Machine-Learning Based Co-adaptive Calibration: Towards a Cure for BCI illiteracy

Carmen Vidaurre[1] **Claudia Sannelli**[1] **Klaus-Robert Müller**[1,2]

Benjamin Blankertz[1,2,3]

[1]Machine Learning Dp, Berlin Institute of Technology, Berlin, Germany
[2]Bernstein Focus: Neurotechnology, Berlin, Germany
[3]Fraunhofer FIRST (IDA), Berlin, Germany
carmen.vidaurreATtu-berlin.de

Abstract

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 (BCI illiteracy problem, generic to all motor imagery based BCIs). We hypothesize that one main difficulty for a BCI-user is the transition from off-line calibration to on-line feedback. In this work we therefore 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 initially starting from a subject-independent classifier operating on simple features to a subject-optimized state-of-the-art classifier within one session, while the user interacts continuously. These initial runs use supervised techniques for robust co-adaptive 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 off-line calibration measurement, good performance was obtained by six users (also one novice) after 3-6 minutes of adaptation. More importantly, this novel guided learning allows also participants suffering from BCI illiteracy to gain significant control with the BCI in less than 60 minutes. Additionally, one volunteer without sensory motor idle rhythm peak in the

beginning of the BCI experiment could develop it during the course of the session and use voluntary modulation of its amplitude to control the feedback application.

Keywords

Co-adaptive Learning, Brain-Computer Interfacing, Electroencephalogram (EEG), Sensorimotor Rhythms (SMR), Machine Learning (ML), Common Spatial Patterns (CSP), Supervised adaptation, Unsupervised adaptation.

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, see (Dornhege et al., 2007; Allison et al., 2007; Birbaumer et al., 2006; Pfurtscheller et al., 2005; Wolpaw et al., 2002; Kübler et al., 2001)) for a summary of EEG-based BCI, refer to (Leuthardt et al., 2004; Pistohl et al., 2008; Schalk et al., 2008; Leuthardt et al., 2009) for ECoG-based BCIs and to (Carmena et al., 2003; Schwartz, 2004; Hochberg et al., 2006; Fetz, 2007; Rizk et al., 2009; Waldert et al., 2009) for a summary in spike-based approaches. However, independently of the system tested, a significant percent 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 non-negligible portion of users (estimated 15% to 30% in EEG-Based BCIs), cf. (Kübler et al., 2004). Based on prior experience, we hypothesized that one main difficulty in Machine-Learning (ML) based BCIs is the transition from off-line calibration to on-line feedback, see e.g. (Shenoy et al., 2006). This is mainly due to different feature distributions that can be observed between off-line calibration and the feedback application, see (Sugiyama et al., 2007). Furthermore, the reasons for BCI illiteracy can be different from user to user (see (Nikulin et al., 2008)) and therefore, some effort has been previously undertaken to categorize participants into classes according to their BCI control, see (Blankertz et al., 2010) for a report on categorization of BCI users. In that work, a screening study was realized, in which 80 volunteers performed motor imagery first in a calibration (i.e. without feedback) measurement and then in a feedback measurement in which they were to control a 1D cursor application. Coarsely, we observed three categories of users: users for whom (I) a classifier could be successfully trained and who performed feedback with good accuracy; (II) a classifier could be successfully trained, but feedback did not work well. 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; Dornhege et al., 2007); (III) no classifier with acceptable accuracy could be trained after the calibration measurement. Whereas users of Cat. II had obvious difficulties with the transition from off-line on on-line operation, participants of Cat. III did not show the expected modulation of sensorimotor rhythms (SMRs): either no SMR idle rhythm was observed over motor areas, or it was not attenuated during motor imagery. In this manuscript we present results of a one-session pilot study in which it was investigated, whether individually guided co-adaptive learning using machine-learning techniques could help users of Cat. II and III to achieve successful

feedback. In previous studies, adaptive BCI systems have been successfully tested, in offline and/or online experiments see (Lu et al., 2009; Sugiyama et al., 2007; Blumberg et al., 2007; Vidaurre et al., 2007; Wang et al., 2007; Buttfield et al., 2006; Shenoy et al., 2006; Vidaurre et al., 2006). Note however that they were not applied to participants suffering from BCI illiteracy, thus the present study for the first time links co-adaptive feedback and improvement of performance in BCI illiteracy. Our results show that adaptive machine learning methods can indeed help participants who suffer from the BCI illiteracy problem, to successfully gain control of the system.

The paper is divided into six sections. The second one is a description of 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 different 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 number five. Finally, there is a paragraph with the conclusions.

2 Experimental setup

The present study consists of a one-day single session of approximately 120 minutes of BCI feedback for each user. It immediately starts with feedback using a pre-trained subject-independent classifier, following (Vidaurre et al., 2007) (see also (Fazli et al., 2009) for other 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 whole 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 (namely 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 multi-channel EEG amplifiers (BrainAmp from Brain Products) using 64 Ag/AgCl electrodes in an extended international 10-20 system sampled at 1000 Hz, with a band-pass from 0.05 to 200 Hz. The EEG is filtered and down-sampled to 100 Hz for the online

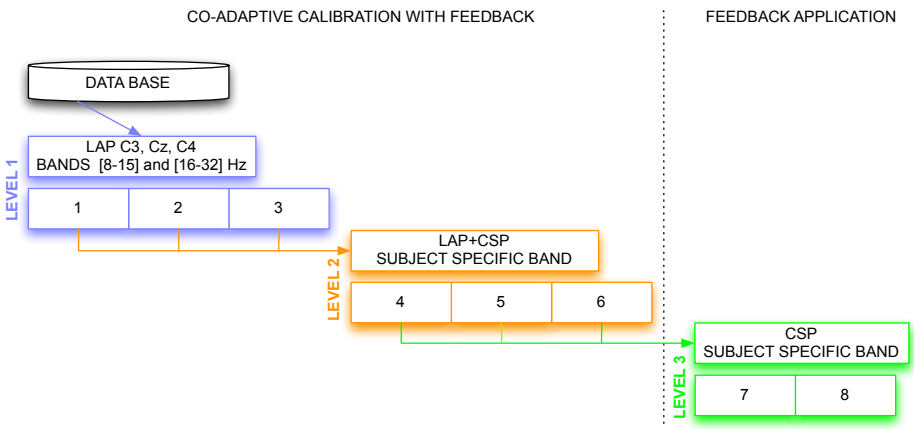

Figure 1: Schema of the experimental setup. Level 1: First, a database is used to calculate an initial subject-independent classifier in the laplacian channels C3, Cz and C4. For three runs the LDA is adapted. Level 2: the collected data of first 3 runs are used to select a subject dependent frequency band and calculate CSP and Laplacian channels. Then, three runs more are performed, where the Laplacian channels are re-selected trial-based 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.

operation of the system. The data is processed using an in-house matlab toolbox and the calculations are performed every 40 ms.

The Categorization of users performed in Blankertz et al. (2010) allows us to select BCI users according to their BCI ability. Our new BCI system is designed to work beyond Cat. I participants, improving thereby the performance of Cat. II and III volunteers. After the selection, eleven volunteers take part in the study and are selected according to their Category (if prior data is available). Five participants belong to Cat. 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 into Cat. I a posteriori. Two further participants belong to Cat. II and three users to Cat. 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 perform eight feedback runs, each of them consisting of 100 trials (50 trials of each class). After each run, there is a short break (1-5 minutes) for the BCI-users to relax. The timing of the trials is as follows: at time 0, the cue is provided in the form of a small arrow over a cross placed in the middle of the screen, one second later, the cross starts to move to provide feedback. Its speed was determined by the classifier output (similar to (Blankertz et al., 2007, 2008a)). The task of the participants is to use motor imagery to make the cursor move into a previously indicated target direction. The feedback lasts for 3 seconds and is followed by a short pause. Two different types of motor imagery, chosen out of three possibilities (motor imagery of left hand, right hand or foot) are selected in advance. For seven

participants, previous data revealing which two motor imagery tasks to use are available. For the other four volunteers (three of Cat. III and one novice) no prior information can be used. Therefore, they are asked to select two out of the three possible motor imagery tasks.

3 Methods

The experimental paradigm consists of three different methodological levels (see figure 1). Each of them is introduced separately.

3.1 Adaptation Level 1, runs 1-3

The experiments start with immediate feedback from a subject-independent binary linear classifier (Vidaurre et al., 2007). The feature of this level are simple and 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 datasets of 48 users (out of the 80 participants screened) whose performance in the pair of classes of interest (any pair of Left - Right hand motor imagery, Left hand - Foot motor imagery or Right hand - Foot motor imagery) is above 70 % of accuracy Blankertz et al. (2010). None of the participants of Cat. II and III (due to their none or low performance) participating in the co-adaptive study are part of this selection, but three Cat. I are. Feedback measurements 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 the linear discriminant analysis (LDA). For LDA the covariance matrices of both classes are assumed to be equal (assumption of linear separability) and denoted by $\mathbf{\Sigma}$ here. Furthermore we denote the sample means of the two classes by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, an arbitrary feature vector by $\boldsymbol{x}$ and define:

$$
\begin{aligned}
D(\boldsymbol{x}) &= [b; \boldsymbol{w}]^\top \cdot [1; \boldsymbol{x}] && (1)\\
\boldsymbol{w} &= \mathbf{\Sigma}^{-1} \cdot (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) && (2)\\
b &= \boldsymbol{w}^\top \cdot \boldsymbol{\mu} && (3)\\
\boldsymbol{\mu} &= (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2 && (4)
\end{aligned}
$$

where $D(\boldsymbol{x})$ is the difference in the distance of the feature vector $\boldsymbol{x}$ to the separating hyperplane, which is described by its normal vector $\boldsymbol{w}$ and bias b and $^\top$ indicates the transpose operator. Note that the covariance matrices and mean values used in this paper are sample covariance matrices and sample means, estimated from the data. In order to simplify the notation and the description of the methods, we will in the following just write covariance matrix instead of sample covariance matrix and mean instead of sample mean. Usually, the covariance matrix employed in Eq. (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 Eq. (2). Similarly, the class-average mean (calculated in Eq. (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 $\boldsymbol{x}$ is classified as class 1, if $D(\boldsymbol{x})$ is less than 0, and otherwise as class 2. But in the cursor control application we use the classifier output $D(\boldsymbol{x})$ as 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 et al., 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 mean (inverse) covariance matrices.

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

$$\boldsymbol{\mu}_x(t) = (1 - UC_\mu) \cdot \boldsymbol{\mu}_x(t-1) + UC_\mu \cdot \boldsymbol{x}(t) \tag{5}$$

whereby UC_μ is the update coefficient, describing an exponential weighting window

$$w_i = UC_\mu \cdot (1 - UC_\mu)^i \tag{6}$$

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 off-line by simulating the adaptation of the mean over 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

$$cov(x) = \boldsymbol{\Sigma}_x = \frac{1}{N} \sum_{t=1}^{N} (\boldsymbol{x}(t) - \boldsymbol{\mu}_x) \cdot (\boldsymbol{x}(t) - \boldsymbol{\mu}_x)^\top \tag{7}$$

For convenience, we define also the so-called *extended covariance matrix* (ECM) $\boldsymbol{E}$ as

$$ECM(x) = \boldsymbol{E}_x = \sum_{t=1}^{N_x} [1, \boldsymbol{x}(t)] \cdot [1, \boldsymbol{x}(t)]^\top = \left[\begin{array}{c|c} a & \boldsymbol{b}^\top \\ \hline \boldsymbol{c} & \boldsymbol{D} \end{array}\right] = N_x \cdot \left[\begin{array}{c|c} 1 & \boldsymbol{\mu}_x^\top \\ \hline \boldsymbol{\mu}_{\boldsymbol{x}} & \boldsymbol{\Sigma}_{\boldsymbol{x}} + \boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top \end{array}\right] \tag{8}$$

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

The adaptive version of $\boldsymbol{E}$ estimator is

$$\boldsymbol{E}_x(t) = (1 - UC_\Sigma) \cdot \boldsymbol{E}_x(t-1) + UC_\Sigma \cdot [1, \boldsymbol{x}(t)] \cdot [1, \boldsymbol{x}(t)]^\top \tag{9}$$

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

LDA relies on the inverse $\mathbf{\Sigma}^{-1}$ of the covariance matrix $\mathbf{\Sigma}$; adaptive classifiers require, therefore, an adaptive estimation of the inverse covariance matrix, in the following it is shown 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 $\boldsymbol{A} = (\boldsymbol{B} + \boldsymbol{UDV})$, its inverse $\boldsymbol{A}^{-1}$ can be determined by

$$\begin{aligned} \boldsymbol{A}^{-1} &= (\boldsymbol{B} + \boldsymbol{UDV})^{-1} = \\ &= \boldsymbol{B}^{-1} - \boldsymbol{B}^{-1}\boldsymbol{U}\left(\boldsymbol{D}^{-1} + \boldsymbol{VB}^{-1}\boldsymbol{U}\right)^{-1}\boldsymbol{VB}^{-1} \end{aligned} \tag{10}$$

To adaptively estimate the inverse of the *extended covariance matrix*, we identify the matrices in (10) as follows (see Eq. (9)):

$$\boldsymbol{A} = \boldsymbol{E}(t) \tag{11}$$

$$\boldsymbol{B} = (1 - UC_\Sigma) \cdot \boldsymbol{E}(t-1) \tag{12}$$

$$\boldsymbol{U}^T = \boldsymbol{V} = \boldsymbol{x}(t)^\top \tag{13}$$

$$\boldsymbol{D} = UC_\Sigma \tag{14}$$

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

$$\boldsymbol{E}(t)^{-1} = \frac{1}{(1 - UC_\Sigma)} \cdot \left(\boldsymbol{E}(t-1)^{-1} - \frac{\boldsymbol{v}(t) \cdot \boldsymbol{v}^\top(t)}{\frac{1-UC_\Sigma}{UC_\Sigma} + \boldsymbol{x}^\top(t) \cdot \boldsymbol{v}(t)} \right) \tag{15}$$

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

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 $\boldsymbol{E}(t)^{-1}$ is $\mathbf{\Sigma}(t)^{-1}$. 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:

$$D(\boldsymbol{x}(t)) = [b(t),\, \boldsymbol{w}(t)]^\top \cdot [1,\, \boldsymbol{x}(t)] \tag{16}$$

$$= b(t) + \boldsymbol{w}(t)^\top \cdot \boldsymbol{x}(t) \tag{17}$$

$$= -\Delta\boldsymbol{\mu}(t)^\top \cdot \mathbf{\Sigma}(t)^{-1} \cdot \boldsymbol{\mu}(t) + \Delta\boldsymbol{\mu}(t)^\top \cdot \mathbf{\Sigma}(t)^{-1} \cdot \boldsymbol{x}(t) \tag{18}$$

$$= [0, \boldsymbol{\mu}(t)_i - \boldsymbol{\mu}(t)_j]^\top \cdot \boldsymbol{E}(t)^{-1} \cdot [1,\, \boldsymbol{x}(t)] \tag{19}$$

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

Accordingly, the output of the adaptive LDA can be estimated with equation (19) using (15) for estimating $\boldsymbol{E}(t)^{-1}$ and (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 the equations (5) and (15). These UCs have to be determined before hand. 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.

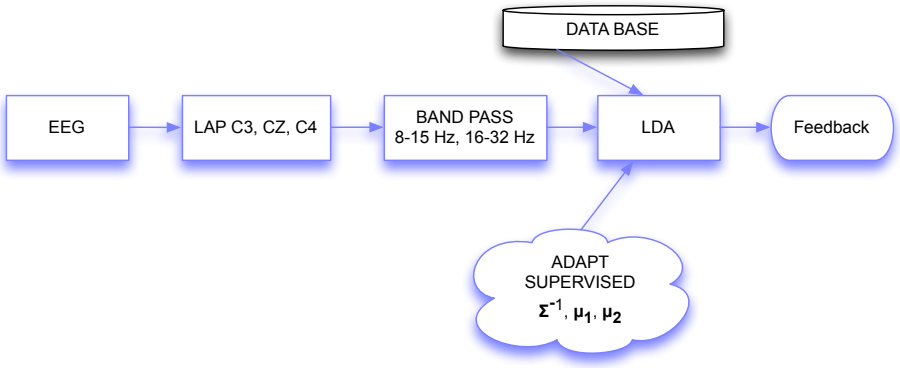

Figure 2: Schema of adaptation level 1. First, a database is used to calculate an initial subject-independent classifier in the laplacian channels C3, Cz and C4. The EEG is filtered in fixed bands (8-15 and 16-32 Hz). During the runs, the mean values and covariance matrices of both classes are trial-based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

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 et al., 2008b). Then, features are extracted using Common Spatial Patterns/Filters (CSP/CSF), which remain fixed along the runs 4 to 6. After every trial, six subject-selected Laplacians derivations are concatenated to the CSF to allow some "spatial adaptation". In the following paragraphs, the methods are explained 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 $\boldsymbol{x}(t)$ parametrized by a matrix $\boldsymbol{W}$ that projects the signal from the original sensor space to a surrogate sensor space $\boldsymbol{x}_{CSP}(t)$, (Blankertz et al., 2008b): $\boldsymbol{x}_{CSP}(t) = \boldsymbol{x}(t) \cdot \boldsymbol{W}$. Each column vector of a $\boldsymbol{W}$ is a spatial filter. CSP filters maximize the variance of the spatially filtered signal under one task while minimizing it for the other task. Since the variance of a band-pass filtered signal is equal to band-power, CSP analysis is applied to band-pass 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 et al., 2008b). For our study CSP filters are individually calculated for each participant using the band-pass filtered signals of runs 1 to 3. The number of filters used is automatically selected and ranges between 2 and 6 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.

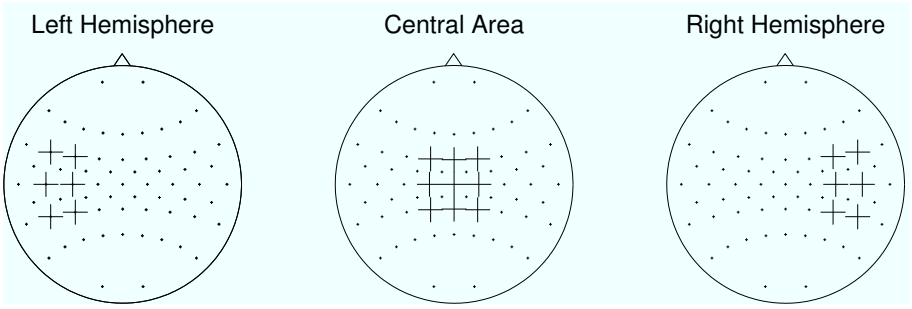

Figure 3: Two Laplacian derivations are selected from each of the scalp areas to complement the CSFs. The selection of the derivations is subject-dependent and performed in a trial basis.

The selection of the Laplacian channels is updated every trial. It is based on the value of a robust Fisher score (see eq. 20) obtained from every channel and after every trial (using the last 100 trials, that can partly originate from a previous run).

$$\text{score} = \frac{(\text{median}(\boldsymbol{x}_1) - \text{median}(\boldsymbol{x}_2))^2}{\text{var}(\boldsymbol{x}_1) + \text{var}(\boldsymbol{x}_2)} \tag{20}$$

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 trial-based selected.

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 16 dimensions, a regularized version of LDA is used. We follow the approach by (Ledoit and Wolf, 2004a,b) and replace $\mathbf{\Sigma}$ in Eq. (2) by a shrinkage estimate of the form

$$\mathbf{\Sigma}_\gamma \quad = \quad (1-\gamma)\,\mathbf{\Sigma} + \gamma\widetilde{\mathbf{\Sigma}},\ \gamma \in [0,1]\,.$$

The matrix $\widetilde{\mathbf{\Sigma}}$ is the sample covariance matrix of a restricted sub-model, and for some sub-models, 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 et al., 2009). In the present study, we use the following sub-model: 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:

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

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

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

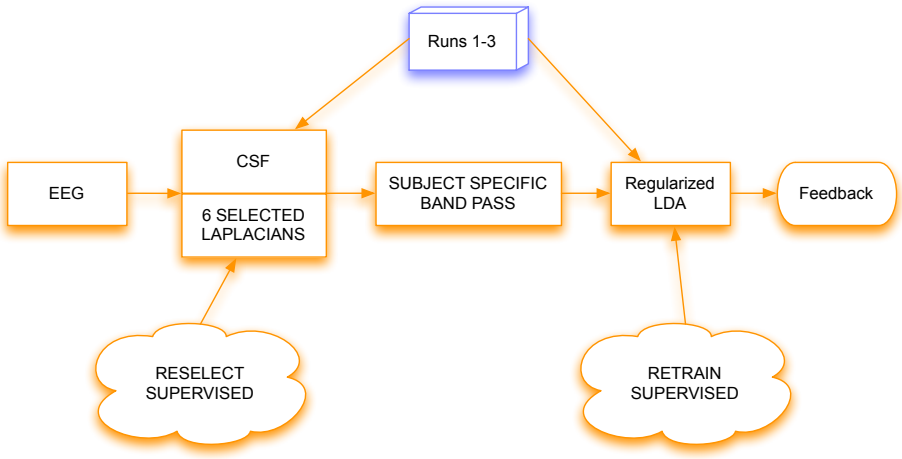

Figure 4: Schema of adaptation level 2. The data recorded in runs 1 to 3 are used to calculate CSF and 6 laplacian channels. The frequency band is subject-optimized. During the runs 4 to 6, the laplacian channels are trial-based re-selected (using the robust Fischer score) and the classifier retrained.

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 the 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 allowing to track features' drifts and at the same time estimate 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 3.2.1).

3.3.1 Unsupervised adaptation of the classifier

The technique used in this paper for the unsupervised adaptation of the linear classifier has been previously studied off-line in (Vidaurre et al., 2008). Here we use the adaptation of the pooled-mean, with the formula in Eq. (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 the present study is 0.05. Figure 5 depicts a schema of this adaptation level.

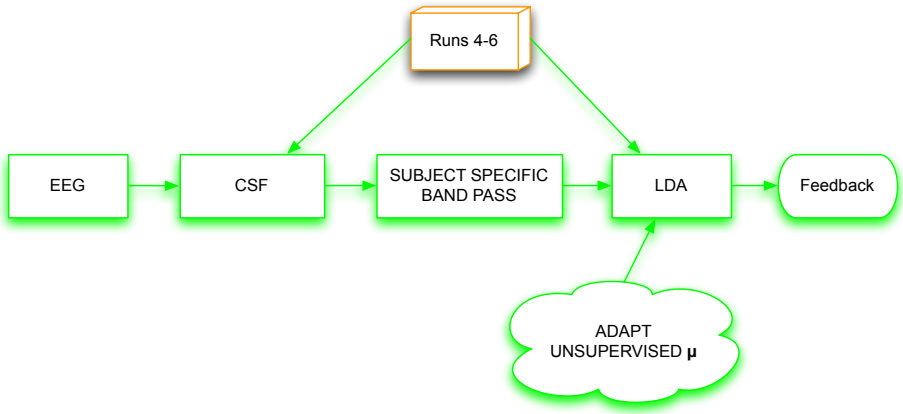

Figure 5: Schema of adaptation level 3. The data recorded in runs 4 to 6 are used to calculate CSF. The frequency band is subject optimized as in runs 4 to 6. During the feedback, the pooled mean is updated without using class information.

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 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 different levels. The black dashed line marks the performance 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 (e.g. (Williamson et al., 2009)). The red dashed line marks the performance level of 50% (random performance in two-class systems). The first 3 runs 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. Additionally, table 1 presents the average and standard deviation of the users' performance in the 3 different levels. Again, the users were grouped according their Category.

Table 1: Mean and Standard deviation of the BCI performance. Subjects are divided in Categories and the performance is divided in levels.

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

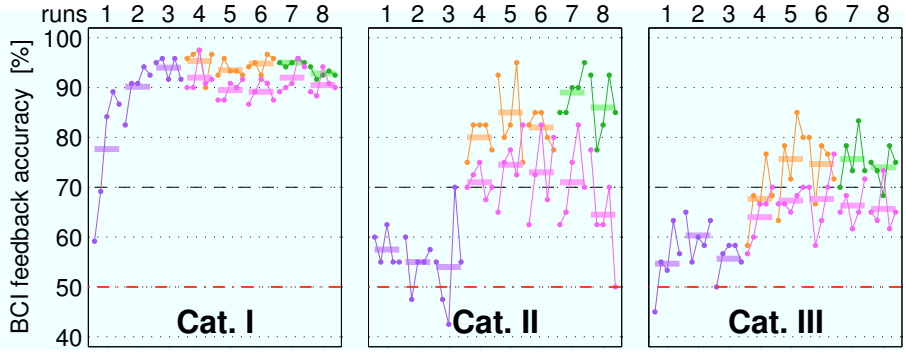

Figure 6: Grand average of BCI performance within each run (horizontal bars and dots for each group of 20 trials) according to adaptation levels for subjects of Cat. I (N=6), Cat. II (N=2) and Cat. III (N=3). An accuracy of 70% is assumed to be a threshold required for BCI applications (dash-point black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

The first finding is that participants of Cat. I (good calibration and feedback) can have very high performance with the novel co-adaptive system which corresponds to level 1, within just 20-40 trials (i.e. around 3-6 minutes). However, volunteers of Cat. II and III are not able to achieve control with this BCI system within the first 3 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 Cat. I users, the simulated performance is slightly worse than the obtained in run 3 (last of 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 Cat. I participants can adjust their performance to the system they receive feedback from and achieve improved results. For Cat. II and (less prominent for) Cat. III volunteers, we can observe that the simulated performance increased notably after the first 3 runs. Consequently, the increase in performance is not only due 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/ERS curves [1] of two participants in runs 1 and 2 (beginning of first level) versus runs 5 and 6 (end of second level). The horizontal bands on the bottom represent the

[1] The definition of ERD and ERS can be found, e.g, in (Pfurtscheller, 1992), "event-related desynchronization (ERD) describes the short lasting (phasic) and regional localized 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".

discriminability between the classes. They are computed as minus the 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 form of a bar next to the ERD/ERS curves. Those laplacian channels were the ERD/ERS discriminability between classes is better visible are displayed, together 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 Cat. II and III users.

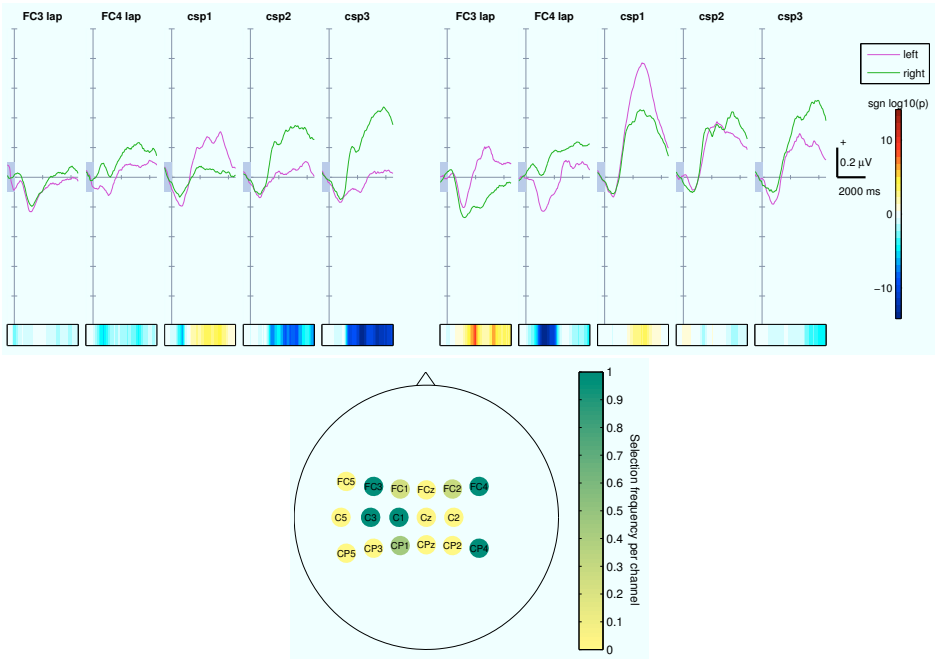

Figure 7: ERD/ERS curves of signals filtered in subject-selected Laplacians and CSP. Those laplacian channels where the effect is best visible are displayed. The top left image displays the curves in the first 2 runs of the experiments. The top right image displays the curves in runs 5 and 6. We can see how ERD/ERS of CSP deteriorates over time and at the same time some Laplacian channels start to be more significant in the performing tasks. 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 2 channels per area are selected (the total frequency per area is 2 and the greatest frequency per channel is 1). This is a Cat. II participant (number 7 in Figure 12), also corresponding to Figure 10.

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. As in this level is adaptive but unsupervised, it also provides an estimate of the performance of the users in the end of the session. As an illustration of the adaptation of the classifier, Figure 9 shows in a low dimensional projection that

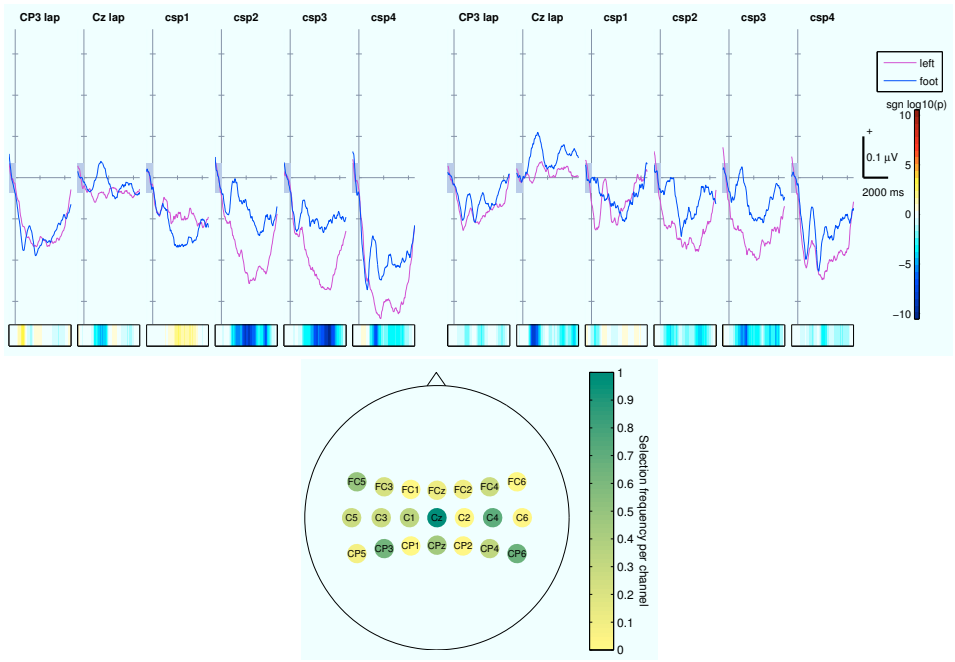

Figure 8: ERD/ERS curves of signals filtered in subject-selected Laplacians and CSP. Those laplacian channels where the effect is best visible are displayed. The top left image displays the curves in the first 2 runs of the experiments. The top right image displays the curves in runs 5 and 6. We can see how ERD/ERS of CSP deteriorates over time and at the same time some Laplacian channels start to be more significant in the performing tasks. 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 2 channels per area are selected (the total frequency per area is 2 and the greatest frequency per channels is 1). This is a Cat. III participant (number 10 in Figure 12), also corresponding to Figure 13.

the feature distributions change over time and that the classifier boundary moves. It also depicts the advantage of using such adaptation in on-going BCI sessions. Note that the unsupervised adaptation method allows only parallel shifts of the separating hyperplane.

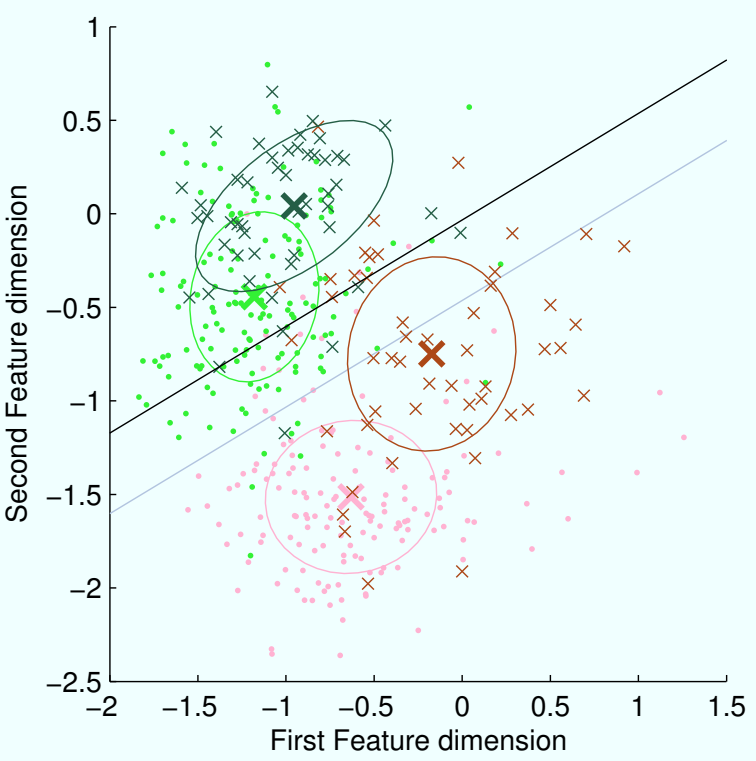

Figure 9: The figure plots the feature distribution of runs 4 to 6 (light green and light red; samples indicated by dots) versus run 8 (dark green and dark red; samples indicated by crosses). The separating lines correspond to the classifiers trained in runs 4 to 6 (gray) and the unsupervised classifier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

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 Cat. II and Cat. 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/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 0% of error. On the right of the same figure,

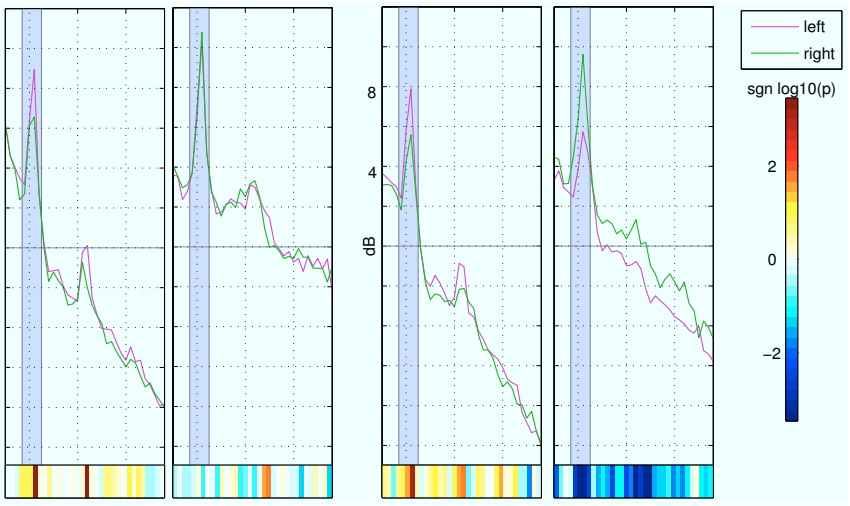

Figure 10: Spectra of subject-selected Laplacians. Left: spectra of the channels were the effect is better visible, in the first 2 runs of the experiments. Right: spectra in the last two runs. We can see how the discriminability has increases over the session. The figure corresponds to a Cat. II participant (number 7 in Figure 12).

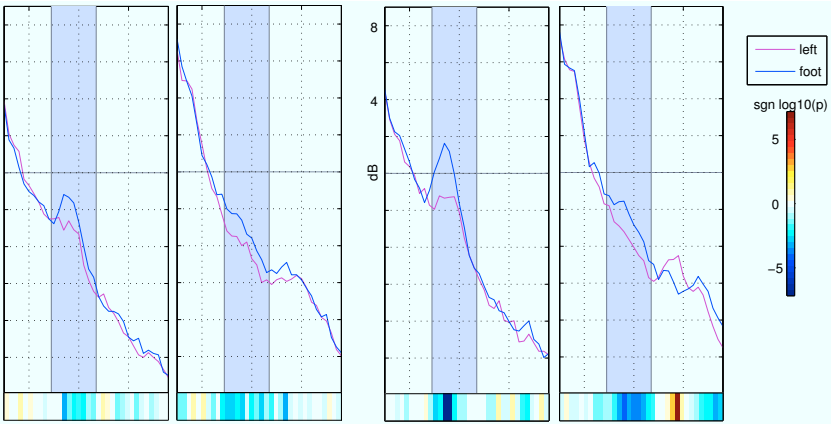

Figure 11: Spectra of subject-selected Laplacians. Left: spectra of the channels were the effect is better visible, in the first 2 runs of the experiments. Right: spectra in the last two runs. We can see how the discriminability has increased over the session. The figure corresponds to a Cat. III participant (number 10 in Figure 12).

there is an example of the scalp maps for a Cat. III participant that we use to estimate the difference between the beginning and end of the experimental session.

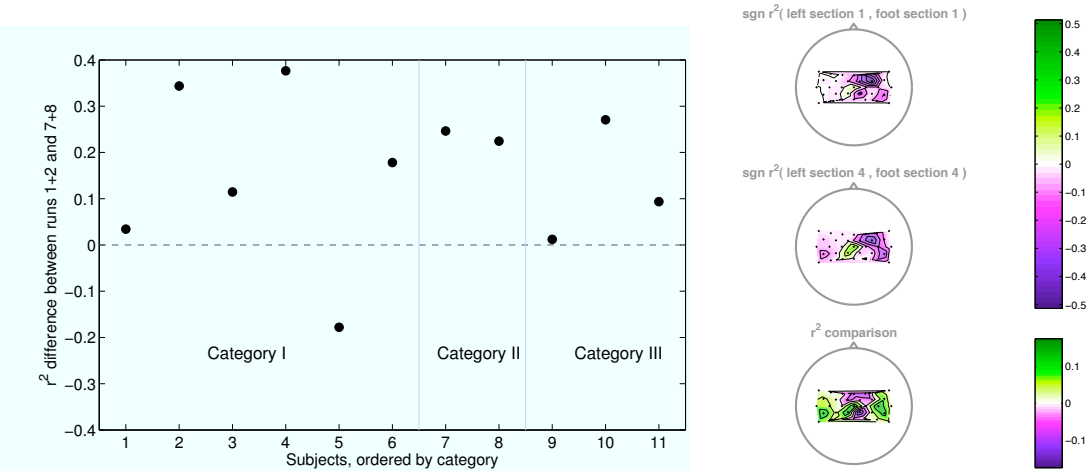

Figure 12: Left: discriminative difference of features at the beginning and end of the session. The features were extracted from laplacian channels over the motor areas of interest that were selected during the second level of adaptation. The frequency band and time window to extract the features correspond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

Finally, Figure 13 illustrates a Cat. 3 participant who could not perform any meaningful feedback in 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 is to find out whether the co-adaptive approach works as well as the state-of-the-art system (with offline calibration). This is clearly confirmed with the performance of Cat. I participants. However, our main goal is to challenge the system with participants of Cat. II and III. The Cat. I participants (except the naive user) selected for this study had reached performances over 85% in previous BCI sessions. Cat. II participants exhibited 20-25% 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 et al., 2010). Finally, none of the Cat. III users had been able to perform feedback before the current study. The selection of participants together with the results presented in the previous section demonstrate 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, Fig. 6 illustrates that "good" BCI users (i.e. Cat. I) obtain accurate control in level 1, after a short period of adaptation (about 6 minutes). Users with less typical patterns of motor imagery (i.e.

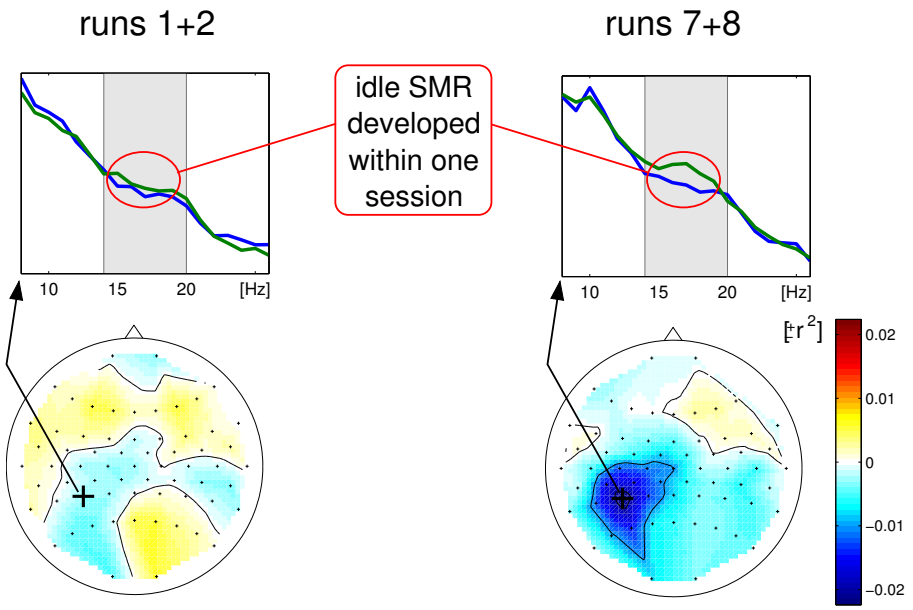

Figure 13: A participant of Cat. III that could develop SMR over the session (user number 10 in Figure 12).

Cat. II) need a more flexible way of adaptation to their specific patterns of brain activity, as provided by the adaptive training of the classifier for level 2. Also, users who lack a suitable strategy to generate discriminable patterns by motor imagery in a classical approach (i.e. Cat. III) can be guided to develop a successful strategy during ML-based co-adaptive learning, i.e. 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 co-adaptation 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 to track class specific changes. We choose to adapt the bias but adapting the global covariance matrix is also possible and allows in principle to account 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, 13 show spectral differences of a Cat. III volunteer. A similar effect can be seen in Figure 11 for a Cat. II user. A general comparison of the spectral discriminative

power over the motor area, further supports our hypothesis that co-adaptive 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 manuscript , cf. (Vidaurre et al., 2006, 2007), this co-adaptive study is the first one to deal with users that 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 non-average users to achieve control within one 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 for any type of SMR modulation at any location of the sensorimotor cortex and in any frequency band, which additionally can change in time.

Finally, from the point of view of the BCI user, the co-adaptive approach is more interesting because feedback starts from the very beginning of the session. Furthermore, it is more motivating because the performance increases, which in turn might yield to a more successful experiment (Nijboer et al., 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, 2008b; Dornhege et al., 2004, 2007; Müller et al., 2003, 2008; Tomioka and 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 non-functional for feedback (see (Dornhege et al., 2007; Sugiyama et al., 2007; Krauledat, 2008; von Bünau et al., 2009) for a discussion of non-stationarities in BCI). Moreover, some users have difficulties to properly perform separable motor imagery patterns in the absence of feedback (offline calibration). Here, we have presented a novel Machine Learning method which helps to 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 very quickly (3-6 mins) to gain accurate BCI control. Other BCI users, who could not gain BCI control with a classic Machine Learning approach (i.e. belonging to Cat. II or II), 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 to cure BCI illiteracy. Further studies will focus on the challenging Cat. III users and also on uncategorized participants to demonstrate that the co-adaptive approach works also for totally BCI-naive users. Also, we intend to support our findings with fMRI analysis to better investigate the change in brain activity caused by the co-adaptive training.

Acknowledgment

The authors would like to 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) Vital-BCI Project, EU TOBI Project ICT-224631. This publication only reflects the authors' views. Funding agencies are not liable for any use that may be made of the information contained herein.

References

Allison B, Wolpaw E and Wolpaw J (2007). Brain-computer interface systems: progress and prospects. Expert Rev Med Devices, 4(4):463–474.

Birbaumer N, Weber C, Neuper C, Buch E, Haapen K and 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 KR, Curio G (2007). The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. NeuroImage 37(2):539–550.

Blankertz B, Losch F, Krauledat M, Dornhege G, Curio G, Müller KR (2008a). The Berlin Brain-Computer Interface: Accurate performance from first-session in BCI-naive subjects. IEEE Trans Biomed Eng 55(10):2452–2462.

Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR (2008b). Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process Mag 25(1):41–56.

Blankertz B, Tangermann M, Vidaurre C, Dickhaus T, Sannelli C, Popescu F, Fazli S, Danóczy M, Curio G and Müller K-R (2009). Non-Invasive and Invasive Brain-Computer Interfaces. In Allison, B., Graimann, B., and Pfurtscheller, G. editors, *The Frontiers Collection.* Springer, in press.

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, in press. DOI:10.1016/j.neuroimage.2010.03.022

Blumberg J, Rickert J, Waldert S, Schulze-Bonhage A, Aertsen A. and Mehring C (2007). Adaptive classification for brain computer interfaces. In Conf Proc IEEE Eng Med Biol Soc 2007, 2536–2539.

von Bünau P, Meinecke FC, Kiraly F, Müller KR (2009). Finding Stationary Subspaces in Multivariate Time Series, Physics Review Letters, 103(21):214101.

Buttfield A, Ferrez PW, and Millán JdR (2006). Towards a robust BCI: Error recognition and online learning, IEEE Trans. Neural Sys. Rehab. Eng., 14:164–168.

Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov DF, Patil PG, Henriquez CS, and Nicolelis MAL (2003). Learning to control a brain-machine interface for reaching and grasping by primates. PLoS Biol., 1:193–208.

Dickhaus T, Sannelli C, Müller KR, Curio G, Blankertz B (2009). Predicting BCI performance to study BCI illiteracy. BMC Neuroscience 2009 10:(Suppl 1):P84.

Dornhege G, Blankertz B, Curio G, Müller KR (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 KR (eds) (2007). Toward Brain-Computer Interfacing. MIT Press, Cambridge, MA.

Fazli S, Popescu F, Danoczy M, Blankertz B, Müller KR, Grozea C (2009). Subject independent mental state classification in single trials. Neural Networks, 22(9):1305–1312

Fetz EE (2007). Volitional control of neural activity: implications for brain-computer interfaces. J Physiol. 579:571–579.

Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH, Branner A, Chen D, Penn RD, and Donoghue JP (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 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 Managament 30:110–119.

Ledoit O. and Wolf M. (2004b). A well-conditioned estimator for largedimensional covariance matrices. Journal of Multivariate Analysis, 88:365–411.

Leuthardt EC, Schalk G, Wolpaw JR, Jemann JG, and Oran DW (2004). A brain-computer interface using electrocorticographic signals in humans. J. Neural Eng., 1:63–71.

Leuthardt EC, Freudenberg 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.

Lu S, Guan C and Zhang H (2009). Unsupervised brain computer interface based on intersubject information and online adaptation. IEEE Trans. Neural Sys. Rehab. Eng, 17:135–145.

McFarland DJ, McCane LM, David SV and Wolpaw JR. (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 KR, Anderson CW, Birch GE (2003). Linear and non-linear methods for Brain-Computer Interfaces. IEEE Trans Neural Sys Rehab Eng, 11(2):165–169.

Müller KR, 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 Andrea (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 VV, Hohlefeld FU, Jacobs AM, Curio G (2008). Quasi-movements: a novel motor-cognitive phenomenon. Neuropsychologia. 46:727–742.

Pistohl T, Ball T, Schulze-Bonhage A, Aertsen A, and 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(1):62–69.

Pfurtscheller G, Neuper C and Birbaumer N (2005). Human Brain-Computer Interface. In Riehle, A. and Vaadia, E., editors, Motor Cortex in Voluntary Movements, chapter 14, pages 367–401. CRC Press, New York.

Rizk M, Bossetti CA, Jochum TA, Callender SH, Nicolelis MA, Turner DA, Wolf PD (2009). A fully implantable 96-channel neural data acquisition system. J Neural Eng. 6:026002.

Schäfer J and Strimmer K (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statistical Applications in Genetics and Molecular Biology:4–32.

Schalk G, Miller KJ, Anderson NR, Wilson JA, Smyth MD, Ojemann JG, Moran DW, Wolpaw JR, Leuthardt EC (2008). Two-dimensional movement control using electrocorticographic signals in humans. Journal of Neural Engineerning, 5:75–84.

Schwartz AB (2004). Cortical neural prosthetics. Annu Rev Neurosci. 27:487–507.

Shenoy P, Krauledat M, Blankertz B, Rao RPN, Müller KR (2006). Towards adaptive classification for BCI. Journal of Neural Engineering, 3(1):R13–R23.

Sugiyama M, Krauledat M, Müller KR (2007). Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research, 8:1027–1061.

Tomioka R, Müller KR (2010) A regularized discriminative framework for EEG based communication. Neuroimage, 49(1):415–432.

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.

Vidaurre C, Schlögl A, Blankertz B, Kawanabe M, Müller KR (2008). Unsupervised adaptation of the LDA classifier for Brain-Computer Interfaces. In: Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course 2008, Verlag der Technischen Universität Graz, pp 122–127.

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. and Blankertz B. (2010). Towards a cure for BCI illiteracy. Brain Topography. 23:194–198.

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, 5059–5062.

Williamson J, Murray-Smith R, Blankertz B, Krauledat M and Müller KR (2009). Designing for uncertain, asymmetric control: Interaction design for brain-computer interfaces. International Journal of Human-Computer Studies, 67(10):827–841.

Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002). Brain-computer interfaces for communication and control. Clinical Neurophysiology 113:767–791.