Open access • Journal Article • DOI:10.1109/TBME.2009.2032162

Nonstationary Brain Source Separation for Multiclass Motor Imagery

— Source link

Cédric Gouy-Pailler, Marco Congedo, Clemens Brunner, Christian Jutten ...+1 more authors

Institutions: Graz University of Technology

Published on: 01 Feb 2010 - IEEE Transactions on Biomedical Engineering (IEEE Trans Biomed Eng)

Topics: Blind signal separation

Non-Stationary Brain Source Separation for Multi-Class Motor Imagery

Cedric Gouy-Pailler, Marco Congedo, Clemens Brunner, Christian Jutten, Gert Pfurtscheller

Non-Stationary Brain Source Separation for Multi-Class Motor Imagery

Cédric Gouy-Pailler, Marco Congedo, Clemens Brunner, Christian Jutten, *Fellow, IEEE,* and Gert Pfurtscheller

***Abstract*—This article describes a method to recover task-related brain sources in the context of multi-class Brain-Computer Interfaces (BCIs) based on non-invasive electroencephalography (EEG). We extend the method Joint Approximate Diagonalization (JAD) for spatial filtering using a maximum likelihood framework. This generic formulation (1) bridges the gap between the Common Spatial Patterns (CSP) and Blind Source Separation (BSS) of non-stationary sources, and (2) leads to a neurophysiologically adapted version of JAD, accounting for the successive activations/deactivations of brain sources during motor imagery trials.**

Using dataset 2a of BCI Competition IV (2008) in which nine subjects were involved in a four-class two-session motor-imagery (MI) based BCI experiment, a quantitative evaluation of our extension is provided by comparing its performance against JAD and CSP in the case of cross-validation as well as session-to-session transfer. Whereas JAD, as already proposed in other works, does not prove to be significantly better than classical one-versus-rest CSP, our extension is shown to perform significantly better than CSP for cross-validated and session-to-session performance. The extension of JAD introduced in this paper yields among the best session-to-session transfer results presented so far for this particular dataset, thus it appears of great interest for real-life BCIs.

***Index Terms*—Brain-Computer Interfaces, Multi-class motor imagery, joint approximate diagonalization.**

I. Introduction

THE aim of non-invasive Brain-Computer Interfaces (BCI) is to establish a direct communication pathway between human intentions and electronic devices [1]. In a medical context, BCIs are conceived to provide people suffering from severe motor disabilities with a tool to rehabilitate communication and movement [2], [3]. They entail central nervous system activity to be measured, usually by electroencephalography (EEG), and then converted into an appropriate device command. Although not every intent of the user can be decoded due to the complexity of the brain, a small subset of mental tasks are sufficiently known from a neurophysiological point of view to be extracted and mapped onto commands. A well-known principle consists in using the somatotopical organization of the motor cortical areas [4]. When subjects are asked to imagine movements of different parts of their body, spatially localized brain activity arises, and this can be associated to commands [5]. The low Signal-to-Noise Ratio (SNR) and low spatial resolution of the recorded data jeopardize the interpretation of the electric brain activity, yielding inaccurate control of the systems. BCIs also have to cope with inter-individual variability, enforcing the parameters of the methods to be adapted to each subject.

To reduce the impact of such obstacles, a crucial part of the EEG processing consists in transforming the signals acquired from a large and equivocal array of sensors into a small number of components focused on task-related brain activity. Such a step is called spatial filtering. The most widely used algorithm for this purpose is the Common Spatial Pattern (CSP), which was introduced in the BCI community in the context of a two-class MI paradigm [6]. In several comparative studies CSP has been found superior to competitive spatial filters [7]. Given the covariance matrices of two different tasks (for example left hand and right hand MI), CSP computes linear spatial filters maximizing the variance difference between the two classes. Many enhancements are still in progress to improve CSP. For instance, the local, sparse and spectral versions of the CSP algorithm have been proposed and proved useful in some cases to increase classification rates [8], [9], [10] (see [11] for a review of the different CSP principles).

Another promising approach to extract unequivocal signals from an array of EEG sensors resides in Blind Source Separation (BSS) [12]. Generally speaking BSS problems relate to situations when an array of sensors is measuring a linear mixture of unobserved sources. The issue of recovering sources from the acquired signals has been attacked by incorporating various assumptions about the generating processes and/or mixing transformation. For example in [12], Independent Component Analysis (ICA) was introduced using two key assumptions; first, a linear relation links generating and observed signals and second, the sources are statistically mutually independent and non-Gaussian. Even if the non-Gaussianity assumption proved useful to remove ocular artifacts from EEG signals [13], other approaches based on time-lagged decorrelation [14] or non-stationarity appeared

Manuscript received February 4, 2009; revised June 4, 2009. *Asterisk indicates corresponding author.*

C. Gouy-Pailler and M. Congedo are with the Department Images-Signal (DIS), Grenoble Images, Speech, Signal and Control Laboratory (GIPSA-lab), Grenoble 38031, France.

C. Jutten is with the Department Images-Signal (DIS), Grenoble Images, Speech, Signal and Control Laboratory (GIPSA-lab), Grenoble 38031, France. He is also a senior member of the Institut Universitaire de France, 75005 Paris, France.

C. Brunner and G. Pfurtscheller are with the Institute for Knowledge Discovery, BCI-Lab, Graz University of Technology, 8010 Graz, Austria.

This Research has been partially supported by the French National Research Agency (ANR) within the National Network for Software Technologies (RNTL), project Open-ViBE, Grant ANR05RNTL01601, by the European COST Action B27 "Electric Neuronal Oscillations and Cognition". During the period of the research the first author has been supported by the French Ministry of Defense (DGA), the second author has been partially supported by Nova Tech EEG, Inc., Knoxville, TN.

profitable for some natural signals [15], [16], [17].

As reviewed in [15], non-stationary source separation is a suitable framework for analyzing EEG data. It relies on rigorous theoretical work [18] and practical as well as computationally-efficient algorithms. In [18], [19], the authors showed that a generic non-stationarity assumption is sufficient to tackle the blind problem of separating non-stationary linearly mixed sources. They also showed in a maximum likelihood framework that second-order statistics (SOS) are sufficient in the case of non-stationary Gaussian sources, while higher-order statistics (HOS) have to be considered when non-Gaussianity is assumed (see [15] for more details on those different assumptions). As initially shown by [18], BSS of non-stationary sources based on SOS amounts to a Joint Approximate Diagonalization (JAD) problem. Considered as a natural extension of CSP to multi-class paradigms, a specific JAD algorithm was used by [20] in the context of BCIs to find efficient task-related spatial filters. The connection between this approach and BSS has been pointed out in [21], where JAD was proved to outperform one-versus-rest CSP.

Two shortcomings of the works devoted to JAD in the context of BCI are addressed in this paper. First, we elucidate the connections between JAD and BSS of non-stationary sources in a maximum likelihood framework. This formulation leads to a new neurophysiologically-adapted JAD formulation, which aims at taking into account the successive activation/deactivation of sources during MI trials. Second, we use a BCI dataset in which nine subjects are involved in a four-class two-session MI experiment to compare our extension against JAD and CSP using cross-validation as well as session-to-session conditions [22]. Whereas the cross-validated performance of JAD methods have already been assessed and compared to one-versus-rest CSP performance in [20], [21], the robustness of the JAD methods have not yet been evaluated for session-to-session transfer.

The rest of this paper is organized as follows. In section II, we elucidate the links between brain source assumptions and corresponding models and obtain source separation in terms of JAD of a set of covariance matrices. The multi-class MI BCI experiment is described in section III. The BCI-related points of the method are detailed in section IV. Results are then presented and discussed in sections V and VI, respectively.

II. Non-Stationary Brain Source Extraction

A. *Non-Stationary Brain Sources*

A random one-dimensional signal $x(t)$ is said to be stationary up to the second order if and only if the two following conditions are fulfilled: 1) its expectation $\mathbb{E}[x]$ is independent of time and 2) its autocorrelation function $\mathbb{E}[x(t_1)x(t_2)]$ only depends on the time difference $t_2 - t_1$. Insomuch, there is a wide range of signals that can be considered as non-stationary. For example, if the variance of a random signal alternatively takes two different values, the signal is non-stationary. Such kinds of signals are common in speech processing. When a speaker is switching between silence and active speech, in first approximation the resulting variance of the signal exhibits high and low variance periods. Even if brain signal dynamics are much more complicated than the previous elementary situation, such nonconstant variance profiles (depending on the observation time scale) can be observed in MI experiments from an inter-task as well as an intra-trial point of view. Generally speaking non-stationarity among real signals always depends on the observation time scale because statistical properties are estimated using fixed-length time windows.

In the BCI context based on MI, we define a "source" to be a cluster of aligned and synchronously activated/deactivated neurons. Neurophysiological studies have shown that the activation of motor cortex displays a somatotopically specific organization [23]. Thus, sources responsible for different types of MI, e. g., right hand versus left hand, can be considered as spatially distinct. Second, as shown in [24], event-related desynchronization and synchronization (ERD/ERS) exhibit some kinds of diversities during a specific MI task. For example when a subject is performing MI a diffuse mu (8–13 Hz) desynchronization is observed, which is often temporally distinct from a high-SNR beta rebound (13–30 Hz). Therefore, two particular kinds of source diversity may be exploited using the non-stationary source separation framework:

- **inter-task diversity**: brain sources involved in different types of MI are spatially distinct. Thus we observe sources in different spatial locations for each task of the experiment;
- **intra-trial diversity**: as specific trials of MI consist in successive activations/deactivations of distinct brain sources localized in different areas, namely the Supplementary Motor Area (SMA) and the primary motor cortex, we can also expect a non-stationary variance profile of each source within a single trial.

The first kind of diversity has been extensively exploited in the case of JAD or CSP. Following the framework proposed in [15], [25], [26], the purpose of this work is to propose a method that can exploit also the second kind of diversity.

B. *Block Gaussian Likelihood Separation*

We consider the random processes $\mathbf{x}(t) \in \mathbb{R}^N$ to represent the EEG time courses, recorded at N sensors. We assume, without loss of generality, that the EEG data can be first transformed using a data-independent linear algorithm. This pre-processing transformation denoted $\mathcal{T}$ yields $\tilde{\mathbf{x}}(t) \in \mathbb{R}^N$. For example, as we are dealing with MI tasks, signals may be filtered in some frequency bands, namely mu or beta. Second, we want to find a spatial transformation of $\tilde{\mathbf{x}}(t)$ helping inferring the intention c of the user (also called a class in machine learning) among a set of M predefined mental tasks $c \in \mathcal{C} = \{c_1, \dots, c_M\}$. The signals resulting from the spatial transformation of $\tilde{\mathbf{x}}(t)$ should yield an increased classification accuracy compared to directly using $\tilde{\mathbf{x}}(t)$.

According to the physical properties of the brain [27], [28], [15], we assume that brain sources are linearly related to $\mathbf{x}(t)$. Brain sources $\mathbf{s}(t) \in \mathbb{R}^L$ (L is the number of selected sources) and EEG measurements on the scalp $\mathbf{x}(t)$ are thus linked by

$$\mathbf{s}(t) = W^T \mathbf{x}(t), \tag{1}$$

where $W^T \in \mathbb{R}^{L \times N}$ represents the spatial filtering for recovering sources from observations. As the pre-processing

$\mathcal{T}$ is linear, it is equivalent to write the relation (1) for preprocessed signals, leading to $\tilde{\mathbf{s}}(t) = W^T \tilde{\mathbf{x}}(t)$. In the following, $W \in \mathbb{R}^{N \times L}$ will be referred to as a set of spatial filters, each column of W stands for the coefficients to be applied to each sensor to recover a specific source. In the context of a BCI experiment, the EEG data will be considered as an interval $[0, T]$ and the key assumption about the non-stationary models of sources is the following:

Assumption The interval $[0, T]$ is divided into K subintervals $T_1, \dots, T_K$ on which the variance of each source is constant over the subinterval.

Note that depending on the kind of non-stationarity that will be used, subintervals could represent whole trial tasks or subdivision of trials, i.e., the following framework will be used to model both inter-task and intra-trial diversities. When only inter-task diversity is used, K will be the number of trials; whereas K is the number of trials times the number of subdivisions in one trial when intra-trial diversity is used. In both case we can express source diversity by equations

$$\forall k \in [1 \,..\, K] \qquad \Sigma_{\tilde{\mathbf{s}},k} = \operatorname{diag}(\sigma^2_{1,k}, \cdots, \sigma^2_{L,k}), \tag{2}$$

where $\Sigma_{\tilde{\mathbf{s}},k}$ denotes the covariance matrix of $\tilde{\mathbf{s}}(t)$ on subinterval T_k and $\forall l \in [1 \,..\, L]$, $\sigma_{l,k} \in \mathbb{R}^+$.

A last assumption posits that the sources are temporally independent Gaussian processes. We stress that these are only working assumptions, sufficiently simple to be rigorously derived into an algorithm and sufficiently complex for recovering sources from sensor measurements [19]. Basically, it means that only second order statistics will be used even if higher order statistics do not vanish, thus the method works also for non-Gaussian sources.

Using the previous assumptions, the probability density of the sources can be written as $\tilde{\mathbf{s}}(t) = [s_1(t), \dots, s_L(t)]^T$ on subinterval T_k such that

$$\forall l \in [1 \,..\, L] \qquad \tilde{s}_l(t) \sim \mathcal{N}(0, \sigma^2_{l,k}), \tag{3}$$

where $\mathcal{N}(.)$ denotes a normal probability density function. We define the empirical covariance matrix $\hat{\Sigma}_{\tilde{\mathbf{x}},k} = \frac{1}{\#T_k} \sum_{t \in T_k} \tilde{\mathbf{x}}(t) \tilde{\mathbf{x}}(t)^T$ and $p_k = \frac{\#T_k}{\sum_k \#T_k}$ is the proportion of time points $\#T_k$ in interval T_k among the total number of time points considered. The likelihood objective as expressed in [19] is (see appendix for details)

$$C^\star_{\text{ML}} = \sum_{k=1}^{K} p_k \text{KL}(W^T \hat{\Sigma}_{\tilde{\mathbf{x}},k} W \parallel \operatorname{diag}(W^T \hat{\Sigma}_{\tilde{\mathbf{x}},k} W)). \tag{4}$$

Note that the measure $\text{KL}(R \parallel \operatorname{diag}(R))$, which is the Kullback-Leibler divergence between R and the diagonal matrix having the same diagonal as R, is a measure of deviation from diagonality. Thus if we define $\text{off}(R) = \text{KL}(R \parallel \operatorname{diag}(R))$, the criterion in (4) can be interpreted as a joint diagonalization criterion:

$$C^\star_{\text{ML}} = \sum_{k=1}^{K} p_k \text{off}(W^T \hat{\Sigma}_{\tilde{\mathbf{x}},k} W). \tag{5}$$

It follows that the problem of recovering the non-stationary sources responsible for some observed signals is: given a set of covariance matrices $\mathcal{C} = \{\hat{\Sigma}_{\tilde{\mathbf{x}},k}\}_{k=[1 \,..\, K]}$, find a joint diagonalizer W^{-T} such that for each k

$$W^T \hat{\Sigma}_{\tilde{\mathbf{x}},k} W = D_k, \tag{6}$$

where each D_k is as close to diagonal form as possible. An implementation of an efficient algorithm to minimize (5) is provided by the authors as an open-source CRAN package[1].

In [29] we find a useful necessary condition to be fulfilled to find the approximate diagonalizer.

Identifiability Principle For each pair of positions (m, n) of the diagonal matrices, there exists a k such that $D_{k_m} \neq D_{k_n}$. In other words, the diagonalization set must provide a source of diversity between source m and source n along interval k.

In the context of non-stationary brain source extraction, the different activation/deactivation profile of sources would be sufficient to recover brain sources.

C. Common Spatial Patterns (CSP)

The idea of CSP [30], [6] for two-class problems is to find the most discriminative spatial filters $\mathbf{w} \in \mathbb{R}^N$ which optimize the Rayleigh quotient

$$\{\min, \max\} \frac{\mathbf{w}^T \Sigma_{\tilde{\mathbf{x}},c_1} \mathbf{w}}{\mathbf{w}^T \left(\Sigma_{\tilde{\mathbf{x}},c_1} + \Sigma_{\tilde{\mathbf{x}},c_2}\right) \mathbf{w}},$$

where $\Sigma_{\tilde{\mathbf{x}},c_1}$ and $\Sigma_{\tilde{\mathbf{x}},c_2}$ are the covariances matrix of the data belonging to class c_1 and c_2, respectively. We use the notation $\{\min, \max\}$ to express the fact that we are equally interested in maximizing or minimizing the previous quotient. The solution of this optimization problem yields an ensemble of eigenvalues, each of them accounts for the explaining variance of the corresponding eigenvector. This problem is equivalent to finding a matrix W and a diagonal matrix D such that:

$$\left\{ \begin{array}{l} W^T \Sigma_{\tilde{\mathbf{x}},c_1} W = D \\ W^T \Sigma_{\tilde{\mathbf{x}},c_2} W = I - D \end{array} \right. , \tag{7}$$

which is solved by generalized eigenvalue decomposition. One can approximately formulate this optimization problem in terms of a non-diagonality criterion:

$$C_{\text{CSP}_2} = \sum_{c_i \in \{c_1, c_2\}} \text{off}(W^T \hat{\Sigma}_{\tilde{\mathbf{x}},c_i} W) \ . \tag{8}$$

For two-class paradigms, CSP is nothing but a method aiming at exploiting the non-stationary sources related to two different classes. Whereas CSP for two-class paradigms is solved by an exact joint diagonalization of two matrices, JAD makes use of an approximate optimization method, which is nothing but the extension of joint diagonalization to more than two matrices. Nevertheless, JAD is considered to be more robust to estimation errors of covariance matrices, while such errors lead to wrong results with CSP [15], [29].

For multi-class paradigms, an extension has been proposed in [30], [8]. The simple idea is to decompose the M-class problem into a set of M binary problems. To that end, each

[1]jointDiag (http://cran.r-project.org/web/packages/jointDiag/index.html) is an R package, which provides an implementation of Pham's algorithm as well as other efficient joint approximate diagonalization algorithms.

problem consists of discriminating one class against the mean of the other ones. In terms of the diagonalization criterion, this can be formulated such that for each class $i \in [1\,..\,M]$ we have

$$\begin{cases} W_i^T \hat{\Sigma}_{\tilde{\mathbf{x}},c_i} W_i = D_i \\ W_i^T \left(\sum_{j \neq i} \hat{\Sigma}_{\tilde{\mathbf{x}},c_j}\right) W_i = I - D_i \end{cases} . \tag{9}$$

In terms of JAD, the diagonalization criterion to be minimized is $\forall i \in [1\,..\,M]$, $C^i_{\text{CSP}_M} = \text{off}(W^T \hat{\Sigma}_{\tilde{\mathbf{x}},c_i} W) + \text{off}(W^T \sum_{j \neq i} \hat{\Sigma}_{\tilde{\mathbf{x}},c_j} W)$.

D. Non-Stationary Source Extraction and JAD

First proposed in [20] and extended in [21], JAD proved useful to find efficient spatial filters in the context of multi-class MI BCIs. We can formulate JAD using the general framework of non-stationary source extraction. The idea is to consider that the variances of the sources are constant over the task interval and the variance of the sources differs from task to task. That is why the JAD of M class-conditional sample covariance matrices yields generating sources. Formally, as explained in [21], the set of covariance matrices to be jointly diagonalized is

$$\mathcal{C} = \{\hat{\Sigma}_{\tilde{\mathbf{x}},c_i}\}_{i=[1\,..\,M]}. \tag{10}$$

In terms of the non-diagonality criterion, this can be rewritten as

$$C_{\text{JAD}} = \sum_{i=1}^{M} \text{off}(W^T \hat{\Sigma}_{\tilde{\mathbf{x}},c_i} W). \tag{11}$$

This model is neurophysiologically plausible because it assumes diversity of the brain sources related to the MI of different body parts. The somatotopical organization of the motor area supports this hypothesis. However, the limits of this model consist in considering that the brain sources responsible for MI must be constantly activated/deactivated during a whole trial of a few seconds.

E. Extension of JAD: MSJAD

As aforementioned, the general JAD method as proposed in [20], [21] may be extended to take into account the activation/deactivation profile of brain sources during the whole trial. A concrete example is the beta rebound. This phenomenon starts at the end of the MI task, but its spatial location and high absolute power is often used to discriminate between different MI tasks. As motor tasks are known to be a succession of activations/deactivations in different brain areas, it can be assumed that sources related to a mental task realization can be activated/deactivated with different energies across the task. Joint diagonalization covariance matrices computed using successive time windows will help recovering task-related sources according to the identifiability principle. The set of covariance matrices that is considered is thus

$$\mathcal{C} = \{\hat{\Sigma}_{\tilde{\mathbf{x}},c_i,\mathcal{I}_j}\}_{i=[1\,..\,M],j=[1\,..\,J]}, \tag{12}$$

where $\{\mathcal{I}_j\}_{j=[1\,..\,J]}$ is a partition of the task interval. Once again in terms of the non-diagonality criterion, we can write the cost function as

$$C_{\text{JAD}} = \sum_{i=1}^{M} \sum_{j=1}^{J} \text{off}(W^T \hat{\Sigma}_{\tilde{\mathbf{x}},c_i,\mathcal{I}_j} W). \tag{13}$$

This method yields the joint approximate diagonalization of $M \cdot J$ covariance matrices. We call this method Multi-Segment Joint Approximate Diagonalization (MSJAD) from now onwards. Note that either inter-task or intra-trial diversities would now suffice to recover MI-related brain sources. For the purpose of this paper, a simple and natural interval partitioning has been used; each trial interval of the training set is partitioned into four subintervals of equal length The number of subintervals has been chosen to cover the whole MI task with 1 s non-overlapping windows, from the beginning of imagery to one second after the end of the task. Reference [18] provides a very efficient algorithm to solve the optimization problem of equation (4), which can be seen as a JAD. Its description is out of the scope of this article.

III. Subjects and Experimental Paradigm

We consider dataset 2a from BCI Competition IV (2008). We here only remind the crucial parts of the paradigm (detailed description can be found on the competition website[2]). Nine subjects were involved in a BCI experiment consisting of four-class MI tasks. Two sessions on different days were recorded for each subject, each session consisted of 288 trials (72 trials of each task). The paradigm is illustrated in Figure 1b.

EEG was acquired at 22 Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm). The setup is depicted in Figure 1a. 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 line noise.

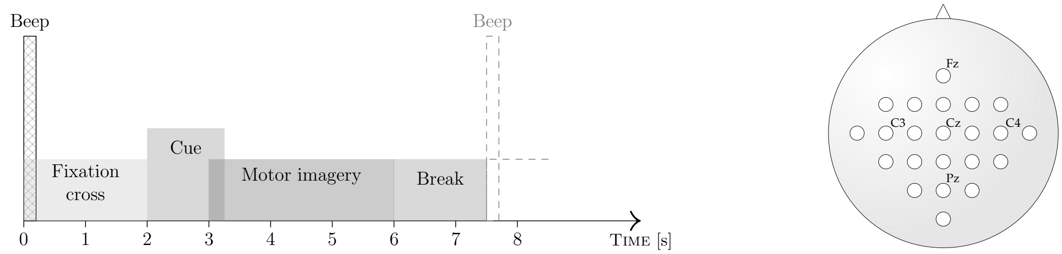
Although a visual inspection of the raw EEG data was performed by an expert, no trials were removed from the subsequent analysis in this study in order to evaluate the robustness and sensitivity to outliers and artifacts of each method. The fraction of artifacteous trials over all subjects was rather low anyway, namely 7.5 % on average (median value of 6.1 %).

IV. Methods

The performance of our framework is assessed using two evaluation methods: cross-validation and session-to-session transfer. We begin by describing the overall method in the case of cross-validation [7]. Figure 2 gives an overview of the processing flow. It details the different operations performed during the training and testing step and elucidates the parameters transferred between these two steps (dashed lines in the figure).

For each subject and each session the EEG signals were filtered with a fifth-order Butterworth filter with 5 and 35 Hz cut-off frequencies. First of all, the set of trials is partitioned

[2]Dataset 2a from http://ida.first.fraunhofer.de/projects/bci/competition_iv/. The results were announced on November 2008 and the competition started on July 2008.



(a) Paradigm (b) Montage

Fig. 1. Timing scheme of the BCI paradigm (a) 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 (b).

into a training and a test set. The total number of trials available for one session of each subject is 72 for each class (right hand, left hand, foot and tongue). The training set is randomly chosen among the whole dataset (1×20 cross-validation). Its size varies between 10 and 60 trials of each task. Thus the total training set size within each session varies from 40 to 240 trials. For each method, the set of covariance matrices is computed using only the training set, and the spatial filter is computed. For JAD and MSJAD this procedure results in a set of N potential spatial filters. We then use an Information Theoretic Feature Extraction procedure [21] to rank the spatial filters and select the $L = 8$ best of them.

In the case of CSP, the procedure yields a set of $M \times N$ spatial filters resulting from M CSP algorithms. As the eigenvalues are ranked according to their importance for explaining variances, we are able to easily select a certain number of components for separating one class among the others. As a matter of fair comparison with JAD algorithms and according to previous studies, we selected the best two spatial filters of each CSP decomposition. This results in a total of $L = 8$ spatial filters. Once the set of spatial filters has been determined, they are applied to the training and test sets, yielding estimated sources $\hat{\mathbf{s}}(t)$. Three seconds of each trial of the training and test sets are then extracted, from $t = 4$ s to $t = 7$ s, covering the whole imagery period. This interval is decomposed into one-second segments with 80 % overlap between two successive segments. For each one-second segment of data features are computed by extracting spectral powers in 15 equally spaced narrow bands of 2 Hz between 5 and 35 Hz. For each segment this results in a 120-dimensional feature vector. The feature vector of the training set is used to train a regularized multinomial regression (generalization of the logistic regression to multi-class problems) [31]. Note that the classifier has been successfully used in much more drastic ill-posed situations in which no over-fitting had been observed [31]. The regularization parameter was set to 0.3. Lastly, we try to infer the class of each segment of each trial of the test set. The output of the classifier is the probability that the segment belongs to a certain class. We attribute to each segment the class yielding the highest probability. This procedure is repeated 20 times for each partitioning size. The global label of a test trial is chosen as the class for which the maximum probability is observed (integrating over the set of successive segments).

In the case of session-to-session transfer, the procedure is greatly simplified. We use the spatial filters and classifier learned with one session and test using the second session. The procedure is therefore applied only once. This evaluation method is of the greatest interest for real BCI applications. A BCI would indeed be user-friendly if the training time needed to use the system is reduced to a few minutes on the first day of use. But this implies robustness of the spatial filters and classifier obtained on the first day.

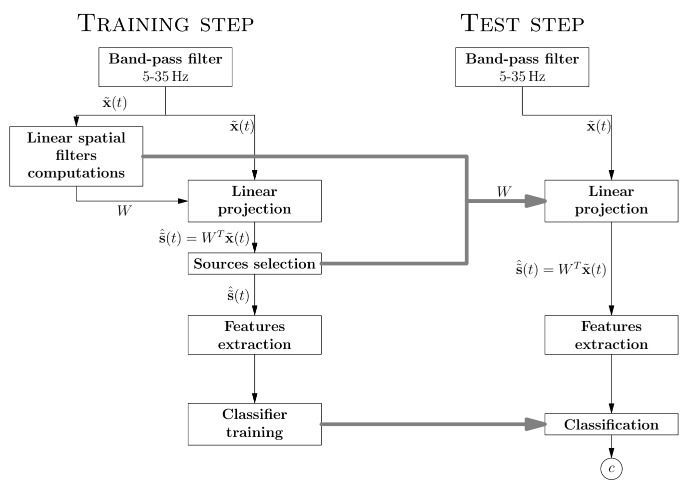


Fig. 2. Overview of the processing method in the case of cross-validated performance evaluation. The broad lines linking some elements between the training and testing steps represent an exchange of information. The spatial filters are computed using the training set, a certain number of them is selected and then reported into the test step to project the data. See the text for details about the notation.

V. Results

A. Cross-Validated Results

In order to evaluate any significant difference between the mean performance of CSP, JAD and MSJAD, we first perform an analysis of variance [32]. Our repeated measure design involves two within-subjects factors: SESSION is a two-level factor and METHOD is a three-level factor. In order to simplify the analysis the training set size has been discarded from the analysis of variance and only the performances for training set sizes of 60 trials per task are considered. In the following, we will see that this simplification does not influence the results because the rankings of methods are not dependent on the training set size. Our experimental design consists also of one random factor: SUBJECT (reasons why it is random can be found in [33]). The dependent variable PERF is the cross-validation performance, which is the average of 20 randomized repetitions. The analysis of variance is summarized in Table I.

	df	F	p
SESSION	(1,8)	0.94	0.360
METHOD	**(2,16)**	**5.99**	**0.011**
METHOD × SESSION	(2,16)	1.16	0.339

TABLE I
SUMMARY OF THE ANALYSIS OF VARIANCE FOR THE CROSS-VALIDATED PERFORMANCE. THE SUCCESSIVE ROWS INDICATES THE INFLUENCE ON THE LINEAR MODEL OF THE DIFFERENT VARIABLES CONSIDERED. THE TRAINING SET SIZE IS SET TO 60 TRIALS PER TASKS.

A significant main effect of METHOD appears ($F(2,16) = 5.99$, $p = 0.011$), whereas no SESSION main effect or METHOD × SESSION interaction are observed. These results show that there is no significant differences between the two sessions, thus no training effect of the subjects can be claimed. It is also important that no METHOD × SESSION effect is found because a significant effect would suggest that the main effect of METHOD is not consistent across sessions. A post-hoc analysis was performed to test all pairwise differences between the mean performances of the three methods (JAD vs. CSP, MSJAD vs. CSP and MSJAD vs. JAD). We used Tukey contrasts to adjust p-values due to multiple comparisons. We observe a significant difference of means between MSJAD and CSP ($t(17) = 2.867$, $p = 0.027$). No significant difference of mean is observed neither between JAD and CSP nor between MSJAD and JAD ($p = 0.120$ and $p = 0.726$).

Table II presents the cross-validated performance obtained for each subject and each session. The performance score is averaged between 20 repetitions for the larger training set size (60 trials per task) given by the best model. The third column of the table gives the corresponding best method. Performances are substantially heterogeneous among subject, ranging from 48.1 % for subject 6 session 1 to 87.5 % for subject 3 session 2. A summary of the cross-validated performances is presented in Table III. The mean value presented in this table is the average among the performances of each subject for a specific method (training set size is 60 trials per task), the third column of the table also indicates the corresponding standard deviation.

Figure 3 shows a detailed view of the performance of the different algorithms as a function of the training set size. The results of two representative subjects are presented here. For some subjects, all methods appear to be quite comparable and differences are slight, whereas for some others, important differences exist among methods, e. g., in subject 3 session 2, where a difference of almost 20 % of classification accuracy is observed. This figure shows that for most subjects the training set size can be reasonably reduced to 20 or 40 trials for each class without any significant loss of performance. We also want to emphasize that the ranking of the methods is not

	Accuracy (%)	Best method
S1 ses1	72.2	JAD
ses2	77.2	MSJAD
S2 ses1	62.7	CSP
ses2	59.3	JAD
S3 ses1	85.4	JAD
ses2	87.5	JAD
S4 ses1	76.0	MSJAD
ses2	70.5	JAD
S5 ses1	75.0	MSJAD
ses2	80.6	MSJAD
S6 ses1	48.1	CSP
ses2	55.3	MSJAD
S7 ses1	58.1	MSJAD
ses2	55.7	MSJAD
S8 ses1	81.7	JAD
ses2	77.9	MSJAD
S9 ses1	72.6	JAD
ses2	79.1	JAD

TABLE II
CLASSIFICATION RATES FOR EACH SESSION (SES1 OR SES2) AND EACH SUBJECT (S1 TO S9) GIVEN BY THE BEST MODEL. TRAINING SET SIZE IS 60 TRIALS PER TASK.

	Mean [%]	Std Dev.
JAD	68.73	13.06
CSP	65.90	9.92
MSJAD	**69.76**	**10.63**

TABLE III
SUMMARY OF THE CROSS-VALIDATED PERFORMANCES FOR EACH METHOD. TRAINING SET SIZE IS SET TO 60 TRIALS PER TASK.

influenced by the training size.

B. Session-to-Session Transfer

As in the case of cross-validation, we performed an analysis of variance to assess any significant superiority of a method over the others. The model considered is the same as the one presented in the cross-validated case. Results are presented in Table IV. This table shows that no significant main effects of factor SESSION and of interaction SESSION × METHOD are observed. On the contrary, we observe a significant main effect of METHOD ($F(2,16) = 5.81$, $p = 0.011$).

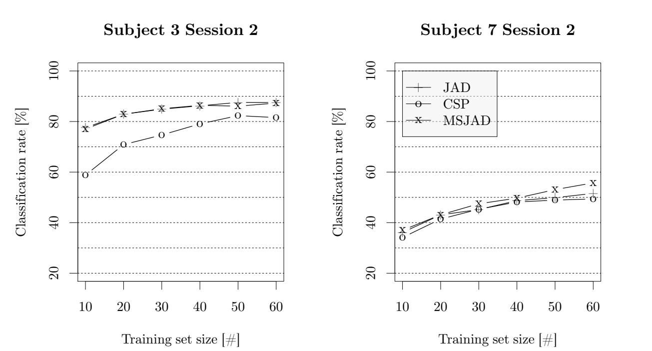


Fig. 3. The cross-validated performance of subjects 3 and 7 during session 2 are presented in this figure. The classification accuracy is plotted versus the training set size for the three methods compared in this paper.

	df	F	p
SESSION	(1,8)	2.45	0.156
METHOD	**(2,16)**	**5.80**	**0.013**
METHOD × SESSION	(2,16)	0.11	0.89

TABLE IV
SUMMARY OF THE ANALYSIS OF VARIANCE FOR THE SESSION-TO-SESSION TRANSFER PERFORMANCE.

We performed a post-hoc analysis to identify the significant differences of means [34]. A significant difference appears between MSJAD and CSP ($t(17) = 2.981$, $p = 0.022$). No significant difference can be found between MSJAD and JAD but a marginal difference is observed between JAD and CSP ($t(17) = 2.545$, $p = 0.052$). These results further corroborate the fact that MSJAD performed better than CSP in this multi-class MI experiment for both the cross-validated and session-to-session results.

Figure 4 presents a detailed view of the performance of each subject for each method and each session. This figure presents the classification accuracy obtained by considering session 1 (session 2) for training and session 2 (session 1) for testing. The performances of a specific subject during session 1 and 2 for a specific method are linked by a straight line. This clearly shows a high heterogeneity between subjects. Table V shows the average performance of each method across subjects and sessions.

	Mean [%]	Std Deviation
JAD	63.3	13.48
CSP	60.5	11.09
MSJAD	**63.8**	**12.28**

TABLE V
SUMMARY OF THE SESSION-TO-SESSION TRANSFER PERFORMANCES FOR EACH METHOD CONSIDERED IN THIS PAPER. EACH SCORE IS THE AVERAGE ACROSS 18 SESSIONS.

C. Comparison with BCI Competition IV (2008)

The dataset used in this paper was proposed during BCI Competition IV. The goal of this competition was to evaluate algorithms for session-to-session transfer between the first and the last session (session 1 was used as training set and session 2 was used as evaluation set). Even if the competitors were clearly asked to remove the artifacts present in the dataset, we can fairly compare the results of the competition with our algorithm by using exclusively session-to-session transfer from session 1 to session 2 and using the same criterion, namely the kappa score as defined in [35]. The results are summarized in table VI. Results obtained by the best three competitors as well as results obtained by the algorithms presented in this paper are reported. The robust measure used here confirms the results obtained in this paper for cross-validation as well as session-to-session transfer performance. It also shows that our MSJAD algorithm, even if results could be improved by first removing artifacts, behave fairly well compared to the results obtained by the best competitors. We indeed see that MSJAD as well as the winner of the competition proved the best method for 4 out of 9 subjects. Nevertheless MSJAD would have been ranked third especially because of subjects S7 and S9.

	1^{st}	2^{nd}	3^{rd}	JAD	CSP	MSJAD
S1	0.68	**0.69**	0.38	0.65	0.52	0.66
S2	**0.42**	0.34	0.18	0.40	0.39	**0.42**
S3	0.75	0.71	0.48	**0.77**	0.67	**0.77**
S4	0.48	0.44	0.33	0.50	0.50	**0.51**
S5	0.40	0.16	0.07	0.44	0.49	**0.50**
S6	**0.27**	0.21	0.14	0.19	0.18	0.21
S7	**0.77**	0.66	0.29	0.25	0.26	0.30
S8	**0.75**	0.73	0.49	0.72	0.57	0.69
S9	0.61	**0.69**	0.44	0.50	0.40	0.46
Mean	0.57	0.52	0.31	0.49	0.41	0.50

TABLE VI
KAPPA SCORES OBTAINED BY THE THREE BEST COMPETITORS AS WELL AS THE THREE METHODS PRESENTED IN THIS PAPER.

D. Does Intra-Trial Diversity add information?

As pointed out by [29], the inclusion of new matrices to a diagonalization set could be sufficient to improve results. Therefore, to be sure that the improvements observed with MSJAD were not the results of an increase of the number of matrices, we performed the same computations as presented with MSJAD, but we broke the time structure of each trial: in MSJAD, each of the 16 covariance matrices of the diagonalization set corresponded to a specific task and a determined time segment, both ranging from one to four. The trial time structure was broken by randomizing the segments indexes before averaging such that the four covariance matrices related to each task do not correspond to a specific time segment position in the trial anymore. Thus in this method the 16 covariance matrices only exploit the inter-task diversity but not the intra-trial diversity. As expected, performance appeared comparable to JAD in both cross-validation and session-to-session transfer (66.8 and 62.1 %, respectively). This demonstrates that inclusion of intra-trial diversity is indeed the key factor allowing improvement in MI BCI classification rate.

VI. DISCUSSION AND CONCLUSION

The method proposed in this paper yields among the best session-to-session results for this particular dataset, as compared to [36] as well as the BCI Competition IV. It is also notable that the difference between cross-validated and session-to-session performance is very low. This is of great interest for real-life BCI systems because it means that parameters set using a first day of training can be used with an acceptable classification accuracy during the following days. We also showed in the cross-validation results that the training set can be considerably reduced without any significant decrease of classification rates. This is a valuable result for real-life BCI systems. People could therefore be able to use the BCI system with a reduced training time. Results are globally satisfying given the fact that subjects were untrained (see [37] for an extended study about naive BCI users). Nevertheless, the results presented in this paper suggest that the intrinsic subject ability to control a BCI is much more crucial than the

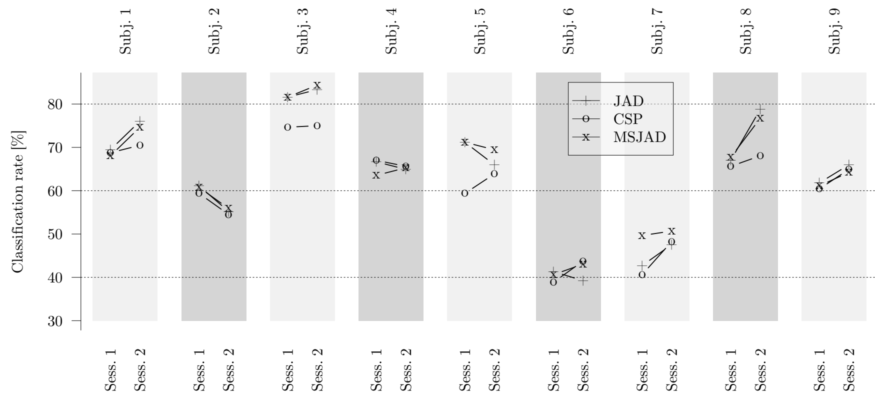


Fig. 4. Session-to-session performance details. The performances of a method for session 1 and 2 are linked by a straight line.

training effect. We did not observe any significant effect of the session factor, however our training session may not be long enough to accumulate long-term training effects.

The computational load of the three methods compared in this paper is approximately the same. Each of them is able to instantaneously compute spatial filters given a set of covariance matrices of reasonable dimensions (up to $N = 200$ remains reasonable). We also noticed that the inclusion of more matrices in the training set does not significantly increase the JAD computational time. Although we did not go into the details of the JAD algorithm used throughout this paper, an important characteristic of this algorithm is that it is deterministic. It means that if you run the algorithm two times with the same data the result will be the same. As mentioned by [38] this is not the case with all ICA algorithms.

a) Relation to previous works: Two preceding papers analyzed in detail the application of ICA algorithms for a subset of this particular dataset [38], [36] (subject S8 was not present in these analysis). In [38], different commonly used ICA algorithms were compared to CSP. The best results were attained by CSP with a score of about 65 % in the cross-validation case and about 61.5 % in the session-to-session transfer case. When we remove subject S8 from the analysis, MSJAD is shown to outperform results from [38] with scores of 68.5 % and 62.8 % for cross-validation and session-to-session transfer respectively.

On the other hand, results presented in [36] using Infomax ICA remains better than the results of this paper. These results have been obtained with the help of an extensive and computationally demanding sequential floating forward selection algorithm (SFFS) considering 1364 features. Yet, even if cross-validation results were largely improved, especially because of Infomax results (about 75 % of classification rate), session-to-session transfer results did not outperform results given in [38] (about 61.5 %). Our work is also closely related to [21]. In this article, connections between CSP and ICA were made and JAD was compared to multi-class CSP in an MI-based BCI experiment involving three subjects. Performance was compared in cross-validation only and showed a clear superiority of JAD over multi-class CSP. This result was not so clear with our dataset. Even if JAD performed marginally better than CSP for session-to-session performance, no significant results were observed between the two methods for cross-validation performance.

b) Multi-Class Spatial Filtering: More generally, we once again proved that linear spatial filtering is of great interest to improve BCI classification rates. Up to now, most BCI experiments have considered two-class paradigms because of a lack of appropriate algorithms to process and classify data. Our paper provides a natural generalization of CSP to multi-class paradigms. We think that it could help considering multi-class paradigms in real-life BCI experiments. Not only does our framework provide efficient ways to find efficient spatial filters, but it also provides well-based theoretical links between source assumptions and algorithms. Insomuch, future works about linear non-stationary source extraction could be facilitated by exhibiting other kinds of diversities in the data. A straightforward way to extend our framework would be to incorporate precise frequency information about the generating processes, thus exploiting spectral source profile diversities in addition to inter-class and non-stationarity [15].

c) Conclusion and Future Directions: In summary, we presented here an efficient framework for increasing classification rates of multi-class BCI paradigms. By formulating CSP by JAD in a maximum likelihood context for separating linear mixtures of non-stationary sources, we bridged the gap between the family of CSP algorithms and BSS. The separation of non-stationary source separation is well grounded on Pham's and Cardoso's theoretical work, and the algorithm provided by Pham is computationally efficient, in particular much more efficient than most ICA-based source separation methods [15]. The general framework exposed here leads to a new method to make use of successive activation/deactivation of brain sources. The results showed that the new method proposed in this paper outperforms the commonly used CSP algorithm. We showed that the CSP by JAD can be formulated in terms of non-stationary source separation and linked the underlying assumptions in terms of neurophysiological phenomena.

Appendix

We here remind the main theoretical aspects of non-stationary source separation given in [18], [19]. Using the mutual independence hypothesis, we can write the log probability density of $\tilde{\mathbf{s}}(t)$ as

$$-\frac{1}{2}\sum_{l=1}^{L}\frac{\tilde{s}_l^2(t)}{\sigma_{l,k}^2}+\log(2\pi\sigma_{l,k}^2) = -\frac{1}{2}\text{tr}\left(\Sigma_{\tilde{\mathbf{s}},k}^{-1}\tilde{\mathbf{s}}(t)\tilde{\mathbf{s}}(t)^T\right) -\frac{1}{2}\text{logdet}\left(2\pi\Sigma_{\tilde{\mathbf{s}},k}\right). \quad (14)$$

The probability density of $\tilde{\mathbf{x}}(t)$, $p_{\tilde{x}}(\tilde{\mathbf{x}})$ is related to the one of $\tilde{\mathbf{s}}(t)$, $p_{\tilde{s}}(\tilde{\mathbf{s}})$ by $p_{\tilde{x}}(\tilde{\mathbf{x}}) = |\text{det}W|p_{\tilde{s}}(W^T\tilde{\mathbf{x}})$. Therefore, the maximum likelihood criterion is

$$C_{\text{ML}} = \frac{1}{\#T_k}\sum_{t\in T_k}\frac{1}{2}\text{tr}(\Sigma_{\tilde{\mathbf{s}},k}^{-1}W^T\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^T W) + \text{logdet}(2\pi\Sigma_{\tilde{\mathbf{s}},k}) + \log|\text{det}W^{-T}|, \quad (15)$$

where $\#T_k$ stands for the number of time points in the subinterval T_k. Defining the matrices $\Sigma_{\tilde{\mathbf{x}},k} = W^{-T}\Sigma_{\tilde{\mathbf{s}},k}W^{-1}$ and $\hat{\Sigma}_{\tilde{\mathbf{x}},k} = \frac{1}{\#T_k}\sum_{t\in T_k}\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^T$, namely the model-based and data-based covariance matrices, and integrating the information over all subintervals, the maximum likelihood can be rewritten:

$$C_{\text{ML}} = \frac{1}{2}\sum_{k=1}^{K}p_k[\text{tr}(\Sigma_{\tilde{\mathbf{x}},k}^{-1}\hat{\Sigma}_{\tilde{\mathbf{x}},k}) - \text{logdet}(\Sigma_{\tilde{\mathbf{x}},k}^{-1}\hat{\Sigma}_{\tilde{\mathbf{x}},k}) - N] + \text{const}, \quad (16)$$

where $p_k = \frac{\#T_k}{\sum_k \#T_k}$. The expression of the Kullback-Leibler divergence for two zero mean N-variate densities of respective covariance matrices R_1 and R_2 is known to be $\text{KL}(R_1 \parallel R_2) = \frac{1}{2}[\text{tr}(R_2^{-1}R_1) - \text{logdet}(R_2^{-1}R_1) - N]$. The Kullback-Leibler divergence is invariant under invertible transformation, thus $\text{KL}(\hat{\Sigma}_{\tilde{\mathbf{x}},k} \parallel \Sigma_{\tilde{\mathbf{x}},k}) = \text{KL}(W^T\hat{\Sigma}_{\tilde{\mathbf{x}},k}W \parallel \Sigma_{\tilde{\mathbf{s}},k})$. The expression of the maximum likelihood becomes then:

$$C_{\text{ML}} = \sum_{k=1}^{K}p_k\text{KL}(W^T\hat{\Sigma}_{\tilde{\mathbf{x}},k}W \parallel \Sigma_{\tilde{\mathbf{s}},k}) + \text{const}. \quad (17)$$

Lastly, we can use a key property of the Kullback-Leibler divergence, which relates a positive matrix R with any diagonal matrix Σ by the Pythagorean decomposition $\text{KL}(R \parallel \Sigma) = \text{KL}(R \parallel \text{diag}(R)) + \text{KL}(\text{diag}(R) \parallel \Sigma)$. From this relation, we learn that the closest diagonal matrix to R is $\Sigma = \text{diag}(R)$. Equation (17) is minimized with respect to $\{\Sigma_{\tilde{\mathbf{s}},k}\}_{k\in[1\,..\,K]}$ by choosing $\Sigma_{\tilde{\mathbf{s}},k} = \text{diag}(W^T\hat{\Sigma}_{\tilde{\mathbf{x}},k}W)$.

Acknowledgment

The authors want to express their gratitude to the anonymous reviewers, whose comments and remarks improved the quality of this paper.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun 2002.

[2] N. Birbaumer and L. G. Cohen, "Brain-Computer-Interfaces (BCI): Communication and restoration of movement in paralysis." *J. Physiol.*, vol. 579, no. 3, pp. 621–636, Jan 2007.

[3] A. Nijholt, D. Tan, G. Pfurtscheller, C. Brunner, J. d. R. Millán, B. Allison, B. Graimann, F. Popescu, B. Blankertz, and K.-R. Müller, "Brain-computer interfacing for intelligent systems," *IEEE Intell. Syst.*, vol. 23, no. 3, pp. 72–79, May 2008.

[4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul 2001.

[5] 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*, vol. 31, no. 1, pp. 153–159, May 2006.

[6] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task." *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, May 1999.

[7] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces." *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, Jun 2007.

[8] 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.*, vol. 53, no. 11, pp. 2274–2281, Nov 2006.

[9] 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.*, vol. 52, no. 9, pp. 1541–1548, Sep 2005.

[10] H. Wang and W. Zheng, "Local Temporal Common Spatial Patterns for Robust Single-Trial EEG Classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 2, pp. 131–139, Apr 2008.

[11] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG Single-Trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[12] C. Jutten and J. Hérault, "Blind separation of sources, Part 1: an adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, 1991.

[13] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings." *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 589–593, May 2000.

[14] A. Ziehe, K. R. Müller, G. Nolte, B. M. Mackert, and G. Curio, "Artifact reduction in magnetoneurography based on time-delayed second-order correlations," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 1, pp. 75–87, 2000.

[15] M. Congedo, C. Gouy-Pailler, and C. Jutten, "On the Blind Source Separation of Human Electroencephalogram by Approximate Joint Diagonalization of Second Order Statistics," *Clin. Neurophysiol.*, vol. 119, no. 12, pp. 2677–2686, Dec 2008.

[16] L. Parra and C. Spence, *Independent Components Analysis, Principles and Practice*. Cambridge University Press, 2001, ch. Separation of non-stationary natural signals, pp. 135–157.

[17] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Netw.*, vol. 8, no. 3, pp. 411–419, Mar 1995.

[18] D.-T. Pham, "Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion," *Signal Process.*, vol. 81, no. 4, pp. 855–870, 2001.

[19] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sept 2001.

[20] 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.*, vol. 51, no. 6, pp. 993–1002, Jun 2004.

[21] M. Grosse-Wentrup and M. Buss, "Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug 2008.

[22] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PLoS ONE*, vol. 3, no. 8, p. e2967, Aug 2008.

[23] R. Hari and N. Forss, "Magnetoencephalography in the study of human somatosensory cortical processing," *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, vol. 354, no. 1387, pp. 1145–1154, Jul 1999.

[24] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles." *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, Nov 1999.

[25] C. Gouy-Pailler, M. Congedo, C. Jutten, C. Brunner, and G. Pfurtscheller, "Model-Based Source Separation for multi-class motor imagery," in *Proc. of the 16th European Sign. Process. Conf. (EUSIPCO-2008), EURASIP*, Lausanne, Switzerland, August 2008.

[26] C. Gouy-Pailler, M. Congedo, C. Brunner, C. Jutten, and G. Pfurtscheller, “Multi-Class Independent Common Spatial Patterns: Exploiting Energy Variations of Brain Sources,” in *Proceedings of the 4th International BCI Workshop*, Graz, Austria, 2008.
[27] P. L. Nuñez, *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford Univ Press, February 2005.
[28] M. Congedo, F. Lotte, and A. Lécuyer, “Classification of movement intention by spatially filtered electromagnetic inverse solutions.” *Phys. Med. Biol.*, vol. 51, no. 8, pp. 1971–1989, Apr 2006.
[29] B. Afsari, “Sensitivity Analysis for the Problem of Matrix Joint Diagonalization,” *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1148–1171, Sep 2008.
[30] K. Fukunaga, *Introduction to statistical pattern recognition (2nd ed.)*. San Diego, CA, USA: Academic Press Professional, Inc., 1990.
[31] J. Friedman, T. Hastie, and R. Tibshirani, “Regularized Paths for Generalized Linear Models via Coordinate Descent,” 2008, the R package glmnet is available from CRAN.
[32] B. S. Everitt, *An R and S-Plus Companion to Multivariate Analysis*, 1st ed., ser. Springer Texts in Statistics. Springer, 2005.
[33] W. Hays, *Statistics*, 5th ed. Wadsworth Publishing, 1994.
[34] T. Hothorn, F. Bretz, and P. Westfall, “Simultaneous Inference in General Parametric Models,” *Biom. J.*, vol. 50, no. 3, pp. 346–363, May 2008.
[35] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.
[36] 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 Recognit. Lett.*, vol. 28, no. 1, pp. 957–964, 2007.
[37] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, “The non-invasive berlin brain-computer interface: fast acquisition of effective performance in untrained subjects,” *Neuroimage*, vol. 37, no. 2, pp. 539–550, Aug 2007.
[38] 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.*, vol. 3, no. 3, pp. 208–216, Sep 2006.

Cédric Gouy-Pailler was born in Grenoble, France, in 1983. In 2006, he received both the engineering degree in electronics and the M.Sc. degree in Cognitive Sciences from the Grenoble University, Grenoble, France. He is currently working toward the Ph.D. degree in biomedical signal processing at the Grenoble University, Grenoble, France.

He has worked at Orange Labs on the use of haptic devices for the inclusion of visually impaired pupils in geometry lessons. His current research interests include statistical signal processing and classification of neurophysiological recordings with applications to brain-computer interfaces.

Marco Congedo was born in Bari, Italy, in 1972. In 2003 he obtained the Ph.D. degree in Biological Psychology with a minor in Statistics from the University of Tennessee, Knoxville. From 2003 to 2006 he has been a post-doc fellow at the French National Institute for Research in Informatics and Control (INRIA) and at France Telecom RD, in France. Since 2007 Dr. Congedo is a Senior Scientist at the “Centre National de la Recherche Scientifique” (cnrs) in the GIPSA Laboratory, Grenoble, France. Dr. Congedo has been the recipient of several awards, scholarships and research grants. He is interested in basic human electroencephalography (EEG) and magnetoencephalography (MEG), real-time neuroimaging (neurofeedback and brain computer interface) and multivariate statistical tools useful for EEG and MEG such as inverse solutions and blind source separation. Dr. Congedo is a Fellow of the International Society for Neurofeedback and Research and a Consulting Editor for the Journal of Neurotherapy.

Clemens Brunner Clemens Brunner received the M.S. and the Ph.D. degrees in electrical engineering from the Graz University of Technology, Graz, Austria, respectively in 2003 and 2007. He is currently a postdoctoral researcher at the Institute for Knowledge Discovery, Graz University of Techonology, Graz, Austria. His research interests include brain-computer interfaces, biomedical signal processing, and software development.

Christian Jutten (M’03-SM’06-F’08) received the Ph.D. degree in 1981 and the Docteur ès Sciences degree in 1987 from the Institut National Polytechnique of Grenoble (France). He taught as associate professor in the Electrical Engineering Department from 1982 to 1989. He was visiting professor in Swiss Federal Polytechnic Institute in Lausanne in 1989, before to become full professor in University Joseph Fourier of Grenoble, more precisely in the sciences and technologies department: Polytech’Grenoble. He is currently associate director of the Grenoble images, speech, signal and control laboratory (GIPSA, 300 people) and head of the Department Images-Signal (DIS) of this laboratory. For 25 years, his research interests are blind source separation, independent component analysis and learning in neural networks, including theoretical aspects (separability, source separation in nonlinear mixtures), applications in signal processing (biomedical, seismic, speech) and data analysis. He is author or co-author of more than 45 papers in international journals, 3 books, 16 invited papers and 140 communications in international conferences. He has been associate editor of IEEE Trans. on Circuits and Systems (1994-95), and co-organizer the 1st International Conference on Blind Signal Separation and Independent Component Analysis (Aussois, France, January 1999). He has been a scientific advisor for signal and images processing at the French Ministry of Research from 1996 to 1998 and for the French National Research Center from 2003 to 2006. He has been associate editor of IEEE Trans. CAS from 1992 to 1994. He is a member of the technical committee "Blind signal Processing" of the IEEE CAS society and of the technical committee "Machine Learning for signal Processing" of the IEEE SP society. He is a reviewer of main international journals (IEEE Trans. on Signal Processing, IEEE Signal Processing Letters, IEEE Trans. on Neural Networks, Signal Processing, Neural Computation, Neurocomputing, etc.) and conferences in signal processing and neural networks (ICASSP, ISCASS, EUSIPCO, IJCNN, ICA, ESANN, IWANN, etc.). He received the EURASIP best paper award in 1992 and Medal Blondel in 1997 from SEE (French Electrical Engineering society) for his contributions in source separation and independent component analysis, and has been elevated as a Fellow IEEE and a senior Member of Institut Universitaire de France in 2008.

Gert Pfurtscheller received the M.S. and the Ph.D. degrees in electrical engineering from the Graz University of Technology, Graz, Austria. He is professor of medical informatics, head of the BCI Lab at the Institute for Knowledge Discovery at the Graz University of Technology. His research interests are focused on the dynamics of brain oscillations, functional brain topography using event-related desynchronization, and the design of brain-computer communication systems.