Multi-Class Independent Common Spatial Patterns: Exploiting Energy Variations of Brain Sources

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

[1]GIPSA-lab, CNRS/UJF-UPMF-Stendhal, Grenoble, FRANCE
[2]Institute for Knowledge Discovery, BCI Lab, Graz University of Technology, Graz, AUSTRIA

cedric.gouypailler@gipsa-lab.inpg.fr

Abstract

This paper presents a method 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 Pattern (CSP) to discriminate between different tasks; 2) the criterion of statistical independence of non-stationary sources used in Independent Component Analysis (ICA). We show that the resulting spatial filters have to be adapted to each subject and that the combined use of intra-trial and inter-class energy variations of brain sources yield an increase of classification rates for four among eight subjects.

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 [1]. A typical example of a BCI is based on movement imagery, which results in somatotopic brain signal variations in specific frequency bands [3].

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 the mixing matrix nor the sources are known) by assuming mutual statistical independence between sources. Such models have proved useful to increase classification rates of BCIs [2, 9], 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 required, which involves higher order statistics or mutual information. The non-gaussianity assumption case can be relaxed, yielding other families of algorithms based on second order statistics and requiring coloration or time-varying energy [10].

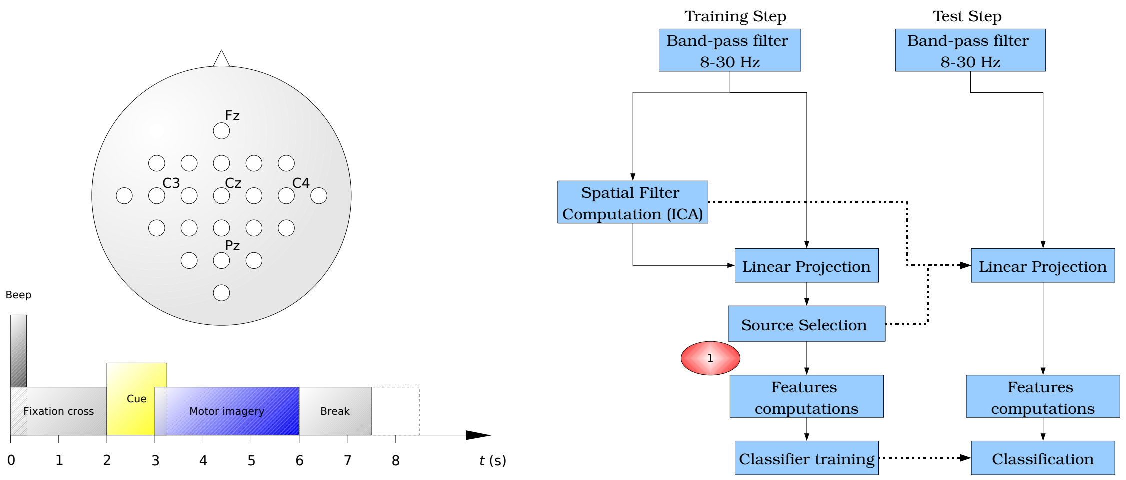
On the other hand, the goal-oriented approach of Common Spatial Pattern (CSP) has been introduced in [8]. The idea of CSP is to find the linear combination optimizing the ratio between within-class scatter and the mixture scatter matrices. From a methodological point of view, it is nothing but an exact joint diagonalization of two matrices, hence very similar to Approximate Joint Diagonalization (AJD). 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 is often used to generalize the approach to multi-class discrimination problems. Following ideas from [4], an extension to multi-class problems has been proposed in [6, 5]. These approach were based on AJD of sample covariances matrices.

Extending the work of [6, 5], this paper presents an approach to use intra-trial energy variations of sources and inter-class diversity. Our method is compared to CSP and the approach proposed in [6]. The 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 detailed description of our method; 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, [9, 2]), 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.

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 (a).



(a) Experimental paradigm (b) Method Overview

Figure 1: (a): Timing scheme of the BCI paradigm and electrode setup of the 22 channels. (b) Method overview

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. 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.

3 Methods

We begin by stating the notations. $x \in \mathbb{R}^N$ represents EEG data, recorded at N electrodes at each time t. In this work, we aim at finding a linear transformation of the data $s = W^T x$ to increase the classification rate. The following section presents different methods for finding W, based on different criteria. Intentions of the users are called classes and indexed $k \in [1..4]$.

Figure 1 (b) shows an overview of the whole processing stage during the training and the test step. 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.

3.1 Method 1: Common Spatial Patterns

The idea of Common Spatial Patterns (CSP, [8]) for two-class problems is to find the more discriminative spatial filters $v \in \mathbb{R}^N$, which optimizes the Rayleigh quotient:

$$\{\min, \max\} \frac{v^T C_{xx|k=1} v}{v^T C_{xx|k=1,2} v}$$

where $C_{xx|k=i}$ is the covariance matrix of the data belonging to class c_i. This optimization problem can be solved by a generalized eigenvalue decomposition method. An advantage of this technique is that spatial filters are ranked according to their discriminative power, thus allowing to select specific features dimension L. Computations are also exact and fast. Although this method is optimal for two-class problems, extensions to multi-class paradigms is not straight-forward. We use in the following a One-Versus-Rest CSP to generalize to multi-class problems.

3.2 Multi-Class Independent Common Spatial Patterns

Whereas some algorithms try to maximize independence, *e.g.* using non-gaussianity of the sources without considering time structure, another way to separate source components is to consider simple time structures within the data. The goal of our work is to study the performance of some simple time structures. We thus look for some kind of non-stationarity in the data.

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 mixtures of sensor measurements $x(t)$:

$$x(t) = As(t) \tag{1}$$

where A is the mixing matrix and $s$are the sources. The separation principle given by Pham and Cardoso in [10] aiming at exploiting slow-varying variances of sources yields the joint diagonalization of covariance matrices.

The observation interval is partitioned into Q parts: $\mathcal{T}_q$, with $q \in [1..Q]$. For each time interval, we define the covariance matrix:

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

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

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

This joint diagonalization may be performed for example by the Pham's algorithm [10]

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 with the advantage of inherently being a multi-class approach.

3.2.1 Model-Based source separation for Spatial Filtering

In the following, $\mathbb{E}_k(.)$ 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.

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 [10].

We want to stress the differences between these approach based on source-separation and the approach based on CSP. First of all, whereas CSP tries to find the quasi-optimal linear combination that optimizes the Rayleigh quotient given above, our methods try to incorporate the best physiological spatial diversity. These approaches not only estimates the most discriminating sources but also allow to recover some independent neurophysiological brain waves (according to the spatial diversity considered). Lastly we highlight the fact that the types of diversity mentioned here are sufficient conditions that can be provided to the joint diagonalization algorithm.

3.2.2 Method 2: Exploiting inter-class diversity

This first model uses ICD and was shown to outperform the classical CSP [6]. 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 $t = 2.5\,\mathrm{s}$ to $t = 7.5\,\mathrm{s}$. 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):

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

3.2.3 Method 3: Exploiting inter-class diversity and time-varying energy

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, $[2.5, 7.5] = \cup_{i=1}^{4} \mathcal{T}_i$. Thus the diagonalization set consists of 16 covariance matrices:

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

3.3 Global Procedure

In order to test the generalization ability of each method, a cross-validation procedure is used. For each subject and each session, we have 72 trials for each of the four classes. We permute the 72 trials of each class to obtain four randomly ordered sets of labeled trials. We then select the first 7 trials of the four randomly ordered sets. They constitute the first test set of the cross-validation, the remaining trials constitute the first training set of the cross-validation. The second test set will be the 7 next trials of each class. One cross-validation is completed when the ten successive test sets and their associated training sets have been considered.

For each test and training sets, we apply the procedure as illustrated in figure 1 (b). We first consider the band-pass filtered signals corresponding to the training set to find the optimal spatial filters according to each methods. In the case of CSP, the best potential spatial filters are naturally ranked in the method, we only keep the first and last spatial filter of each of the one-versus-rest CSP, thus resulting in 8 spatial filters. In the case of method 2 and 3, the potential components are not ranked, we thus select relevant sources using the same method as [6], based on an approximation of the mutual information between the label and the sources. In order to achieve a fair comparison between the three methods, we also select the 8 best ranked sources and their associated spatial filters. Thus at the end of step one in figure 1, we have 8 sources for each method. They correspond to the linear projections of the data onto the source space, depicting the same time courses as the EEG measurements.

The next step consists in computing the features related to each trial of the training set. In line with neurophysiological considerations, we computed the energy of each sources in the μ and β band. This estimation is made by computing the Discrete Wavelet Transform of each sources. Thus the number of total features to be classified is 16 for each method. The features are gathered in a 65×16 matrix to train a LDA [9]. Parameters of the LDA are conserved for the test step.

Lastly, as depicted in 1, the procedure is applied on the test set using the selected spatial filters to project the data, the same method to compute features for each trial. The LDA is used to classify features gathered in a 7×16 matrix.

This procedure is in fact applied 100 times, which corresponds to 10 cross-validations, a cross-validation consisting of 10 disjoint test sets.

4 Results

The mean classification accuracies across subjects and sessions are not significantly different: 70.7 %, 70.7 % and 70.6 % for respectively the CSP, ICD and ICD&TVE. Furthermore, we found a strong inter-subject variability. Overall, performances of our methods are satisfying considering the difficulty of the task. Our methods differ from the one employed in [9] because they did not select features according to some qualitative criterion. A slight increase of classification rates is thus not surprising. The best result was achieved with Infomax and was about 65 %. Moreover, Infomax was used in a completely blind manner and did not use any a priori information about the performed task to achieve the separation. Results obtained in [2] outperforms the results presented here (ranging from 65 to 75 %) but used a numerically demanding 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 1 presents results for each subject and each session. The classification rate (percentage and standard deviation) is considered in the second column of the table. All pairwise t-test comparing the three models for each subject separately using the cross-validations as observation units reveals that the best model outperforms the other two ($p < 0.5$) for five out of eight subjects (S2, S4, S5, S6, S7).

	Correct [%] (Std Dev)	Best Model
S1	80.6 (0.9)	ICD, TVE
S2	53.9 (2.2)	ICD, TVE
S3	86.4 (1.5)	ICD
S4	84.5 (2.2)	CSP

	Correct [%] (Std Dev)	Best Model
S5	82.1 (7.3)	CSP
S6	62.8 (5.6)	CSP
S7	43.3 (2.6)	ICD, TVE
S8	86.0 (3.5)	ICD, TVE

Table 1: Classification rates for each subject (S1 to S8) given by the best model.

4.1 Discussion

Different a priori information were considered in this paper, namely we used Inter-Class Diversity and Time-Varying Energy. First of all, we showed in section 3 that finding multi-class spatial filters can benefit from the use of simple a priori knowledge. 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. This consideration has to be tackled to make such signal processing algorithm available for daily life use: an automatic procedure should be designed to select subject-specific methods.

5 Conclusion

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 and Common Spatial Patterns.

References

[1] 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.

[2] 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.

[3] 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.

[4] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans Biomed Eng*, 51(6):993–1002, Jun 2004.

[5] C. Gouy-Pailler, M. Congedo, C. Jutten, C. Brunner, and G. Pfurtscheller. Model-based source separation for multi-class motor imagery. In *Proceedings of the 16th European Signal Processing Conference (EUSIPCO-2008), EURASIP*, Lausanne, Switzerland, August 2008.

[6] M. Grosse-Wentrup and M. Buss. Multi-class common spatial pattern and information theoretic feature extraction. *IEEE Transactions on Biomedical Engineering (accepted)*, 2008.

[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] 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.

[9] 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.

[10] 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.