Speeding up classification of multi-channel Brain-Computer Interfaces: Common spatial patterns for slow cortical potentials

Guido Dornhege[1], Benjamin Blankertz[1], Gabriel Curio[2]
[1]Fraunhofer FIRST.IDA, Berlin, Germany
[2]Neurophysics Group, Dept. of Neurology, Klinikum Benjamin Franklin, Freie Universität Berlin, Germany
{guido.dornhege, benjamin.blankertz}@first.fraunhofer.de, curio@zedat.fu-berlin.de

Abstract—**During the last years interest has been growing to find an effective communication channel which translates human intentions into control signals for a computer, the so called Brain-Computer Interface (BCI). One main goal of research is to help patients with severe neuromuscular disabilities by substituting normal motor outputs. Various cortical processes were identified which are suitable for implementing such a system on basis of scalp recorded electroencephalogram (EEG), e.g., slow cortical potentials (SCP) and event-related desynchronisation (ERD) of 10–20 Hz brain rhythms. Until quite recently BCI systems used only few EEG channels but by use of advanced machine learning techniques it became possible to exploit the spatial information provided by multi-channel EEG. While the use of such high density spatial sampling increases the accuracy of the system it may—depending on the computational effort of the signal processing—pose a problem for the implementation of the feedback in real-time. Here we propose a method that offers a substantial speed-up for classification of SCP features as used in the Berlin Brain computer interface (BBCI) [1]. Instead of applying the time consuming low-pass filtering to all, say 120, EEG channels a suitable spatial projection extracts only 2 or 4 new channels which can be used without any loss of classification accuracy in our experiments. Our approach is based on the technique of common spatial patterns (CSP) which were suggested in [2] to extract ERD features from EEG. While in its original form CSP is only applicable to oscillatory features we present a new variant which allows to use CSP for SCP features without regularisation even in case of large channel numbers or few training samples.**

Keywords—**multi-channel electroencephalogram, brain-computer interface, slow cortical potentials, movement related potential, Bereitschaftspotential, common spatial patterns, single-trial classification.**

I. Introduction

A brain-computer interface (BCI) is a system for controlling a device, e.g, a computer, a wheelchair or a neuroprothesis by human intentions. According to the definition agreed upon at the first international meeting for BCI technology the system must *not depend on the brain's normal output pathways of peripheral nerves and muscles* [3]. The present approach is applied to non-invasively measured electroencephalogram (EEG) data but it applies equally well to electrocorticogram (ECoG) data. This Berlin Brain-Computer Interface (BBCI) uses neuronal signatures of well-established ('overlearned') motor competences, such as keyboard typing. Hence, it provides the chance to work without time-comsuming training of subjects or patients for efficient BCI performances which in some other BCI variants could eventually take weeks. In the current BBCI version healthy subjects do execute such overlearned movements rather than just imagine them because this instruction appears as to correspond as closely as possible to the movement intentions of a paralyzed patient. In contrast, if the 'no-motor-output' condition of a patient were to be mimicked with imagined movements, the vetoing of the actual movement required to cancel the motor output would face healthy subjects with a paradoxical task which has no correspondence in paralyzed patients. On the other hand, in studies with real movements one has to be careful to exclude any possible use of afferent nerve signals which the sensorimotor areas could receive during actual movement execution. Accordingly, the basic rationale of the BBCI paradigm requires a capability to predict the laterality of imminent hand movements prior to any EMG activity. This *movement intention* is expected to safely transfer from BCI studies with healthy subjects to true BCI control of paralyzed patients.

The studies were supported by a grant of the *Bundesministerium für Bildung und Forschung* (BMBF), FKZ 01IBB02A and FKZ 01IBB02B.

In [1] we suggested a preprocessing method that extracts features of the *Bereitschaftspotential* (BP, readiness potential) from multi-channel EEG data. In the meantime we carried out several experiments with the paradigm of self-paced keyboard typing using up to 128 EEG channels. Also a realtime BCI feedback system was established. While the use of a large number of electrodes increases classification performance in offline studies, for the feedback system a smaller number of channels has to be selected due to the computational demand of the processing algorithms.

In this paper we suggest to extract few discriminative spatial patterns during the calibration of the system (training period) and to use this information to effectively reduce the number of channels that have to be processed for feedback control. The method is based on a technique known from statistical pattern recognition called common spatial patterns (CSP) which allows to determine directions that maximize the variance that can be explained by one condition and at the same time minimize the variance explained by another condition. This technique was suggested by [4] for spatial analyses of EEG signals, and more specifically by [2] to find spatial structures of event-related desynchronization (ERD) in a BCI context. Since ERD is directly reflected by the variance of band-pass filtered signals this can be done by a straight forward application of CSP. Here we show that by a slight but crucial modification of the CSP algorithm it can be used also for slow cortical potential (SCP) variations like the *Bereitschaftspotential*.

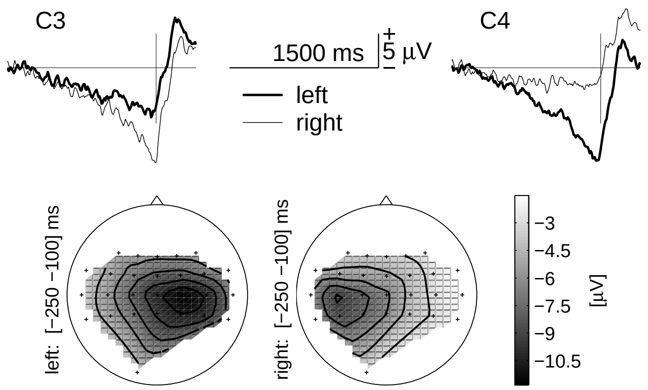

Fig. 1. Event-related potentials (ERPs) for subject *aa* at channels C3 and C4, thin line for right, thick line for left events. The scalp maps show the potential distributions for left and right hand finger movements between -250 and -100 ms.

II. Neurophysiological Background

Concerning the selection of brain signals we focussed here on one variant of SCPs which are specifically related to the preparation of motor commands. Using multi-channel EEG-mapping it has been repeatedly demonstrated that several highly localized brain areas contribute to cerebral motor command processes, [5]. A negative *Bereitschaftspotential* precedes the voluntary initiation of movements, and actual task-performance (finger flexions and extensions over a period of six seconds) are accompanied by a negative DC shift called a performance-related negativity (PN). A differential scalp potential distribution can be reliably demonstrated in a majority of experimental subjects with larger BP at lateral scalp positions (C3, C4) positioned over the left or right hemispherical primary motor cortex, respectively, consistenly correlating with the performing (right or left) hand, see Fig. 1.

III. Experiments

In this paper we analyze EEG data from 13 experiments with 6 healthy subjects. The persons sat in a normal chair with fingers in the standard typing position at the computer keyboard. They were instructed to press keys with index and little fingers in self-chosen order and timing matching approximately a predefined speed (one tap every 1 s, 1.5 s, 2 s or 5 s).

Brain activity was measured with 27, 52 or 120 Ag/AgCl electrodes at positions of the extended international 10-20 system referenced to nasion and sampled at 1000 Hz, band-pass filtered to 0.05–200 Hz, and down-sampled to 100 Hz for further offline analyses. Besides EEG a bilateral electromyogram (EMG) of the *musculus flexor digitorum* and a horizontal and vertical electrooculogram (EOG) were recorded to check for muscle activation and eye movements. All processing is solely based on the EEG channels. An important characteristics of our present analysis was to refrain from any trial rejection because of eventual artifacts so as to enforce robust classification.

IV. Single-Trial Processing

A. Feature Extraction and Aim of Classification

Starting point is our approach for the *Bereitschaftpotential* presented in [1]. It acts on segments of 128 samples in time (i.e. 1.28 s). A low-pass filter which puts emphasis on the late part of the time window is implemented by means of FFT: The windowed ($w(n) := 1 - \cos(n\pi/128)$) signals are transformed to the spectral domain where those coefficients are set to 0 that do not fall into the pass-band 0.4–3.5 Hz. The inverse FFT gives filtered signals from which the last 150 ms are downsampled to 20 Hz by calculating means in 3 non-overlapping 50 ms windows. This leads to 3 sample values per channel. The concatenation of those values for all chosen channels is taken as feature vector.

The aim of classification is to predict the laterality of upcoming movements (left vs. right hand) before EMG activity starts. The exact start of EMG activity is different in each trial. Since in this paper we do not aim at an absolute evaluation of the BCI approach, but only to assess the effectiveness of the new method we simply compare classification at -100 ms before keypress, which is in most cases before EMG onset.

B. Classification and Validation

There are neurophysiological indications that the assumption of gaussianity is realistic for EEG signals, cf. [6]. As this gaussianity is preserved by our linear preprocessing steps the classification problem in feature space should still be linear, cf. also [7], [8]. Therefore we classify by linear discriminant analysis (LDA). In case one has high-dimensional features and only comparatively few samples, regularization [9] is needed to avoid overfitting. The regularization parameter is fitted to the data by cross-validation on each training set.

We evaluate the performance of classification algorithms by 10×10-fold cross-validation. To this end the set of all trials is divided randomly in 10 parts. Classifiers are trained on nine of those parts and evaluated on the left-out part. This is repeated 10 times with different random divisions. This procedure yields 100 test error results. The mean of those is an estimation for the generalization error of the algorithm ('cross-validation error').

C. Feedback

For BCI feedback every 40 ms the most recent block of EEG data is processed and the continuous output of the previously trained classifier is translated into a cursor movement on the computer screen. While we do not report on feedback experiments because this goes beyond the scope of this paper, the proposed method is capable to enhance feedback generation.

V. Common spatial patterns for SCPs

In [2] the CSP algorithm is applied with covariance matrices that are calculated from band-pass filtered EEG signals. The resulting patterns are directions of the most pronounced differences in variance between the two classes. Since the variance of a band-pass filtered signal is a measure for the energy in the corresponding frequency band, the patterns reflect the spatial distributions of event-related (de)synchronization effects.

For the SCPs this approach is not appropriate since variance is not a measure for the (negative) shift in the EEG signals. We need a viable concept of variance that reflects the extent of an EEG shift relative to a baseline of event unrelated brain activity. After FFT filtering, see Section IV-A, the baseline is at 0 μV. Therefore the *non-centered* variance $\mathrm{var}(z) = z^\top z$ (i.e. assuming zero mean) is a measure for the deflection relative to the baseline. Let $\hat{x}_i \in \mathbb{R}^n$ (n: # of channels) be the mean over the 150 ms window that we extract after FFT filtering and $y_i \in \{\text{'L'}, \text{'R'}\}$ be the label for each trial $i \in \{1, \ldots, N\}$. The non-centered class covariance matrices are defined as follows:

$$\Sigma_l = \frac{1}{\#\{i : y_i = l\}} \sum_{i:y_i=l} x_i x_i^\top \quad \text{for } l \in \{\text{'L'}, \text{'R'}\}. \tag{1}$$

The choice of calculating the mean in the 150 ms window was made to extract spatial patterns that are valid for the whole time of the analysis window, see also the discussion on stationarity in Section VII. The following steps are similar to [2]. First of all we make a whitening transformation, i.e., we determine a matrix P such that

$$P(\Sigma_{\text{'L'}} + \Sigma_{\text{'R'}})P^\top = I. \tag{2}$$

By spectral theory there is an orthogonal matrix R (columns are eigenvectors) and a diagonal matrix D (eigenvalues) satisfying

$$P\Sigma_{\text{'R'}}P^\top = RDR^\top. \tag{3}$$

Subtracting equation (3) from (2) yields

$$P\Sigma_{\text{'L'}}P^\top = R(1-D)R^\top. \tag{4}$$

Obviously the eigenvalues of corresponding eigenvectors of the transformed covariances matrices sum to one: A direction that has much variance in events of one class (high eigenvalue) has little variance in events of the other class (low eigenvalue). For discriminability between the two classes one should extract from R those eigenvectors which correspond to the eigenvalues close to 0 or 1, i.e. which guarantee a large difference in variance for the two classes. We take 1 or 2 eigenvectors with lowest and 1 or 2 with highest eigenvalue (i.e. 2 or 4 eigenvectors in total). Finally, the projection on the relevant spatial patterns is performed by

$$\text{Signal}_{\text{new}} = \tilde{R}^\top P \ \text{Signal}_{\text{old}},$$

where signals are row vectors and $\tilde{R}$ is the matrix of relevant eigenvectors (columns) extracted from R. Due to the channel-wise linearity of all processing steps this projection on the CSPs can be performed at any position in the processing chain without changing the output. Choice (3) of Fig. 2 makes clear that using

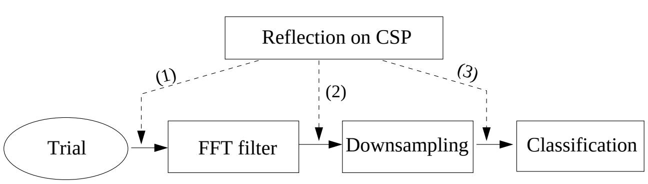

Fig. 2. The schema shows the processing of one single-trial. The projection on the CSPs can be performed in any place of that chain without changing the output because all transformations are linear in each channel.

the CSP projection cannot increase classification performance. This is at least true in the ideal case where the two classes are distributed according to known gaussian distributions: In this case LDA determines a linear projection w of the CSP-projected features on a hyperplane that minimizes the misclassification risk. Since a projection on the same hyperplane can be obtained from the original features (namely by $w\tilde{R}^\top P$) the application of LDA on the original features will attain the same misclassification risk or less. But the dimensionality of the features vectors is

TABLE I

THE TABLE SHOWS THE CROSS-VALIDATION ERRORS FOR ALL EXPERIMENTS. 'UNREG' AND 'REG' DENOTE THE FORMER PROCESSING WITH UNREGULARIZED RESP. REGULARIZED CLASSIFICATION, AND CSP-SCP IS THE PROPOSED NEW APPROACH WITH UNREGULARIZED CLASSIFICATION. THE NUMBER OF USED PATTERNS ARE GIVEN IN THE LAST ROW MARKED #. SUPERSCRIPTS ON THE SUBJECT CODE INDICATE THE NUMBER OF EEG CHANNELS IF DIFFERENT FROM 27.

Subj	Rate	Unreg	Reg	CSP-SCP	#
aa	2 s	6.1 ± 1.2	5.2 ± 0.8	3.5 ± 0.8	2
aa	1 s	15.1 ± 1.8	15.2 ± 0.9	14.2 ± 0.8	2
aa^{52}	1.5 s	12.3 ± 0.7	11.9 ± 0.6	12.3 ± 0.4	2
aa^{52}	1.5 s	7.7 ± 0.4	7.9 ± 0.8	6.4 ± 0.5	4
aa^{120}	2 s	14.5 ± 2.1	7.5 ± 0.7	7.8 ± 0.7	4
al^{52}	2 s	31.2 ± 1.6	31.8 ± 1.4	33.4 ± 0.6	2
ai	1 s	19.4 ± 1.0	18.4 ± 1.0	19.6 ± 0.8	4
ai	2 s	19.3 ± 1.4	18.8 ± 1.0	17.6 ± 1.0	4
ad	1 s	19.7 ± 1.0	20.5 ± 1.0	20.5 ± 0.6	4
ad	2 s	25.1 ± 2.6	22.2 ± 1.8	18.5 ± 1.3	4
ab	2 s	26.4 ± 1.1	25.9 ± 1.5	26.4 ± 0.5	4
af	2 s	20.1 ± 1.3	19.7 ± 0.9	19.3 ± 0.9	2
af	5 s	17.5 ± 1.2	14.7 ± 0.6	15.8 ± 0.5	2
mean		18.0 ± 7.1	16.9 ± 7.6	16.6 ± 8.1	

reduced to 6 or 12 (for 2 resp. 4 CSPs times 3 samples in time) independent of the number of EEG channels. That means regularization is not needed even in a setup with a large number of EEG electrodes or cases where only very few training samples are available. On the other hand it could happen that the few chosen CSPs do not capture all discriminative information and classification results degrade. This issue will be addressed in the results Section VI-A.

The advantage of the CSP method becomes apparent when the projection on the spatial patterns is put in position (1) of Fig. 2. Then the number of channels that has to be sent through the time consuming FFT filtering is substantially reduced, e.g., from 120 to 4. Of course the spatial projection also takes some time. Section VI-B discusses the overall gain of the proposed method in this respect.

VI. RESULTS

A. Classification Accuracy

In table I the results for our previous approach for the regularized and unregularized case can be compared to the results for the new CSP-SCP approach, where a suitable number of patterns (2 or 4) was chosen. The only significant difference in the table is that in the 120 channel setup unregularized classification breaks down. The good news is that as little as 2 or 4 CSP channels are sufficient for a competitive classification.

As example, in Fig. 3 the first spatial pattern is shown for each class (subject *ad* with the 27 channel cap and rate of 2 s). Note that the patterns do not have a direction/polarity. The numbers below the patterns are the corresponding eigenvalues. Due to equation (2) these values range between 0.5 ('non decisive') and 1 ('decisive').

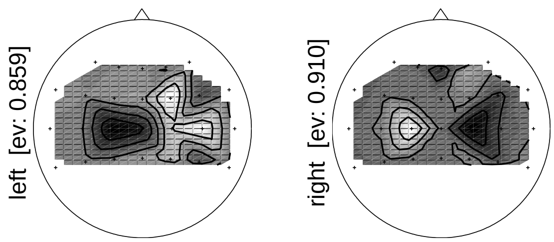

Fig. 3. The first spatial pattern for left resp. right finger events (subject *ad* at a tapping rate of 2 s, 'ev'=eigenvalue). White and black locations are brain areas which give most contribution to the discrimination between the two classes.

B. Computational Load

Having demonstrated that the new approach reaches the same classification accuracy even when using only a few (new) channels we have to assess the gain of the method. Let N be the number of channels, T the number of samples and n the number of spatial patterns we project on. In our previous approach we have to do the FFT on each channel twice, which can be done in $O(NT\log T)$. The following steps of our old method are included in this asymptotic behavior. In our new approach we have to do the projection first, which can be done in $O(nTN)$, and together with the FFT on the remaining channels we obtain $O(nTN + nT\log T)$. In our case we have $n \ll \min(\log T, N)$. Furthermore the hidden constant factors in the asymptotic complexity are much higher for the FFT as for the matrix multiplication. Therefore the new method is expected to be considerably faster than the old one. This was experimentally confirmed (factor of about 10) by comparing the time needed to process and to classify one EEG segment according to the new and to the old method, cf. Table II.

VII. Conclusion

Starting point of the investigations was the BBCI approach of [1] for detecting lateralized slow pre-movement potentials in self-paced finger movements. In that study machine learning techniques were used to analyze the full spatio-temporal structure of multi-channel EEG trials. In the approach presented here, spatial and temporal analysis are split in two separate parts. In the first step of machine training spatial patterns are extracted that are stationary during the time of the analysis window, i.e., from -250 to -100 ms relative to keypress. In the second step the temporal structure is analyzed with respect to the few stationary patterns that we extracted before. This procedure reduces the complexity of the problem considerably. At first sight the methodology might seem unappropriate since it is known from neurophysiological studies that there are different components during the time course of movement related potentials (MRPs) that have distinct spatial patterns, [5]. But the classification algorithm focuses only on the negative slope which immediately preceeds the motor potential. During this time interval that belongs to only one MRP component, the late preparation or BP2, the assumption of stationarity holds to a sufficient degree [5]. This is confirmed for our data by the fact the the CSP-SCP classification results are equally good compared to the full spatio-temporal approach, cf. Table I. The advantange of the new approach is that after having determined the spatial structures the calculation of the BCI feedback signal can be done in a fractional amount of time, cf. Table II. Due to the reduction of the problem complexity it is no longer neccessary to use regularization even when using a large number of channels or having only few training examples.

Furthermore the patterns which are determined by the CSP-SCP technique reveal the spatial structures that hold the most discriminative information of the slow cortical potential variations which can be interpreted neurophysiologically.

TABLE II

The table shows the ratio between the time needed for the previous method and the new CSP-SCP approach for N =27, 52, 120 channels and n =2 or 4 spatial patterns.

$N=27$	2	4
	8.4	4.9
$N=52$	2	4
	11.8	7.3
$N=120$	2	4
	15.4	7.5

VIII. Further Research

The next step is to perform feedback experiments using a larger number of EEG electrodes which became possible with the presented algorithm and verify if the classification accuracy benefits in the same way as it was true for offline analysis. The choice of calculating the mean of the 150 ms window, which was described in Section V is heuristic. Probably the algorithm can be enhanced, e.g., by using a weighted mean.

Furthermore the CSP approach in it present version is only suitable for two-class experiments. A simple way to use CSP for multi-class problems is to combine all pairwise binary classifications or to combine all one-against-the-rest classifications. More sophisticated extensions for multi-class cases are subject of ongoing research.

Acknowledgment

We thank Klaus-Robert Müller for helpful discussions.

References

[1] B. Blankertz, G. Curio, and K.-R. Müller, "Classifying Single Trial EEG: Towards Brain Computer Interfacing", in: T. G. Diettrich, S. Becker, and Z. Ghahramani, eds., *Advances in Neural Inf. Proc. Systems (NIPS 01)*, vol. 14, 157–164, 2002.

[2] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement", *IEEE Trans. Rehab. Eng.*, 8(4): 441–446, 2000.

[3] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-Computer Interface Technology: A review of the First International Meeting", *IEEE Trans. Rehab. Eng.*, 8(2): 164–173, 2000.

[4] Z. J. Koles and A. C. K. Soong, "EEG soure localization: implementing the spatio-temporal decomposition approach", *Electroencephalogr. Clin. Neurophysiol.*, 107: 343–352, 1998.

[5] R. Q. Cui, D. Huter, W. Lang, and L. Deecke, "Neuroimage of voluntary movement: topography of the Bereitschaftspotential, a 64-channel DC current source density study", *Neuroimage*, 9(1): 124–134, 1999.

[6] J. Knuutila and M. Hämäläinen et al., "Characterization of brain noise using a high sensitivity 7-channel magnetometer", *Proc. 6th Inst. Conf. Biomag.*, 186–189, 1987.

[7] L. Parra, C. Alvino, A. C. Tang, B. A. Pearlmutter, N. Yeung, A. Osman, and P. Sajda, "Linear spatial integration for single trial detection in encephalography", *NeuroImage*, 2002, to appear.

[8] K.-R. Müller, C. W. Anderson, and G. E. Birch, "Linear and Non-Linear Methods for Brain-Computer Interfaces", *IEEE Trans. Neural Sys. Rehab. Eng.*, 2003, submitted.

[9] J. H. Friedman, "Regularized Discriminant Analysis", *J. Amer. Statist. Assoc.*, 84(405): 165–175, 1989.