. Hill, B. Schölkopf, A. Gharabaghi, and M. Grosse-Wentrup, "Closing the sensorimotor loop: haptic feedback facilitates decoding of motor imagery," *J. neur. eng.*, vol. 8, no. 3, 2011.
- [6] K. A. McCreadie, D. H. Coyle, and G. Prasad, "Is sensorimotor BCI performance influenced differently by mono, stereo, or 3-D auditory feedback?" *IEEE Trans. Neur. Syst. Rehab.*, vol. 22, no. 3, 2014.
- [7] F. Nijboer, I. Gunst, S. Von Hartlieb, D. McFarland, N. Birbaumer, and A. Kübler, "A comparison between auditory and visual feedback of sensorimotor rhythms (SMR) for a brain-computer interface (BCI) in healthy participants," in *Psychophysiology*, vol. 43, 2006, pp. S71–S71.
- [8] F. Lotte, J. Faller, C. Guger, Y. Renard, G. Pfurtscheller, A. Lécuyer, and R. Leeb, "Combining BCI with Virtual Reality: Towards New Applications and Improved BCI," in *Towards Practical Brain-Computer Interfaces*. Springer, 2013, pp. 197–220.
- [9] H.-J. Hwang, K. Kwon, and C.-H. Im, "Neurofeedback-based motor imagery training for brain-computer interface (BCI)," *J. neuro. meth.*, vol. 179, no. 1, pp. 150–156, 2009.
- [10] M. Grosse-Wentrup, B. Schölkopf, and J. Hill, "Causal influence of gamma oscillations on the sensorimotor rhythm," *NeuroImage*, 2011.
- [11] M. Grosse-Wentrup and B. Schölkopf, "High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces," *J. neur. eng.*, vol. 9, no. 046001, 2012.
- [12] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain-computer interface: A brief review," *J. neurosci. meth.*, 2015.
- [13] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu, "The predictive role of pre-cue EEG rhythms on MI-based BCI classification performance," *J. neurosci. meth.*, vol. 235, pp. 138–144, 2014.
- [14] T. Kaufmann, J. Williamson, E. Hammer, R. Murray-Smith, and A. Kübler, "Visually multimodal vs. classic unimodal feedback approach for SMR-BCIs: a comparison study," *Int. J. Bioelectromagn.*, 2011.
- [15] C. Jeunet, B. N'Kaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting mental-imagery based BCI performance from personality, cognitive profile and neurophysiological patterns," *PLOS One*, 2015.
- [16] E. V. Friedrich, C. Neuper, and R. Scherer, "Whatever works: A systematic user-centered training protocol to optimize brain-computer interfacing individually," *PloS one*, vol. 8, no. 9, p. e76214, 2013.
- [17] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 441–446, 2000.
- [18] E. M. Hammer, S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, K.-R. Müller, and A. Kübler, "Psychological predictors of SMR-BCI performance," *Biol. psych.*, vol. 89, no. 1, pp. 80–86, 2012.
- [19] M. Grosse-Wentrup and B. Schölkopf, "A brain-computer interface based on self-regulation of gamma-oscillations in the superior parietal cortex," *J. neur. eng.*, vol. 11, no. 5, p. 056015, 2014.
- [20] S. Weichwald, T. Meyer, O. Özdenizci, B. Schölkopf, T. Ball, and M. Grosse-Wentrup, "Causal interpretation rules for encoding and decoding models in neuroimaging," *NeuroImage*, vol. 110, pp. 48–59, 2015.
- [21] I. Goncharova, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "EMG contamination of EEG: spectral and topographical characteristics," *Clin. Neurophys.*, vol. 114, no. 9, pp. 1580–1593, 2003.