

Machine-Learning-Based Coadaptive Calibration for Brain-Computer Interfaces

Carmen Vidaurre

carmen.vidaurre@tu-berlin.de

Claudia Sannelli

claudia.sannelli@tu-berlin.de

*Machine Learning Department, Berlin Institute of Technology,
Berlin 10587, Germany*

Klaus-Robert Müller

klaus-robert.mueller@tu-berlin.de

*Machine Learning Department, Berlin Institute of Technology, Berlin 10587,
Germany, and Bernstein Focus: Neurotechnology, Berlin 10115, Germany*

Benjamin Blankertz

benjamin.blankertz@tu-berlin.de

*Machine Learning Department, Berlin Institute of Technology, Berlin 10587,
Germany; Bernstein Focus: Neurotechnology, Berlin 10115, Germany; and
Fraunhofer FIRST (IDA), Berlin 12489, Germany*

Brain-computer interfaces (BCIs) allow users to control a computer application by brain activity as acquired (e.g., by EEG). In our classic machine learning approach to BCIs, the participants undertake a calibration measurement without feedback to acquire data to train the BCI system. After the training, the user can control a BCI and improve the operation through some type of feedback. However, not all BCI users are able to perform sufficiently well during feedback operation. In fact, a non-negligible portion of participants (estimated 15%–30%) cannot control the system (a BCI illiteracy problem, generic to all motor-imagery-based BCIs). We hypothesize that one main difficulty for a BCI user is the transition from offline calibration to online feedback. In this work, we investigate adaptive machine learning methods to eliminate offline calibration and analyze the performance of 11 volunteers in a BCI based on the modulation of sensorimotor rhythms. We present an adaptation scheme that individually guides the user. It starts with a subject-independent classifier that evolves to a subject-optimized state-of-the-art

Color versions of figures in this letter are presented in the online version, available at http://www.mitpressjournals.org/doi/abs/10.1162/NECO_a.00089.

classifier within one session while the user interacts continuously. These initial runs use supervised techniques for robust coadaptive learning of user and machine. Subsequent runs use unsupervised adaptation to track the features' drift during the session and provide an unbiased measure of BCI performance. Using this approach, without any offline calibration, six users, including one novice, obtained good performance after 3 to 6 minutes of adaptation. More important, this novel guided learning also allows participants with BCI illiteracy to gain significant control with the BCI in less than 60 minutes. In addition, one volunteer without sensorimotor idle rhythm peak at the beginning of the BCI experiment developed it during the course of the session and used voluntary modulation of its amplitude to control the feedback application.

1 Introduction

Brain-computer Interface (BCI) systems aim to provide users control over a computer application by means of their brain activity. Efforts toward this goal have been reported over more than 30 years (for a summary of EEG-based BCI, see Blankertz, Tangermann et al., 2010; Dornhege, Millán, Hinterberger, McFarland, & Müller, 2007; Allison, Wolpaw, & Wolpaw, 2007; Birbaumer et al., 2006; Millán, Renkens, Mouriño, & Gerstner, 2004; Pfurtscheller, Neuper, & Birbaumer, 2005; Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002; Kübler, Kotchoubey, Kaiser, Wolpaw, & Birbaumer, 2001). Refer to Leuthardt, Schalk, Wolpaw, Jemann, and Oran (2004), Pistohl, Ball, Schulze-Bonhage, Aertsen, and Mehring (2008), Schalk et al. (2008), and Leuthardt, Freudenberger, Bundy, and Roland (2009) for ECoG-based BCIs and to Carmena et al. (2003), Schwartz (2004), Hochberg et al. (2006), Fetz (2007), Rizk et al. (2009), and Waldert et al. (2009) for a summary in spike-based approaches. However, independent of the system tested, a significant percentage of users cannot gain control over BCI systems. Hence, in BCI research and specifically in EEG-based BCIs, one of the biggest research challenges is to understand and solve the problem of BCI illiteracy, which is that BCI control cannot be established for a nonnegligible portion of users, estimated at 15% to 30% in EEG-based BCIs (see Kübler, Neumann, Wilhelm, Hinterberger, & Birbaumer, 2004). Based on prior experience, we hypothesized that one main difficulty in machine learning (ML)-based BCIs is the transition from offline calibration to online feedback (see e.g., Shenoy, Krauledat, Blankertz, Rao, & Müller, 2006). This is mainly due to different feature distributions that can be observed between offline calibration and the feedback application (see Sugiyama, Krauledat, & Müller, 2007). Furthermore, the reasons for BCI illiteracy can differ from user to user (see Nikulin, Hohlefeld, Jacobs, & Curio, 2008), and therefore some effort has been previously undertaken to categorize participants into classes

according to their BCI control (see Blankertz, Sannelli, et al., 2010, for a report on categorization of BCI users). In a screening study by Blankertz, Sannelli, et al. (2010) 80 volunteers performed motor imagery first during calibration (i.e., without feedback) and then during feedback in which they had to control a 1D cursor application. Coarsely, we observed three categories of users. First there were users for whom (I) a classifier could be successfully trained and who performed feedback with good accuracy. Second, there were users for whom a classifier could be successfully trained but feedback did not work well (II). It is known that there are changes between the calibration and the feedback step that can affect the EEG signals, making the feedback fail. In this categorization study, the bias of the classifier was updated in a supervised fashion using the first 20 feedback trials, as in Shenoy et al. (2006) and Dornhege et al. (2007). Third, there were users for whom no classifier with acceptable accuracy could be trained after the calibration (III). Whereas users of category II had obvious difficulties with the transition from offline on online operation, participants of category III did not show the expected modulation of sensorimotor rhythms (SMRs): either no idle SMR was observed over motor areas, or it was not attenuated during motor imagery.

In this letter, we present results of a one-session pilot study in which we investigated whether individually guided coadaptive learning using machine learning techniques could help users of categories II and III to achieve successful feedback. In previous studies, adaptive BCI systems have been successfully tested in offline and online experiments (Lu, Guan, & Zhang, 2009; Sugiyama et al., 2007; Blumberg et al., 2007; Vidaurre, Schlögl, Cabeza, Scherer, & Pfurtscheller, 2007; Wang, Hong, Gao, & Gao, 2007; Buttfeld, Ferrez, & Millán, 2006; Shenoy et al., 2006; Vidaurre, Schlögl, Cabeza, Scherer, & Pfurtscheller, 2006). Note, however, that they were not applied to participants suffering from BCI illiteracy, thus, the study we report links coadaptive feedback and improvement of performance in BCI illiteracy for the first time. Our results show that adaptive machine learning methods can indeed help participants who suffer from the BCI illiteracy problem to gain control of the system.

The letter is divided into six sections. The second one describes the experimental setup with details about the hardware, software, paradigm, experimental session, and users. The third section explains the methods (algorithms) used in the experiments. We have divided the methods into three levels, each with an individual subsection. The fourth section presents the results of the study, and the discussion of these results is done in section 5. Section 6 concludes.

2 Experimental Setup

This study consists of a single one-day session of approximately 120 minutes of BCI feedback for each user. It immediately starts with feedback using a

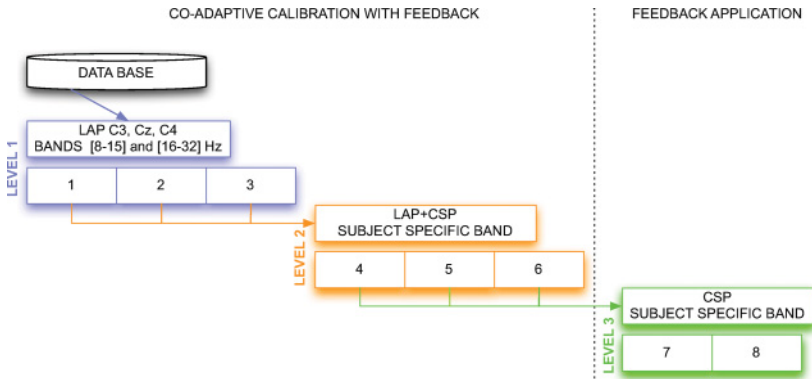


Figure 1: Schema of the experimental setup. Level 1: A database is used to calculate an initial subject-independent classifier in Laplacian channels C3, Cz, and C4. For three runs, the LDA is adapted. Level 2: The collected data of the first three runs are used to select a subject-dependent frequency band and calculate CSP and Laplacian channels. Then three runs more are performed, where the Laplacian channels are reselected after every trial and the classifier retrained. Level 3: Data are used to recalculate CSP and perform the last two runs, with an unsupervised adaptation of the classifier that allows tracking the features' drift and estimate an unbiased BCI performance.

pretrained subject-independent classifier, following Vidaurre et al. (2007); see also Fazli et al., 2009, for another subject-independent classification approach). Using supervised and unsupervised techniques, the BCI system is continuously adapted to the specific brain signals of the BCI user during the session. This change is guided by the algorithms that control which part of the system is adapted in each level.

Adaptation is performed in three incremental levels, starting from a simple system in level 1 and increasing the complexity in levels 2 and 3. While the feedback application itself stays the same for the entire experiment, the features on which the classifier operates and the classifiers itself can change in a trial-based manner. Figure 1 summarizes the experimental setup. Roughly, in level 1, a subject-independent classifier is computed from a database of users and used as starting point for BCI calibration with online feedback. Three runs are performed with a simple and robust system consisting of three Laplacian channels where the signals are filtered in fixed, subject-independent frequency bands. A simple linear discriminant classifier (LDA) is adapted in a supervised manner after every trial during the runs. In level 2, three more runs are recorded using a more complex and subject-specific set of features—common spatial patterns (CSP) and subject-selected Laplacian channels—to provide feedback. Here, the LDA classifier

is retrained in a supervised way after each trial. The last two runs (level 3) use subject-optimized features, and the LDA is adapted in an unsupervised manner to track the features' possible drift during the feedback. These last two runs serve as well to estimate the performance of the participant in a feedback application because no class information is used.

During the session, the subjects are sitting in a comfortable chair with arms lying relaxed on armrests. Brain activity is recorded from the scalp with multichannel EEG amplifiers (BrainAmp from Brain Products) using 64 Ag/AgCl electrodes in an extended international 10–20 system sampled at 1000 Hz, with a bandpass from 0.05 to 200 Hz. The EEG is filtered and downsampled to 100 Hz for the online operation of the system. The data are processed using an in-house Matlab toolbox, and the calculations are performed every 40 ms.

The categorization of users performed in Blankertz, Sannelli, et al. (2010) allows us to select BCI users according to their BCI ability. Our new BCI system is designed to work beyond category I participants, thereby improving the performance of category II and III volunteers. After the selection, 11 volunteers take part in the study and are selected according to their category (if prior data are available). Five participants belong to category I and serve to confirm the validity of the system for users with good performance. For one novice user, no prior categorization is possible, but this participant has almost perfect performance and is included in category I *a posteriori*. Two further participants belong to category II and three users to category III. Out of those 11 volunteers, 5 had participated in the screening study described in section 1. The other 5 users had participated in other (but similar) studies with an offline calibration step, which allowed us to perform their categorization.

All participants performed eight feedback runs, each of them consisting of 100 trials (50 trials of each class). After each run, there was a short break (1–5 minutes) for the BCI users to relax. The timing of the trials was as follows. At time 0, the cue was provided in the form of a small arrow over a cross placed in the middle of the screen; 1 second later, the cross started to move to provide feedback. Its speed was determined by the classifier output (similar to Blankertz, Dornhege, Krauledat, Müller, & Curio, 2007; Blankertz, Losch, Krauledat, Dornhege, Curio, & Müller, 2008). The task of the participants was to use motor imagery to make the cursor move into a previously indicated target direction. The feedback lasted for 3 seconds and was followed by a short pause. Two types of motor imagery, chosen out of three possibilities (motor imagery of left hand, right hand or foot) were selected in advance. For seven participants, previous data revealing which two motor imagery tasks to use were available. For the other four volunteers (three of category III and one novice), no prior information could be used. Therefore, they were asked to select two out of the three possible motor imagery tasks.

3 Methods

The experimental paradigm consists of three methodological levels (see Figure 1).

3.1 Adaptation Level 1, Runs 1 to 3. The experiments start with immediate feedback from a subject-independent binary linear classifier (Vidaurre et al., 2007). The features of this level are simple; they allow fast adaptation to the user and increase rapidly the quality of the feedback.

3.1.1 The Subject-Independent Classifier. This classifier is calculated prior to the study from data sets of 48 users (out of the 80 participants screened) whose performance in the pair of classes of interest (any pair of left- or right-hand motor imagery, left hand-foot motor imagery, or right hand-foot motor imagery is above 70% of accuracy; Blankertz, Sannelli, et al., 2010). None of the participants of categories II and III (due to their low performance or none at all) participating in the coadaptive study are part of this selection, but three in category I are. Feedback data of this pool of users are used to calculate the three binary classifiers, out of which only one is used to provide feedback for a particular participant of our experiments. The initial classifier is based on linear discriminant analysis (LDA). For LDA, the covariance matrices of both classes are assumed to be equal (assumption of linear separability) and denoted by Σ here. Furthermore, we denote the sample means of the two classes by μ_1 and μ_2 , an arbitrary feature vector by x , and define

$$D(x) = [b; \mathbf{w}]^\top \cdot [1; x] \quad (3.1)$$

$$\mathbf{w} = \Sigma^{-1} \cdot (\mu_2 - \mu_1) \quad (3.2)$$

$$b = \mathbf{w}^\top \cdot \mu \quad (3.3)$$

$$\mu = (\mu_1 + \mu_2) / 2, \quad (3.4)$$

where $D(x)$ is the difference in the distance of the feature vector x to the separating hyperplane, which is described by its normal vector \mathbf{w} and bias b and $^\top$ indicates the transpose operator. Note that the covariance matrices and mean values used in this letter are sample covariance matrices and sample means, estimated from the data. In order to simplify the notation and the description of the methods, in the following, we write only the covariance matrix instead of the sample covariance matrix and the mean instead of the sample mean. Usually the covariance matrix employed in equation 3.2 is the class-average covariance matrix. But it can be shown that using the pooled covariance matrix (which can be estimated without using label information, just by aggregating the features of all classes) yields the same separating hyperplane. In this study, we used the pooled covariance matrix in equation 3.2. Similarly, the class-average mean (calculated in equation 3.4)

can be replaced by the pooled mean (average over all feature vectors of all classes).

If LDA is to be used as a classifier, the observation \mathbf{x} is classified as class 1 if $D(\mathbf{x})$ is less than 0 and otherwise as class 2. But in the cursor control application, we use the classifier output $D(\mathbf{x})$ as a real number to determine the speed of the cursor.

3.1.2 Data Filtering. For runs 1 to 3, the EEG is preprocessed as follows. Three small Laplacian derivations (McFarland, McCane, David, & Wolpaw, 1997) over C3, Cz, and C4 are calculated from four surrounding channels, equally weighted, that are subtracted from the central one. After this spatial filter, the EEG is frequency-filtered in the μ (8–15 Hz) and β (16–32 Hz) bands using two Butterworth filters of order 10. The dimension of the resulting feature vector is 6.

3.1.3 Adaptive Classifier in Level 1. In this level, the LDA classifier is iteratively updated using the class information of the cue-based BCI system. The inverse of the covariance matrix and class mean values are updated after every trial using the class label (type of motor imagery task) only for the mean values of the past trial. Before describing the method used, we explain some concepts for the adaptive estimation of the mean and the (inverse) covariance matrices.

Adaptive mean estimation Given a feature vector \mathbf{x} , the mean can be estimated in the following way, which does not need memory for its past sample values:

$$\boldsymbol{\mu}_x(t) = (1 - UC_\mu) \cdot \boldsymbol{\mu}_x(t-1) + UC_\mu \cdot \mathbf{x}(t), \quad (3.5)$$

whereby UC_μ is the update coefficient describing an exponential weighting window

$$w_i = UC_\mu \cdot (1 - UC_\mu)^i \quad (3.6)$$

with a time constant $\tau = 1/(UC_\mu \cdot F_s)$ if the sampling rate is F_s . For these experiments, the update coefficient is determined offline by simulating the adaptation of the mean over the data of 80 users, and the optimal value is 0.05.

Adaptive inverse covariance matrix estimation The stationary covariance matrix of a multivariate process is defined as

$$\text{cov}(\mathbf{x}) = \boldsymbol{\Sigma}_x = \frac{1}{N} \sum_{t=1}^N (\mathbf{x}(t) - \boldsymbol{\mu}_x) \cdot (\mathbf{x}(t) - \boldsymbol{\mu}_x)^\top. \quad (3.7)$$

For convenience, we define the so-called extended covariance matrix (ECM) E as

$$\begin{aligned} ECM(x) = E_x &= \sum_{t=1}^{N_x} [1, \mathbf{x}(t)] \cdot [1, \mathbf{x}(t)]^\top \\ &= \left[\begin{array}{c|c} a & \mathbf{b}^\top \\ \hline \mathbf{c} & \mathbf{D} \end{array} \right] = N_x \cdot \left[\begin{array}{c|c} 1 & \boldsymbol{\mu}_x^\top \\ \hline \boldsymbol{\mu}_x & \boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top \end{array} \right]. \end{aligned} \quad (3.8)$$

Note that one can obtain from E_x the number of samples $N = a$, the mean $\boldsymbol{\mu} = \mathbf{b}/a$, as well as the variance-covariance matrix $\boldsymbol{\Sigma} = \mathbf{D}/a - (\mathbf{c}/a) \cdot (\mathbf{b}^\top/a)$.

The adaptive version of E estimator is

$$E_x(t) = (1 - UC_\Sigma) \cdot E_x(t-1) + UC_\Sigma \cdot [1, \mathbf{x}(t)] \cdot [1, \mathbf{x}(t)]^\top, \quad (3.9)$$

where t is the sample time and UC_Σ is the update coefficient.

LDA relies on the inverse $\boldsymbol{\Sigma}^{-1}$ of the covariance matrix $\boldsymbol{\Sigma}$; adaptive classifiers require an adaptive estimation of the inverse covariance matrix. In the following, we show how to update the inverse without an explicit matrix inversion.

The matrix inversion lemma (also know as Woodbury matrix identity) states that for a given matrix $A = (\mathbf{B} + \mathbf{UDV})$, its inverse A^{-1} can be determined by

$$\begin{aligned} A^{-1} &= (\mathbf{B} + \mathbf{UDV})^{-1} = \\ &= \mathbf{B}^{-1} - \mathbf{B}^{-1} \mathbf{U} (\mathbf{D}^{-1} + \mathbf{V} \mathbf{B}^{-1} \mathbf{U})^{-1} \mathbf{V} \mathbf{B}^{-1}. \end{aligned} \quad (3.10)$$

To adaptively estimate the inverse of the extended covariance matrix, we identify the matrices in equation 3.10 (see equation 3.9):

$$\mathbf{A} = E(t) \quad (3.11)$$

$$\mathbf{B} = (1 - UC_\Sigma) \cdot E(t-1) \quad (3.12)$$

$$\mathbf{U}^T = \mathbf{V} = \mathbf{x}(t)^\top \quad (3.13)$$

$$\mathbf{D} = UC_\Sigma, \quad (3.14)$$

where UC_Σ is the update coefficient and $\mathbf{x}(t)$ is the current sample vector. Accordingly, the inverse of the covariance matrix is

$$E(t)^{-1} = \frac{1}{(1 - UC_\Sigma)} \cdot \left(E(t-1)^{-1} - \frac{\mathbf{v}(t) \cdot \mathbf{v}^\top(t)}{\frac{1 - UC_\Sigma}{UC_\Sigma} + \mathbf{x}^\top(t) \cdot \mathbf{v}(t)} \right), \quad (3.15)$$

with $\mathbf{v}(t) = \mathbf{E}(t-1)^{-1} \cdot \mathbf{x}(t)$. Note that the term $\mathbf{x}^\top(t) \cdot \mathbf{v}(t)$ is a scalar, and no explicit matrix inversion is needed, except for the calculation of the initial value $\mathbf{E}(0)^{-1}$.

In practice, this adaptive estimator can become numerically unstable. This problem can be avoided if the symmetry is enforced.

The matrix obtained by eliminating the first row and column of $\mathbf{E}(t)^{-1}$ is $\boldsymbol{\Sigma}(t)^{-1}$. The update of the inverse of the extended covariance matrix has the advantage that the classification output of the LDA can be computed with a simple operation:

$$D(\mathbf{x}(t)) = [\mathbf{b}(t), \mathbf{w}(t)]^\top \cdot [1, \mathbf{x}(t)] \quad (3.16)$$

$$= \mathbf{b}(t) + \mathbf{w}(t)^\top \cdot \mathbf{x}(t) \quad (3.17)$$

$$= -\Delta\boldsymbol{\mu}(t)^\top \cdot \boldsymbol{\Sigma}(t)^{-1} \cdot \boldsymbol{\mu}(t) + \Delta\boldsymbol{\mu}(t)^\top \cdot \boldsymbol{\Sigma}(t)^{-1} \cdot \mathbf{x}(t) \quad (3.18)$$

$$= [0, \boldsymbol{\mu}(t)_i - \boldsymbol{\mu}(t)_j]^\top \cdot \mathbf{E}(t)^{-1} \cdot [1, \mathbf{x}(t)], \quad (3.19)$$

with $\Delta\boldsymbol{\mu}(t) = \boldsymbol{\mu}(t)_i - \boldsymbol{\mu}(t)_j$, $\mathbf{b} = -\Delta\boldsymbol{\mu}(t)^\top \cdot \boldsymbol{\Sigma}(t)^{-1} \cdot \boldsymbol{\mu}(t)$, and $\mathbf{w} = \Delta\boldsymbol{\mu}(t) \cdot \boldsymbol{\Sigma}(t)^{-1}$.

Accordingly, the output of the adaptive LDA can be estimated with equation 3.19, using equation 3.15 for estimating $\mathbf{E}(t)^{-1}$ and equation 3.5 for estimating the class-specific adaptive mean $\boldsymbol{\mu}(t)_i$ and $\boldsymbol{\mu}(t)_j$. The adaptation speed is determined by the two update coefficients UC_μ and UC_Σ used in equations 3.5 and 3.15. These UCs have to be determined beforehand. In this study, recorded feedback data of 80 subjects is used to find optimal coefficients: $UC_\mu = 0.05$ for the mean values and $UC_\Sigma = 0.015$ for the covariance matrices.

Figure 2 summarizes level 1 in a schema.

3.2 Adaptation Level 2, Runs 4 to 6. In this level, feedback with optimized features is provided to the user because data for subject-dependent optimization can be gathered from runs 1 to 3. First, these data are used to estimate the frequency band in which the classes are better discriminated. This procedure is described in detail in the tutorial (Blankertz, Tomioka, Lemm, Kawanabe, & Müller, 2008). Then features are extracted using common spatial patterns/filters (CSP/CSF), which remain fixed along runs 4 to 6. After every trial, six subject-selected Laplacian derivations are concatenated to the CSF to allow some spatial adaptation. We explain the methods in detail.

3.2.1 Common Spatial Patterns (CSP). CSP is a technique to analyze multichannel data based on recordings from two classes (tasks). It yields a data-driven supervised decomposition of the signal $\mathbf{x}(t)$ parameterized by a matrix \mathbf{W} that projects the signal from the original sensor space to a surrogate sensor space $\mathbf{x}_{CSP}(t)$ (Blankertz, Tomioka, et al., 2008): $\mathbf{x}_{CSP}(t) = \mathbf{x}(t) \cdot \mathbf{W}$. Each column vector of \mathbf{W} is a spatial filter. CSP filters maximize the

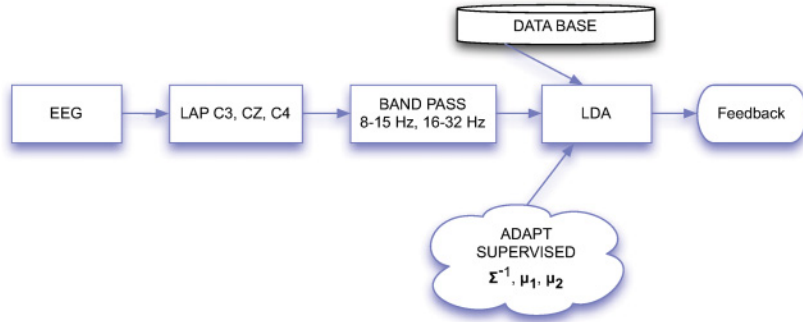


Figure 2: Schema of adaptation level 1. First, a database is used to calculate an initial subject-independent classifier in the Laplacian channels C3, Cz, and C4. The EEG is filtered in fixed bands (8–15 and 16–32 Hz). During the runs, the mean values and covariance matrices of both classes are trial-based updated and an LDA calculated.

variance of the spatially filtered signal under one task while minimizing it for the other task. Since the variance of a bandpass filtered signal is equal to band-power, CSP analysis is applied to bandpass filtered signals to obtain an effective discrimination of mental states that are characterized by ERD/ERS (even related desynchronization/synchronization) effects. Detailed information about this technique can be found in Blankertz, Tomioka, et al. (2008). For our study, CSP filters are individually calculated for each participant using the bandpass filtered signals of runs 1 to 3. The number of filters used is automatically selected and ranges between two and six filters. These filters are maintained fixed during the experiment.

3.2.2 Subject-Dependent Laplacian Channels. Six subject-selected Laplacian derivations are concatenated to fixed CSF. The scalp is divided into three different areas, as shown in Figure 3. Two Laplacian channels are selected from each of them.

The selection of the Laplacian channels is updated for every trial. It is based on the value of a robust Fisher score (see equation 3.20) obtained from every channel and after every trial (using the last 100 trials, that can partly originate from a previous run):

$$\text{Score} = \frac{(\text{median}(x_1) - \text{median}(x_2))^2}{\text{var}(x_1) + \text{var}(x_2)}. \quad (3.20)$$

The difference from the normal Fisher score is the use of medians instead of means. For each area, the two channels with greater Fisher score are reselected after every trial.

Figure 4 summarizes level 2 in a schema.

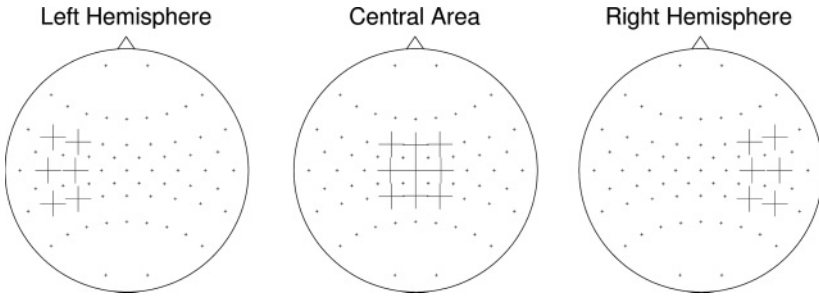


Figure 3: Two Laplacian derivations are selected from each of the scalp areas to complement the CSFs. The selection of the derivations is subject dependent and performed on a trial basis.

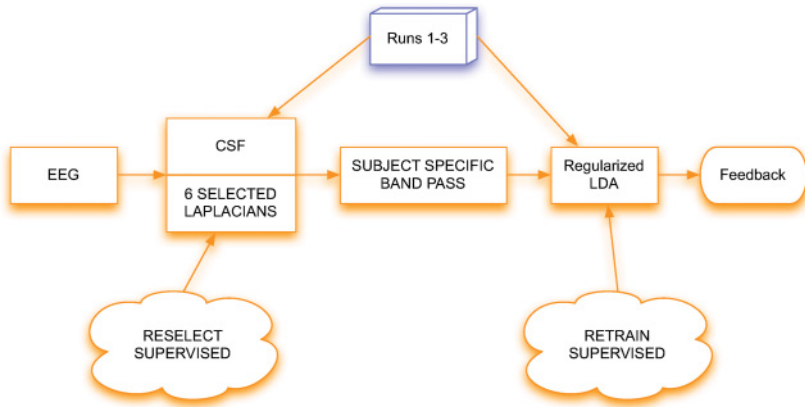


Figure 4: Schema of adaptation level 2. The data recorded in runs 1 to 3 are used to calculate CSF and six Laplacian channels. The frequency band is subject-optimized. During runs 4 to 6, the Laplacian channels are reselected after every trial (using the robust Fischer score) and the classifier retrained.

3.2.3 Recalculation of the Classifier. The classifier is recalculated every trial to account for the adaptation of the features. In order to cope with the increased dimensionality of the feature vector, which varies depending on the subject between 8, and 12 dimensions, a regularized version of LDA is used. We follow the approach of Ledoit and Wolf (2004a, 2004b) and replace Σ in equation 3.2 by a shrinkage estimate of the form

$$\Sigma_{\gamma} = (1 - \gamma) \Sigma + \gamma \tilde{\Sigma}, \gamma \in [0, 1].$$

The matrix $\tilde{\Sigma}$ is the sample covariance matrix of a restricted submodel, and for some submodels, the optimal shrinkage intensity γ can be analytically

estimated from the data. This is an important advantage because the time restrictions of the system (the calculations must be concluded before the next trial starts) would render it impossible to find a regularization parameter using cross-validation. This approach was used for EEG data in Vidaurre, Krämer, Blankertz, and Schlögl (2009). In this study, we use the following submodel: all variances (i.e., all diagonal elements) are equal, and all covariances (i.e., all off-diagonal elements) are zero. The analytical γ can be calculated as follows:

$$t_{ij} = \begin{cases} v = \text{avg}(s_{ii}) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$\gamma = \frac{\sum_{i \neq j} \widehat{\text{Var}}(s_{ij}) + \sum_i \widehat{\text{Var}}(s_{ii})}{\sum_{i \neq j}},$$

where t_{ij} is the target matrix and $\widehat{\text{Var}}(s_{ij})$ is the variance of the individual entries of the empirical covariance matrix. (See Schäfer, & Strimmer, 2005, for other alternatives, and their corresponding optimal γ .)

3.3 Adaptation Level 3, Runs 7 and 8. Since the techniques presented in sections 3.1.3 and 3.2 are supervised, it is not possible to estimate the real performance of the participants during runs 1 to 6 because additional information of the subjects' intention, which is not available in real applications, is used to update the system. Therefore, the last two runs, 7 and 8, used static features and an unsupervised adaptation of the classifier that allows tracking the drifts of features and at the same time estimating the performance. The features are CSF calculated from runs 4 to 6. The frequency band is also obtained from those runs. The number of filters used is optimized for each user (see section 3.2.1).

3.3.1 Unsupervised Adaptation of the Classifier. The technique used in this letter for the unsupervised adaptation of the linear classifier has been previously studied offline in Vidaurre, Schlögl, Blankertz, Kawanabe, and Müller (2008). Here we use the adaptation of the pooled mean with the formula in equation 3.5. 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. The selection of the update coefficient is performed in recorded data of 19 users, described in Vidaurre et al. (2008). The UC value in this study is 0.05. Figure 5 depicts a schema of this adaptation level.

4 Results

The online performance of the users is depicted in Figure 6. A hit is achieved when the participant places the cross on the correct side of the screen at the

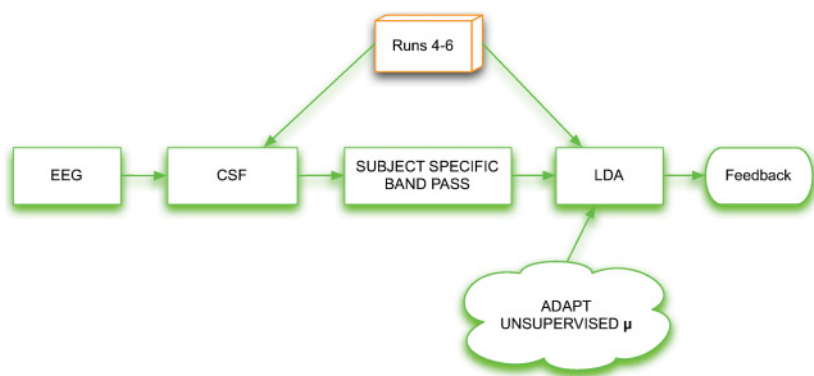


Figure 5: Schema of adaptation level 3. The data recorded in runs 4 to 6 are used to calculate CSF. The frequency band is subject optimized as in runs 4 to 6. During the feedback, the pooled mean is updated without using class information.

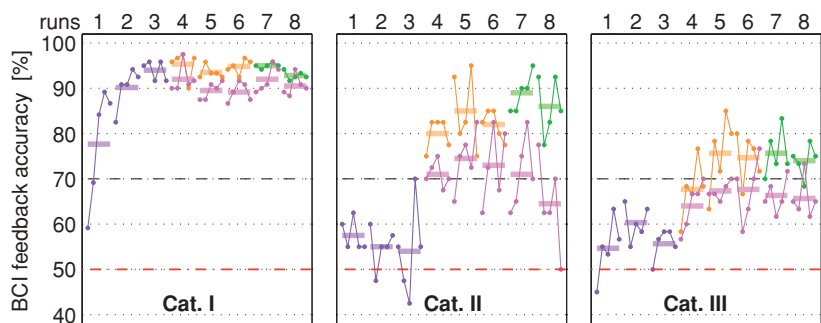


Figure 6: Grand average of BCI performance within each run (horizontal bars and dots for each group of 20 trials) according to adaptation levels for subjects of category I ($N = 6$), category II ($N = 2$), and category III ($N = 3$). An accuracy of 70% is assumed to be a threshold required for BCI applications (dash-point black line), and 50% is random performance (in dash-point red line). Violet corresponds to level 1, orange to level 2, and green to level 3. Pink from runs 4 to 8 corresponds to the performance obtained by simulating the methods of level 1 until the end of the experimental session.

end of the trial. Otherwise the trial is considered a miss. The performance is computed in percentage of correct trials. The participants are grouped according to their category. Figure 6 reflects the grand average performance in the three levels. The black dashed line marks the performance

Table 1: Mean and Standard Deviation of the BCI Performance.

	Level 1	Level 2	Level 3
Category I	82.3±7.8	94.6±3.6	93.9±5.6
Category II	55.5±6.5	82.3±5.0	87.5±1.5
Category III	56.9±4.7	72.7±0.3	74.9±7.5

level of 70%. According to Kübler et al. (2004), an accuracy of 70% is assumed to be a threshold required for BCI applications related to communication, such as using a BCI-driven text entry system (Williamson, Murray-Smith, Blankertz, Krauledat, & Müller, 2009). The red dashed line marks the performance level of 50% (random performance in two-class systems). Runs 1 to 3 are coded in violet, runs 4 to 6 are orange, and the last two are green. An additional pink line over runs 4 to 8 is added to include a simulation using the adaptation methods of level 1 (3 Laplacian channels plus an adaptive LDA classifier) in all experimental runs. Its interpretation will be discussed later. In addition, Table 1 presents the average and standard deviation of the users’ performance in the three different levels.

The first finding is that participants of category I (good calibration and feedback) can have very high performance with the novel coadaptive system, which corresponds to level 1, within just 20 to 40 trials (i.e., around 3–6 minutes). However, volunteers of categories II and III are not able to achieve control with this BCI system within the first three runs.

During level 2 (with supervised adaptation), the performance of all users of all categories improves over the 70% level; this occurs after less than 60 minutes from the beginning of the session. We hypothesize that this improvement is partly due to the optimized feedback quality: the use of CSF, selected Laplacian channels, a specific frequency band, and the adaptation procedure itself. The pink values over runs 4 to 8 in Figure 6 correspond to a simulated baseline of level 1 (3 Laplacians). For category I users, the simulated performance is slightly worse than that obtained in run 3 (the last of the level 1 feedback) with the same system. This performance is also worse than the one obtained online (with the methods of levels 2 and 3), which suggests that category I participants can adjust their performance to the system they receive feedback from and achieve improved results. For category II and (less prominent for) category III volunteers, we can observe that the simulated performance increased notably after the first three runs. Consequently, the increase in performance is due not only to the optimized BCI system used, but better feedback clearly improved the quality of the mental strategy and, accordingly, the signals.

In Figures 7 and 8, we see the ERD and ERS curves of two participants in runs 1 and 2 (beginning of the first level) versus runs 5 and 6

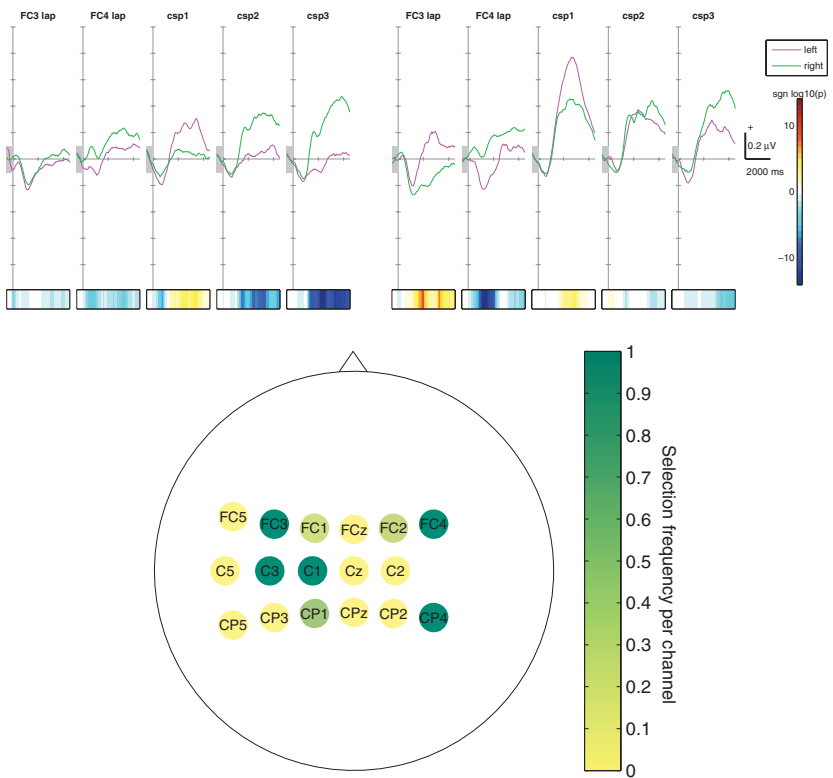


Figure 7: ERD/ERS curves of signals filtered in subject-selected Laplacians and CSP. Those Laplacian channels where the effect is best visible are displayed. The top-left image displays the curves in the first two runs of the experiments. The top-right image displays the curves in runs 5 and 6. The horizontal bars at the bottom are computed as minus the logarithm (basis 10) of the probability (p -values of a two-sided test for nonzero correlation between class label and feature) and multiplied by the sign of the correlation coefficient. Values greater or smaller than 1.30 represent a p -value under 0.05 (scale on the right side). The bottom row displays the frequency of selection of each Laplacian channel during runs 5 and 6, where two channels per area are selected (the total frequency per area is two and the greatest frequency per channel is one). This is a category II participant (number 7 in Figure 12), also corresponding to Figure 10.

(end of the second level).¹ The horizontal bands on the bottom represent the discriminability between the classes. They are computed as minus the

¹The definition of ERD and ERS can be found in Pfurtscheller (1992): “Event-related desynchronization (ERD) describes the short lasting (phasic) and regional localized

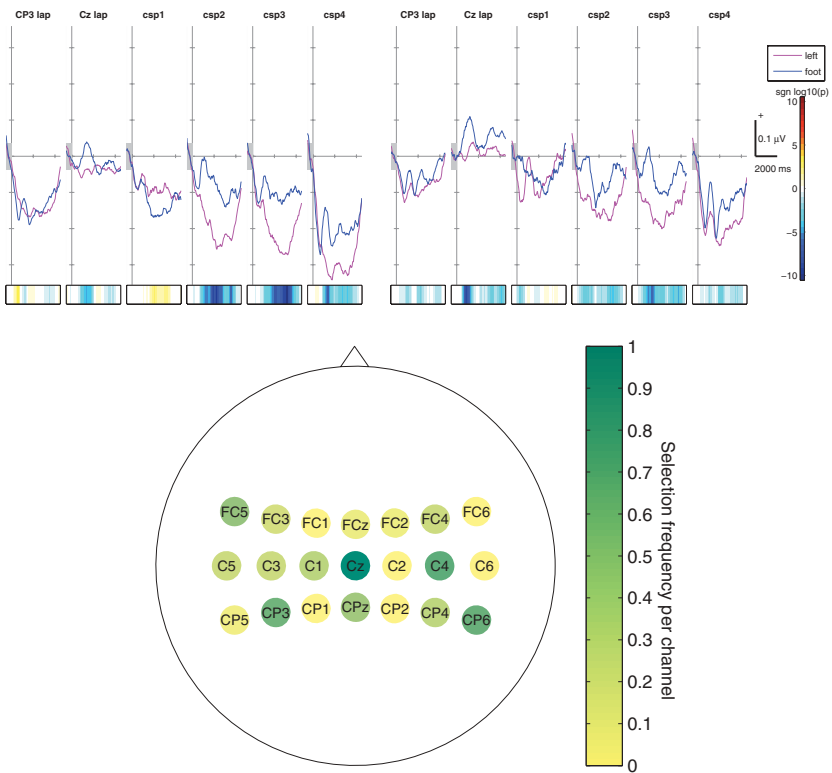


Figure 8: ERD/ERS curves of signals filtered in subject-selected Laplacians and CSP. Those Laplacian channels where the effect is best visible are displayed. The top-left image displays the curves in the first two runs of the experiments. The top-right image displays the curves in runs 5 and 6. The horizontal bars at the bottom are computed as minus the logarithm (basis 10) of the probability (p -values of a two-sided test for nonzero correlation between class label and feature) and multiplied by the sign of the correlation coefficient. Values greater or smaller than 1.30 represent a p -value under 0.05 (scale on the right side). The bottom row displays the frequency of selection of each Laplacian channel during runs 5 and 6, where two channels per area are selected (the total frequency per area is two, and the greatest frequency per channels is one). This is a category III participant (number 10 in Figure 12), also corresponding to Figure 13.

amplitude attenuation or blocking of oscillations in the alpha and beta bands that occurs in direct relation to an event” and “the opposite phenomenon, event-related synchronization (ERS), describes the phasic and regional localized increase of alpha and beta band activity in the form of bursts or spindles.”

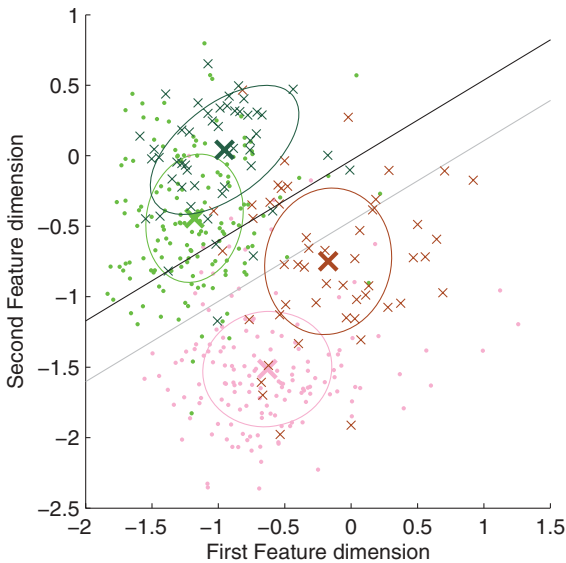


Figure 9: The figure plots the feature distribution of runs 4 to 6 (light green and light red; samples indicated by dots) versus run 8 (dark green and dark red; samples indicated by crosses). The separating lines correspond to the classifiers trained in runs 4 to 6 (gray) and the unsupervised classifier of run 8 (black).

logarithm (basis 10) of the p -values of a two-sided test for nonzero correlation (between class label and feature) and multiplied by the sign of the correlation coefficient. A value greater (or smaller) than 1.30 means that the p -value is under 0.05. The discrimination power is shown in the form of a bar next to the ERD or ERS curves. Those Laplacian channels where the ERD or ERS discriminability between classes is more visible are displayed with the CSP filtered data. The scalp plots show the selection frequency of each Laplacian channel. The discriminability of the CSP filters decreases over the runs, whereas the selected Laplacian channels improve their performance. The pictures demonstrate that allowing adaptation in the spatial domain is a suitable strategy for helping category II and III users.

During level 3, all categories maintain the performance reached in level 2. In this case, the features are fixed, and the classifier tracks the shifting of the features in the space. Because this level is adaptive but unsupervised, it also provides an estimate of the performance of the users at the end of the session. As an illustration of the adaptation of the classifier, Figure 9 shows in a low-dimensional projection that the feature distributions change over time and that the classifier boundary moves. It also depicts the advantage of using such adaptation in ongoing BCI sessions.

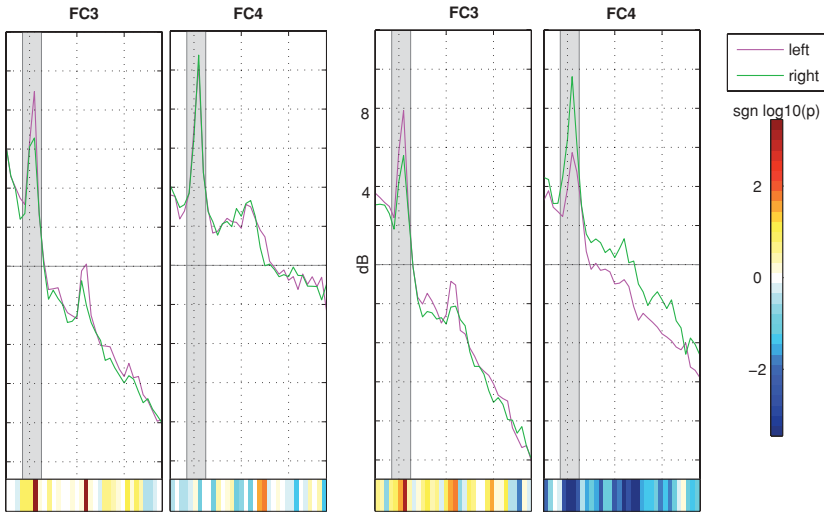


Figure 10: Spectra of subject-selected Laplacians. (Left) Spectra of the channels where the effect is better visible in the first two runs of the experiments. (Right) spectra in the last two runs. We can see how the discriminability has increased over the session. The figure corresponds to a category II participant (number 7 in Figure 12).

Note that the unsupervised adaptation method allows only parallel shifts of the separating hyperplane.

Finally, it is interesting to compare the beginning and end of the session and look at differences in the spectral information of the subjects. In Figures 10 and 11, we can quantitatively observe how the discriminative power at the band selected for feedback increases when comparing the first two and the last two runs. It corresponds to a category II and category III participant (numbers 7 and 10 in Figure 12), respectively.

A quantitative comparison between the beginning and end of the session can be found calculating the difference in the discriminability of the signals between runs 1 + 2 and 7 + 8. The features are extracted from the Laplacian channels over the motor areas of interest, which change depending on the pair of classes selected (right hemisphere, central, left hemisphere). The frequency band and time window to extract the features correspond to those of level 3, because we suppose that at this stage, the user learned to adjust his or her strategy, yielding to stable parameters. For each user, the value displayed on the left of Figure 12 is the total r^2 . One can observe that the discriminative power of the signals is higher at the end of the session for all users except one. This volunteer is category I, and, interestingly, his performance does not deteriorate over the session but stays the same, at

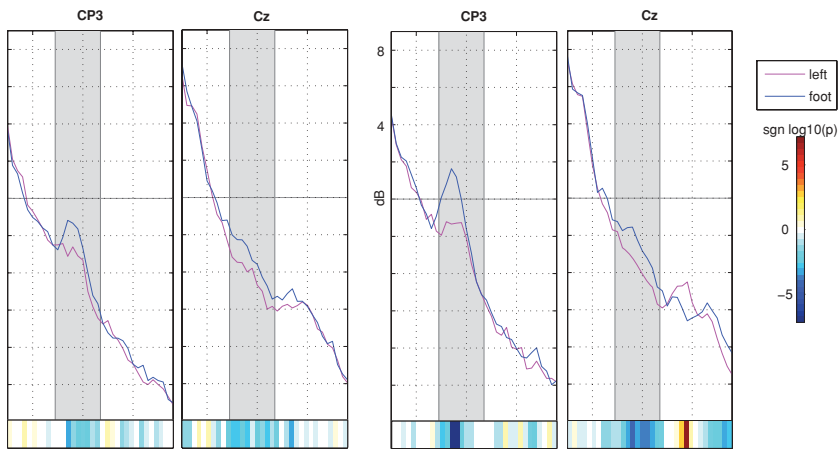


Figure 11: Spectra of subject-selected Laplacians. (Left) Spectra of the channels where the effect is more visible in the first two runs of the experiments. (Right) Spectra in the last two runs. We can see how the discriminability has increased over the session. The figure corresponds to a category III participant (number 10 in Figure 12).

0% of error. At the right of the same figure there is an example of the scalp maps (for a category III participant) that we use to estimate the difference in the discriminability of the signals between the beginning and end of the session.

Finally, Figure 13 illustrates a category 3 participant who could not perform any meaningful feedback at the beginning but was able to develop a proper motor imagery strategy by the end of the session and develop SMR that was not present before.

5 Discussion

In this study, our initial goal was to find out whether the coadaptive approach works as well as the state-of-the-art system (with offline calibration). This is clearly confirmed with the performance of category I participants. However, our main goal was to challenge the system with participants of categories II and III. The category I participants (except the naive user) selected for this study had reached performances over 85% in previous BCI sessions. Category II participants exhibited a 20% to 25% difference between the expected (cross-validation with the offline calibration data) and the feedback application performance, in a session performed with our previous BCI approach as in Blankertz, Sannelli, et al. (2010). Finally, none of the category III users had been able to perform feedback before this study.

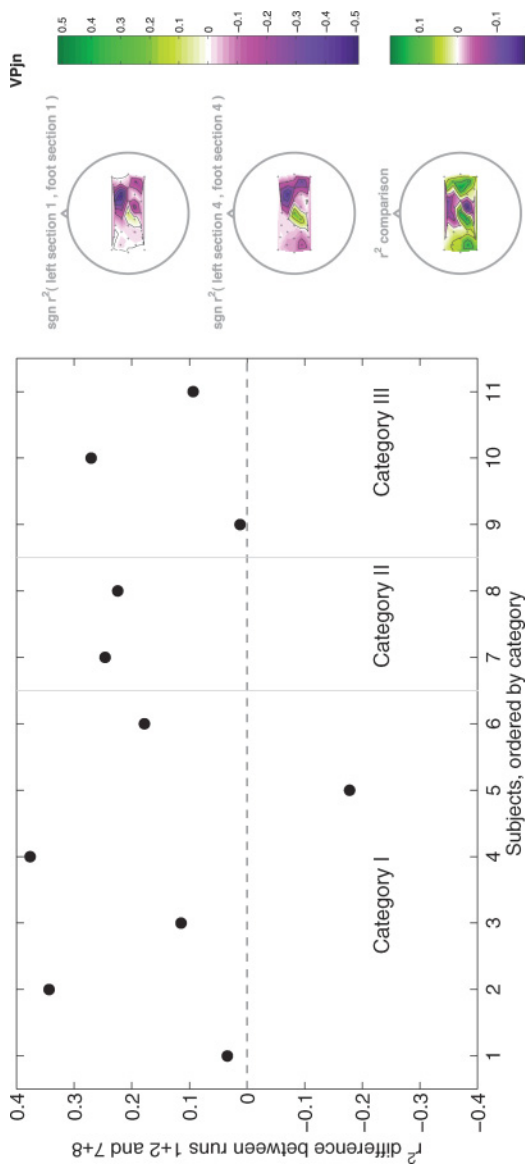


Figure 12: (Left) Discriminative difference of features at the beginning and end of the session. The features were extracted from Laplacian channels over the motor areas of interest that were selected during the second level of adaptation. The frequency band and time window to extract the features correspond to those of the last level, because it is supposed that at this stage, the user learned to adjust his or her strategy, yielding to stable parameters. (Right) r^2 difference on the same features for a category III user (participant number 10).

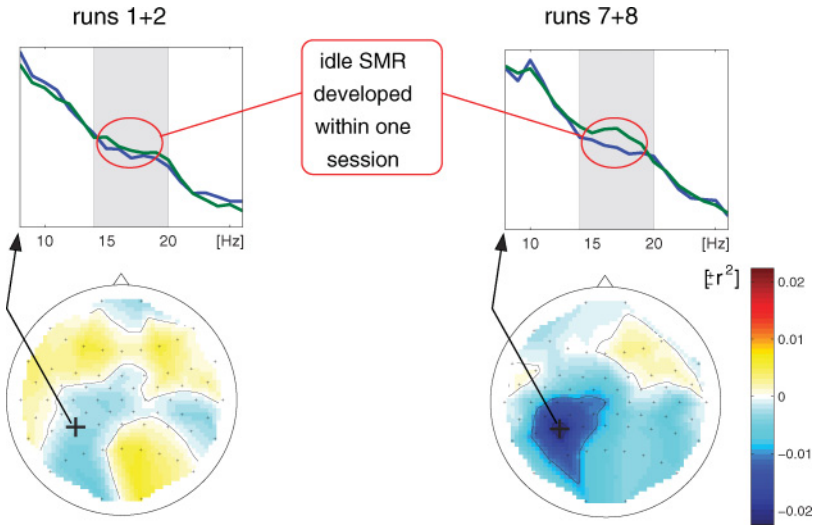


Figure 13: A participant of category III who could develop SMR over the session (user number 10 in Figure 12).

The selection of participants, together with the results presented in the previous section, demonstrates in detail the feasibility of our novel ML-based approach with co-adaptive learning; clearly a broad range of BCI users can benefit from it.

In particular, Figure 6 illustrates that “good” BCI users (category I) obtain accurate control in level 1 after a short period of adaptation (about 6 minutes). Users with less typical patterns of motor imagery (category II) need a more flexible way of adaptation to their specific patterns of brain activity, as provided by the adaptive training of level 2. Also, users who lack a suitable strategy to generate discriminable patterns by motor imagery in a classical approach (category III) can be guided to develop a successful strategy during ML-based coadaptive learning during level 2.

Figures 7 and 8 show that adaptation in level 2 allows better feedback when CSP filters start to fail due to changes in the patterns of task-related brain activity. The improvement obtained in level 2 is caused by two elements: the increased feedback quality through careful subject optimization and the increase in the quality of the subject’s signals, as demonstrated by the simulated performance of level 1 depicted in Figure 6. This is the striking consequence of coadaptation of the system to the subject, and vice versa. Regarding level 3, Figure 9 shows the suitability of the unsupervised adaptation (shifting of the classifier by bias tracking) to the problem by comparing the feature distributions of the runs recorded with levels 2 and 3. This unsupervised approach can track general shifts in the features that affect both classes in the same way, while the supervised method (applied

in earlier levels) allows tracking class-specific changes. We choose to adapt the bias but adapting the global covariance matrix is also possible and in principle allows accounting for more general changes in the task-unrelated signals (noise). However, our previous analyses show that the increased number of parameters to estimate is not advantageous (Vidaurre et al., 2008).

Additionally, an extensive comparison between the beginning (of level 1) and end (after level 3) of the session has been performed. Figures 10 and 13 show spectral differences of a category III volunteer. A similar effect can be seen in Figure 11 for a category II user. A general comparison of the spectral discriminative power over the motor area further supports our hypothesis that coadaptive learning is an ideal tool for alleviating the BCI illiteracy problem (see Figure 12).

Although previous approaches of adaptive BCI systems showed the feasibility of some elements presented in this letter (Vidaurre et al., 2006, 2007), this coadaptive study is the first to deal with users who previously could not achieve BCI control. In the previous applications, the BCI system was designed assuming that users can rapidly develop average SMR related features, that is, at specific locations (C3, Cz, and C4) and at specific frequency bands. Although this was the case for many users, some of them needed several sessions to learn to control the system. However, this new approach is much more flexible and clearly helps nonaverage users achieve control within a single session. This is accomplished by allowing changes not only in time, but also in frequency (after every level) and in space (CSP in levels 2 and 3 and selected Laplacian channels in level 2). The new BCI system seeks any type of SMR modulation at any location of the sensorimotor cortex and in any frequency band, which can change in time.

Finally, from the point of view of the BCI user, the coadaptive approach is more interesting because feedback starts from the very beginning of the session. Furthermore, it is more motivating because the performance increases, which might yield to a more successful experiment (Nijboer, Birbaumer, & Kübler, 2010).

6 Conclusion

Usually machine learning-based BCIs use EEG features of large complexity that are fitted better to the individual characteristics of brain patterns of each user (see Blankertz et al., 2007; Blankertz, Tomioka, et al., 2008; Dornhege, Blankertz, Curio, & Müller, 2004; Dornhege et al., 2007; Müller et al., 2008; Müller, Anderson, & Birch, 2003; Tomioka & Müller, 2010) during an initial offline calibration. However, since users are in a different mental state during offline calibration and online feedback (cf. Shenoy et al., 2006), a classifier that is optimized on the calibration data will turn out to be suboptimal and sometimes even nonfunctional for feedback (see Dornhege et al., 2007; Sugiyama et al., 2007; Krauledat, 2008,

and von Bünau et al., 2009 for a discussion of nonstationarities in BCI). Moreover, some users have difficulty properly performing separable motor imagery patterns in the absence of feedback (offline calibration). Here, we have presented a novel machine learning method that helps overcome these problems. It replaces the offline calibration used in our classic machine learning-based BCI (Dornhege et al., 2004) by a coadaptive calibration in which the mental strategy of the user and the algorithm of the BCI system are jointly optimized. This novel concept can lead some users to gain accurate BCI control very quickly (in 3 to 6 minutes). Other BCI users, who could not gain BCI control with a classic machine learning approach (belonging to category II or III), can now gain BCI control within one session, and even develop and modulate a SMR peak in this short time.

This important finding gives rise to the development of neurofeedback training procedures that might help cure BCI illiteracy. Further studies will focus on the challenging category III users and also on uncategorized participants to demonstrate that the coadaptive approach also works for BCI naive users. Also, we intend to support our findings with fMRI analysis to better investigate the change in brain activity caused by the coadaptive training.

Acknowledgments

We thank Alois Schögl for his support with the methods in sections 3.1.3 and 3.3. This work was partly founded by EU project 040666 Multiadaptive BCI, DFG (MU 987/3-1) VitalBCI Project, and EU TOBI Project ICT-224631. This publication reflects only our views. Funding agencies are not liable for any use that may be made of the information contained in this letter.

References

- Allison, B., Wolpaw, E., & Wolpaw, J. (2007). Brain-computer interface systems: Progress and prospects. *Expert Rev. Med. Devices*, 4, 463–474.
- Birbaumer, N., Weber, C., Neuper, C., Buch, E., Haapen, K., & Cohen, L. (2006). Physiological regulation of thinking: Brain-computer interface (BCI) research. *Prog. Brain Res.*, 159, 369–391.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K-R., & Curio, G. (2007). The noninvasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37, 539–550.
- Blankertz, B., Losch, F., Krauledat, M., Dornhege, G., Curio, G., & Müller, K-R. (2008). The Berlin brain-computer interface: Accurate performance from first-session in BCI-naïve subjects. *IEEE Trans. Biomed. Eng.*, 55, 2452–2462.
- Blankertz, B., Sannelli, C., Halder, S., Hammer, E-M., Kübler, A., Müller, K-R., et al. (2010). Neurophysiological predictor of SMR-based BCI performance. *Neuroimage*, 51, 1303–1309.

- Blankertz, B., Tangermann, M., Vidaurre, C., Dickhaus, T., Sannelli, C., Popescu, F., et al. (2010). Non-invasive and invasive brain-computer interfaces. In B. Allison, B. Graimann, & G. Pfurtscheller (Eds.), *Detecting mental states by machine learning techniques* (pp. 113–135). New York: Springer.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Müller, K.-R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process Mag.* 25, 41–56.
- Blumberg, J., Rickert, J., Waldert, S., Schulze-Bonhage, A., Aertsen, A., & Mehring, C. (2007). Adaptive classification for brain computer interfaces. In *Proceedings of the IEEE Eng. Med. Biol. Soc. 2007* (pp. 2536–2539). Piscataway, NJ: IEEE.
- Buttfield, A., Ferrez, P. W., & Millán, J. del R. (2006). Towards a robust BCI: Error recognition and online learning, *IEEE Trans. Neural Sys. Rehab. Eng.*, 14, 164–168.
- Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., et al. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.*, 1, 193–208.
- Dornhege, G., Blankertz, B., Curio, G., & Müller, K.-R. (2004). Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Transactions on Biomedical Engineering*, 51(6), 993–1002.
- Dornhege, G., Millán, J. del R., Hinterberger, T., McFarland, D., & Müller, K.-R. (Eds.). (2007). *Toward brain-computer interfacing*. Cambridge, MA: MIT Press.
- Fazli, S., Popescu, F., Danoczy, M., Blankertz, B., Müller, K.-R., & Grozea, C. (2009). Subject independent mental state classification in single trials. *Neural Networks*, 22(9), 1305–1312.
- Fetz, E. E. (2007). Volitional control of neural activity: Implications for brain-computer interfaces. *J. Physiol.*, 579, 571–579.
- Hochberg, L. R., Serruya, M. D., Friebs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442, 164–171.
- Krauledat, M. (2008). *Analysis of nonstationarities in EEG signals for improving brain-computer interface performance*. Berlin: Technische Universität Berlin, Fakultät IV—Elektrotechnik und Informatik.
- Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J., & Birbaumer, N. (2001). Brain-computer communication: Unlocking the locked in. *Psychol. Bull.*, 127(3), 358–375.
- Kübler, A., Neumann, N., Wilhelm, B., Hinterberger, T., & Birbaumer, N. (2004). Predictability of brain-computer communication. *Journal of Psychophysiology*, 18, 121–129.
- Ledoit, O., & Wolf, M. (2004a). Honey, I shrunk the sample covarianc matrix. *Journal of Portfolio Management*, 30, 110–119.
- Ledoit, O., & Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88, 365–411.
- Leuthardt, E. C., Freudenberger, Z., Bundy, D., & Roland, J. (2009). Microscale recording from human motor cortex: Implications for minimally invasive electrocorticographic brain-computer interfaces. *Neurosurg. Focus.*, 27(1), E10.

- Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Jemann, J. G., & Oran, D. W. (2004). A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.*, 1, 63–71.
- Lu, S., Guan, C., & Zhang, H. (2009). Unsupervised brain computer interface based on intersubject information and online adaptation. *IEEE Trans. Neural Sys. Rehab. Eng.*, 17, 135–145.
- McFarland, D. J., McCane, L. M., David, S. V., & Wolpaw, J. R. (1997). Spatial filter selection for EEG-based communication. *Electroencephalogr. Clin. Neurophysiol.*, 103, 386–394.
- Millán, J. del R., Renkens, F., Mouriño, J., & Gerstner, W. (2004). Brain-actuated interaction. *Artificial Intelligence*, 159, 241–259.
- Müller, K-R., Anderson, C. W., & Birch, G. E. (2003). Linear and non-linear methods for brain-computer interfaces. *IEEE Trans. Neural. Sys. Rehab. Eng.*, 11, 165–169.
- Müller, K-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial analysis: From brain-computer interfacing mental state monitoring. *Journal of Neuroscience Methods*, 167, 82–90.
- Nijboer, F., Birbaumer, N., & Kübler, A. (2010). The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis—a longitudinal study. *Frontiers in Neuroprosthetics*, doi=10.3389/fnins.2010.00055.
- Nikulin, V. V., Hohlefeld, F. U., Jacobs, A. M., & Curio, G. (2008). Quasi-movements: A novel motor-cognitive phenomenon. *Neuropsychologia*, 46, 727–742.
- Pistohl, T., Ball, T., Schulze-Bonhage, A., Aertsen, A., & Mehring C. (2008). Prediction of arm movement trajectories from ECoG-recordings in humans. *J. Neurosci. Methods*, 167, 105–114.
- Pfurtscheller, G. (1992). Event-related synchronization (ERS): An electrophysiological correlate of cortical areas at rest. *Electroencephalography and Clinical Neurophysiology*, 83, 62–69.
- Pfurtscheller, G., Neuper, C., & Birbaumer, N. (2005). Human brain-computer interface. In A. Riehle & E. Vaadia (Eds.), *Motor cortex in voluntary movements* (pp. 367–401). Boca Raton, FL: CRC Press.
- Rizk, M., Bossetti, C. A., Jochum, T. A., Callender, S. H., Nicoletis, M. A., Turner, D. A., et al. (2009). A fully implantable 96-channel neural data acquisition system. *J. Neural. Eng.*, 6, 026002.
- Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4, article 32.
- Schalk, G., Miller, K. J., Anderson, N. R., Wilson, J. A., Smyth, M. D., Ojemann, J. G., et al. (2008). Two-dimensional movement control using electrocorticographic signals in humans. *Journal of Neural Engineering*, 5, 75–84.
- Schwartz, A. B. (2004). Cortical neural prosthetics. *Annu. Rev. Neurosci.*, 27, 487–507.
- Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P. N., & Müller, K-R. (2006). Towards adaptive classification for BCI. *Journal of Neural Engineering*, 3, R13–R23.
- Sugiyama, M., Krauledat, M., & Müller, K-R. (2007). Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8, 1027–1061.

- Tomioka, R., & Müller, K.-R. (2010). A regularized discriminative framework for EEG based communication. *Neuroimage*, 49, 415–432.
- Vidaurre, C., Krämer, N., Blankertz, B., & Schlögl, A. (2009). Time domain parameters as a feature for EEG-based brain-computer interfaces. *Neural Networks*, 22, 1313–1319.
- Vidaurre, C., Schlögl, A., Blankertz, B., Kawanabe, M., & Müller, K.-R. (2008). Un-supervised adaptation of the LDA classifier for brain-computer interfaces. In *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course 2008* (pp. 122–127). Graz: TU-Graz Verlag.
- Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R., & Pfurtscheller, G. (2006). A fully on-line adaptive BCI. *IEEE Trans on Biomed Eng*, 53, 1214–1219.
- Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R., & Pfurtscheller, G. (2007). Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. *IEEE Trans on Biomed Eng*, 54, 550–556.
- von Büna, P., Meinecke, F. C., Kiraly, F., & Müller, K.-R. (2009). Finding stationary subspaces in multivariate time series. *Physics Review Letters*, 103(21), 214101.
- Waldert, S., Pistohl, T., Braun, C., Ball, T., Aertsen, A., & Mehring, C. (2009). A review on directional information in neural signals for brain-machine interfaces. *J. Physiol. Paris.*, 103, 244–254.
- Wang, Y., Hong, B., Gao, X., & Gao, S. (2007). Implementation of a brain-computer interface based on three states of motor imagery. *Conf. Proc. IEEE Eng. Med. Biol. Soc. 2007* (pp. 5059–5062).
- Williamson, J., Murray-Smith, R., Blankertz, B., Krauledat, M., & Müller, K.-R. (2009). Designing for uncertain, asymmetric control: Interaction design for brain-computer interfaces. *International Journal of Human-Computer Studies*, 67, 827–841.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113, 767–791.