16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

MODEL-BASED SOURCE SEPARATION FOR MULTI-CLASS MOTOR IMAGERY

C. Gouy-Pailler[1], M. Congedo[1], C. Jutten[1], C. Brunner[2] and G. Pfurtscheller[2]

[1]GIPSA-lab/CNRS/INPG-UJF-UPMF-Stendhal
46 Avenue Félix Viallet, 38000 Grenoble, FRANCE
phone: + (33) 476574353, fax: + (33) 476570000
email: Cedric.Gouy-Pailler@gipsa-lab.inpg.fr

[2]Institute for Knowledge Discovery, BCI Lab
Krenngasse 37, 8010 Graz, AUSTRIA
Graz University of Technology
email: clemens.brunner@tugraz.at

ABSTRACT

This paper presents a general framework to recover task-related sources from a multi-class Brain-Computer Interface (BCI) based on motor imagery. Our method gathers two common approaches to tackle the multi-class problem: 1) the supervised approach of Common Spatial Patterns and Sparse and/or Spectral variants (CSP, CSSP, CSSSP) to discriminate between different tasks; 2) the criterion of statistical independence of non-stationary sources used in Independent Component Analysis (ICA). Our method can exploit different properties of the signals to find the best discriminative linear combinations of sensors. This yields different models of separation. This work aims at comparing these models. We show that the use of a priori knowledge about the sources and the performed task increases classification rates compared to previous studies. This work gives a general framework to improve Brain-Computer Interfaces and to adapt spatial filtering methods to each subject.

1. INTRODUCTION

The ultimate goal of Brain-Computer Interfaces (BCIs) is to provide disabled people suffering from severe motor diseases with a tool to restore communication and movement [2]. Although substantial progress has been made in the field, some improvements could help BCIs to move out of the laboratories. Brain-computer systems are always an association between an electrophysiological phenomenon and a signal processing algorithm. A typical example of a BCI is based on the imagination of movement, which results in somatotopic brain signal variations in specific frequency bands [11, 5].

On the one hand, Independent Component Analysis (ICA) has been widely used for analyzing and cleaning brain signals in electroencephalography (EEG). This approach, initiated in the early 90's by Jutten and Hérault [7], aims at tackling the Blind Source Separation (BSS) problem (neither a priori information about the sources nor the mixing process is provided) by assuming mutual statistical independence between sources. Such models have proved useful to increase classification rates of BCIs [4, 10], but do not use a priori information about the tasks, namely the labels of tasks during the training step. Different separation principles can be used to tackle the BSS problem. They depend on the statistical properties of sources, and on how statistical independence is evaluated. When sources are assumed to be independent and identically distributed (iid), non-gaussianity of sources is used, which involves higher order statistics or mutual information. The non-gaussianity assumption case can be released, yielding other families of algorithms based on coloration or time-varying energy [13]. The efficient algorithm provided by Pham [13] to jointly diagonalize matrices allows to use different kinds of diversity in the time and frequency domains, therefore it is suitable for comparing different separation models.

On the other hand, the goal-oriented approach of Common Spatial Patterns (CSPs) has been developed and improved since its introduction in [9]. The idea of CSP is to find the linear combination optimizing the ratio between intra and inter-class covariance matrices. Many improvements were achieved: 1) by considering frequency-specific covariance matrices [8] (CSSP); 2) the Sparse-Spatio Spectral version (CSSSP) aimed at jointly optimizing a temporal filter [6]; 3) and the invariant CSP (iCSP) was able to reduce the negative effects introduced by non-stationary signals [3]. This approach proved useful to discriminate two motor imagery tasks but suffers from a lack of generalization to multi-class problems. A one-versus-rest (OVR) approach has been proposed [14] to generalize the approach to multi-class discrimination problems, but it suffers from some obvious limitations. For example, if one class is the exact barycenter of the two others, a one-versus-rest approach would fail.

This paper presents an approach to use both the supervised idea of CSP to exploit our knowledge about performed tasks and the theoretical framework of ICA for non-stationary sources to recover task-related sources. Different properties of the signal can be used to perform the separation. The resulting different models are compared and quality of separation is assessed by classification rates in a 8-subject 4-class motor imagery experiment (left hand, right hand, foot and tongue). The remainder of this paper is organized as follows: in section 2, we present the experimental paradigm, section 3 provides the reader with the different models and methods we compared; finally we present and discuss results.

2. SUBJECTS AND EXPERIMENTAL PARADIGM

In this study, the EEG data of eight subjects (three females and five males with a mean age of 23.8 years and a standard deviation of 2.5 years, [10, 4]), recorded during a cue-based four-class motor imagery task, was analyzed. Two sessions on different days were recorded for each subject, each session consisting of six runs separated by short (a couple of minutes) breaks. One run consisted of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session.

The subjects were sitting in a comfortable armchair in front of a computer screen. As mentioned above, the paradigm consisted of four different tasks, namely the imagination of movement (motor imagery) of the left hand, right hand, foot, and tongue, respectively. At the beginning of each trial ($t = 0$ s), a fixation cross appeared on the black screen. In addition, a short acoustic warning tone was presented at this time instant. After two seconds (at $t = 2$ s), a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes left hand, right hand, foot or tongue) appeared for 1.25 s, prompting the subjects to perform the target motor imagery task. No feedback (neither visual nor acoustic) was provided. The subjects were asked to carry out the mental imagination until the fixation cross disappeared from the screen at $t = 6$ s. A short break followed, lasting at least 1.5 s. After that, the next trial started. The paradigm is illustrated in Figure 1 (left).

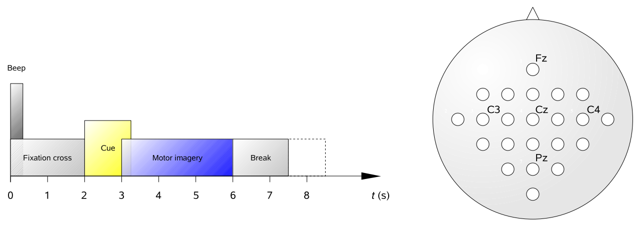

Figure 1: Timing scheme of the BCI paradigm (left) and electrode setup of the 22 channels with inter-electrode distances of 3.5 cm. Some locations corresponding to the international 10-20 system are labeled (right).

22 Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm) were used to record the EEG, the setup is depicted in Figure 1 (right). Monopolar derivations were used throughout all recordings, where the left mastoid served as reference and the right mastoid as ground. The signals were sampled at 250 Hz and bandpass-filtered between 0.5 and 100 Hz. An additional 50 Hz notch filter was enabled to suppress power line noise.

Although a visual inspection of the raw EEG data was performed by an expert, no trials were removed from the subsequent analysis in this study in order to evaluate the robustness and sensitivity to outliers and artifacts of each model. Three EOG channels and one ECG channel were also used to measure electrophysiological activity of the subjects. Those channels were added to the analysis of the data according to the method described below.

3. METHODS

3.1 Method Overview

Figure 2 shows an overview of the whole processing stage during the training and the test step. The wavelet box in the training step processing overview is represented in gray because its use depends on the model employed for source separation (see below). The spatial filter computation is done according to the different methods described below. The training step is used to fix some of the parameters of the method whereas the test step consists in applying the procedure to unseen data with previously fixed parameters. The dashed lines represent information, which is shared between learning and test steps. Trainings are dependent from subjects and sessions.

3.2 Preprocessing

3.2.1 General Source Separation Framework

The main goal of this work is to find the best spatial filters using some statistical models of source separation and a priori knowledge about the tasks performed. We know that source

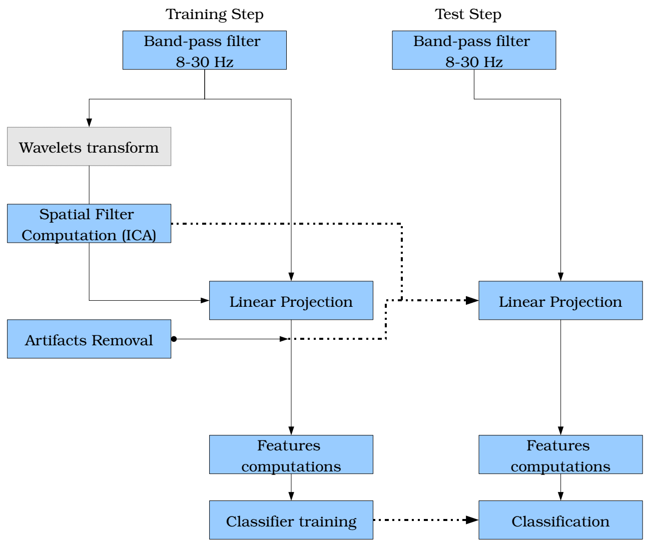

Figure 2: Training and test steps. Overview of the method.

separation is impossible if sources are modeled as Gaussian, identically and independently distributed. Whereas some algorithms like Infomax try to maximize independence, therefore using non-gaussianity of the sources; one other way to move from the Gaussian iid case is to consider simple time structures within the data. The goal of our work is to study the performance of some simple time structures by assuming that, in some basis (Fourier, wavelets, time), the coefficient of the process are uncorrelated with a smoothly varying variance. Such a general model leads to separation principles based only on second-order statistics.

The general framework is that we are trying to recover sources $s(t)$ related to each task by assuming the simplest source separation model for linear or convolutive mixtures of sensor measurements $x(t)$:

$$x(t) = As(t) \quad \leftrightarrow \quad X(t,f) = AS(t,f) \tag{1}$$
$$x(t) = A \star s(t) \quad \leftrightarrow \quad X(t,f) = A(f)S(t,f) \tag{2}$$

with A unknown and independent rows in s. The arrow means that a linear transform is applied, typically a short Fourier transform or a wavelet transform to take into account the time-frequency or time-scale properties of the signal. The separation principle given by Pham and Cardoso in [13] aiming at exploiting slow-varying variances of sources yields the joint diagonalization of covariance matrices either in the time or frequency/scale domain.

In the time domain, this principle leads to consider a partition of the observation interval into $\mathscr{T}_q$, with $q \in [1..Q]$. For each time interval, we define the covariance matrix:

$$C_{xx}(\mathscr{T}_q) = \mathbb{E}_{t \in \mathscr{T}_q}(x(t)x(t)^T)$$

Then, the estimation of the separation matrix $B = A^{-1}$ is done by approximately jointly diagonalizing the set

$$\mathscr{S} = \{C_{xx}(\mathscr{T}_q) | q \in [1..Q]\}$$

This joint diagonalization is done by the Pham's algorithm [13] by minimizing the following criterion:

$$C(\mathbf{B}) = \sum_{q=1}^{Q} w_q \left[\log\det\operatorname{diag}\left(\mathbf{B}C_{XX}(\mathscr{T}_q)\mathbf{B}^T\right) - \log\det\left(\mathbf{B}C_{XX}(\mathscr{T}_q)\mathbf{B}^T\right)\right]$$

where w_q is used to normalize covariance matrices to unit traces. Using the Hadamart inequality [12], it can be shown that this criterion is zero if and only if every matrices are diagonal.

The same procedure can be applied in the frequency/scale domain by previously projecting the signals onto a Fourier or wavelet basis. In this case it exploits the time-frequency diversity of sources. The model remains the same in such a basis due to linearity of the transform (wavelet basis).

In such a case covariance matrices are defined as:

$$C_{XX}(\mathscr{T}_q, f) = \mathbb{E}_{t\in\mathscr{T}_q}(X(t,f)X(t,f)^T)$$

Then, the estimation of the separation matrix B is done by approximately jointly diagonalizing the set

$$\mathscr{S} = \{C_{XX}(\mathscr{T}_q, f) | q \in [1..Q]\}$$

This joint diagonalization is done by the Pham's algorithm [13]. The criterion is the same as above except that the sum is over q and f.

Thus the general procedure of this framework is summarized by:

1. for a partition $\{\mathscr{T}_1, \cdots, \mathscr{T}_Q\}$ of the observation interval in Q subintervals, compute local sample covariance matrices $C_{xx}(\mathscr{T}_q)$,
2. jointly diagonalize the local normalized covariance matrices using the Pham's algorithm:

A priori knowledge about the performed tasks during training is included by considering only task-specific covariance matrices. This makes our approach close to CSPs but our advantage is that the multi-class generalization problem due to the Rayleigh quotient in CSP is avoided by our approach based on independence.

3.2.2 *Model-Based source separation for Spatial Filtering*

In the following, k will represent the index of the mental tasks $k \in [1..4]$. Frequency domains will be denoted as mu or beta (μ or β), which represent wavelet transforms of the signals at scales 4 and 3 respectively. Those bands are known to be involved in motor imagery [11]. Discrete wavelet transform was used as frequency basis (Daubechies). $\mathbb{E}_k(.)$ will denote the average across all trials related to class k. $C_{xx}(t \in [t_1, t_2], k)$ will denote the set of covariance matrices for every trial of one session of a subject for task k computed with EEG in the time domain between t_1 and t_2. And finally $C_{XX}(t \in [t_1, t_2], k, f)$ will denote the set of covariance matrices for every trials of one session of a subject for task k computed with EEG in the wavelet basis domain between t_1 and t_2.

Different kinds of diversities are to be considered in the following models:

1. Inter-class diversity (ICD): sources related to motor imagery have a varying energy among classes. We exploit the fact that a source active for one mental task is active with a different energy (or not active at all) for another mental task. This kind of *diversity* is exploited by considering task-specific covariance structures, it is used by CSP to find discriminative linear transforms of sensors.
2. Time-varying energy (TVE): as motor tasks are known to be a succession of activations in different brain areas, it can be assumed that sources related to a mental task realization can be active with different energies across the task. Joint diagonalization covariance matrices computed using successive time windows will help recovering sources [13].
3. Coloration (Col): sources are known to be non-white in the brain. This information can be used in the frequency domain by jointly diagonalizing matrices in the frequency domain. Pham showed [12] that sources can be recovered using this criterion if they have non-proportional power-spectra.

Lastly signals can be considered either in the time domain (TD) or in the frequency domain (FD).

3.2.3 *Exploiting inter-class diversity in the time domain*

This first model uses ICD. We recall that this kind of *diversity* is exploited by considering task-specific covariance structures. For each trial of one specific task, we compute the covariance matrix of the EEG from $t = 2.5$ s to $t = 7.5$ s. Then we average across every trials of one specific task. As this is done for every mental task, the procedure leads to a joint diagonalization of 4 covariance matrices (one for each task):

$$\mathscr{S} = \{\mathbb{E}_k(C_{xx}(t \in [2.5, 7.5], k)) \quad | \quad k \in [1..4]\}$$

3.2.4 *Exploiting inter-class diversity and time-varying energy in the time-domain*

This second model aims at exploiting the idea that sources are active with different energies between different tasks and/or that the energy of a source is time-varying inside one task. This information is used by partitioning the previous interval to 4 subintervals, $[2.5, 7.5] = \cup_{i=1}^{4} \mathscr{T}_i$. Thus the diagonalization set consists of 16 covariance matrices:

$$\mathscr{S} = \{\mathbb{E}_k(C_{xx}(t \in \mathscr{T}_i, k)) \quad | \quad i \in [1..4], k \in [1..4]\}$$

3.2.5 *Exploiting inter-class diversity and coloration in the frequency domain*

As brain waves variations resulting from mental tasks are frequency-specific, we also designed a model to exploit diversity of frequency distribution of sources. This assumes that sources have non-proportional power spectra and/or a coloration diversity. The discrete wavelet transform is used to project signals onto wavelet basis (Daubechies). The set of diagonalization consists of 8 covariance matrices in the frequency domain, 2 scales for each task:

$$\mathscr{S} = \{\mathbb{E}_k(C_{XX}(t \in [2.5, 7.5], k, f)) \quad | k \in [1..4], f \in \{\mu, \beta\}\}$$

3.2.6 *Exploiting separately inter-class diversity in the frequency domain, convolutive model*

The last method consists in assuming that the structure of the mixing process has to be considered separately in each band (different mixing models in μ and β). This assumption implies that sources in mu band have to be found separately from sources in beta band. This yields to 2 joint diagonalizations of 4 frequency-specific covariance matrices (4 tasks):

$$\mathscr{S}_\mu = \{\mathbb{E}_k(C_{XX}(t \in [2.5, 7.5], k, \mu)) \quad | k \in [1..4]\}$$

$$\mathscr{S}_\beta = \{\mathbb{E}_k(C_{XX}(t \in [2.5, 7.5], k, \beta)) \quad | k \in [1..4]\}$$

Such a model results in two separation matrices B_μ and B_β, and thus two different sets of frequency-specific sources.

3.2.7 *Artifacts removal*

As EOG and ECG channels was available for this experiment, we added those channels into the separation and rejected the sources that were highly correlated with non-EEG channels (correlation coefficient greater than 0.5). Only columns of the mixing matrix corresponding to clean channels were kept for the test step.

3.3 Features Extraction

Our convolutive model yields frequency-specific sources, whereas the sources related to the other models are not frequency-specific. Due to the elimination of contaminated sources (high correlation with EOG channel), the number of sources can vary. We detected that, whereas the convolutive model could theoretically result in twice the number of sources we get for other models, our elimination step removes much more sources for the convolutive model than for any other model. Thus we observed that the number of viable sources remains almost the same for each model (around 20 sources).

In line with physiological knowledge of movement imagination, our feature extraction evaluates frequency-specific band power in the μ and β band. Power spectral density estimations were computed using Discrete Wavelet Transform (3 seconds of signal is used to compute the discrete wavelet transform). As the frequency sampling rate of the data is 250 Hz, only scales four and three (corresponding to the μ and β bands) are considered. Features from our frequency-specific sources (convolutive model) were considered differently: μ band power is computed from μ-specific sources and β band power is computed from β-specific sources. This finally results in around 40 features for every model.

3.4 General Procedure

For each subject and each session, a 10×10-fold cross-validation is performed. For each cross-validation, 90 % of the trials are used as training trials and the remaining of the data is used as test trials. A linear discriminant analysis (LDA, [10]) is used to classify features.

4. RESULTS

4.1 Cross-Validated Results

We present in Table 1 the mean classification accuracy across subjects and sessions for cross-validated data. These results do not show any significant superiority of one model among the others. A high diversity of skills among subjects for controlling a BCI accounts for this result. Nevertheless, methods based on frequency domain seems to perform slightly worse than methods based on time domain. Overall performance of our methods is very satisfying regarding the difficulty of the considered task. Our method do not benefit from any feature selection algorithm, therefore it can be compared with [10] in which the best classification rate is achieved with Infomax (about 65 %). Results obtained in [4] outperforms the results presented here (ranging from 65 to 75 %) but used a feature selection (sequential forward selection) to range about 1300 features from the feature extraction step.

High variability of classification rates across subjects (ranging from 40 to 80 percents) leads us to consider subject-specific results. Table 2 presents results for each subject and each session. The classification rate (percentage and standard deviation) is considered in the second column of the table. The third column of the table gives the corresponding best model (notations are those introduced in section 3). From basic probability calculus we can derive our specific significance test: probability that one method appears 7 or more times as the best method if methods are randomly ranked is 0.08 (cumulative binomial distribution with parameters $p = 0.25$ and $n = 16$). Thus the table 2 shows that our method based on the use of inter-class diversity and time-varying energy in the time-domain is significantly best than the others ($p = 0.08$). It appears clear that the method based on inter-class diversity and coloration in the frequency domain performs worse than the others (best method for only one session).

Lastly we note an overall superiority of the time domain based methods over the frequency domain based methods, respectively the two first and two last methods. This fact appears obvious by considering the number of session for which time based methods appears as best models (12 over 16) against the number of frequency-based methods ranked as best (4 over 16 sessions).

	Mean (%)	Std Deviation
ICD, TD	69.3	17.4
ICD, TVE, TD	70.2	16.2
ICD, col., FD	66.8	16.8
Convolutive	68.8	17.2

Table 1: Summary of the cross-validated performance

	Correct (%)	Best Model
S1 ses1	77.2 (1.27)	ICD, TD
S1 ses2	78.8 (1.32)	ICD, TD
S2 ses1	52.3 (2.00)	ICD, TVE, TD
S2 ses2	56.0 (2.08)	ICD, TVE, TD
S3 ses1	82.3 (1.08)	ICD, TVE, TD
S3 ses2	84.3 (0.86)	ICD, TD
S4 ses1	83.4 (1.24)	ICD, col., FD
S4 ses2	80.7 (2.11)	ICD, ,TVE, TD
S5 ses1	77.1 (1.38)	ICD, TD
S5 ses2	88.4 (1.31)	Convolutive
S6 ses1	55.8 (2.07)	ICD, TVE, TD
S6 ses2	66.1 (1.83)	Convolutive
S7 ses1	39.1 (1.44)	ICD, TVE, TD
S7 ses2	43.6 (2.31)	ICD, TD
S8 ses1	81.4 (0.84)	Convolutive
S8 ses2	86.7 (0.90)	ICD, TVE, TD

Table 2: Classification rates for each session (ses1 or ses2) and each subject (S1 to S8) given by the best model. The standard deviation is given between round brackets.

5. DISCUSSION

Different a priori information were considered in this paper, namely we used Inter-Class Diversity, Time-Varying Energy and Coloration either in the time (for Time-Varying Energy) or in the frequency (Coloration) domain. First of all, we showed in section 3 that finding multi-class spatial filters can benefit from the use of simple a priori knowledge. Then some simple tests of significance showed that exploiting Inter-Class Diversity and Time-Varying energy in the time domain proved better than the other methods considered here. It was quite obvious that using a priori knowledge about the tasks performed would improve classification rates. But improvements due to a priori knowledge about time-varying energy was quite surprising. This result supports the hypothesis that different sources appears during the performance of the tasks and that their time course is not constant. Time interval partitioning was very simple and we think that some refined partitioning of intervals could result in significant improvements of the classification rates.

We pointed out a disadvantage of such a refined framework by showing that none of the presented methods could be considered as best for every subjects. Unsurprisingly, the design of optimal spatial filters have to cope with inherent difficulties of studying brains and real subjects: methods have to be subject-dependent to yield optimal results. The set of methods presented here raises the classical issue of Occam's Razor: fitting data with the most refined model could be dangerous if subjects' imagined tasks are significantly modified by training. These two considerations have to be tackled in two different ways to make such signal processing algorithm available for daily life use: 1) an automatic procedure has to be designed to select subject-specific methods; 2) adaptive algorithms will be necessary to adapt systems to subjects.

In summary, we presented here an efficient framework for increasing classification rates of multi-class BCI paradigms. Our framework is well grounded on the Pham's theoretical work about joint approximate diagonalization and provides natural a priori knowledge that can be used to gather advantages of both Independent Component Analysis (we assume statistical independence of sources) and Common Spatial Patterns.

6. CONCLUSION

We presented different models to use a priori knowledge about the performed task. Those models are based on the mathematical theory of non-stationary source separation. Our models do not suffer from any multi-class limitations and yield good results regarding the complexity of our treatment.

Acknowledgments

This work was supported by the French Ministry of Defense (Délégation Générale pour l'Armement – DGA), the OpenViBE project (Open-Source platform for a Virtual Brain Environment, Agence Nationale de la Recherche – ANR), and by the European program of Cooperation in the Field of Scientific and Technical Research (COST B27 – Electric Neuronal Oscillations and Cognition).

REFERENCES

[1] J. Anemüller, T. J. Sejnowski, and S. Makeig. Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Netw*, 16(9):1311–1323, Nov 2003.

[2] N. Birbaumer and L. G. Cohen. Brain-computer-interfaces (BCI): Communication and restoration of movement in paralysis. *J Physiol*, 579(3):621–636, Jan 2007.

[3] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K.-R. Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008.

[4] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller. Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. *Pattern Recognition Letters*, 28:957–964, 2007.

[5] M. Congedo, F. Lotte, and A. Lécuyer. Classification of movement intention by spatially filtered electromagnetic inverse solutions. *Phys Med Biol*, 51(8):1971–1989, Apr 2006.

[6] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Trans Biomed Eng*, 53(11):2274–2281, Nov 2006.

[7] C. Jutten and J. Hérault. Blind separation of sources, part 1: an adaptive algorithm based on neuromimetic architecture. *Signal Process.*, 24(1):1–10, 1991.

[8] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans Biomed Eng*, 52(9):1541–1548, Sep 2005.

[9] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin Neurophysiol*, 110(5):787–798, May 1999.

[10] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller. Seperability of four-class motor imagery data using independent components analysis. *J Neural Eng*, 3(3):208–216, Sep 2006.

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

[12] D.-T. Pham. Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion. *Signal Processing*, 81(4):855–870, 2001.

[13] D.-T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Transactions on Signal Processing*, 49(9):1837–1848, Sept. 2001.

[14] W. Wu, X. Gao, and S. Gao. One-Versus-the-Rest (OVR) algorithm: An extension of common spatial patterns(CSP) algorithm to multi-class case. *Conf Proc IEEE Eng Med Biol Soc*, 3:2387–2390, 2005.