

Ensembles of adaptive spatial filters increase BCI performance: an online evaluation

Claudia Sannelli¹, Carmen Vidaurre^{1,2}, Klaus-Robert Müller*^{1,3} and Benjamin Blankertz*⁴

¹ Machine Learning Group, Technische Universität Berlin, Marchstr. 23, 10587 Berlin, Germany

² Mathematics Dp., Public University of Navarra, Calle Arrosadia s/n, 31006 Pamplona, Spain

³ Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea

⁴ Neurotechnology Group., Technische Universität Berlin, Marchstr. 23, 10587 Berlin, Germany.

E-mail: klaus-robert.mueller AT tu-berlin.de, benjamin.blankertz AT tu-berlin.de

Abstract.

Objective In electroencephalographic (EEG) data, signals from distinct sources within the brain are widely spread by volume conduction and superimposed such that sensors receive mixtures of a multitude of signals. This reduction of spatial information strongly hampers single-trial analysis of EEG data as, e.g., required for brain-computer interfacing (BCI) when using features from spontaneous brain rhythms. Spatial filtering techniques are therefore greatly needed to extract meaningful information from EEG. Our goal is to show, in online operation, that Common Spatial Patterns Patches (CSPP) are valuable to counteract this problem.

Approach Even though the effect of spatial mixing can be encountered by spatial filters, there is a trade-off between performance and requirement of calibration data. Laplacian derivations do not require calibration data at all, but their performance for single-trial classification is limited. Conversely, data-driven spatial filters such as Common Spatial Patterns (CSP) can lead to highly distinctive features, however they require a considerable amount of training data. Recently, we showed in an offline analysis that CSPP can establish a valuable compromise. In this paper, we confirm these results in an online BCI study. In order to demonstrate the paramount feature of CSPP to require little training data, we used them in an adaptive setting with 20 participants and focused on users who did not have success with previous BCI approaches.

Main results The results of the study show that CSPP adapts faster and thereby allows users to achieve better feedback within shorter time than previous approaches performed with Laplacian derivations and CSP filters. The success of the experiment highlights that CSPP has the potential to further reduce BCI inefficiency.

Significance CSPP are a valuable compromise between CSP and Laplacian filters. They allow attaining better feedback within shorter time and thus reduce BCI inefficiency to one fourth in comparison to previous non-adaptive paradigms.

1. Introduction

Common Spatial Patterns (CSP) is the most widely used and efficient spatial filter algorithm for Electroencephalography (EEG) Brain-Computer Interfaces (BCI) based on the modulation of sensorimotor rhythms (SMR), see e.g (Blankertz et al., 2008; Ramoser et al., 2000). A big drawback of CSP is the considerable amount of training data required in order to obtain robust CSP filters and avoid overfitting. In fact, the CSP algorithm is supervised and the number of parameters to estimate by CSP increases quadratically with the number of EEG channels involved, bearing a risk of overfitting when using little training samples or in the presence of outliers.

In Sannelli et al. (2010) it was shown that, in absence of previous subject-specific knowledge, reducing the number of channels does not help to improve the performance of CSP. This is especially the case with users suffering from BCI inefficiency, whose data are highly non-stationary and contain little class relevant information. In fact, a certain spatial resolution is required in order to cover the potentially relevant brain areas, which often differ from user to user. Based on a data set of 80 participants performing binary SMR-based BCI by the imagination of limb movements, a number of 32 channels and at least 120 trials were recommended as optimal configuration for training the CSP algorithm (Sannelli et al., 2010).

Due to practical constraints, long multi-channel recordings might be difficult to realize. In the case of BCI medical applications, the time needed to setup a BCI system is extremely important, since the patients become tired very soon and long BCI sessions might be unfeasible (Ang and Guan, 2015). For this reason, researchers have investigated ways to reach good performance with little calibration data (Vidaurre et al., 2007; Kayagil et al., 2009; Krauledat et al., 2008; Kang et al., 2009; Lu et al., 2009; Lotte and Guan, 2010; Kindermans et al., 2014) or in absence of it, i.e. with subject-independent spatial filters and classifiers previously trained on datasets of other users (Lotte et al., 2009; Fazli et al., 2009; Samek et al., 2014). Other solutions concentrate on training CSP on “stationary” properties of the data ((Samek et al., 2012; Arvaneh et al., 2013)). Alternatively, and especially when also the number of recording channels is limited, a simple solution is offered by Laplacian filters, LAP (McFarland et al., 1997). They are widely used in the BCI context (Babiloni et al., 2001; He et al., 2001; Solis-Escalante et al., 2008; Pfurtscheller et al., 2009; Faller et al., 2012, 2014) and are often preferred over complex features. Band power features calculated from LAP filtered channels represent a robust signal to control a BCI whose calculation does not involve any class information and may run less frequently into the risk of overfitting. On the other hand, often the performance obtained with LAP features is poor in comparison to subject-specific optimized spatial filters. For example in (Vidaurre et al., 2011c,b), it was shown that the performance of people who do not experience very good BCI control is significantly worse using Laplacian channels than CSP filters.

In Sannelli et al. (2011), a novel spatial filter that can be considered as a compromise between LAP and CSP filters was proposed. Since it consists of an ensemble of local

CSP patches, the method is called CSPP. CSPP were developed to offer a solution to the overfitting problem of CSP and as a substitute for LAP for all those cases where the application of CSP is not feasible. CSPP are expected to work better than LAP, since they use the class information and are therefore optimized on the user’s brain activity.

In Sannelli et al. (2011), CSPP is offline evaluated with several training sample sizes (from 10 to 75 trials) and compared with Laplacian derivations, classical CSP and the regularized CSP (R-CSP) presented in Lu et al. (2010).

Here, the evaluation of CSPP in online operation is presented in a study with 20 participants in order to effectively assess the suitability of the algorithm to BCI experiments and its performance. All participants except for three, previously participated either in a study consisting of a calibration and feedback session (which from now on will be called “standard approach”) or in one of two co-adaptive calibration studies with Laplacian filters (Vidaurre et al., 2011c,b) (from now on the first and the second co-adaptive study respectively), or in both, but not in more than two studies out of these three. After these sessions, the BCI systems tested could not offer but poor performance to those 17 users. Additionally, the last motor imagery (MI) session for each user was at least two years earlier and many of them did not even remember the task.

2. Methods

In section 2.1 and 2.2 we describe the method and general approach of CSPP-based co-adaptive calibration. The specific settings of the current study are presented in section 2.3.

2.1. Common Spatial Patterns Patches

CSPP is obtained by applying CSP analysis to small channel sets (patches) and combining the resulting features by some type of feature selection method. In brief, Common Spatial Patterns (CSP) (Ramoser et al., 2000) is a discriminative algorithm that determines the spatial filters \mathbf{W} from band-pass filtered EEG data such that the difference between the variances of the filtered data for the two classes is maximized.

This is done by a simultaneous diagonalization of the matrices $\mathbf{\Sigma}_1 = \mathbf{X}_1\mathbf{X}_1^\top$ and $\mathbf{\Sigma}_2 = \mathbf{X}_2\mathbf{X}_2^\top$. These matrices are related to the covariance matrices of the two classes (centering is not necessary as the signals are band-pass filtered and therefore have zero-mean):

$$\mathbf{W}^\top \mathbf{\Sigma}_1 \mathbf{W} = \mathbf{\Lambda}_1 \tag{1}$$

$$\mathbf{W}^\top \mathbf{\Sigma}_2 \mathbf{W} = \mathbf{\Lambda}_2, \tag{2}$$

$$\text{s.t. } \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2 = \mathbf{I} \tag{3}$$

where $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are diagonal matrices and each generalized eigenvalue λ on the diagonal corresponds to an eigenvector \mathbf{w}^\top . In this way, the eigenvectors are the same for both

decompositions (Eq. (1) and (2)), while the corresponding Eigenvalues are complementary due to Eq. (3): First note, that the eigenvalues are bounded between 0 and 1 due to Eq. (3). If an eigenvalue of Σ_1 is close to 1, i.e., relatively large, the corresponding eigenvalue of Σ_2 is close to 0, i.e., relatively small. Since large/small eigenvalues correspond to high/low band-power, these CSP filters provide well discriminable band-power features, see Blankertz et al. (2008) for more information.

The the formulas in Eqs. (1) to (3) form the generalized eigenvalue problem:

$$\Sigma_2 \mathbf{W} = (\Sigma_1 + \Sigma_2) \mathbf{W} \Lambda \quad (4)$$

Putting D chosen filters corresponding to extreme eigenvalues (either close to 1 or close to 0) in a matrix \mathbf{W}_D , the filtered data $\hat{\mathbf{s}}(t) = \mathbf{W}_D^\top \mathbf{X}$ will have smaller dimensionality $D < N$ and the two classes will be maximally separated by their variance. A CSP feature is the log-variance of the band-pass and CSP filtered data.

CSP bears the risk of over-fitting because the data covariance matrices usually has high dimensionality, what makes it difficult to estimate. By applying CSP analysis on just a few channels it is possible to reduce this risk because those matrices have significantly less dimensions. Also, CSPP can be interpreted as a Laplacian filter where the weights of all channels are optimized in a data-driven manner. We will call *center* of the patch to the central channel of a Laplacian filter, i.e. the one that is weighted by 1 in Laplacian filtering.

CSPP is also inspired by the fact that many users exhibit non conventional SMR activity, regarding not just the size of the involved area, but also the location. To take into account this problem and maximize the flexibility of these filters, several forms of patches and several centers are possible. In this way, each patch can include a different number of surrounding channels and the position of the centers of the patches can be chosen, depending on the number of channels available, on the task and, when training data are already acquired, also depending on the subject-specific SMR activity. Here, eight patch forms in combination with a number of centers going from three to 18 are evaluated. The analyzed patch forms using the channel C3 as center are shown on the left of Fig. 1 and the centers of the configurations are shown on the right of Fig. 1. Similarly to

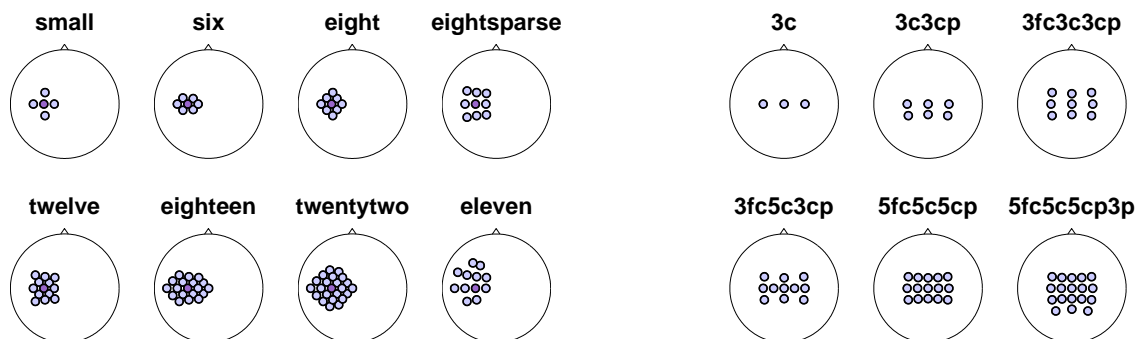


Figure 1. Left: patch configurations centered on C3. Right: channel sets used as center of the patches/Laplacian filters.

CSP, CSPP are applied on band-pass and time filtered data, see Blankertz et al. (2008);

Lemm et al. (2011) . By computing CSP on one patch with some n channels (being N the total number of channels and $n \leq N$), n filters are obtained. With N_c patch centers, for each patch p with $p = 1, \dots, N_c$, one filter per class is selected by choosing the extreme eigenvalues, i.e. two filters per patch are obtained. A sparse filter matrix \mathbf{W} with dimension $N_c \times N$ results from the concatenation of all CSPP filters, where N is the number of all available electrodes:

$$\mathbf{W} = [\mathbf{w}_{11}, \mathbf{w}_{12}, \dots, \mathbf{w}_{p1}, \mathbf{w}_{p2}, \dots, \mathbf{w}_{N_c1}, \mathbf{w}_{N_c2}] \quad (5)$$

The matrix \mathbf{W} is sparse since each column \mathbf{w}_{pi} contains just n non zero elements. From the resulting ensemble of filters \mathbf{W} , the most informative ones are chosen using the *ratio-of-medians* score, see (Blankertz et al., 2008). The sparsity of matrix \mathbf{W} is represented in Fig. 2.

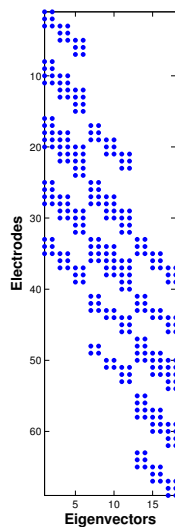


Figure 2. Sparse matrix \mathbf{W} of spatial filters resulting from the concatenation of all CSPP filters.

Fig. 3 is an example of CSPP optimized weights of the 13 channels necessary to compute *small* patches over C3, Cz and C4. In comparison to Laplacian filters, CSPP filters can change and provide features with better class discriminability. The top row presents the classical Laplacian derivations with 4 neighboring channels centered in C3, Cz and C4. The lower three rows depict the CSPP involving the same channels for the class combinations *Left/Right*, *Left/Foot* and *Foot/Right*. The CSPP were calculated on the calibration data of good performing users.

2.2. Co-adaptive calibration approach with CSPP features

The co-adaptive calibration design consists of three different subsequent levels that can be summarized as follows:

- **Level 1:** Use of subject-independent features and classifier. The classifier is adapted in a supervised manner.

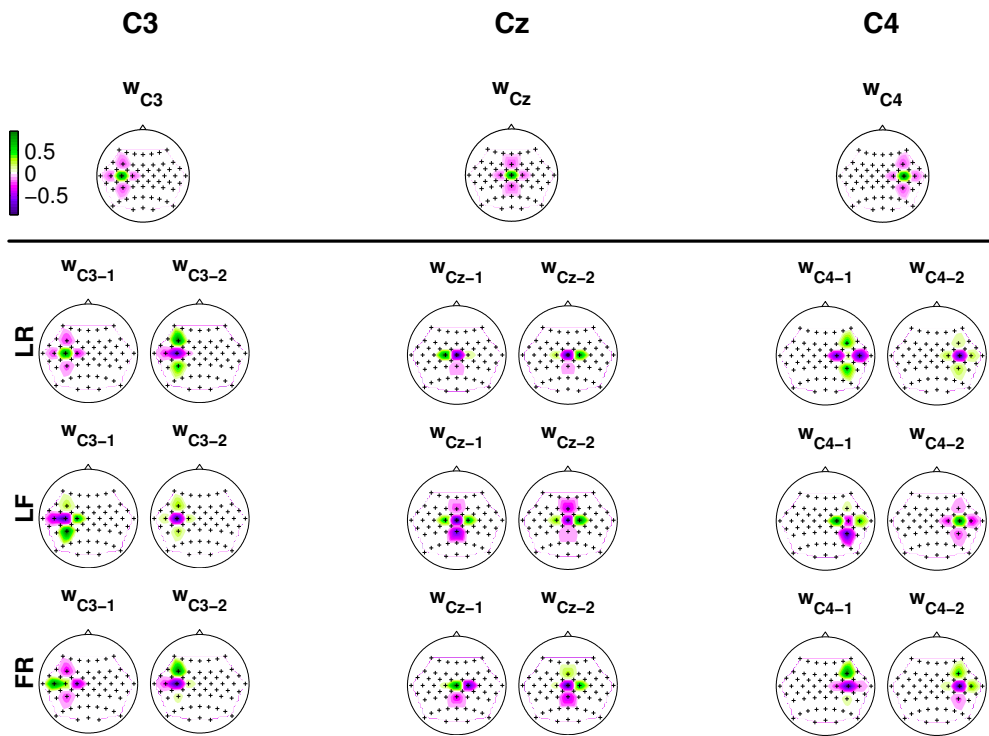


Figure 3. Top: *Small* Laplacian filters for C3, Cz and C4. Bottom: *Small* CSPP filters centered in C3, Cz and C4, for the class combination *Left/Right* (first row), *Left/Foot* (second row) and *Foot/Right* (third row), calculated on the data of three good performing users.

- **Level 2:** Use of subject-dependent features and classifier. The classifier and the features are adapted in a supervised manner.
- **Level 3:** Use of subject-dependent features and classifier. The classifier is adapted in an unsupervised manner, features are not changed.

During the whole experiment the feedback application employed was cursor control.

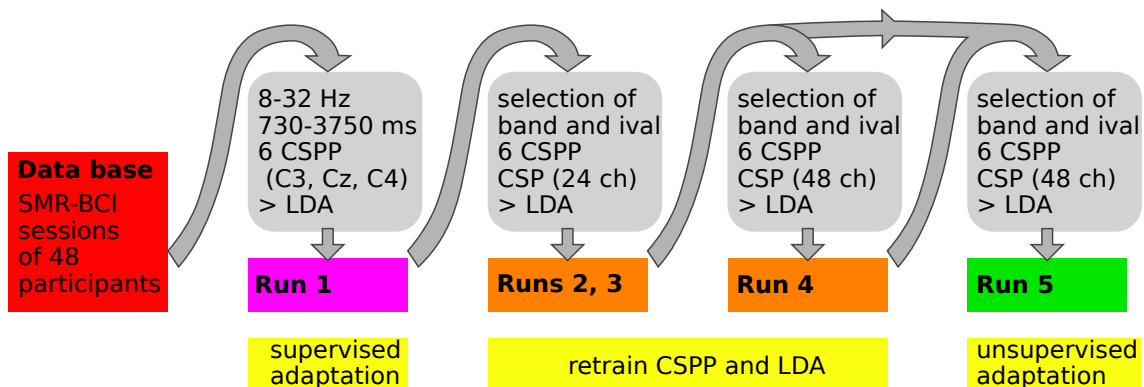


Figure 4. Summary of methods employed in the co-adaptive framework with CSPP.

2.2.1. Offline analysis before level 1 Before starting level 1 there was no previous subject data available. Spatial filters, features and classifiers were computed using a data base

of recorded data from 48 successful BCI participants (feedback performance over 70% of accuracy). Their EEG data were concatenated before processing it to subsequently obtain features and classifiers. For each binary combination of MI classes (left/right, left/foot, right/foot), six CSPP features were used. They were obtained by CSP analysis on three *small* patches centered on C3, Cz and C4 (see bottom panel of Fig. 3). To compute the CSPP, the data was band-pass filtered in a broad band between 8-32 Hz using only the post-stimulus interval 750-3750 ms (during effective MI). CSPP features are the log-variance of the six CSPP filtered signals. The subject independent classifiers were also calculated for each binary combination of MI classes using Linear Discriminant Analysis (LDA) classifier. Each LDA was trained on the aforementioned CSPP features. Before starting the first run, the binary class combination to be used was determined and stayed fixed throughout the session. The BCI system was co-adaptive with the goal of guiding the user to modulate the SMR activity in the areas covering C3, Cz and C4 as done by the successful participants whose data were used to prepare the subject-independent spatial filters and classifier.

2.2.2. Online operation of level 1 During the online operation, the spatial filters and features stayed fixed whereas the initial subject-independent classifier was adapted after each trial t to the current brain signals of the user. This was done by updating the class means and covariance matrices. For more details about this algorithm, refer please to (Vidaurre et al., 2006). The adaptation speed of the classifier was determined by two update coefficients, one for the class average and another one for the covariance matrices and they were determined before hand. In particular a pool of 80 datasets was used to choose the best update coefficient for each subject category (I, II and III). The 20 subjects of this CSPP study were assigned to one category depending on their previous performance and the corresponding update coefficient was used. The selected coefficient was greater for worse performing subjects. Note that this adaptation method is supervised.

2.2.3. Offline analysis before level 2 Before the first run of level 2, data acquired in the adaptation level 1 was used to initialize all necessary parameters, i.e. to select subject-specific frequency band, time interval, spatial filters (CSP and CSPP) and classifier. Between the runs of level 2 this procedure was repeated but using data from the previous run(s) of level 1 and level 2, with a maximum of two runs. A semi-automatic procedure selected a subject-specific frequency band and time interval where the classes were maximally separated (see (Blankertz et al., 2008) for details). Then the data were band-pass filtered and epoched using these subject-specific parameters. Twenty-four centers distributed in the motor areas were used (see channels in red of Fig. 5). Also, for each patch form in Fig. 1, a generalization error was calculated by 4-fold cross validation where for each fold: 1) two to six CSP features and two to six CSPP features (out of 48) were calculated on the training set and used to train a rLDA corrected by

shrinkage (regularized LDA, rLDA, (Vidaurre et al., 2009; Blankertz et al., 2011)), 2) the test set was spatially filtered by the up to twelve selected CSP and CSPP filters and 3) the resulting features were classified by the trained rLDA. The patch form with the best generalization error was selected. In the first run of level 2 just 24 channels concentrated on the motor area were used for CSP training (see Fig. 5, channels in red). For subsequent runs 48 channels were used (see Fig. 5, channels in blue), as in Sannelli et al. (2010) they were found to be the optimal locations. Finally, CSPP (with the chosen patch form), CSP and rLDA were trained using all data.

2.2.4. Online operation of level 2 The second phase of the co-adaptive design (level 2) is also supervised. During the online operation in level 2, the patch centers were reselected and the classifier re-trained after each trial, but the patch form stayed unchanged. The selection of CSPP filters was performed as follows: 18 *small* patches were analyzed, resulting in 36 CSPP features (2 per patch). The dimension was reduced by automatically choosing a maximum of six (and minimum of one per class) CSPP features using the *ratio-of-medians* score Blankertz et al. (2008). CSPP was recalculated using the last 60 trials, resulting in new spatial filters and eventually new positions and number of features. The selected CSPP features were concatenated to the fixed CSP filters and the rLDA classifier was re-trained on these new features.

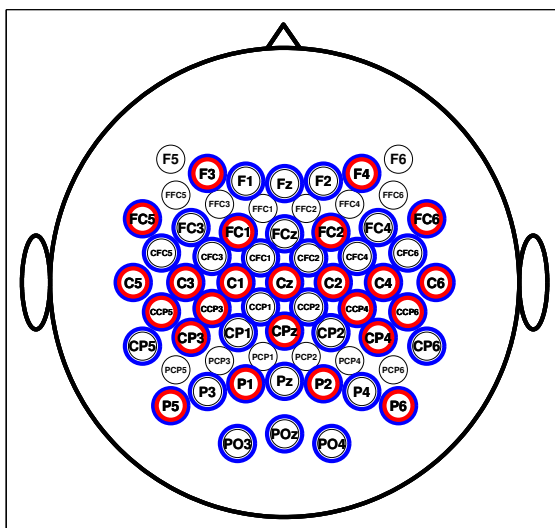


Figure 5. Scalp outline of recorded electrodes. 24 channels over the motor areas marked in red were used to compute CSP in runs 2 and 3. Subsequent runs used 48 channels for CSP (marked in blue).

2.2.5. Offline analysis before level 3 The offline analysis carried out was similar to that in level 2 after run 2, that is, using 48 channels to compute CSP filters.

2.2.6. Online operation of level 3 The last adaptation level, uses static features and an unsupervised adaptation of the classifier to track the features' drifts Vidaurre et al.

(2011a). During the runs, the calculated CSP, CSPP and LDA are used to calculate the features and for online classification. However, part of the LDA classifier is adapted after each trial. In particular, the machine-learning training is carried out using the data of the adaptation level 2, where the users have experienced BCI feedback for longer time. During the runs, the pooled mean of the features $\hat{\boldsymbol{\mu}} = (\hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2)/2$ in the LDA is adapted, where $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$ are the mean of the features of classes 1 and 2 respectively. If both classes have equal a priori probability, one can use the following formula to update the pooled mean in an unsupervised fashion:

$$\boldsymbol{\mu}(t) = (1 - \eta_\mu) \cdot \boldsymbol{\mu}(t - 1) + \eta_\mu \cdot \mathbf{x}(t) \quad (6)$$

whereby η_μ is an update coefficient. In this way each past sample is exponentially weighted. The updated pooled mean modifies the bias of the linear classifier producing the shifting of the hyperplane and tracking the position of the mean of the features. This adaptation is possible in an unsupervised way, i.e. without knowing the label of new trials under the assumption that the mean of the two class means can be estimated by the pooled mean, i.e. the mean of the features of all trials. This technique is called PMean and, among other unsupervised techniques for LDA, it has been previously published in Vidaurre et al. (2011a).

2.3. CSPP Online Study

In this section, the application of CSPP in online operation is presented. The goal is to effectively assess the suitability of the algorithm for BCI experiments and its performance. The parameters, such as adaptation coefficients and number of trials on which CSPP were re-selected, were offline estimated using a large data set of 80 users as in Vidaurre et al. (2011c). Different values of adaptation coefficients were used depending on the type of BCI user. The participants were categorized in three different groups depending on their accuracy in previous studies. Their category refers to the last study in which each user participated. They have a slightly different definition depending on whether the user's performance was calculated in a co-adaptive study or not. Non-coadaptive studies are divided in two phases: calibration (cb., offline performance obtained by cross-validation) and feedback (fb., online performance). Coadaptive studies are divided in levels (1, 2 and 3, as already introduced). The details for each category are summarized in Table 1.

Category I (Cat. I) users exhibit very good performance in all situations and were not invited to this study. All participants (except for the three novice) were ranked Category II or III in previous studies. The classifier adaptation parameters for Cat. III users were chosen slightly larger such that adaptation was faster than for the rest. This is also (and in more detail) explained below.

The result is a flexible BCI system that adapts to the user's brain activity. Performance results are compared with the previous performances of the same users obtained with the standard and/or the co-adaptive approach with Laplacian filters.

Category	Non-coadaptive	Coadaptive
I	cb. > 70% fb. > 70%	all levels > 70%
II	cb. > 70% fb. < 70%	level 1 < 70% level 3 > 70%
III	cb. < 70% fb. not possible	level 3 < 70%

Table 1. User categorization overview. Non-coadaptive studies are divided in two phases, calibration (cb., offline performance obtained by cross-validation) and feedback (fb., online performance). For coadaptive studies the categories are defined depending on performances reached in each level.

2.3.1. Experimental Setup Twenty volunteers participated in a single session of MI-BCI. All participants gave written consent and the study was performed in accordance with the Declaration of Helsinki. Three were novice BCI users, all the others participated either in a large study with calibration and feedback following the standard approach (Blankertz et al., 2010), or in one co-adaptive study (reported in (Vidaurre et al., 2011c) and (Vidaurre et al., 2011b)) or in both co-adaptive studies, but never in more than two sessions. The data for these three studies were acquired respectively 3, 3 and 2 years before.

Only participants belonging to Cat. II and III were invited to take part in this session, since the main goal was to improve the performance of those users. The categorization was assessed based on their spectra at rest Blankertz et al. (2010) and taking into account the previous studies. The three novice BCI users resulted to belong to one category each (I, II and III) by analysis of their achieved performance. The class combination used was chosen either depending on the previous studies, or by the participants.

The trial design was as follows: feedback started at $t = 2s$ until $t = 6s$, i.e. it lasted $t = 4s$. The trial was earlier interrupted if the cursor touched the border of the feedback window and the cursor movement was programmed to be faster for larger classifier outputs and slower for small ones. Since this cursor control resulted to be suboptimal, especially for the beginning of the experiment (see the Section 3.1), a cursor control with fixed trial length and fixed speed was introduced in the experiments with the three BCI novices.

The overview of the experiment design is shown in Fig. 6. Five runs were acquired in total, run 1 with CSPP fixed features and adaptation level 1 (subject-independent), runs 2-4 with fixed CSP features, reselect CSPP features and adaptation level 2 (re-training), and the last run with fixed CSP and CSPP features and unsupervised adaptation (pooled mean adaptation), to obtain a realistic estimation of performance. The first run had 80 trials, the others 100. From run to run the system became more complex and flexible. The settings used for each run are explained in the next section.

For online classification, every 40 ms features were calculated as the log-variance of the band-pass and spatially filtered data of the last 750 ms and classified by rLDA. At the

beginning, after the first three runs and at the end of the experiment, a relax recording was carried out, where the user had to relax and open or close the eyes depending on a vocal instruction. This recording contained 10 trials per condition (eyes open/closed) of 15 seconds.

EEG was recorded by 62 Ag/AgCl electrodes concentrated on the central areas and mastoid references were used (see Figure 5). Figure 4 displays a summary of the methods employed in each level.

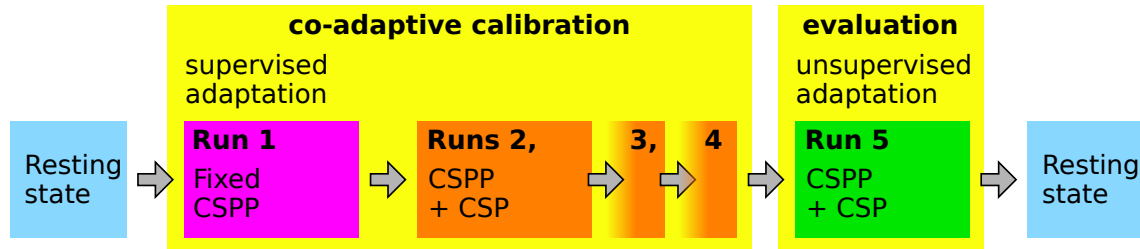


Figure 6. Design of the experiment. Run 1: fixed subject-independent *small* CSPP centered on C3, Cz and C4 and rLDA with supervised adaptation. Runs 2-4: subject-dependent CSPP, CSP and rLDA in combination with supervised adaptation by re-training of CSPP and rLDA. Run 5: fixed CSPP and CSP with unsupervised adaptation. After the first three runs there was a relax recording, not depicted here.

2.3.2. Settings The subject-independent CSPP filters and classifiers for the three class combinations were calculated as explained in section 2.2. After each run, the offline machine learning training was repeated to re-calculate CSP. Subject-dependent parameters, i.e. time interval, frequency band and patch form were eventually reselected as well. This was done in the case of very low performing users, where previous simulations showed that 80 or even 180 trials from the first runs were not enough to estimate all parameters properly. For runs 2 and 3, all previous trials were used (80 and 180 respectively), while for run 4 just trials from run 2 and 3 were used (200). CSP filters with 24 channels were calculated for run 2, with 48 channels for runs 3 to 5. For the last run with unsupervised adaptation, algorithms were trained using the data from runs 2-4. During the feedback, both CSP and CSPP stayed fixed and the classifier was adapted. The experimenter monitored that the subject did not produce artifacts during the runs, and gave reminders and instructions between the runs. The experimenter was sitting behind a desk situated on one side of the user, around 2 m away from the participant. Additionally, after each run a quality check of the previously recorded data was performed. It consisted of an artifact visualization and rejection procedure included in the BBCI toolbox (https://github.com/bbci/bbci_public). This method uses a heuristic based on the variance of the broadband filtered EEG signals. Trials and/or channels showing excessive variance are rejected, which is typically the case due to muscular artifacts (rejected trials) or problems with the EEG sensors (rejected channels).

3. Results

3.1. Analysis of online performance

Fig. 7 presents the online performance averaged across user categories. On the top, the data of the 17 users with previous BCI experience are used, on the bottom, the performance of the three BCI novice participants is depicted separately. The color convention is pink for level 1, orange for level 2, green for level 3 and each horizontal bar represents the mean value for one run, while each point is the mean value across 20 trials each. The Cat. III users are considered to have control if they reach an average performance of 70% in the last run. This criterion of control is represented as a dashed black line in Fig. 7 and has been selected according to the results of (Kübler et al., 2004)) for a two-class BCI. In every subgroup of people, the performance increased from run to run and most users reached the criterion threshold already in run 2, i.e. as soon as the features and parameters used for the feedback were subject-dependent. All non-naive users except for three reached this criterion level of control (top row in Fig. 7). For each user, binomial tests on the hit/miss of the last run were performed to compare to the accuracy reached in the first run. The difference was significant for all users of Cat. II and III with control (for Cat. II participants all $p < 0.002$ and for Cat. III with control $p < 0.04$). For the users of Cat. III without control, there was not a significant difference. The last three participants were novice and are depicted in the bottom row of Fig. 7. They are shown separately because they used another cursor control feedback, with fixed trial length. For all of them the performance significantly improved from first to last run. Again, a binomial test was performed for each user separately. The test was applied to the hits/misses of last run and compared to the accuracy reached in the first run ($p = 0.005$, $p = 0.003$ and $p < 0.001$ respectively). The third user achieved an averaged accuracy of 68% in the last run (that is, not significantly different from 70% using again a binomial test).

In Fig. 8, the performance obtained in adaptation level 2 (left) and 3 (center and right) is compared with the one obtained in previous studies by the same users. Each asterisk is one user and the performance of the new study is better for all points above the black diagonal line. In the legend, the methods used to give feedback in the previous studies and the runs used to extract the performance within the study are shown. Thus, the average over runs 2-4 in the CSPP study (adaptation by re-train) is compared against 1) in red, the average across runs 4-6 of feedback with CSP without online adaptation (performed after 3 calibration runs), 2) in light blue, the average across runs 4-6 of the first co-adaptive study, where feedback was provided by LAP+CSP features and 3) in black, the average over runs 2-4 of the second co-adaptive study, where feedback was provided by LAP features. In the center, the performance in the last run of CSPP is compared against the last run of feedback for the three previous studies (red for standard approach, green for the first co-adaptive study and black for the second co-adaptive study). Since one may argue that just one run with unsupervised adaptation is performed in CSPP study, immediately after the machine learning algorithm training,

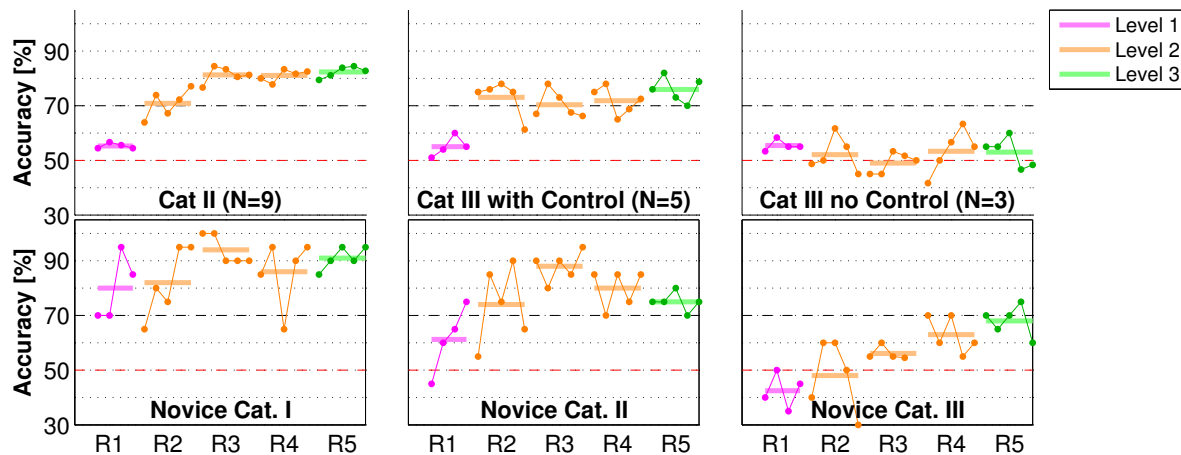


Figure 7. Online performance of the CSPP study for different user groups. Group and number of users are written on the bottom of each plot. Each bar is the mean of the accuracy in one run (from R1 to R5, indicated on the x-label). Each point is the mean across 20 trials. Pink, orange and green are used for the three adaptation levels. The black dashed line indicates the threshold criterion of 70%.

the comparison against the first run after the algorithm training for the other studies is also shown on the right of Fig. 8. Again, we performed Wilcoxon paired, one-tailed significance tests to find out whether the comparisons in Fig. 8 were significant. And this is indeed the case ($p=0.0015$, $p=0.0009$ and $p=0.0043$ respectively, see title of subplots in Fig. 8).

These results demonstrate that CSPP adapts faster and provides better feedback in a shorter time. Some users who could improve their performance from the standard to co-adaptive approach, performed in this study earlier well and better. One user who had difficulties in both standard and previous co-adaptive approach, learned with CSPP in the first run to modulate his SMR activity.

3.2. Analysis of the EEG at rest

The SMR-predictor was described in (Blankertz et al., 2010) and it allows predicting how likely a subject can achieve BCI control by SMR modulation. It was developed under the assumption that users with more prominent peaks in the spectrum at rest have a higher potentiality to suppress them. This predictor has been recently confirmed in a large-scale study (cf. Acqualagna et al. (2016)) and also by another group (Bauer et al. (2015)). To calculate the SMR-predictor, the difference between the PSD at rest and the estimated noise curve at a particular scalp location is estimated by modeling both the PSD and the noise curve. The maximum distance between the peaks in μ and β and the noise for each channel can be considered as the *SMR-strength* over that channel location. Accordingly, an increase of these peaks indicates an improvement on the ability to perform MI tasks. To investigate whether the users learnt to better control

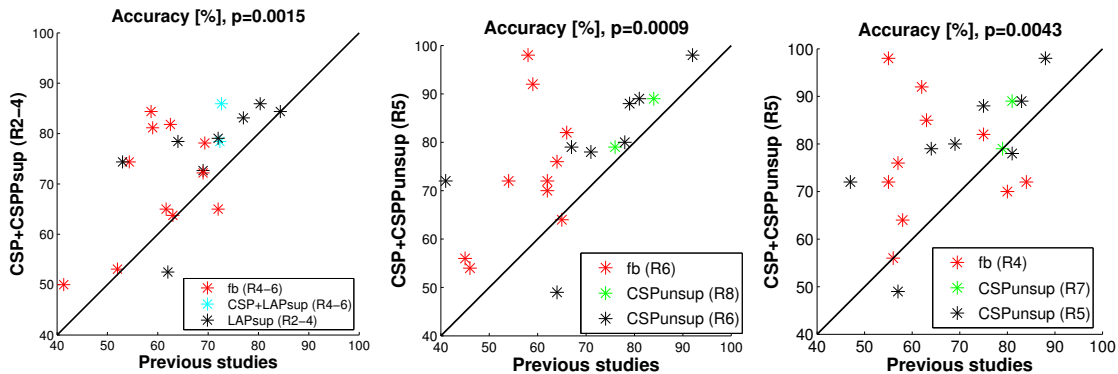


Figure 8. Comparison of performance obtained by CSPP study (y-axis) and previous studies (x-axis). Each point is a subject and the performance in the CSPP study is better for all points above the diagonal line. Left: CSPP+CSPP in the runs 2-4 with supervised adaptation. Center and Right: CSPP+CSPP in run 5 with unsupervised adaptation. Red: CSP feedback runs of study with standard approach (no adaptation). Light Blue: CSP+LAP with supervised adaptation in the first co-adaptive study. Green: CSP with unsupervised adaptation in the first co-adaptive study. Black: LAP in the second co-adaptive study (left panel) and CSP with unsupervised adaptation in the second co-adaptive study (center and right panels). The probability shown in the title was obtained using non-parametric, paired, one-sided tests.

their SMR modulation, the PSD of the relax recordings before and after the BCI session was modeled and the SMR-strength calculated. For that, the Laplacian channel with the highest r^2 -value in the last MI run was chosen. The r^2 -value is a measure of how much variance of the joint distribution can be explained by class membership. The results are depicted in Figure 9 for all participants.

Figure 9 shows the spectra of the relax recordings before (blue) and after (purple) the MI session with the corresponding noise fit (dashed lines) and SMR-predictor values.

4. Discussion

4.1. First run performance

Despite the good performance achieved, the experimental results of the first run could have been better. The cursor control selected for feedback (see Section 2.3.1) was heavily influenced by the initial bias of the classifier, which takes around 10 trials to vanish. The trials were abruptly interrupted and the users did not have enough time to concentrate on the task, which further influenced the subsequent trials. In this first run, the experienced 17 users could not reach the minimum control threshold of 70% accuracy (the maximum performance reached by one user was 66%) even if some of them could reach up to 80% accuracy in single blocks. Due to this problem the last three (novice) participants used another cursor control feedback. In contrast to the previous results, two out of the three showed a clear learning curve reaching respectively 95% and 75% performance peaks already in the first run (bottom of Fig. 7, left and middle panels). Even the third user (bottom of Fig. 7, right panel) who belonged to Cat. III, showed a continuous

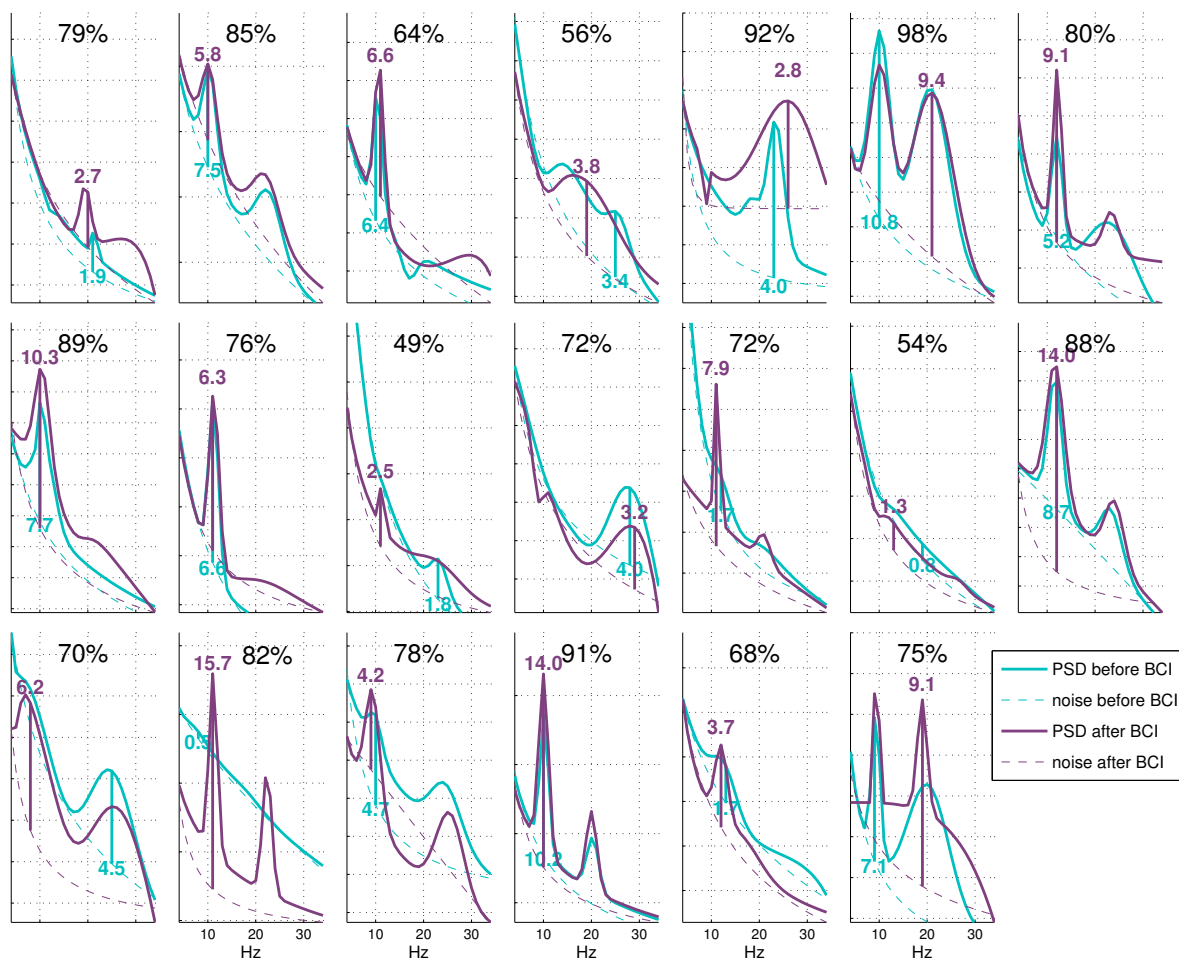


Figure 9. SMR-strength before and after the BCI session and the user's accuracy in the last run (in percentage).

performance improvement and was able to reach 75% accuracy in one block of the last run.

4.2. Session performance

In section 3.1 the binomial significance tests performed for each participant showed that the improvement from run 1 to run 5 was significant for the users in Categories II and III with control and also for the three naive participants, whereas the difference was not significant for the users in Category III without control. These results corroborate the subgrouping performed according to Table 1. Additionally, binomial significance tests individually computed for each participant (on the hit/miss of level 3 in comparison to the accuracy of level 2) showed that the performance achieved by the users in level 2 was maintained (10 users) or improved (4 users and 2 in the limit of significance $p < 0.06$) during level 3, except for one user of Cat. II. For the naive participants this difference was not significant for the users of Cat. I and II, and the participant in Cat. III significantly improved in the last run ($p = 0.007$). This shows that the unsupervised

adaptation does not hinder the results achieved during (the supervised) level 2 and that the users could reach a stable performance.

Regarding the comparison of this session to previous studies, Figure 8 shows that the design presented in this manuscript significantly outperforms all previous attempts (adaptive and non-adaptive). In particular, the performance of users in runs 2-4 (retrain) of the current experiment was compared to previous performance in runs 4-6 (retrain with changing Laplacian channels and fixed CSP or non-adaptive CSP feedback) and runs 2-4 (retrain with changing Laplacian channels) depending on the study. This shows that feedback with CSPP allows users to achieve significantly better performance in shorter time (with the need of less runs). A similar analysis performed for run 5 (unsupervised adaptation with CSPP and CSP filters fixed) of the current design and runs 6 and 8 of previous ones (last non-adaptive feedback run and last unsupervised adaptation with CSP fixed filters) allows us to reach the same conclusion. Finally, it could be argued that better performance might be reached when BCI sessions are shorter, like the one presented here, because the participants might not be so tired. A further comparison taking earlier runs of the previous studies (runs 4, 7 and 5, corresponding to first feedback non-adaptive run, and first unsupervised adaptation runs of previous adaptive studies) was also performed. The conclusions are once again the same, the present study significantly outperforms those.

4.3. Spectrum analysis

The analysis of the Power Spectral Density (PSD) during the runs allows us also to evaluate the progress of the users regarding their ability to produce and suppress SMR, the necessary condition to obtain BCI control with the system described in this manuscript. In comparison to the previous studies, and despite the suboptimal feedback given in the first run, the appearance of a weak or strong SMR modulation could be observed already in the first run, except for two of the three users who did not reach the BCI control. Here, four particular cases are reported.

In Fig. 10, the spectra of one user who previously participated in the study with standard approach and in the second co-adaptive study are visualized. The discriminability of the two classes is visualized in form of r^2 -values (color bars on the x-axis of each plot). For each of the previous studies, the channel with the strongest r^2 -value between the class membership and the band power was calculated using the last run and the reactive frequency band and time interval used to give feedback in the same run. Then the PSD of that channel, averaged across trials belonging to the same class, is plotted for each run of the study from left to right. At the bottom of each plot, a bar indicates the r^2 -values for each frequency, the scale of which is on the right of the figure (same colorbar for all runs of the same study). The three rows of Fig. 10 depict the sequence of spectra respectively for the standard approach study, the second co-adaptive study, and the CSPP study. The same y-axis range is used for the three studies. The date of the experiment is also reported in the title.

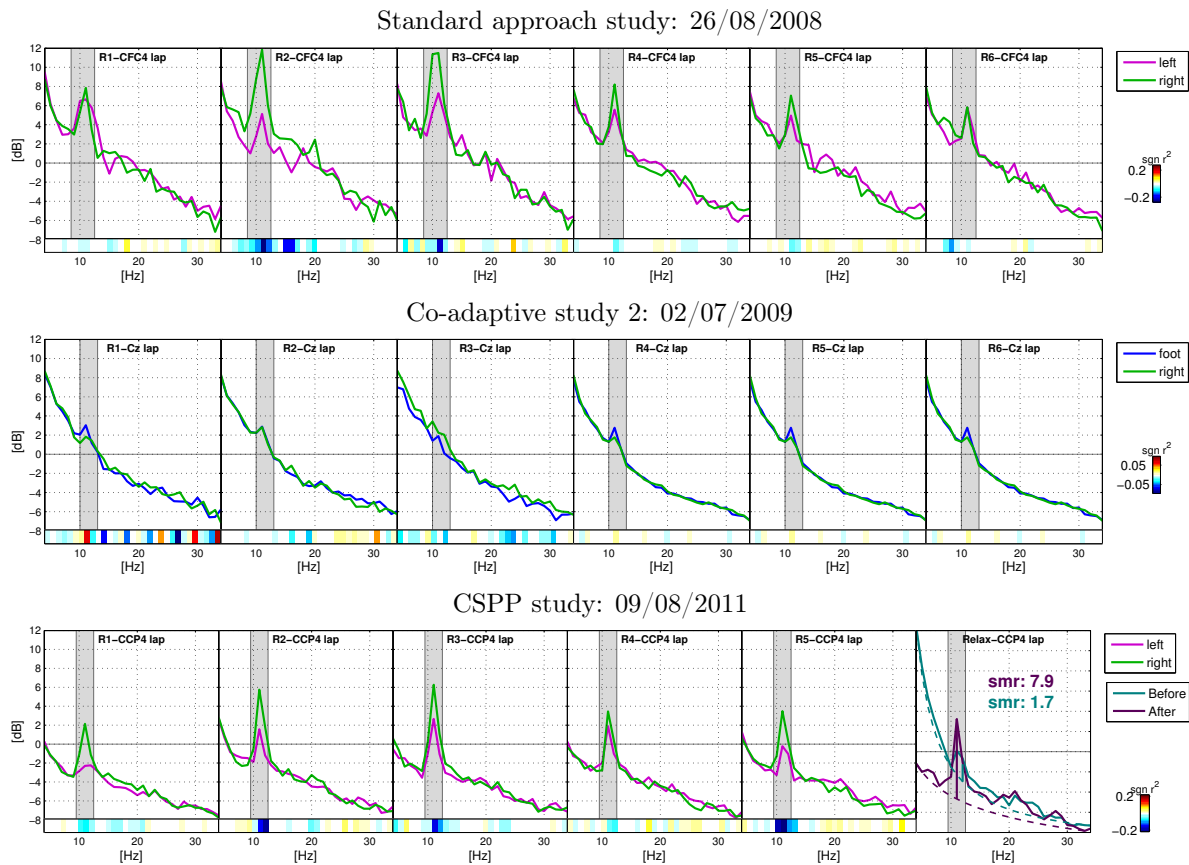


Figure 10. PSD of one representative channel selected on the data of the last run and averaged for the two classes. The frequency band used to give feedback in the last run is marked in gray. From top to bottom, three studies in which the user participated. Type of study and date of recording is reported in the title. From left to right, the PSD of each run. For the CSPP study, the last plot on the right depicts the PSD of the relax recording before (blue) and after (purple) the MI session. SMR-predictor values are written in the corresponding colors. This user exhibits a PSD at rest with a peak in μ .

In order to complete the discussion, we again provide SMR-predictor values of these users in the last subplot of the CSPP study (bottom right of Fig. 10). The plots are slightly different than in Figure 9, where they were smoothed to reduce noise, but the resulting SMR-strength values are the same. The reactive frequency band (which might be different for each study as well as the representative channel) is marked in gray in all plots. In the standard approach study, this user exhibited a proper SMR modulation at 10 Hz during the second and third run of calibration session, but he was not able to deal with the feedback. He was expected to obtain very soon BCI control using a co-adaptive BCI, but the system did not guide the user fast enough to find the proper strategy until the last runs. In the CSPP study, starting with PSD at rest with a small peak at 10 Hz, this user was able to modulate the SMR activity already in the first run. The relax recording prior to the MI session exhibits a very small (SMR-predictor value 1.7) but clearly visible peak in the μ -band. The relax recording performed after the BCI session exhibited a much higher peak and consequently higher SMR-predictor value (7.9).

In Fig. 11 the spectra of one user with almost flat PSD at rest are presented. It is

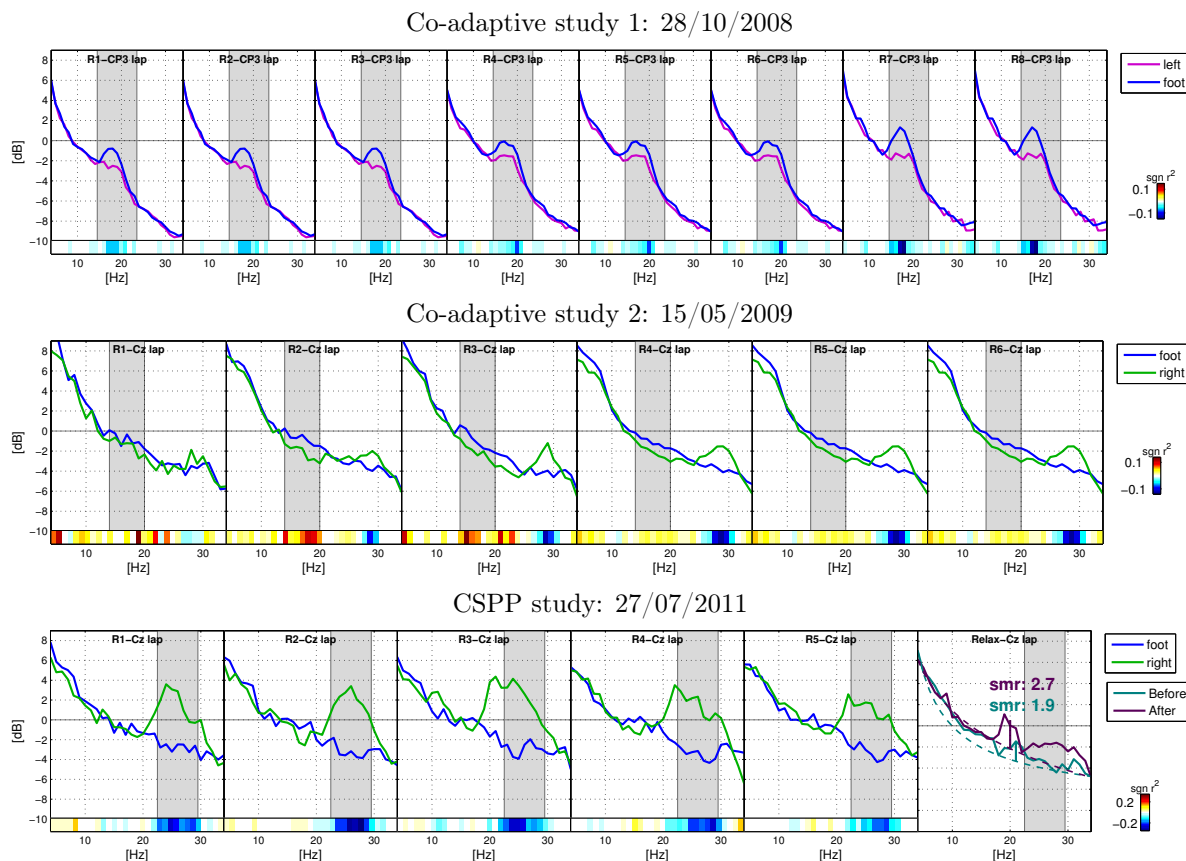


Figure 11. PSD as described in Fig. 10 for a second user with almost flat PSD at rest.

evident that in the previous studies, the user learned from run to run how to modulate his SMR activity and found a MI strategy that allowed him at some point to reach control. In the CSPP study, a strong ERD (Event Related Desynchronization) in Cz for the foot imagination was visible in the first run. It allowed him to obtain a good BCI control already in the second run. This seems not to depend on a better starting state, since the PSD at rest is flat (as flat was also the PSD at rest in the other studies, not shown here). After the MI session, a small peak appeared around 20 Hz and another “hump” in the reactive frequency band. While in the first co-adaptive study the modulation appeared in a centro-parietal area, in the second co-adaptive study and in the CSPP study a foot pattern centered in Cz was visible. Reactive frequency bands also change, but it is clear that the earlier appearance of a peak in the late β band in the CSPP study allows performing a better band selection, what was not possible in the second study, where the band selection was done only after run 1 whereas the SMR modulation in the late β band appears more clearly only in run 4. Finally, Fig. 12 shows the spectra of one user whose performance remained stable without any improvement in both co-adaptive and CSPP study. The spectra at rest is completely flat. Still, it can be observed that in the CSPP study a small peak around 20 Hz appears already after the first run, even if there is not desynchronization between the two classes. In run 4 a second small peak at 13 Hz is visible. In the last run this user started to modulate the SMR activity although not

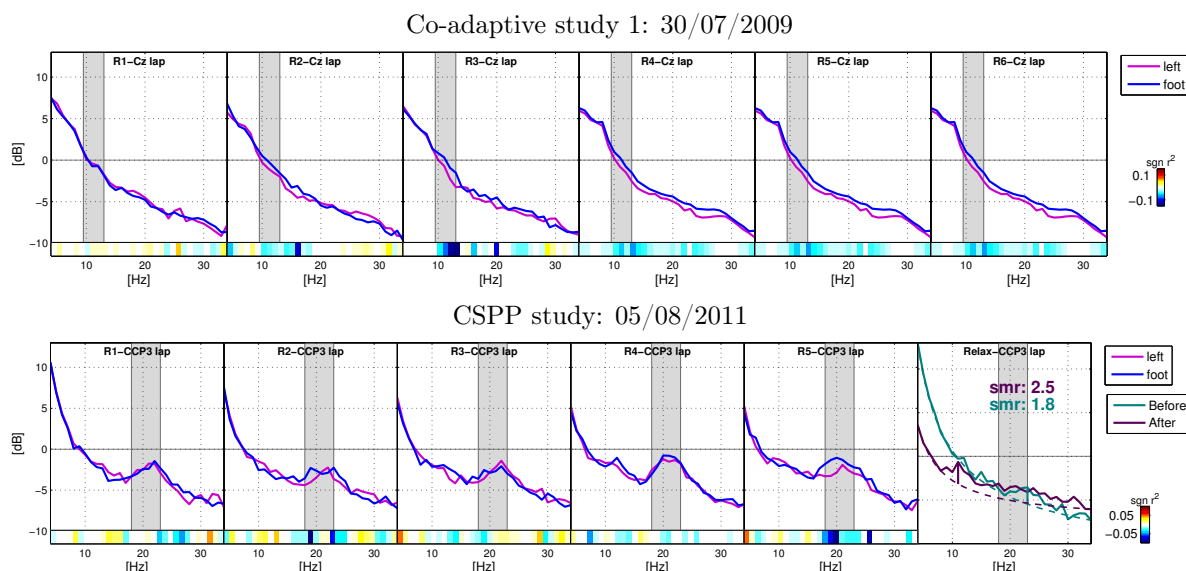


Figure 12. PSD as described in Fig. 10 for a third user with completely flat PSD at rest.

strong enough to obtain BCI control. The PSD at rest is flat also after the MI session.

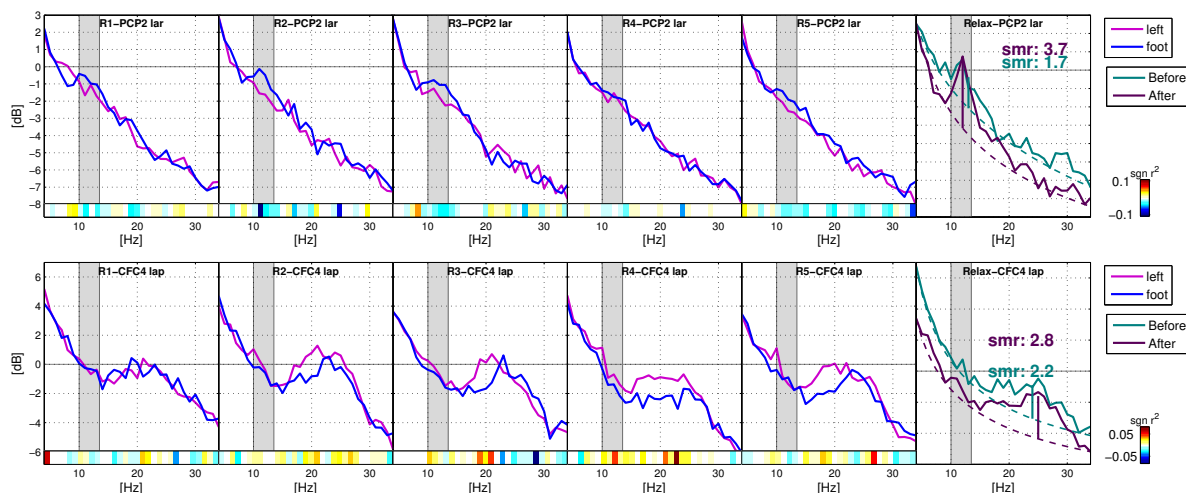


Figure 13. PSD as described in Fig. 10 for a BCI novice of Cat. III for two different channels.

In Fig. 13, the PSD of the BCI novice of Cat. III who reached the threshold of 70% in the last run is presented (performance is depicted in Fig. 7, bottom right panel). On the top panel, the channel PCP2, on the bottom panel the channel CFC4 are visualized. This user found the task extremely difficult and had the tendency to visualize the movement instead of feeling it kinesthetically. Based on run 1, the frequency band 25-29 Hz was selected, but due to the successive poor performance in run 2, the band was changed in run 3 and μ was preferred. As we can see from the figure, in both bands a desynchronization between the two classes appears, but it is difficult to say which one better represents the SMR activity, since the position and the band of PCP2 let us think of visual imagination and the direction of the ERD in CFC4 is actually opposite to the

expected one (desynchronization for the class *Left*). Nevertheless, an increase of the SMR-predictor can be observed and the BCI performance improved from run to run. Probably, the analysis of the SMR at rest could have helped to choose the frequency band.

On the left panel of Fig. 14, a comparison between the AUC (roc score, in the title) obtained by the same CSPP feature in run 1 (x-axis) and in run 5 (y-axis) is shown. In green, the only Cat. I user, who was also a BCI novice, is indicated. The CSPP feature was chosen (depending on the highest score) among those calculated on runs 2-4 to give feedback in the unsupervised run 5. Blue is used for Cat. II, red for Cat. III and an additional red circle for the Cat. III users who did not show any BCI performance improvement. This plot shows that for most of users, the SMR activity captured by this feature was lower in run 1 and increased during the BCI session. A Wilcoxon paired one-tailed significance test confirmed that the score in run 5 is significantly higher than in run 1 ($p=0.006$).

Finally, on the right panel of Fig. 14 the SMR-predictor values calculated on the subject-depending representative channel for the relax recordings before and after the MI session are compared. The improvement is higher for Cat. II users, whereas Cat. III users SMR-indicator stays almost the same (see Blankertz et al. (2010); Suk et al. (2014)). However, the SMR-strength values of some Cat. III users for whom the EEG curve is flat or almost flat might not be very reliable. This is because in these situations the model fitting of the SMR-indicator might be inaccurate. As well, a small improvement of the SMR-indicator might also mean that a performance improvement is due to the algorithm adaptation, the above mentioned flexibility of the system and the design. Nevertheless, it is possible to assess that the SMR-indicator values after the MI session became higher for most users. The group analysis shows that this increase is significant (non-parametric, paired, one-tailed test, $p=0.015$).

4.4. Evolution of CSPP locations over time

Figure 15 shows the patch centers chosen for each of the users discussed in the previous section and the location of the main feature changing over time in the middle runs of the experiment (runs 2 to 4). The small windows show in detail where and when the most discriminative feature was chosen. The x-axis is the trial number, up to 300 trials, with 100 trials per run in chronological order. The y-axis is the frequency of selection of each selected channel. In order to ease the visualization, a few channels with spurious selections were not plotted. The two scalp plots on the top row correspond to the first two users discussed in Section 4.3. The scalp plots on the bottom row correspond to the last two users. The first user performed left versus right hand MI. The plot on the top left corner of Figure 15 shows that the main feature appears in both hemispheres, changing from the left one in runs 2 and 3 to the right one in run 4. The second user executed MI of right hand versus feet. As one can see in the top right corner of Figure

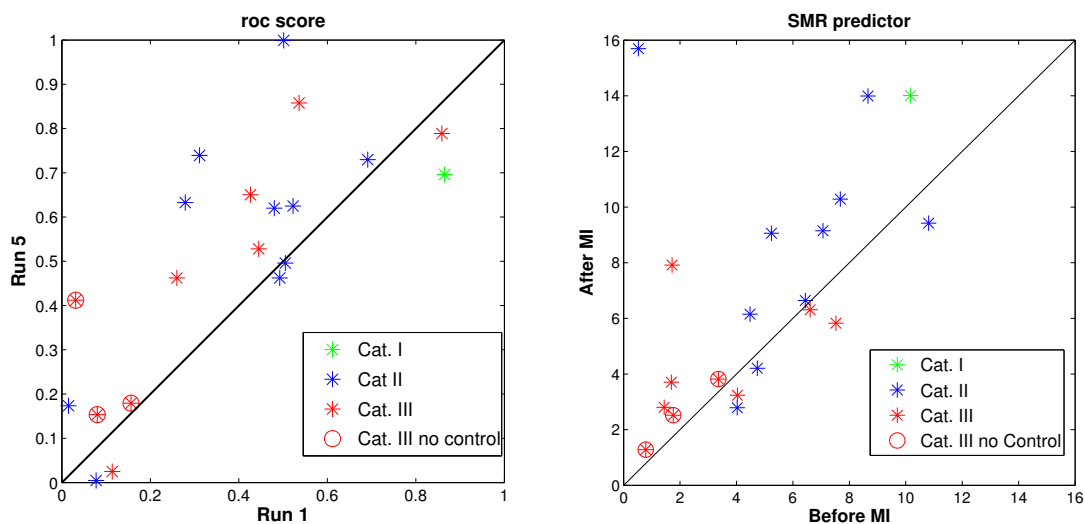


Figure 14. Left: Comparison of the ROC scores obtained by the same CSPP feature on data of run 1 (x-axis) and run 5 (y-axis). The CSPP feature was calculated from the data of runs 2-4 and used, between others, to give feedback in run 5. A different color is used for each user category. Right: Comparison between SMR-predictor values before (x-axis) and after (y-axis) the MI session.

15, the selections start in the central area and over time they shift to the left hemisphere. However, C3 is selected consistently over the three runs. The third user (bottom left) imagined the movement of the left hand versus feet but did not obtain BCI control. The locations in the scalp mainly correspond to the left hemisphere, the contrary to what would be expected. But in run 4, the main feature is most often chosen over the right hemisphere. This is in line with the findings shown in Section 4.3. Finally, the naive user (bottom right) also performed the imagination of the left hand versus the feet. The location of the most discriminative feature changes often in time but is always located over the central area and right hemisphere, as expected from the MI tasks. A slight trend is observed from the central area in the first two runs to the right hemisphere in the last one.

5. Conclusion

The CSPP was presented in Sannelli et al. (2011) as a new spatial filter optimized to reduce the BCI inefficiency. In the offline evaluation, it was shown that in a classical BCI machine learning approach with training and test set, CSPP, unlike CSP, only needs very few data to be tuned. In comparison with previous approaches to the problem of small training sets (Krauledat et al., 2008; Ang et al., 2008; Kayagil et al., 2009; Kang et al., 2009; Lu et al., 2009), which use 128 channels, this study evaluates also the performance of the methods depending on the number of channels and trials, giving a hint about the minimum number of channels necessary to obtain a robust classification result. Moreover, a large data base of 80 users with feedback performances ranging from chance level to 100% was used for the validation. Results show that CSPP always

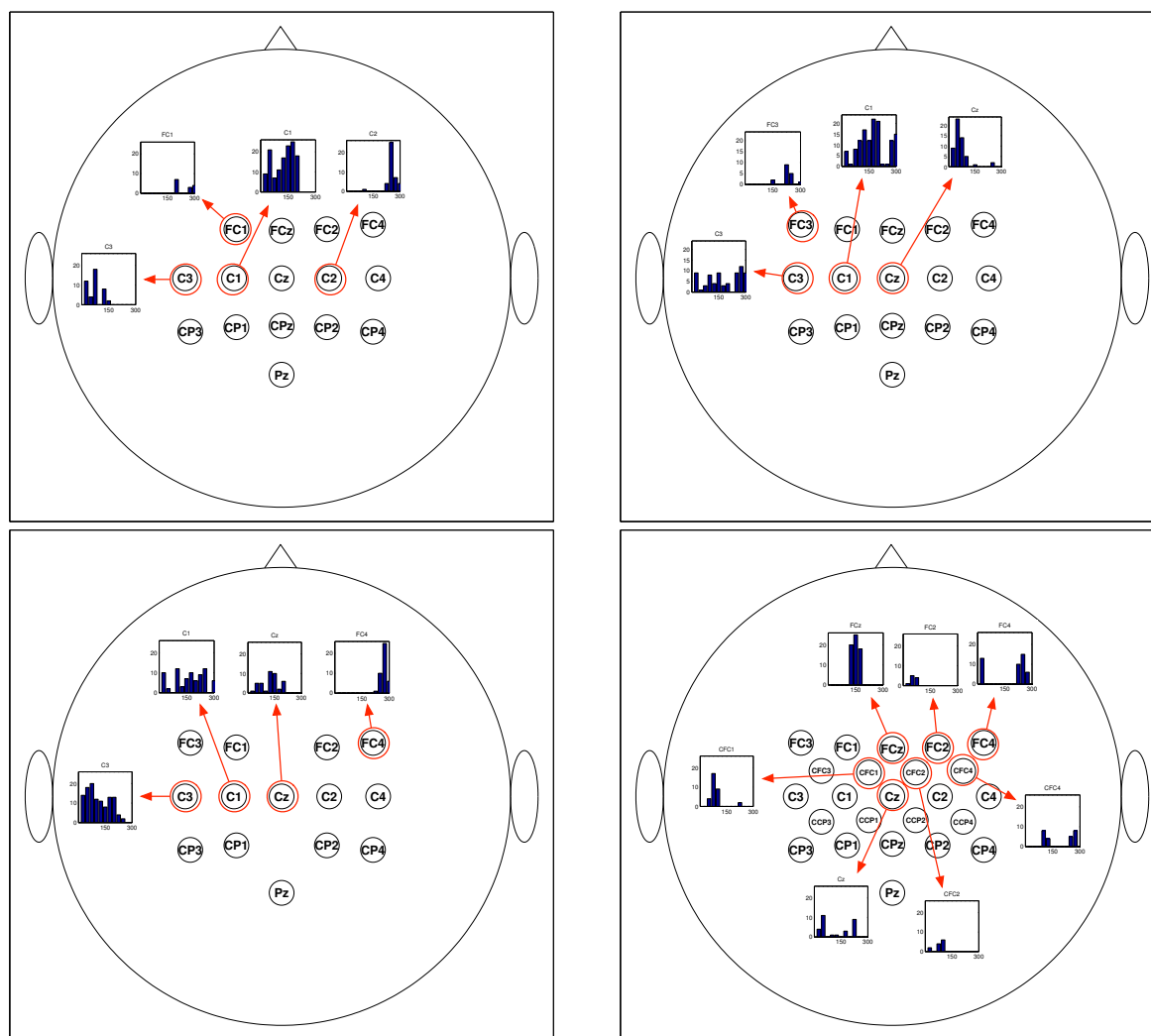


Figure 15. Scalp plot showing the location of the most important feature over runs 2 to 4 for the users presented in Section 4.3. Each plot also depicts the patch centers chosen for each user. The windows plot the frequency of selection of a channel versus the trial number.

outperforms Laplacian filters, at the little expense of recording about 2 minutes of data. In larger multi-channel recordings, it also outperforms CSP and R-CSP which can be considered the state-of-the-art methods.

Here, CSPP is assessed in an online study with 20 participants. Seventeen were selected from Cat. II and III, because they had difficulties or were not able to reach BCI control in previous studies. The other three were BCI novice, one of them resulted to be a Cat. I user, one was Cat. II and the other Cat. III. The first 17 participants could improve their performance in comparison to previous studies, except for three who, as in the previous studies, were not able to reach the threshold criterion of control (70% performance in a two-class BCI). In general, participants could obtain BCI control earlier and maintain it until the last run with unsupervised adaptation.

Moreover, the only Cat. I user started already with 70% accuracy in the first 20 trials. In comparison, the average initial performance of good performing users (already BCI

experienced) in Vidaurre et al. (2011c) was 60% and BCI novice (and good performing) participants in Vidaurre et al. (2011b) did not perform on average so well in the first run (first block accuracy was 62%, first run accuracy was 73% against 80% in this study). Neither in Vidaurre et al. (2011c) nor in Vidaurre et al. (2011b) Cat. II users were able to reach 70% accuracy in the first run as in this study happened for the Cat. II novice user. This good start was not shown by the other Cat. II participants, however, this was unfortunately the effect of a bad the feedback selection in the first run. Finally, the clear learning effect visible in the performance of the third BCI novice user was shown before for a BCI novice with no peak in the μ band over the motor areas at rest. It is important to note that this performance increase of the third user is probably due to the continuous adaptation of the algorithms, since the SMR improvement cannot be considered enough to explain it.

Considering that 19 of these participants belonged to the 40% of users (17.5% Cat. II and 22.5% Cat. III) who could not reach BCI control using the standard non-adaptive approach (i.e. calibration and feedback runs), one can estimate that the CSPP can reduce BCI inefficiency from 40% to only 6%. Apart from this strong decrease, one important observation should be added about the rhythmic activity of these users. The analysis of the PSD in Section 4.3 and the comparison with previous studies and with the PSD of the relax recording reveals that users with strong difficulties have in common the complete absence of the μ -rhythm at rest. They might have irregular rhythmic activity which produces on average a small “hump” (i.e., not a peak) in the β band, or exhibit a completely flat spectrum. Some users exhibit a small but visible peak in the μ -rhythm. Based on the experience collected in this study, they can achieve good performance by finding a new MI strategy, improving their concentration and receiving stable feedback. With the standard approach, this was not possible because of the absence of feedback in the first runs (calibration runs). In the co-adaptive approach with Laplacian channels, this became possible for most users and in the CSPP study, an improvement in the features led to better feedback and allowed the participants to find a successful MI strategy earlier. The users without μ -rhythm, are strongly helped by the co-adaptive approach to develop new rhythms and the CSPP and the previous co-adaptive studies show that this is possible within one session. These new rhythms are not around 12 Hz, but more often in the β band and generally change from run to run before becoming stable, probably because they are reinforced by proper feedback. More interestingly, one of these users with flat PSD at rest, who developed a foot pattern in the last runs of the second co-adaptive study, exhibited again the same pattern (but with stronger modulation) in the same frequency band (late β) in the CSPP study two years later and already in the first run. This phenomenon requires further experiments and analysis to understand whether it is primarily caused by the CSPP feedback, but it anyway demonstrates that also these users can achieve stable BCI control after a few co-adaptive BCI sessions and that points to the fact that MI feedback could have high capacity to induce neuroplasticity Vidaurre et al. (2011c). This has been also recently observed in a study with patients after stroke (Varkuti et al., 2012), where the employment of MI in the

upper-extremity rehabilitation treatment produced better results than a motor-assisted rehabilitation. In the same study, the functional connectivity at rest was measured by fMRI before and after the treatment (which lasts 12 weeks) and correlation with the intra-subject improvement variability was found.

A future direction for further improvement is the online detection of artifactual trials in order to adapt the algorithms only on artifact-free trials. This is a general problem of any co-adaptive system that may be resolved by artifact trial rejection or robust CSP methods Samek et al. (2014, 2012).

A second, but not less important point, is the better exploration of the EEG at rest, maybe by long recordings in a separate session (von Bünaeu et al., 2009). This would allow understanding how stationary the EEG rhythmic activity is and to select in advance the reactive frequency band. The BCI session can then start providing subject specific feedback in the stable (respectively stationary) band and, in this way, facilitate the modulation of the SMR in that band. Finally, multiple sessions for the same user and the comparison with long recordings of EEG at rest would allow analyzing the learning (plasticity) phenomena that happen during co-adaptive training and/or across years between BCI sessions and to further understand differences among several users.

Acknowledgments

The work of Claudia Sannelli, Carmen Vidaurre and Klaus-Robert Müller was funded by was funded by the German Ministry for Education and Research (BMBF) under Grant 01IS14013A-E and Grant 01GQ1115, as well as by the Deutsche Forschungsgesellschaft (DFG) under Grant MU 987/19-1, MU987/14-1 and DFG MU 987/3-2. Additionally, the work of Klaus-Robert Müller was funded by the Brain Korea 21 Plus Program. The work of Benjamin Blankertz was funded by the BMBF contract 01GQ0850. Correspondence to KRM and BB.

References

- Acqualagna, L., Botrel, L., Vidaurre, C., Kübler, A., Blankertz, B., 2016. Large-scale assessment of a fully automatic co-adaptive motor imagery-based brain computer interface. *PLoS ONE*, Public Library of Science 11 (2), 1–19.
- Ang, K. K., Chin, Z. Y., Zhang, H., Guan, C., 2008. Filter bank common spatial pattern (FBCSP) in Brain-Computer interface. In: *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence)*. IEEE International Joint Conference on. pp. 2390–2397.
- Ang, K. K., Guan, C., 2015. Brain-computer interface for neurorehabilitation of upper limb after stroke. *Proceedings of the IEEE* 103, 944–953.
- Arvaneh, M., Guan, C., Ang, K. K., Quek, C., 2013. Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain-computer interface. *Neural Networks and Learning Systems, IEEE Transactions on* 24, 610–619.

- Babiloni, F., Cincotti, F., Bianchi, L., Pirri, G., del R. Millán, J., Mouriño, J., Salinari, S., Marciani, M. G., 2001. Recognition of imagined hand movements with low resolution surface laplacian and linear classifiers. *Medical Engineering & Physics* 23 (5), 323–328.
- Bauer, R., Fels, M., Vukelic, M., Ziemann, U., Gharabaghi, A., 2015. Bridging the gap between motor imagery and motor execution with a brainDrobot interface. *Neuroimage* 108, 319–327.
- Blankertz, B., Lemm, S., Treder, M., Haufe, S., Müller, K.-R., 2011. Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56 (2), 814–825.
- Blankertz, B., Sannelli, C., Halder, S., Hammer, E. M., Kübler, A., Müller, K.-R., Curio, G., Dickhaus, T., 2010. Neurophysiological predictor of SMR-based BCI performance. *NeuroImage* 51 (4), 1303–1309.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Müller, K., 2008. Optimizing spatial filters for robust EEG Single-Trial analysis. *Signal Processing Magazine, IEEE* 25 (1), 41–56.
- Faller, J., Scherer, R., Costa, U., Opisso, E., Medina, J., Müller-Putz, G., 2014. A co-adaptive brain-computer interface for end users with severe motor impairment. *PLoS One* 9, e101168.
- Faller, J., Vidaurre, C., Solis-Escalante, T., Neuper, C., Scherer, R., 2012. Autocalibration and recurrent adaptation: towards a plug and play online erd-bci. *EEE Trans Neural Syst Rehabil Eng* 20, 313–319.
- Fazli, S., Popescu, F., Danóczy, M., Blankertz, B., Müller, K.-R., Grozea, C., 2009. Subject-independent mental state classification in single trials. *Neural Networks* 22 (9), 1305–1312.
- He, B., Lian, J., Li, G., 2001. High-resolution EEG: a new realistic geometry spline laplacian estimation technique. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* 112 (5), 845–852.
- Kang, H., Nam, Y., Choi, S., 2009. Composite common spatial pattern for Subject-to-Subject transfer. *Signal Processing Letters, IEEE* 16 (8), 683–686.
- Kayagil, T. A., Bai, O., Henriquez, C. S., Lin, P., Furlani, S. J., Vorbach, S., Hallett, M., 2009. A binary method for simple and accurate two-dimensional cursor control from EEG with minimal subject training. *Journal of NeuroEngineering and Rehabilitation* 6 (1), 14.
- Kindermans, P., Tangermann, M., Müller, K. R., Schrauwen, B., 2014. Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training erp speller. *Journal of neural engineering* 11 (3), 035005.
- Krauledat, M., Tangermann, M., Blankertz, B., Müller, K.-R., 2008. Towards zero training for Brain-Computer interfacing. *PLoS ONE* 3 (8), e2967.
- Kübler, A., Neumann, N., Wilhelm, B., Hinterberger, T., Birbaumer, N., 2004. Predictability of brain-computer communication. *Journal of Psychophysiology* 18 (2-3), 121–129.

- Lemm, S., Blankertz, B., Dickhaus, T., Müller, K.-R., May 2011. Introduction to machine learning for brain imaging. *NeuroImage* 56 (2), 387–399.
- Lotte, F., Guan, C., 2010. Learning from other subjects helps reducing Brain-Computer interface calibration time. In: *Acoustics Speech and Signal Processing (ICASSP)*, 2010 IEEE International Conference on. pp. 614–617.
- Lotte, F., Guan, C., Ang, K. K., 2009. Comparison of designs towards a subject-independent brain-computer interface based on motor imagery. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. pp. 4543–4546.
- Lu, H., Eng, H., Guan, C., Plataniotis, K. N., Venetsanopoulos, A. N., 2010. Regularized common spatial pattern with aggregation for EEG classification in small-sample setting. *IEEE Transactions on Bio-Medical Engineering* 57 (12), 2936–2946.
- Lu, H., Plataniotis, K., Venetsanopoulos, A., 2009. Regularized common spatial patterns with generic learning for EEG signal classification. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. pp. 6599–6602.
- McFarland, D. J., McCane, L. M., David, S. V., Wolpaw, J. R., 1997. Spatial filter selection for EEG-based communication. *Electroencephalography and Clinical Neurophysiology* 103 (3), 386–394.
- Pfurtscheller, G., Linortner, P., Winkler, R., Korisek, G., Müller-Putz, G., 2009. Discrimination of motor Imagery-Induced EEG patterns in patients with complete spinal cord injury. *Computational Intelligence and Neuroscience* 2009, 1–6.
- Ramoser, H., Müller-Gerking, J., Pfurtscheller, G., 2000. Optimal spatial filtering of single trial EEG during imagined hand movement. *Rehabilitation Engineering, IEEE Transactions on* 8 (4), 441–446.
- Samek, W., Kawanabe, M., Müller, K.-R., 2014. Divergence-based framework for common spatial patterns algorithms. *Biomedical Engineering, IEEE Reviews in* 7, 50–72.
URL <http://dx.doi.org/10.1109/RBME.2013.2290621>
- Samek, W., Vidaurre, C., Müller, K.-R., Kawanabe, M., 2012. Stationary common spatial patterns for brain-computer interfacing. *Journal of Neural Engineering* 9 (2), 026013.
- Sannelli, C., Dickhaus, T., Halder, S., Hammer, E.-M., Müller, K.-R., Blankertz, B., 2010. On optimal channel configurations for SMR-based Brain-Computer interfaces. *Brain Topography* 23 (2), 186–193.
- Sannelli, C., Vidaurre, C., Müller, K.-R., Blankertz, B., 2011. CSP patches: an ensemble of optimized spatial filters. an evaluation study. *Journal of Neural Engineering* 8 (2), 025012.
- Solis-Escalante, T., Müller-Putz, G., Pfurtscheller, G., 2008. Overt foot movement detection in one single laplacian EEG derivation. *Journal of Neuroscience Methods* 175 (1), 148–153.

- Suk, H.-I., Fazli, S., Mehnert, J., Müller, K.-R., Lee, S.-W., 2014. Predicting bci subject performance using probabilistic spatio-temporal filters. *PLoS ONE* 9, e87056.
- Varkuti, B., Guan, C., Pan, Y., Phua, K. S., Ang, K. K., Kuah, C. W. K., Chua, K., Ti Ang, B., Birbaumer, N., Sitaram, R., May 2012. Resting state changes in functional connectivity correlate with movement recovery for BCI and robot-assisted upper-extremity training after stroke. *Neurorehabilitation and Neural Repair*.
- Vidaurre, C., Kawanabe, M., von Bünau, P., Blankertz, B., Müller, K. R., 2011a. Toward unsupervised adaptation of LDA for brain-computer interfaces. *IEEE Transactions on Bio-Medical Engineering* 58 (3), 587–597.
- Vidaurre, C., Krämer, N., Blankertz, B., Schlögl, A., 2009. Time domain parameters as a feature for EEG-based Brain-Computer interfaces. *Neural Networks* 22 (9), 1313–1319.
- Vidaurre, C., Sannelli, C., Müller, K.-R., Blankertz, B., 2011b. Co-adaptive calibration to improve BCI efficiency. *Journal of Neural Engineering* 8 (2), 025009.
- Vidaurre, C., Sannelli, C., Müller, K.-R., Blankertz, B., 2011c. Machine-learning-based coadaptive calibration for brain-computer interfaces. *Neural Computation* 23 (3), 791–816.
- Vidaurre, C., Scherer, R., Cabeza, R., Schlögl, A., Pfurtscheller, G., 2007. Study of discriminant analysis applied to motor imagery bipolar data. *Med Bio Eng Comput* 45, 61–68.
- Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R., Pfurtscheller, G., 2006. A fully on-line adaptive BCI. *Biomedical Engineering, IEEE Transactions on* 53 (6), 1214–1219.
- von Bünau, P., Meinecke, F. C., Király, F., Müller, K.-R., 2009. Finding stationary subspaces in multivariate time series. *Physical Review Letters* 103, 214101.