Optimizing Spatial Filters for Robust EEG Single-Trial Analysis

Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, Klaus-Robert Müller

***Abstract*—Due to the volume conduction multi-channel electroencephalogram (EEG) recordings give a rather blurred image of brain activity. Therefore spatial filters are extremely useful in single-trial analysis in order to improve the signal-to-noise ratio. There are powerful methods from machine learning and signal processing that permit the optimization of spatio-temporal filters for each subject in a data dependent fashion beyond the fixed filters based on the sensor geometry, e.g., Laplacians. Here we elucidate the theoretical background of the Common Spatial Pattern (CSP) algorithm, a popular method in Brain-Computer Interface (BCI) research. Apart from reviewing several variants of the basic algorithm, we reveal tricks of the trade for achieving a powerful CSP performance, briefly elaborate on theoretical aspects of CSP and demonstrate the application of CSP-type preprocessing in our studies of the Berlin Brain-Computer Interface project.**

I. Introduction

Noninvasive Brain-Computer Interfacing (BCI) has in the recent years become a highly active research topic in neuroscience, engineering and signal processing. One of the reasons for this development is the striking advances of BCI systems with respect to usability, information transfer and robustness for which modern machine learning and signal processing techniques have been instrumental [2], [14], [15], [4]. Invasive BCIs ([46]), in particular intracranial signals, require completely different signal processing methods and are therefore not discussed here.

The present paper will review a particularly popular and powerful signal processing technique for EEG-based BCIs called common spatial patterns (CSP) and discusses recent variants of CSP. Our goal is to provide comprehensive information about CSP and its application. Thus we address both the BCI expert who is not specialized in signal processing and to the BCI novice who is an expert in signal processing.

Consequently the present paper will mainly focus on CSP filtering (Sec. III), but we will also briefly discuss BCI paradigms and the neurophysiological background thereof (Sec. II). Finally we will report on recent results achieved with the Berlin Brain-Computer Interface using advanced signal processing and machine learning techniques (Sec. IV-A).

This work was supported in part by grants of the *Bundesministerium für Bildung und Forschung* (BMBF), FKZ 01IBE01A/B, and by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors' views. Author RT was supported by MEXT, Grant-in-Aid for JSPS fellows, 17-11866.

BB and KRM are with the Machine Learning Laboratory, Technical University of Berlin, Germany. BB and KRM are also with Fraunhofer FIRST (IDA), Berlin, Germany. E-mail: {blanker,krm}@cs.tu-berlin.de.

RT is with Dept. Mathematical Informatics, IST, The University of Tokyo, Japan and Fraunhofer FIRST (IDA), Berlin, Germany.

SL and MK are with Fraunhofer FIRST (IDA), Berlin Germany.

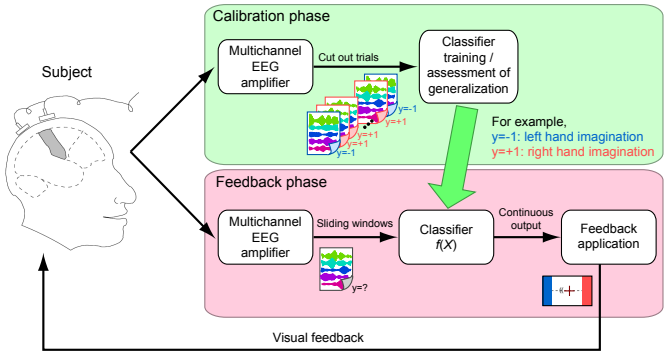

Fig. 1. Overview of the machine-learning-based BCI system. The system runs in two phases. In the calibration phase, we instruct the subjects to perform certain tasks and collect short segments of labeled EEG (trials). We train the classifier based on these examples. In the feedback phase, we take sliding windows from continuous stream of EEG; classifier outputs a real value that quantifies the likeliness of class membership; we run a feedback application that takes the output of the classifier as an input. Finally the subject receives the feedback on the screen as, e.g., cursor control.

II. Background

A. Overview of a BCI system

An overview of a BCI system based on machine learning is shown in Fig. 1. The system operates in two phases, namely the calibration phase and the feedback phase. The feedback phase is the time the users can actually transfer information through their brain activity and control applications; in this phase, the system is composed of the classifier that classifies between different mental states and the user interface that translates the classifier output into control signals, e.g., cursor position or selection from an alphabet. In the calibration phase, we collect examples of EEG signals in order to train the classifier. Here we describe a typical experiment as performed in the Berlin BCI (BBCI) project. We use three types of imaginary movements, namely, left hand (L), right hand (R) and right foot (F) as the mental states to be classified. Other paradigms based on, e.g., modulation of attention to external stimulation can be found in [55]. The subjects are instructed to perform one of the three imaginary movements[1] indicated on the screen for 3.5 seconds at the interval of 5.5 seconds. We obtain 420 trials of imaginary movement (140 for each

[1]For more effective performance it is important to instruct the subjects to concentrate on the kinesthetic aspect rather than the visual ([37]).

class) in a randomized order for each subject (less is sufficient for feedback performance). The data is then used for the training of the classifier and assessment of generalization error by cross-validation. In particular, we compare three pair-wise classifiers and select the combination of two classes that yields the best generalization performance.

After the calibration measurement subjects perform 5 *feedback sessions* consisting of 100 runs. Here the output of the binary classifier is translated into the horizontal position of a cursor. Subjects are instructed to move the cursor to that one of the two vertical bars at the edges of the screen which was indicated as target by color. The cursor is initially at the center of the screen; it starts to follow the classifier output based on the brain signal 750 ms after the indication of the target. A trial ends when the cursor touches one of the two bars; the bar that the cursor reached is colored green if correct and red otherwise. The next trial starts after 520 ms (see [2], [4], [7] for more details).

The performance of the classifier is measured by the accuracy of the prediction in percent. The performance of the overall system is measured by the information transfer rate (ITR, [54]) measured in bits per minute (bpm):

$$\mathrm{ITR} = \frac{\#\text{ of decisions}}{\text{duration in minutes}} \cdot \left(p\log_2(p) + (1-p)\log_2\left(\frac{1-p}{N-1}\right) + \log_2(N) \right) \qquad (1)$$

where p is the accuracy of the subject in making decisions between N targets, e.g., in the feedback explained above, $N = 2$ and p is the accuracy of hitting the correct bars. ITR measures the capacity of a symmetric communication channel that makes mistake with the equal probability $(1-p)/(N-1)$ to all other $N-1$ classes divided by the time required to communicate that amount of information. The ITR depends not only on the accuracy of the classifier but also on the design of the feedback application that translates the classifier output into command. Note that the *duration in minutes* refers to the total duration of the run including all inter-trial intervals. In contrast to the accuracy of the decision, the ITR takes different duration of trials and different number of classes into account. Note that the communication channel model can be generalized to take the nonsymmetric or nonuniform errors into account [44].

Brain activity was recorded from the scalp with multi-channel EEG amplifiers (BrainAmp by Brain Products, Munich, Germany) using 55 Ag/AgCl electrodes in an extended 10-20 system.

B. Neurophysiological Background

Macroscopic brain activity during resting wakefulness comprises distinct 'idle' rhythms located over various cortical areas, e.g. the occipital α-rhythm (8–12 Hz) can be measured over the visual cortex [1]. The perirolandic sensorimotor cortices show rhythmic macroscopic EEG oscillations (μ-rhythm, sensori motor rhythm, SMR) ([24], [20]), with spectral peak energies of about 8–14 Hz (localized predominantly over the postcentral somatosensory cortex) and around 20 Hz (over the precentral motor cortex). The occipital α-rhythm is quite prominent and can be seen in the raw EEG with the naked eye if the subject closes the eyes (idling of the visual cortex). In contrast the μ-rhythm has a much weaker amplitude and can only be observed after appropriate signal processing. In some subjects no μ-rhythm can be observed in scalp EEG.

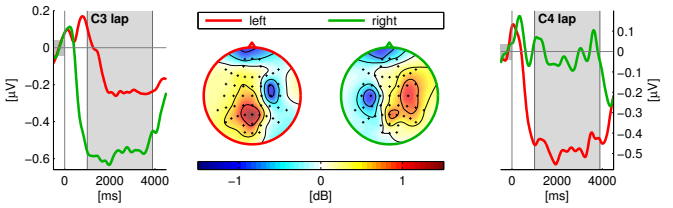

Fig. 2. Event-Related Desynchronization (ERD) during motor imagery of the left and the right hand. Raw EEG signals of one subject have been band-pass filtered between 9 and 13 Hz. For the time courses, the envelope of the signals has been calculated by Hilbert transform (see e.g., [9]) and averaged over segments of -500 to 4500 ms relative to each cue for left or right hand motor imagery. ERD curves are shown for Laplace filtered channels at C3 and C4, i.e. over left and right primary motor cortex. The topographical maps of ERD were obtained by performing the same procedure for all (non Laplace filtered) channels and averaging across the shaded time interval 1000 to 4000 ms.

Our system is based on the modulation of the SMR. In fact, motor activity, both actual and *imagined* [25], [42], [45], as well as somatosensory stimulation [38] have been reported to modulate the μ-rhythm. Processing of motor commands or somatosensory stimuli causes an attenuation of the rhythmic activity termed event-related desynchronization (ERD) [42], while an increase in the rhythmic activity is termed event-related synchronization (ERS). For BCIs the important fact is that the ERD is caused also by *imagined* movements (healthy users, see Fig. 2) and by *intented* movements in paralyzed patients ([30]).

For 'decoding' of different motor intentions from brain activity, the essential task is to distinguish different spatial localization of SMR modulations. Due to the topographical arrangement in the motor and somatosensori cortex, these locations are related to corresponding parts of the body, cf. Fig. 3. For example, left hand and right hand have corresponding areas in the contralateral, i.e., right and left motor cortex, respectively; see Fig. 2.

C. Why Spatial Filtering is Important

Raw EEG scalp potentials are known to have a poor spatial resolution owing to volume conduction. In a simulation study in [39] only half the contribution to each scalp electrode came from sources within a 3 cm radius. This is in particular a problem if the signal of interest is weak, e.g. sensorimotor rhythms, while other sources produce strong signals in the same frequency range like the α-rhythm of the visual cortex or movement and muscle artifacts.

The demands are carried to the extremes when it comes to single-trial analysis as in BCI. While some approaches try to achieve the required signal strength by *training the subjects* ([53], [30]) an alternative is to *calibrate the system*

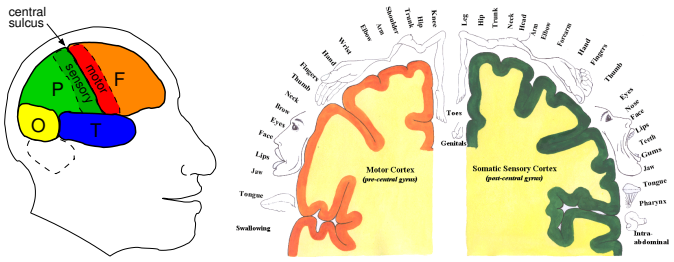

Fig. 3. *Left.* Lobes of the brain: Frontal, Parietal, Occipital, and Temporal (named after the bones of the skull beneath which they are located). The central sulcus separates the frontal and parietal lobe. *Right.* Geometric mapping between body parts and motor/somatosensory cortex. The motor cortex and the somatosensory cortex are shown at the left and right part of the figure, respectively. Note, that in each hemisphere there is one motor area (frontal to the central sulcus) and one sensori area (posterior to the central sulcus). The part which is not shown can be obtained by mirroring the figure folded at the center.

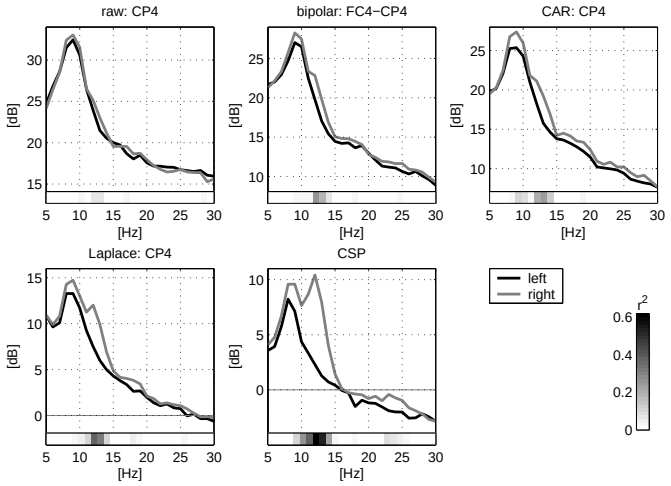

Fig. 4. Spectra of left vs. right hand motor imagery. All plots are calculated from the same dataset but using different spatial filters. The discrimination between the two conditions is quantified by the r^2-value. CAR stands for common average reference.

to the specific characteristics of each user ([19], [2], [7]). For the latter data-driven approaches to calculate subject-specific spatial filters have proven to be useful.

As a demonstration of the importance of spatial filters, Fig. 4 shows spectra of left vs. right hand motor imagery at the right hemispherical sensorimotor cortex. All plots are computed from the same data but using different spatial filters. While the raw channel only shows a peak around 9 Hz that provides almost no discrimination between the two conditions, the bipolar and the common average reference filter can improve the discrimination slightly. However the Laplace filter and even more the CSP filter reveal a second spectral peak around 12 Hz with strong discriminative power. By further investigations the spatial origin of the non-discriminative peak could be traced back to the visual cortex, while the discriminative peak originates from sensorimotor rhythms. Note that in many subjects the frequency ranges of visual and sensorimotor rhythms overlap or completely coincide.

III. Methods

A. General framework

Here we overview the classifier we use. Let $X \in \mathbb{R}^{C\times T}$ be a short segment of EEG signal[2], which corresponds to a trial of imaginary movement; C is the number of channels and T is the number of sampled time points in a trial. A classifier is a function that predicts the label of a given trial X. For simplicity let us focus on the binary classification e.g., classification between imagined movement of left and right hand. The classifier outputs a real value whose sign is interpreted as the predicted class. The classifier is written as follows:

$$f(X;\{\boldsymbol{w}_j\}_{j=1}^J,\{\beta_j\}_{j=0}^J) = \sum_{j=1}^J \beta_j \log\left(\boldsymbol{w}_j^\top X X^\top \boldsymbol{w}_j\right) + \beta_0. \quad (2)$$

The classifier first projects the signal by J spatial filters $\{\boldsymbol{w}_j\}_{j=1}^J \in \mathbb{R}^{C\times J}$; next it takes the logarithm of the power of the projected signal; finally it linearly combines these J dimensional features and adds a bias β_0. In fact, each projection captures different spatial localization; the modulation of the rhythmic activity is captured by the log-power of the band-pass filtered signal. Note that various extensions are possible (see Sec. V-D). A different experimental paradigm might require the use of nonlinear methods of feature extraction and classification respectively [33]. Direct minimization of discriminative criterion [17] and marginalization of the classifier weight [22] are suggested. On the other hand, methods that are linear in the second order statistics $XX^\top$, i.e., Eq. (2) without the log, are discussed in [49], [48] and shown to have some good properties such as convexity.

The coefficients $\{\boldsymbol{w}_j\}_{j=1}^J$ and $\{\beta_j\}_{j=1}^J$ are automatically determined statistically ([21]) from the training examples i.e., the pairs of trials and labels $\{X_i, y_i\}_{i=1}^n$ we collect in the calibration phase; the label $y \in \{+1,-1\}$ corresponds to, e.g., imaginary movement of left and right hand, respectively, and n is the number of trials.

We use Common Spatial Pattern (CSP) [18], [27] to determine the spatial filter coefficients $\{\boldsymbol{w}_j\}_{j=1}^J$. In the following, we discuss the method in detail and present some recent extensions. The linear weights $\{\beta_j\}_{j=1}^J$ are determined by Fisher's linear discriminant analysis (LDA).

B. Introduction to Common Spatial Patterns Analysis

Common Spatial Pattern ([18], [27]) is a technique to analyze multi-channel data based on recordings from two classes (conditions). CSP yields a data-driven supervised decomposition of the signal parameterized by a matrix $W \in \mathbb{R}^{C\times C}$ (C being the number of channels) that projects the signal $\boldsymbol{x}(t) \in \mathbb{R}^C$ in the original sensor space to $\boldsymbol{x}_{\mathrm{CSP}}(t) \in \mathbb{R}^C$, which lives in the surrogate sensor space, as follows:

$$\boldsymbol{x}_{\mathrm{CSP}}(t) = W^\top \boldsymbol{x}(t).$$

[2]In the following, we also use the notation $\boldsymbol{x}(t) \in \mathbb{R}^C$ to denote EEG signal at a specific time point t; thus X is a column concatenation of $\boldsymbol{x}(t)$'s as $X = [\boldsymbol{x}(t), \boldsymbol{x}(t+1), \ldots, \boldsymbol{x}(t+T-1)]$ for some t but the time index t is omitted. For simplicity we assume that X is already band-pass filtered, centered and scaled i.e., $X = \frac{1}{\sqrt{T}} X_{\mathrm{band-pass}}(I_T - \boldsymbol{1}_T \boldsymbol{1}_T^\top)$, where I_T denotes $T \times T$ identity matrix and $\boldsymbol{1}_T$ denotes a T-dimensional vector with all one.

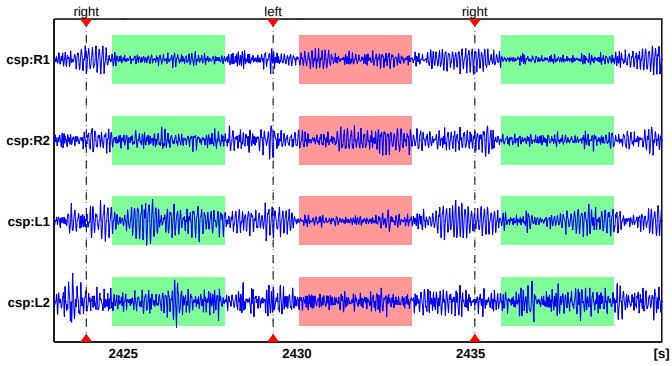

Fig. 5. Effect of spatial CSP filtering. CSP analysis was performed to obtain 4 spatial filters that discriminate left from right hand motor imagery. The graph shows continuous band-pass filtered EEG after applying the CSP filters. The resulting signals in filters CSP:L1 and CSP:L2 have larger variance during right hand imagery (segments shaded in green) while signals in filters CSP:R1 and CSP:R2 have larger variance during left hand imagery (segment shaded red).

In this paper, we call each column vector $\boldsymbol{w}_j \in \mathbb{R}^C$ $(j = 1,\ldots,C)$ of W a *spatial filter* or simply a filter; moreover we call each column vector $\boldsymbol{a}_j \in \mathbb{R}^C$ $(j = 1,\ldots,C)$ of a matrix $A = (W^{-1})^\top \in \mathbb{R}^{C\times C}$ a *spatial pattern* or simply a pattern. In fact, if we think of the signal spanned by A as $\boldsymbol{x}(t) = \sum_{j=1}^{C} \boldsymbol{a}_j s_j(t)$, each vector $\boldsymbol{a}_j$ characterizes the spatial pattern of the j-th activity; moreover, $\boldsymbol{w}_j$ would filter out all but the j-th activity because the orthogonality $\boldsymbol{w}_j^\top \boldsymbol{a}_k = \delta_{jk}$ holds, where δ_{jk} is the Kronecker delta ($\delta_{jk} = 1$ for $j = k$ and $= 0$ for $j \neq k$). The matrices A and W are sometimes called the mixing and de-mixing matrix or the forward and backward model ([41]) in other contexts.

The optimization criterion that is used to determine the CSP filters will be discussed in detail in the subsequent Sec. III-C. In a nutshell, CSP filters maximize the variance of the spatially filtered signal under one condition while minimizing it for the other condition. Since variance of band-pass filtered signals is equal to band-power, CSP analysis is applied to approximately band-pass filtered signals in order to obtain an effective discrimination of mental states that are characterized by ERD/ERS effects (Sec. II-B). Fig. 5 shows the result of applying 4 CSP filters to continuous band-pass filtered EEG data. Intervals of right hand motor imagery are shaded green and show larger variance in the CSP:L1 and CSP:L2 filters, while during left hand motor imagery (shaded red) variance is larger in the CSP:R1 and CSP:R2 filters. See also the visualization of spatial maps of CSP analysis in Sec. IV-B.

C. Technical Approaches to CSP Analysis

Let $\Sigma^{(+)} \in \mathbb{R}^{C\times C}$ and $\Sigma^{(-)} \in \mathbb{R}^{C\times C}$ be the estimates of the covariance matrices of the band-pass filtered EEG signal in the two conditions (e.g., left hand imagination and right hand imagination):

$$\Sigma^{(c)} = \frac{1}{|\mathscr{I}_c|} \sum_{i\in\mathscr{I}_c} X_i X_i^\top \qquad (c \in \{+,-\}) \tag{3}$$

where $\mathscr{I}_c$ $(c \in \{+,-\})$ is the set of indices corresponding to trials belonging to each condition and $|\mathscr{I}|$ denotes the size of a set $\mathscr{I}$. The above expression gives a pooled estimated of covariance in each condition because each X is centered and scaled. Then CSP analysis is given by the simultaneous diagonalization of the two covariance matrices

$$\begin{aligned} W^\top \Sigma^{(+)} W &= \Lambda^{(+)}, \\ W^\top \Sigma^{(-)} W &= \Lambda^{(-)}, \qquad (\Lambda^{(c)} \text{ diagonal}) \end{aligned} \tag{4}$$

where the scaling of W is commonly determined such that $\Lambda^{(+)} + \Lambda^{(-)} = I$ ([18]). Technically this can simply[3] be achieved by solving the generalized eigenvalue problem

$$\Sigma^{(+)} \boldsymbol{w} = \lambda \Sigma^{(-)} \boldsymbol{w}. \tag{5}$$

Then Eq. (4) is satisfied for W consisting of the generalized eigenvectors $\boldsymbol{w}_j$ $(j = 1,\ldots,C)$ of Eq. (5) (as column vectors) and $\lambda_j^{(c)} = \boldsymbol{w}_j^\top \Sigma^{(c)} \boldsymbol{w}_j$ being the corresponding diagonal elements of $\Lambda^{(c)}$ $(c \in \{+,-\})$, while λ in Eq. (5) equals $\lambda_j^{(+)}/\lambda_j^{(-)}$. Note that $\lambda_j^{(c)} \geq 0$ is the variance in condition c in the corresponding surrogate channel and $\lambda_j^{(+)} + \lambda_j^{(-)} = 1$. Hence a large value $\lambda_j^{(+)}$ $\left(\lambda_j^{(-)}\right)$ close to one indicates that the corresponding spatial filter $\boldsymbol{w}_j$ yields high variance in the positive (negative) condition and low variance in the negative (positive) condition, respectively; this contrast between two classes is useful in the discrimination. Koles [27] explained that the above decomposition gives a *common* basis of two conditions because the filtered signal $\boldsymbol{x}_{\mathrm{CSP}}(t) = W^\top \boldsymbol{x}(t)$ is uncorrelated in both conditions, which implies 'independence' for Gaussian random variables. Figure 6 explains how CSP works in 2D. CSP maps the samples in Fig. 6(a) to those in Fig. 6(b); the strong correlation between the original two axes is removed and both distributions are simultaneously decorrelated. Additionally the two distributions are maximally dissimilar along the new axes. The dashed lines in Fig. 6 denote the direction of the CSP projections. Note that the two vectors are *not* orthogonal to each other; in fact they are rather almost orthogonal to the direction that the *opponent class* has the maximum variance.

A *generative view* on CSP was provided by [40]. Let us consider the following linear mixing model with nonstationary sources:

$$\boldsymbol{x}_c = A\boldsymbol{s}_c, \qquad \boldsymbol{s}_c \sim \mathscr{N}(0, \Lambda^{(c)}) \qquad (c \in \{+,-\}),$$

where the sources $\boldsymbol{s}_c \in \mathbb{R}^C$ $(c \in \{+,-\})$ are assumed to be uncorrelated Gaussian distributions with covariance matrices $\Lambda^{(c)}$ $(c \in \{+,-\})$ for two conditions respectively. If the empirical estimates $\Sigma^{(c)}$ are reasonably close to the true covariance matrices $A\Lambda^{(c)}A^\top$, the simultaneous diagonalization gives the maximum likelihood estimator of the backward model $W = \left(A^{-1}\right)^\top$.

A *discriminative view* is the following (see also the paragraph *Connection to a discriminative model* in Sec. V-D). Let

[3] In Matlab this can be done by `» W= eig(S1, S1+S2).`

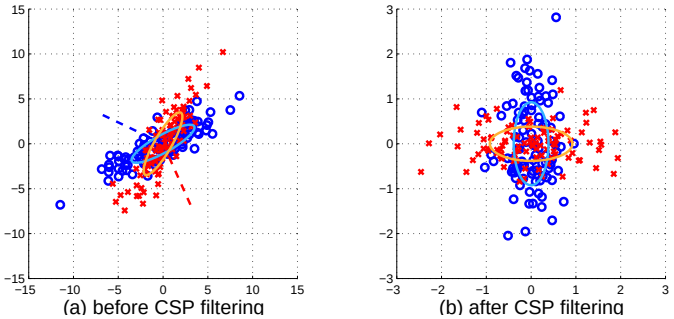

Fig. 6. A toy example of CSP filtering in 2D. Two sets of samples marked by red crosses and blue circles are drawn from two Gaussian distributions. In (a), the distribution of samples before filtering is shown. Two ellipses show the estimated covariances and dashed lines show the direction of CSP projections $\boldsymbol{w}_j$ $(j = 1,2)$. In (b), the distribution of samples after the filtering is shown. Note that both classes are uncorrelated at the same time; the horizontal (vertical) axis gives the largest variance in the red (blue) class and the smallest in the blue (red) class, respectively.

us define S_d and S_c as follows:

$$S_d = \Sigma^{(+)} - \Sigma^{(-)} \quad : \text{discriminative activity}, \tag{6}$$
$$S_c = \Sigma^{(+)} + \Sigma^{(-)} \quad : \text{common activity},$$

where S_d corresponds to the discriminative activity, i.e., the band-power modulation between two conditions and S_c corresponds to the common activity in the two conditions that we are not interested in. Then a solution to the following maximization problem (Rayleigh coefficient) can be obtained by solving the same generalized eigenvalue problem,

$$\underset{\boldsymbol{w}\in\mathbb{R}^C}{\text{maximize}} \quad \frac{\boldsymbol{w}^\top S_d \boldsymbol{w}}{\boldsymbol{w}^\top S_c \boldsymbol{w}}. \tag{7}$$

It is easy to see that every generalized eigenvector $\boldsymbol{w}_j$ corresponds to a local stationary point with the objective value $\lambda_j^{(+)} - \lambda_j^{(-)}$ (assuming $\lambda_j^{(+)} + \lambda_j^{(-)} = 1$ as above). The large positive (or negative) objective value corresponds to large response in the first (or the second) condition. Therefore, the common practice in a classification setting is to use several eigenvectors from both ends of the eigenvalue spectrum as spatial filters $\{\boldsymbol{w}_j\}_{j=1}^J$ in Eq. (2). If the number of components J is too small, the classifier would fail to fully capture the discrimination between two classes (see also the discussion in Sec. V-B on the influence of artifacts); on the other hand, the classifier weights $\{\beta_j\}_{j=1}^J$ could severely overfit if J is too large. In practice we find $J = 6$, i.e., three eigenvectors from both ends, often satisfactory. Alternatively one can choose the eigenvectors according to different criterion (see Sec. A) or use cross-validation to determine the number of components.

D. Feedback with CSP Filters

During BCI feedback the most recent segment of EEG is processed and translated by the classifier into a control signal, see Fig. 1. This can be done according to Eq. (2), where X denotes the band-pass filtered segment of EEG. Due to the linearity of temporal (band-pass) and spatial filtering, these two steps can be interchanged in order. This reduces the computation load (number of signals that are band-pass filtered), since the number of selected CSP filters is typically low (2–6) compared to the number of EEG channels (32–128). Furthermore it is noteworthy, that the length of segment which is used to calculate one time instance of the control signal can be changed during feedback. Shorter segments result in more responsive but also more noisy feedback signal. Longer segments give a smoother control signal, but the delay from intention to control gets longer. This trade-off can be adapted to the aptitude of the subject and the needs of the application. As a caveat, we remark that for optimal feedback the bias of the classifier (β_0 in Eq. (2)) might need to be adjusted for feedback. Since the mental state of the user is very much different during the feedback phase compared to the calibration phase, also the non task related brain activity differs. For a thorough investigation of this issue cf. [29], [47].

TABLE I

RESULTS OF A FEEDBACK STUDY WITH 6 HEALTHY SUBJECTS (IDENTIFICATION CODE IN THE FIRST COLUMN). FROM THE THREE CLASSES USED IN THE CALIBRATION MEASUREMENT (SEE SEC. II-A) THE TWO CHOSEN FOR FEEDBACK ARE INDICATED IN SECOND COLUMN (L: LEFT HAND, R: RIGHT HAND, F: RIGHT FOOT). COLUMNS 3 AND 4 COMPARE THE ACCURACY AS CALCULATED BY CROSS-VALIDATION ON THE CALIBRATION DATA WITH THE ACCURACY OBTAINED ONLINE IN THE FEEDBACK APPLICATION 'RATE CONTROLLED CURSOR'. THE AVERAGE DURATION ± STANDARD DEVIATION OF THE FEEDBACK TRIALS IS PROVIDED IN COLUMN 5 (DURATION FROM CUE PRESENTATION TO TARGET HIT). SUBJECTS ARE SORTED ACCORDING TO FEEDBACK ACCURACY. COLUMNS 6 AND 7 REPORT THE INFORMATION TRANSFER RATES (ITR) MEASURED IN BITS PER MINUTE AS OBTAINED BY SHANNON'S FORMULA, CF. (1). HERE THE COMPLETE DURATION OF EACH RUNS WAS TAKEN INTO ACCOUNT, I.E., ALSO THE INTER-TRIAL BREAKS FROM TARGET HIT TO THE PRESENTATION OF THE NEXT CUE. THE COLUMN *overall ITR* REPORTS THE AVERAGE ITR OF ALL RUNS (OF 25 TRIALS EACH), WHILE COLUMN *peak ITR* REPORTS THE PEAK ITR OF ALL RUNS. FOR SUBJECT *au* NO REASONABLE CLASSIFIER COULD BE TRAINED (CROSS-VALIDATION ACCURACY BELOW 65% IN THE CALIBRATION DATA), SEE [2] FOR AN ANALYSIS OF THAT SPECIFIC CASE.

		calibration	feedback			
		accuracy	accuracy	duration	oITR	pITR
sbj	**classes**	**[%]**	**[%]**	**[s]**	**[b/m]**	**[b/m]**
al	LF	98.0	98.0 ± 4.3	2.0 ± 0.9	24.4	35.4
ay	LR	97.6	95.0 ± 3.3	1.8 ± 0.8	22.6	31.5
av	LF	78.1	90.5 ± 10.2	3.5 ± 2.9	9.0	24.5
aa	LR	78.2	88.5 ± 8.1	1.5 ± 0.4	17.4	37.1
aw	RF	95.4	80.5 ± 5.8	2.6 ± 1.5	5.9	11.0
au	—	–	–	–	–	–
mean		89.5	90.5 ± 7.6	2.3 ± 0.8	15.9	27.9

IV. RESULTS

A. Performance in two BBCI feedback studies

Here we summarize the results of two feedback studies with healthy subjects. The first was performed to explore the limits of information transfer rates in BCIs system not relying on user training or evoked potentials and the objective of the second was to investigate for what proportion of naive subjects our system could provide successful feedback in the very first session. One of the keys to success in this study was the proper application of CSP analysis. Details can be found in [3], [2], [7].

Table I summarizes performance, in particular the information transfer rates that were obtained in the first study. Note that calibration and feedback accuracy refer to quite different measures. From the calibration measurement, trials of

TABLE II

PERFORMANCE RESULTS FOR ALL 14 SUBJECTS OF THE SECOND STUDY. THE FIRST COLUMN SHOWS THE SUBJECT CODE AND THE SECOND COLUMN A TWO LETTER CODE WHICH INDICATES THE CLASSES WHICH HAVE BEEN USED FOR FEEDBACK. THE THIRD COLUMN SHOWS THE AVERAGE ACCURACY DURING THE FEEDBACK ± THE STANDARD ERROR OF INTRA-RUN AVERAGES. THE AVERAGE DURATION ± STANDARD DEVIATION OF THE FEEDBACK TRIALS IS PROVIDED IN THE FOURTH COLUMN (DURATION FROM CUE PRESENTATION TO TARGET HIT). SUBJECTS ARE SORTED ACCORDING TO FEEDBACK ACCURACY. FOR SUBJECT *cq* NO REASONABLE CLASSIFIER COULD BE TRAINED

		calibration	feedback	
		accuracy	accuracy	duration
subject	**classes**	**[%]**	**[%]**	**[s]**
cm	LR	88.9	93.2 ± 3.9	3.5 ± 2.7
ct	LR	89.0	91.4 ± 5.1	2.7 ± 1.5
cp	LF	93.8	90.3 ± 4.9	3.1 ± 1.4
zp	LR	84.7	88.0 ± 4.8	3.6 ± 2.1
cs	LR	96.3	87.4 ± 2.7	3.9 ± 2.3
cu	LF	82.6	86.5 ± 2.8	3.3 ± 2.7
ea	FR	91.6	85.7 ± 8.5	3.8 ± 2.2
at	LF	82.3	84.3 ± 13.1	10.0 ± 8.3
zr	LF	96.8	80.7 ± 6.0	3.1 ± 1.9
co	LF	87.1	75.9 ± 4.8	4.6 ± 3.1
eb	LF	81.3	73.1 ± 5.6	5.9 ± 4.8
cr	LR	83.3	71.3 ± 12.6	4.9 ± 3.7
cn	LF	77.5	53.6 ± 6.1	3.9 ± 2.4
cq	—	–	–	–
mean		87.3	82.6 ± 11.4	4.3 ± 1.9

approx. 3 s after each cue presentation have been taken out and the performance of the processing/classification method was validated by cross-validation. The feedback accuracy refers to the actual hitting of the correct target during horizontal cursor control. This involves integration of several classifier outputs to consecutive sliding windows of 300 to 1000 ms length, see Sec. III-D.

As a test of practical usability, subject *al* operated a mental typewriter based on horizontal cursor control. In a free spelling mode he spelled 3 German sentences with a total of 135 characters in 30 minutes, which is a 'typing' speed of 4.5 letters per minutes. Note that the subject corrected all errors using the deletion symbol. For details, see [11]. Recently, using the novel mental typewriter Hex-o-Spell that is based on principles of human-computer interaction the same subject achieved a typing speed of more than 7 letters per minute, cf. [6], [34].

Table II summarizes the performance obtained in the second study. It demonstrates that 12 out of 14 BCI novices were able for control the BCI system in their very first session. In this study, the feedback application was not optimized for fast performance, which results in longer trial duration times.

B. Visualization of the spatial filter coefficients

Let us visualize the spatial filter coefficients and the corresponding pattern of activation in the brain and see how they correspond to the neurophysiological understanding of ERD/ERS for motor imagination. Figure 7 displays two pairs of vectors $(\boldsymbol{w}_j, \boldsymbol{a}_j)$ that correspond to the largest and the smallest eigenvalues for one subject topographically mapped onto a scalp and color coded. $\boldsymbol{w}_j$ and $\boldsymbol{a}_j$ are the j-th columns of W and $A = (W^{-1})^\top$, respectively. The plot shows the interpolation of the values of the components of vectors $\boldsymbol{w}_j$ and $\boldsymbol{a}_j$ at electrode positions. Note that we use a colormap that has no direct association to signs because the signs of the vectors are irrelevant in our analysis.

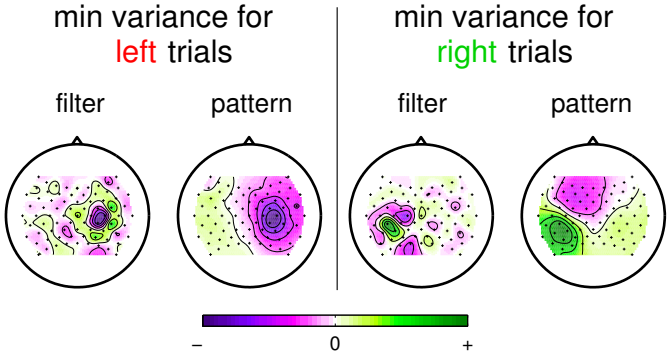

Fig. 7. Example of CSP analysis. The patterns $(\boldsymbol{a}_j)$ illustrate how the presumed sources project to the scalp. They can be used to verify neurophysiological plausibility. The filters $(\boldsymbol{w}_j)$ are used to project the original signals. Here they resemble the patterns but their intricate weighting is essential to obtain signals that are optimally discriminative with respect to variance. See Sec. III-B for the definition of the terms *filter* and *pattern*.

V. DISCUSSION

A. Dependence of linear spatial filtering prior to CSP

The question arises whether the results of CSP-based classification can be enhanced by preprocessing the data with a linear spatial filter (like PCA, ICA or re-referencing like Laplace filtering). The question is difficult to answer in general, but two facts can be derived. Let $B \in \mathbb{R}^{C \times C_0}$ be the matrix representing an arbitrary linear spatial filter while using notions X_i, $\Sigma^{(+)}$, $\Sigma^{(-)}$, S_d, and S_c as in Sec. III-C. Denoting all variables corresponding to the B-filtered signals by $\tilde{\cdot}$, the signals are $\tilde{X} = B^\top X$. This implies $\tilde{\Sigma}^{(+)} = B^\top \Sigma^{(+)} B$, $\tilde{\Sigma}^{(-)} = B^\top \Sigma^{(-)} B$, $\tilde{S}_d = B^\top S_d B$, and $\tilde{S}_c = B^\top S_c B$. The filter matrices calculated by CSP are denoted by W and $\tilde{W}$.

(1) If matrix B is invertible, the classification results will *exactly* be identical, regardless of applying filter B before calculating CSP or not. Let us consider the CSP solution characterized by simultaneous diagonalization of $\Sigma^{(+)}$ and $\Sigma^{(-)}$ in Eq. (4) with constraint $\Lambda^{(+)} + \Lambda^{(-)} = I$. This implies

$$(B^{-1}W)^\top \tilde{\Sigma}^{(+)} B^{-1} W = \Lambda^{(+)}$$
$$(B^{-1}W)^\top \tilde{\Sigma}^{(-)} B^{-1} W = I - \Lambda^{(+)}$$

which means that $B^{-1}W$ is a solution to the simultaneous diagonalization of $\tilde{\Sigma}^{(+)}$ and $\tilde{\Sigma}^{(-)}$. Since the solution is unique up to the sign of the columns, we obtain

$$\tilde{W} D = B^{-1} W \quad \text{with diagonal } D\text{:} \quad (D)_{j,j} = \operatorname{sign}(\boldsymbol{w}_j^\top B \tilde{\boldsymbol{w}}_j).$$

Accordingly, the filtered signals are identical up to the sign: $W^\top X = D\tilde{W}^\top B^\top X = D\tilde{W}^\top \tilde{X}$, so the features, the classifier and the classification performance does not change.

(2) If matrix B is not invertible, the objective of CSP analysis (on the training data) can only get worse. This can easily be seen in terms of the objective of the CSP-maximization in the formulation of the Rayleigh coefficient,

Eq. (7). Then the follwing holds

$$\max_{\tilde{w}\in\mathbb{R}^{C_0}} \frac{\tilde{w}^\top \tilde{S}_d \tilde{w}}{\tilde{w}^\top \tilde{S}_c \tilde{w}} = \max_{\tilde{w}\in\mathbb{R}^{C_0}} \frac{\tilde{w}^\top B^\top S_d B \tilde{w}}{\tilde{w}^\top B^\top S_c B \tilde{w}} \leq \max_{w\in\mathbb{R}^{C}} \frac{w^\top S_d w}{w^\top S_c w}$$

since every term on the left hand side of the inequality is covered on the right hand side for $w = B\tilde{w}$. That means, the CSP-optimum for the unfiltered signals (right hand side) is greater than or equal to the CSP-optimum for the signals filtered by B (left hand side). However, this result holds only for the training data, i.e., it may be affected by overfitting effects. If the prefiltering reduces artifacts, it is well possible that the generalization performance of CSP improves. On the other hand the prefiltering could also discard dicriminative information which would be detrimental for performance.

B. Merits and Caveats

The CSP technique is very successfully used in on-line BCI systems ([2], [19]), see Sec. IV-A. Also in the BCI Competition III many of the successful methods involved CSP type spatial filtering ([8]). Apart from the above results, an advantage of CSP is the interpretability of its solutions. Far from being a black-box method, the result of the CSP optimization procedure can visualized as scalp topographies (filters and patterns). These maps can be used to check plausibility and to investigate neurophysiological properties, cf. Sec. IV-B and also Fig.. 8.

It is important to point out that CSP is not a source separation or localization method. In contrary, each filter is optimized for two effects: maximization of variance for one class while minimizing variance for the other class. Let us consider, e.g., a filter that maximizes variance for class *foot* and minimizes it for *right*: A strong focus on the left hemispherical motor area (corresponding to the right hand) can have two plausible reasons. It can either originate from an ERD during right hand imagery, or from an ERS during foot imagery (hand areas are more relaxed if concentration focuses on the foot, therefore the idle rhythm may increase; lateral inhibition [36], [43]). Or it can be a mixture of both effects. For the discrimination task, this mixing effect is irrelevant. However this limitation has to be kept in mind for neurophysiological interpretation.

Several parameters have to be selected before CSP can be used: the band-pass filter and the time intervals (typically a fixed time interval relative to all stimuli/responses) and the subset of CSP filters that are to be used. Often some general settings are used (frequency band 7–30 Hz ([35]), time interval starting 1000 ms after cue, 2 or 3 filters from each side of the eigenvalue spectrum). But there is report that on-line performance can be much enhanced by subject-specific settings ([2]). In the Appendix we give a heuristic procedure for selection of CSP hyperparameters and demonstrate its favorable impact on classification. A practical example where parameters are selected manually is given in [15].

In addition, one should keep in mind that the discriminative criterion (Eq. (6)) tells only the separation of the mean power of two classes. The mean separation might be insufficient to tell the discrimination of samples around the decision boundary. Moreover, the mean might be sensitive to outliers.

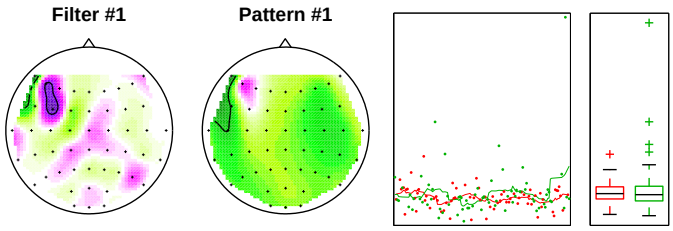

Fig. 8. CSP filter/pattern corresponding to the 'best' eigenvalue in the data set of subject *cr*. This CSP solution is highly influenced by one single-trial in which channel FC3 has a very high variance. The panel on the right shows the variance of all single-trials of the training data (x-axis: number of trial in chronological order, y-axis: log variance of the trial in the CSP surrogate channel; green: left hand imagery, red: right hand imagery). The trial which caused the distorted filter can be identified as the point in the upper right corner. Note that the class-specific box-plots on the right show no difference in median of the variances (black line).

Artifacts, such as blinking and other muscle movements can dominate over EEG signals giving excessive power in some channels. If the artifact happens to be unevenly distributed in two conditions (due to its rareness), one CSP filter will likely to capture it with very high eigenvalue. Taking one specific data set from our database as an example, the CSP filter/pattern corresponding to the best eigenvalue shown in Fig. 8 is mainly caused by one single trial. This is obviously a highly undesirable effect. But it has to be noted that the impact on classification is not as severe as it may seem on the first sight; typically the feature corresponding to such an artifact CSP filter component gets a near-zero weight in the classification step and is thereby neglected.

Finally we would like to remark that the evaluation of CSP-based algorithms needs to take into account that this technique uses label information. This means that CSP filters may only be calculated from training data (of course the resulting filters need then to be applied also to the test set). In a cross-validation, CSP filters have to be calculated repeatedly on the training set within each fold/repetition. Otherwise severe underestimation of the generalization error may occur.

C. Application of CSP to Source Projection

Here we report a novel application of CSP with a different flavor than above. Instead of single trial classification of mental states, CSP is used in the analysis of event-related modulations of brain rhythms. We show that CSP can be used to enhance the signal of interest while suppressing the background activity.

Conventionally event-related (de-)synchronization is defined as the relative difference in signal power of a certain frequency band, between two conditions, for instance a pre-stimulus or reference period and an immediate post-stimulus period [42]:

$$\mathrm{ERD}(t) = \frac{\mathrm{Power}(t) - \mathrm{Reference\ power}}{\mathrm{Reference\ power}}.$$

Thus ERD and ERS describe the relative power modulation of the ongoing activity, induced by a certain stimulus or event. Typically the sensor (possibly after Laplace filtering) that exhibit the strongest ERD/ERS effect at a certain frequency

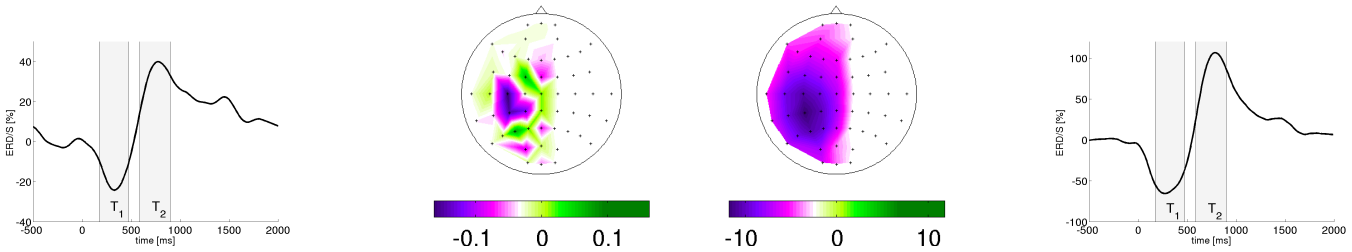

Fig. 9. Illustration of an improved source projection using the CSP technique. *Left panel*: the time course of the averaged band-power (10 Hz) at the channel (CP3) with the most prominent ERD/ERS following a median nerve stimulation at the right wrist. The gray-shaded areas indicate the two selected virtual classes for the CSP-algorithm, where T_1 corresponds to the ERD phase, while T_2 reflects the ERS interval. *Central panel*: depicts the CSP-filter that minimizes the variance for T_1, along with the projection of the corresponding source to the scalp. See main text for the reason to constrain the filter to the left hemisphere. *Right panel*: time course of the averaged band-power of the projected signal. Note that this source projection procedure has yielded ERD and ERS that are much more accentuated as they have almost tripled in magnitude.

band is used for the analysis. Nevertheless the CSP technique can help to further improve on the signal-to-noise ratio, by optimizing the spatial filters focusing on rhythmic cortical generators, that undergo the rhythmic perturbation.

We briefly outline how the CSP algorithm can be used for this purpose in an illustrative example of somatosensory stimulation. In particular, we use single trial EEG recordings of electrical stimulations of the median nerve at the right wrist. Such somatosensory stimulation typically causes modulations of the μ-rhythm, yielding a sequence of ERD followed by a rebound (ERS), overshooting the pre-event baseline level. The left panel of Fig. 9 depicts the time course of the averaged ERD/ERS for the α-band at approximately 10 Hz obtained from the best sensor. Based on this averaged band power modulations, we determine two disjoint temporal intervals T_1 and T_2, associated with the desynchronization and the hyper-synchronization phase, respectively. These two intervals serve as the opposed conditions (classes) in the conventional CSP framework. We estimate covariance matrices $\Sigma^{(+)}$ and $\Sigma^{(-)}$ as in Eq. (3) pooling covariance matrices in the two intervals separately. Solving the CSP problem according to (5), yields a set of spatial filters. The filter that minimized the variance for the desynchronization period, while simultaneously maximizing those of the synchronization period constitutes the optimal spatial projection onto the cortical generator under consideration, i.e., onto the contralateral μ-rhythm. Here we restrict our CSP analysis only to the hemisphere that is contralateral to the stimulation in order to obtain unilateral spatial filter that has no cross talk with the other hemisphere. Fig 9 depicts the obtained spatial CSP filter, along the time course of ERD/ERS of the projected signal.

Note, in case the modulation of rhythmic activity comprises only of an ERD or an ERS response, the same approach can be used by simply contrasting a pre-stimulus reference interval against the period of modulation. In other words CSP should be thought as a general tool for contrasting different brain states that yields a spatial filter solution that can be used to enhance the signal-to-noise ratio and can be interpreted from the physiological viewpoint.

D. Variants and Extensions of the Original CSP algorithm

a) Multi-class: In its original form CSP is restricted to binary problems. A general way to extend this algorithm to the multi-class case is to apply CSP to a set of binary subproblems (all binary pairs or, preferably, in a one-vs-rest scheme). A more direct approach by approximate simultaneous diagonalization was proposed in [12].

b) Automatic selection of spectral filter: The Common Spatio-Spectral Pattern (CSSP) algorithm ([31]) solves the standard CSP problem on the EEG time series augmented by delayed copies of the original signal, thereby obtaining simultaneously optimized spatial filters in conjunction with simple frequency filters. More specifically, CSP is applied to the original $\boldsymbol{x}$ concatenated with its off τ ms delayed version $\boldsymbol{x}(t-\tau)$. This amounts to an optimization in an extended spatial domain, where the delayed signals are treated as new channels $\tilde{\boldsymbol{x}}(t) = (\boldsymbol{x}(t)^\top, \boldsymbol{x}(t-\tau)^\top)^\top$. Consequently this yields spatial projections $\tilde{\boldsymbol{w}} = \left(\boldsymbol{w}^{(0)^\top}, \boldsymbol{w}^{(\tau)^\top}\right)^\top$, that correspond to vectors in this extended spatial domain. Any spatial projection in state space can be expressed as a combination of a pure spatial and spectral filter applied to the original data x, as follow:

$$\begin{aligned}\tilde{\boldsymbol{w}}^\top \tilde{\boldsymbol{x}}(t) &= \sum_{c=1}^{C} w_c^{(0)} x_c(t) + w_c^{(\tau)} x_c(t-\tau) \\ &= \sum_{c=1}^{C} \gamma_c \left(\frac{w_c^{(0)}}{\gamma_c} x_c(t) + \frac{w_c^{(\tau)}}{\gamma_c} x_c(t-\tau) \right), \end{aligned} \quad (8)$$

where $\{\gamma_c\}_{c=1}^{C}$ defines a pure spatial filter, whereas $(\frac{w_c^{(0)}}{\gamma_c}, \overbrace{0, \ldots, 0}^{\tau-1}, \frac{w_c^{(\tau)}}{\gamma_c})$ defines a Finite Impulse Response (FIR) filter at each electrode c. Accordingly this technique automatically neglects or emphasizes specific frequency bands at each electrode position in a way that is optimal for the discrimination of two given classes of signals. Note that individual temporal filters are determined for each input channel.

The Common Sparse Spectral Spatial Pattern (CSSSP) algorithm [13] eludes the problem of manually selecting the frequency band in a different way. Here a temporal FIR filter

is optimized simultaneously with a spatial filter. In contrast to CSSP only one temporal filter is used, but this filter can be of higher complexity. In order to control the complexity of the temporal filter, a regularization scheme is introduced which favors sparse solutions for the FIR coefficients. Although some values of the regularization parameter seem to give good results in most cases, for optimal performance a model selection has to be performed.

In [50] an iterative method (SPEC-CSP) is proposed which alternates between spatial filter optimization in the CSP sense and the optimization of a spectral weighting. As result one obtains a spatial decomposition and a temporal filter with are jointly optimized for the given classification problem.

c) Connection to a discriminative model: Here we show how CSP analysis is related to a discriminative model. This connection is of theoretical interest in itself, and can also be used to further elaborate new variants of CSP. See [49], [48] for related models.

The quantity $S_d = \Sigma^{(+)} - \Sigma^{(-)}$ in Eq. (6) can be interpreted as the empirical average $\hat{\mathbb{E}}_{X,y}\left[yXX^\top\right]$ of the sufficient statistics $yXX^\top$ of a linear logistic regression model:

$$P(y|X,V,b) = \frac{\exp\left(yf(X;V,b)\right)}{Z(X,V,b)}$$
$$f(X;V,b) = \mathrm{Tr}\left[V^\top XX^\top\right] + b,$$

where $y \in \{+1,-1\}$ is the label corresponding to two classes, $V \in \mathbb{R}^{C\times C}$ is the regression coefficient, b is the bias, and $Z(X,V,b) = e^{f(X;V,b)} + e^{-f(X;V,b)}$. In fact, given a set of trials and labels $\{X_i, y_i\}$ the log-likelihood of the above problem can be written as follows:

$$\begin{aligned}
&\log\prod_{i=1}^{n} P(y_i|X_i,V,b) \\
&= \mathrm{Tr}\left[V^\top\left(\sum_{i=1}^{n} y_i X_i X_i^\top\right)\right] + b\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\log Z(X_i,V,b) \\
&= \frac{n}{2}\mathrm{Tr}\left[V^\top S_d\right] - \sum_{i=1}^{n}\log Z(X_i,V,b),
\end{aligned}$$

where for simplicity we assumed that each condition contains equal number $(n/2)$ of trials. Unfortunately, because of the log-normalization $Z(X,V,b)$ term, the maximum likelihood problem cannot be solved as simple as the simultaneous diagonalization. One can upper bound the $\log Z(X,V,b)$ under the following condition:

$$\sum_{i=1}^{n}\left|\mathrm{Tr}\left[V^\top X_i X_i^\top\right]\right| \le 1,$$

and maximize the lower bound of the likelihood as follows:

$$\begin{aligned}
\underset{V\in\mathbb{R}^{C\times C}}{\text{maximize}} \quad & \frac{n}{2}\mathrm{Tr}\left[V^\top S_d\right], \\
\text{subject to} \quad & \sum_{i=1}^{n}\left|\mathrm{Tr}\left[V^\top X_i X_i^\top\right]\right| \le 1.
\end{aligned}$$

Indeed this yields the first generalized eigenvector of the CSP problem (Eq. (5)) when V is rank=1 matrix $V = \boldsymbol{w}\boldsymbol{w}^\top$.

d) Regularizing CSP: In practical BCI applications, the smaller the number of electrodes, the smaller the effort and time to set up the cap and also the smaller the stress of patients would be. CSP analysis can be used to determine where the electrodes should be positioned; therefore it would be still useful for experiments with a small number of electrodes. In [16], ℓ_1 regularization on the CSP filter coefficients was proposed to enforce a sparse solution; that is, many filter coefficients become numerically zero at the optimum. Therefore it provides a clean way of selecting the number and the positions of electrodes. Their results have shown that the number of electrodes can be reduced to 10-20 without significant drop in the performance.

e) Advanced techniques towards reducing calibration data: Because there exists substantial day-to-day variability in EEG data, the calibration session (15-35 min) is conventionally carried out every time before day-long experiments even for an experienced subject. Thus, in order to increase the usability of BCI systems, it is desirable to make use of previous recordings so that we can reduce the calibration measurement as small as possible (cf. also data set IVa of the BCI competition III, [8]). For experienced BCI users whose EEG data were recorded more than once, [28] proposed a procedure to utilize results from the past recordings. They extracted prototypical filters by a clustering algorithm from the data recorded before and use them as an additional prior information for the current new session learning problem.

Recently [32] proposed an extended EM algorithm, where the extraction and classification of CSP features are performed jointly and iteratively. This method can be applied to the cases where either only a small number of calibration measurements (semisupervised) or even no labeled trials (unsupervised) are available. Basically, their algorithm repeats the following steps until a stable result is obtained: (i) constructing an expanded training data which consists of calibration trials with observed labels and a part of unlabeled (feedback) data with labels estimated by the current classifier, (ii) reextracting the CSP feature and updating the classifier based on the current data sets. They analyzed the data IVa of BCI competition III ([8]) and reported that because of the iterative reextraction of the CSP features, they could achieve satisfactory performance from only 30 labeled and 120 unlabeled data or even from 150 unlabeled trials (off-line analysis). Note that only results of selected subjects of the competition data set IVa were reported. Although there was no experimental result presented, it was claimed that the extended EM procedure can also adapt to nonstationarity in EEG signals.

f) Dealing with the nonstationary of EEG signals: Another practical issue is nonstationarity in EEG data. There are various suggestions how to handle the nonstationarity in BCI systems ([52], [51], [26], [10]). With respect to CSP-based BCIs, the result of [29], [47] was that a simple adaptation of the classifier bias can compensate nonstationarity astonishingly well. Further changes like retraining LDA and recalculating CSP contributed only slightly or sometimes increased the error rate.

The question whether the CSP filter W or the pattern A should generalize to a new recording was raised by [23].

From a source separation point of view, the j-th column $\boldsymbol{w}_j$ of the filter W tries to capture the j-th source denoted by the j-th column $\boldsymbol{a}_j$ of the pattern A while trying to suppress all other sources that are irrelevant to the motor-imagination task. Therefore, if the disturbances change while the relevant source remains unchanged the optimal filter should adaptively change to cancel out the new disturbances while still capturing the relevant source. In [23] the Fixed Spatial Pattern (FSP) approach was proposed; that is to keep the spatial pattern of the relevant source, i.e., subset of the columns of A unchanged while changing the spatial filter adaptively in a new recording. The true labels (i.e., the actual intension of a subject) are not required when the FSP is applied because only the irrelevant sources, which are assumed to be common to two classes, are re-estimated.

A novel approach to make CSP more robust to nonstationarities during BCI feedback was proposed in [5]. In this work a short measurement of non task related disturbances is used to enforce spatial filters which are invariant against those disturbances. In invariant CSP (iCSP) the covariance matrix of the disturbance is added to the denominator in the Rayleigh coefficient representation of CSP, cf. Eq. (7).

VI. Concluding discussion

We have reviewed a spatial filtering technique that often finds its successful use in BCI: Common Spatial Patterns (CSP). The method is based on the second order statistics of the signal between electrodes and the solution is obtained by solving a generalized eigenvalue problem. We have shown a generative and a discriminative interpretation of the method. We have applied the method to two motor imagination based BCI studies. In the first study, we have reported the peak information transfer rate from one subject of 35.4 bits/min. In the second study we have shown that 12 out of 14 naive subjects could perform BCI control on their first BCI experiments. We have pointed out not only the advantage of the method, such as low computation cost and interpretability but also some caveats such as model selection and pre-processing issues or deterioration under outliers. We showed subsequently that CSP can be extended and robustified in order to alleviate these critical aspects. In this review we have focused our attention to applications of CSP for single trial EEG analysis in the context of BCI. Note however that CSP-filtering and extensions thereof can be applied to extract general discriminative spatio-temporal structure from multivariate data streams beyond EEG. Future work will continue the quest to develop novel spatio-temporal filtering methods that allow more accurate and interpretable classification even for nonstationary, noisy, interacting data sources. Special attention will be placed on the construction of probabilistically interpretable nonlinear modeling that allows the integration of feature extraction and classification steps within a one step procedure in the spirit of, e.g., [49], [48], [17], [22]).

Appendix

A. How to Select Hyperparameters for CSP

Here we give a heuristic procedure to automatically select all parameters that are needed for successful CSP application. There is no claim whatsoever that these heuristics are close to optimal or natural in any sense. But we have found them practically working and evaluate them here in comparison to the general setting and to manual selection by the experimenter.

a) Selection of a Frequency Band: We provide our heuristic for the selection of a discriminative frequency band in pseudo code, see Algorithm 1. The EEG trials X should be spatially filtered by a Laplacian or bipolar filter. In our experience the algorithm works best if only few channels are used. A good choice is, e.g., to choose $C = \{c_1, c_2, c_3\}$ with c_i being one from each area of the left hand, right hand and feet with $\max \sqrt{\sum_f (\mathrm{score}_c(f))^2}$.

Algorithm 1 Selection of a discriminative frequency band

Let $X_{(c,i)}$ denote trial i at channel c with label y_i and let C denote the set of channels.

$\mathrm{dB}_c(f,i) \leftarrow$ log band-power of $X_{(c,i)}$ at frequency f (f from 5 to 35Hz)
$\mathrm{score}_c(f) \leftarrow \mathrm{corrcoef}\left(\mathrm{dB}_c(f,i), y_i\right)_i$
$f_{\max} \leftarrow \mathrm{argmax}_f \sum_{c\in C} \mathrm{score}_c(f)$
$\mathrm{score}^*_c(f) \leftarrow \begin{cases} \mathrm{score}_c(f) & \text{if } \mathrm{score}_c(f_{\max}) > 0 \\ -\mathrm{score}_c(f) & \text{otherwise} \end{cases}$
$\mathrm{fscore}(f) \leftarrow \sum_{c\in C} \mathrm{score}^*_c(f)$
$f^*_{\max} \leftarrow \mathrm{argmax}_f \mathrm{fscore}(f)$
$f_0 \leftarrow f^*_{\max};\ f_1 \leftarrow f^*_{\max}$
while $\mathrm{fscore}(f_0 - 1) \geq \mathrm{fscore}(f^*_{\max}) * 0.05$ **do**
  $f_0 \leftarrow f_0 - 1$
while $\mathrm{fscore}(f_1 + 1) \geq \mathrm{fscore}(f^*_{\max}) * 0.05$ **do**
  $f_1 \leftarrow f_1 + 1$
return frequency band $[f_0, f_1]$

b) Selection of a Time Interval: The heuristic selection of a time interval proceeds similar to the selection of the frequency band, see Algorithm 2.

Algorithm 2 Selection of a discriminative time interval

Let $X_{(c,i)(t)}$ denote time sample t of trial i at channel c with label y_i and let C denote the set of channels.

$\mathrm{env}_c(t,i) \leftarrow$ envelope of $X_{(c,i)(t)}$, calculated by Hilbert transform (e.g. [9]) and smoothed
$\mathrm{score}_c(t) \leftarrow \mathrm{corrcoef}\left(\mathrm{env}_c(t,i), y_i\right)_i$
$t_{\max} \leftarrow \mathrm{argmax}_t \sum_{c\in C} |\mathrm{score}_c(t)|$
$\mathrm{score}^*_c(t) \leftarrow \begin{cases} \mathrm{score}_c(t) & \text{if } \sum_{t_{\max}-100\mathrm{ms}<t'<t_{\max}+100\mathrm{ms}} \mathrm{score}_c(t') > 0 \\ -\mathrm{score}_c(t) & \text{otherwise} \end{cases}$
$\mathrm{tscore}(t) \leftarrow \sum_{c\in C} \mathrm{score}^*_c(t)$
$t^*_{\max} \leftarrow \mathrm{argmax}_t \mathrm{tscore}(t)$
$\mathrm{thresh} \leftarrow 0.8 * \sum_t \mathrm{tscore}^+(t)$ (with $f^+(x) = f(x)$ if $f(x) > 0$ and $= 0$ otherwise)
$t_0 \leftarrow t^*_{\max};\ t_1 \leftarrow t^*_{\max}$
while $\sum_{t_0 \leq t \leq t_1} \mathrm{tscore}(t) < \mathrm{thresh}$ **do**
  if $\sum_{t<t_0} \mathrm{tscore}^*(t) > \sum_{t>t_1} \mathrm{tscore}^*(t)$ **then**
    $t_0 \leftarrow t_0 - 1$
  else
    $t_1 \leftarrow t_1 + 1$
return time interval $[t_0, t_1]$

TABLE III

COMPARISON OF CSP-BASED CLASSIFICATION PERFORMANCE WHEN THE HYPERPARAMETERS ARE FIXED A-PRIORI, SELECTED AUTOMATICALLY BY THE PROPOSED HEURISTICS, OR SELECTED MANUALLY. EVALUATION BY A CHRONOLOGICAL SPLIT OF THE CALIBRATION DATA (FIRST HALF FOR TRAINING, SECOND HALF FOR TESTING). NOTE THAT 'AUTO' USES ONLY THE FIRST HALF FOR HYPERPARAMETER SELECTION, WHEREAS 'MANUAL' USES THE WHOLE CALIBRATION DATA.

sbj	fixed	auto	manual
zq	2.5	*0.5*	**0.1**
zp	*11.9*	14.8	**8.1**
zr	*0.8*	**0.2**	**0.2**
cs	9.6	*4.1*	**1.3**
at	6.9	*6.7*	**6.7**
ct	20.7	*8.9*	**5.2**
zk	9.9	*6.0*	**1.5**
cm	14.9	*6.5*	**5.0**
cm	15.1	*6.4*	**2.1**
cm	*18.2*	18.2	**6.9**
cm	13.7	*8.2*	**5.0**
ea	5.7	*1.7*	**1.6**
eb	*25.0*	27.1	**12.1**
mean	11.9	*8.4*	**4.3**

c) Selection of a Subset of Filters: The classical measure for the selection of CSP filters is based on the eigenvalues in (5). Each eigenvalue is the relative variance of the signal filtered with the corresponding spatial filter (variance in one condition divided by the sum of variances in both conditions). This measure is not robust to outliers because it is based on simply pooling the covariance matrices in each condition (Eq. (3)). In fact, one single trial with very high variance can have a strong impact on the CSP solution (see also Fig. 8). A simple way to circumvent this problem is to calculate the variance of the filtered signal within each trial and then calculate the corresponding ratio of medians:

$$\text{score}(\boldsymbol{w}_j) = \frac{\text{med}_j^{(+)}}{\text{med}_j^{(+)} + \text{med}_j^{(-)}}$$

where $\text{med}_j^{(c)} = \text{median}_{i \in \mathscr{I}_c} \left(\boldsymbol{w}_j^\top X_i X_i^\top \boldsymbol{w}_j \right) \quad (c \in \{+,-\})$. As with eigenvalues, a 'ratio-of-medians' score near 1 or near 0 indicates good discriminability of the corresponding spatial filter. These scores are more robust with respect to outliers than the eigenvalue score, e.g., the filter shown in Fig. 8 would get a minor (i.e., near 0.5) ratio-of-medians score.

d) Evaluation of Heuristic Selection Procedure: Here we compare the impact of individually choosing the hyperparameters for CSP-based classification. We compare the method 'fixed' which uses a broad frequency band 7–30 Hz and the time window 1000 to 3500 ms post stimulus. The method 'auto' uses the heuristics presented in this Section to select frequency band and time interval. In 'manual' we use the settings that were chosen by an experienced experimenter by hand for the actual feedback (see [15] for a practical example with manual selection). Note there is a substantial improvement of performance in most of the data sets. Interestingly in one feedback data set (subject *ct*) the 'auto' method performs badly, although the selected parameters were reasonable.

TABLE IV

COMPARISON OF PERFORMANCE ANALOG TO TABLE III, BUT WITH EVALUATION BY TRAINING ON THE WHOLE CALIBRATION MEASUREMENT AND TESTING ON THE FEEDBACK DATA (WINDOWS OF 1000 MS DURING CURSOR MOVEMENT). NOTE THAT THESE ERROR RATES DO NOT REFLECT THE ERRORS IN HITTING THE CORRECT BARS; A SUCCESSFUL TRIAL OFTEN INCLUDES ERRONEOUS INTERMEDIATE STEPS.

sbj	fixed	auto	manual
zq	17.4	*13.1*	**12.5**
zp	*24.4*	24.6	**22.8**
zr	25.3	**18.6**	*23.1*
cs	26.1	*23.0*	**21.8**
at	39.6	*34.9*	**33.6**
ct	*12.1*	31.0	**10.9**
zk	*28.2*	**27.3**	28.8
cm	19.9	*8.8*	**7.4**
cm	6.2	*2.5*	**2.0**
cm	7.7	*6.6*	**6.1**
cm	27.7	*7.0*	**5.9**
ea	21.6	*20.4*	**19.1**
eb	50.3	*42.3*	**39.1**
mean	23.6	*20.0*	**17.9**

REFERENCES

[1] H. Berger. Über das Elektroenkephalogramm des Menschen. *Arch. Psychiat. Nervenkr.*, 99(6):555–574, 1933.

[2] Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Gabriel Curio, and Klaus-Robert Müller. The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550, 2007.

[3] Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Klaus-Robert Müller, Volker Kunzmann, Florian Losch, and Gabriel Curio. The Berlin Brain-Computer Interface: EEG-based communication without subject training. *IEEE Trans. Neural Sys. Rehab. Eng.*, 14(2):147–152, 2006.

[4] Benjamin Blankertz, Guido Dornhege, Steven Lemm, Matthias Krauledat, Gabriel Curio, and Klaus-Robert Müller. The Berlin Brain-Computer Interface: Machine learning based detection of user specific brain states. *J. Universal Computer Sci.*, 12(6):581–607, 2006.

[5] Benjamin Blankertz, Motoaki Kawanabe, Ryota Tomioka, Friederike Hohlefeld, Vadim Nikulin, and Klaus-Robert Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008. accepted.

[6] Benjamin Blankertz, Matthias Krauledat, Guido Dornhege, John Williamson, Roderick Murray-Smith, and Klaus-Robert Müller. A note on brain actuated spelling with the Berlin Brain-Computer Interface. In C. Stephanidis, editor, *Universal Access in HCI, Part II, HCII 2007*, volume 4555 of *LNCS*, pages 759–768, Berlin Heidelberg, 2007. Springer.

[7] Benjamin Blankertz, Florian Losch, Matthias Krauledat, Guido Dornhege, Gabriel Curio, and Klaus-Robert Müller. The Berlin Brain-Computer Interface: Accurate performance from first-session in BCI-naive subjects. *IEEE Trans. Biomed. Eng.*, 2007. to be submitted.

[8] Benjamin Blankertz, Klaus-Robert Müller, Dean Krusienski, Gerwin Schalk, Jonathan R. Wolpaw, Alois Schlögl, Gert Pfurtscheller, José del R. Millán, Michael Schröder, and Niels Birbaumer. The BCI competition III: Validating alternative approachs to actual BCI problems. *IEEE Trans. Neural Sys. Rehab. Eng.*, 14(2):153–159, 2006.

[9] Ronald N. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill, 1999. 3rd ed.

[10] J. del R. Millán. On the need for on-line learning in brain-computer interfaces. In *Proceedings of the International Joint Conference on Neural Networks*, Budapest, Hungary, July 2004. IDIAP-RR 03-30.

[11] Guido Dornhege. *Increasing Information Transfer Rates for Brain-Computer Interfacing*. PhD thesis, University of Potsdam, 2006.

[12] Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Trans. Biomed. Eng.*, 51(6):993–1002, June 2004.

[13] Guido Dornhege, Benjamin Blankertz, Matthias Krauledat, Florian Losch, Gabriel Curio, and Klaus-Robert Müller. Optimizing spatio-temporal filters for improving brain-computer interfacing. In *Advances*

in Neural Inf. Proc. Systems (NIPS 05), volume 18, pages 315–322, Cambridge, MA, 2006. MIT Press.

[14] Guido Dornhege, José del R. Millán, Thilo Hinterberger, Dennis McFarland, and Klaus-Robert Müller, editors. *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA, 2007.

[15] Guido Dornhege, Matthias Krauledat, Klaus-Robert Müller, and Benjamin Blankertz. General signal processing and machine learning tools for BCI. In *Toward Brain-Computer Interfacing*, pages 207–233. MIT Press, Cambridge, MA, 2007.

[16] J. Farquhar, N. J. Hill, T. N. Lal, and B. Schölkopf. Regularised CSP for sensor selection in BCI. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006*, pages 14–15. Verlag der Technischen Universität Graz, 09 2006.

[17] Jason Farquhar, Jeremy Hill, and Bernhard Schölkopf. Learning optimal EEG features across time, frequency and space, 2006. In NIPS 2006 workshop *Current Trends in Brain-Computer Interfacing*.

[18] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, Boston, 2nd edition edition, 1990.

[19] Christoph Guger, H. Ramoser, and Gert Pfurtscheller. Real-time EEG analysis with subject-specific spatial patterns for a Brain Computer Interface (BCI). *IEEE Trans. Neural Sys. Rehab. Eng.*, 8(4):447–456, 2000.

[20] R. Hari and R. Salmelin. Human cortical oscillations: a neuromagnetic view through the skull. *Trends in Neuroscience*, 20:44–9, 1997.

[21] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: data mining, inference and prediction*. Springer series in statistics. Springer, New York, N.Y., 2001.

[22] Jeremy Hill and Jason Farquhar. An evidence-based approach to optimizing feature extraction in eeg signal classification. Technical report, Max Planck Institute for Biological Cybernetics, 2007. Under preparation.

[23] N. J. Hill, J. Farquhar, T. N. Lal, and B. Schölkopf. Time-dependent demixing of task-relevant EEG signals. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006*, pages 20–21. Verlag der Technischen Universität Graz, 09 2006.

[24] H. Jasper and H.L. Andrews. Normal differentiation of occipital and precentral regions in man. *Arch. Neurol. Psychiat. (Chicago)*, 39:96–115, 1938.

[25] H. Jasper and W. Penfield. Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus. *Arch. Psychiatrie Zeitschrift Neurol.*, 183:163–74, 1949.

[26] Motoaki Kawanabe, Matthias Krauledat, and Benjamin Blankertz. A bayesian approach for adaptive BCI classification. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006*, pages 54–55. Verlag der Technischen Universität Graz, 2006.

[27] Z. J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalogr. Clin. Neurophysiol.*, 79(6):440–447, 1991.

[28] Matthias Krauledat, Michael Schröder, Benjamin Blankertz, and Klaus-Robert Müller. Reducing calibration time for brain-computer interfaces: A clustering approach. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 753–760, Cambridge, MA, 2007. MIT Press.

[29] Matthias Krauledat, Pradeep Shenoy, Benjamin Blankertz, Rajesh P. N. Rao, and Klaus-Robert Müller. Adaptation in CSP-based BCI systems. In *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA, 2007. in press.

[30] A. Kübler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer, and J. R. Wolpaw. Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–1777, 2005.

[31] Steven Lemm, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. Spatio-spectral filters for improving classification of single trial EEG. *IEEE Trans. Biomed. Eng.*, 52(9):1541–1548, 2005.

[32] Yuanqing Li and Cuntai Guan. An extended EM algorithm for joint feature extraction and classification in brain-computer interfaces. *Neural Comput.*, 18:2730–2761, 2006.

[33] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Neural Networks*, 12(2):181–201, May 2001.

[34] Klaus-Robert Müller and Benjamin Blankertz. Toward noninvasive brain-computer interfaces. *IEEE Signal Proc. Magazine*, 23(5):125–128, September 2006.

[35] Johannes Müller-Gerking, Gert Pfurtscheller, and Henrik Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.*, 110:787–798, 1999.

[36] C. Neuper and G. Pfurtscheller. Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates. *Int. J. Psychophysiol.*, 43:41–58, 2001.

[37] C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller. Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. *Brain Res. Cogn. Brain Res.*, 25(3):668–677, 2005.

[38] V. Nikouline, K. Linkenkaer-Hansen, Wikström; H., M. Kesäniemi, E. Antonova, R. Ilmoniemi, and J. Huttunen. Dynamics of mu-rhythm suppression caused by median nerve stimulation: a magnetoencephalographic study in human subjects. *Neurosci. Lett.*, 294, 2000.

[39] Paul L. Nunez, Ramesh Srinivasan, Andrew F. Westdorp, Ranjith S. Wijesinghe, Don M. Tucker, Richard B. Silberstein, and Peter J. Cadusch. EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.*, 103:499–515, 1997.

[40] Lucas Parra and Paul Sajda. Blind source separation via generalized eigenvalue decomposition. *Journal of Machine Learning Research*, 4:1261–1269, 2003.

[41] Lucas C. Parra, Clay D. Spence, Adam D. Gerson, and Paul Sajda. Recipes for the linear analysis of EEG. *NeuroImage*, 28(2):326–341, 2005.

[42] G. Pfurtscheller and A. Arabibar. Evaluation of event-related desynchronization preceding and following voluntary self-paced movement. *Electroencephalogr. Clin. Neurophysiol.*, 46:138–46, 1979.

[43] G. Pfurtscheller, C. Brunner, A. Schlögl, and F.H. Lopes da Silva. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159, 2006.

[44] Alois Schlögl, Julien Kronegg, Jane Huggins, and Steve G. Mason. Evaluation Criteria for BCI Research. In Guido Dornhege, Jose del R. Millán, Thilo Hinterberger, Dennis McFarland, and Klaus-Robert Müller, editors, *Towards Brain-Computer Interfacing*, pages 297–312. MIT press, Cambridge, MA, 2007.

[45] A. Schnitzler, S. Salenius, R. Salmelin, V. Jousmäki, and R. Hari. Involvement of primary motor cortex in motor imagery: a neuromagnetic study. *NeuroImage*, 6:201–8, 1997.

[46] Stephen H. Scott. Converting thoughts into action. *Nature*, 442:141–142, 2006.

[47] Pradeep Shenoy, Matthias Krauledat, Benjamin Blankertz, Rajesh P. N. Rao, and Klaus-Robert Müller. Towards adaptive classification for BCI. *J. Neural Eng.*, 3:R13–R23, 2006.

[48] Ryota Tomioka and Kazuyuki Aihara. Classifying Matrices with a Spectral Regularization. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 895–902. ACM Press, 2007.

[49] Ryota Tomioka, Kazuyuki Aihara, and Klaus-Robert Müller. Logistic regression for single trial EEG classification. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1377–1384. MIT Press, Cambridge, MA, 2007.

[50] Ryota Tomioka, Guido Dornhege, Kazuyuki Aihara, and Klaus-Robert Müller. An iterative algorithm for spatio-temporal filter optimization. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006*, pages 22–23. Verlag der Technischen Universität Graz, 2006.

[51] C. Vidaurre, A. Schlögl, R. Cabeza, R. Scherer, and G. Pfurtscheller. A fully on-line adaptive BCI. *IEEE Trans. Biomed. Eng.*, 6, 2006. In Press.

[52] C. Vidaurre, A. Schlögl, R. Cabeza, R. Scherer, and G. Pfurtscheller. Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. *IEEE Trans. Biomed. Eng.*, 54(3):550–556, 2007.

[53] J. R. Wolpaw and D. J. McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. USA*, 101(51):17849–17854, 2004.

[54] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan. Brain-computer interface research at the Wadsworth Center. *IEEE Trans. Rehab. Eng.*, 8(2):222–226, 2000.

[55] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.*, 113(6):767–791, 2002.