[ Amar Kachenoura, Laurent Albera, Lotfi Senhadji, and Pierre Comon ]

**BRAIN-COMPUTER INTERFACES**

© STOCKBYTE & DIGITAL VISION

# ICA: A Potential Tool for BCI Systems

## [ An in-depth look ]

**S**everal studies dealing with independent component analysis (ICA)-based brain-computer interface (BCI) systems have been reported. Most of them have only explored a limited number of ICA methods, mainly FastICA and INFOMAX. The aim of this article is to help the BCI community researchers, especially those who are not familiar with ICA techniques, to choose an appropriate ICA method. For this purpose, the concept of ICA is reviewed and different measures of statistical independence are reported. Then, the application of these measures is illustrated through a brief description of the widely used algorithms in the ICA community, namely SOBI, COM2, JADE, ICAR, FastICA, and INFOMAX. The implementation of these techniques in the BCI field is also explained. Finally, a comparative study of these algorithms, conducted on simulated electroencephalography (EEG) data, shows that an appropriate selection of an ICA algorithm may significantly improve the capabilities of BCI systems.
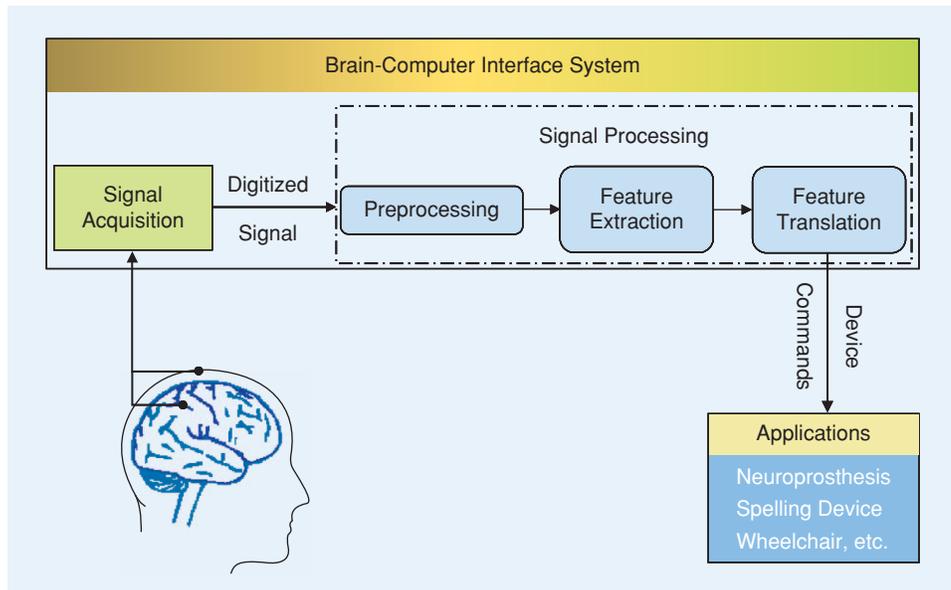
## INTRODUCTION

BCI technology is a research field that has emerged and grown rapidly over the past 15 years (see [8], [31], and [44] for a review). The BCI system is a set of sensors and signal processing components that allows acquiring and analyzing brain activities with the goal of establishing a reliable communication channel directly between the brain and an external device such as a computer, neuroprosthesis, etc. More precisely, the basic design and functioning of any BCI system are depicted in Figure 1. The brain activity is recorded by means of electrodes located on the scalp (noninvasive BCI systems) or by implanted electrodes placed, in general, in the motor cortex (invasive BCI systems) [8]. A preprocessing step is applied to enhance the signal-to-noise ratio (SNR) and to remove artifacts, such as power line noise, electrode movements, and broken wire contacts, but also interfering physiological signals as those related to ocular, muscular, and cardiac activities. Then the feature extraction step is conducted to detect the specific patterns in brain activity that encode the user's commands or reflect the patient's motor intentions [31], [44]. The last step is aimed at translating (i.e. associating) specific features into useful control signals to be sent to an external device.



[FIG1] Basic design and operation of any BCI system.

Several existing brain monitoring technologies have been tested in BCI fields for acquiring data. They can be divided in two subcategories: i) noninvasive procedures such as the EEG, magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and near infrared spectroscopy (NIRS) and ii) invasive approaches such as the electrocorticography (ECoG) where the signal is recorded from intracranial microelectrodes [31].

Until now, a majority of practical BCI systems exploit EEG signals and ECoG signals [31], [44]. Indeed, since MEG, fMRI, and PET are expensive and bulky, and as fMRI, PET, and NIRS present long time constants (they do not measure neural activity directly but rely on the hemodynamic coupling between neural activity and regional changes in blood flow), they cannot be deployed as ambulatory BCI systems. Several varieties of neurological phenomena are used by BCI systems. They include EEG rhythms such as Mu, Alpha, Beta, event-related synchronization/desynchronization (ERS/ERD) phenomena, P-300 component of the evoked-related potentials (ERPs), slow cortical potentials (SCPs), steady-state visual evoked potentials (SSVEPs), etc. (see [31, Table 3] for details). Fast and reliable signal processing tools for preprocessing the recorded data and for

extracting significant features are crucial in the development of practical BCI systems. ICA [14] is one of the popular signal processing tools, which has been widely studied during the last twenty years. Indeed, a great number of algorithms is available and ICA received a broad attention in various fields such as biomedical signal analysis and processing [33], image recognition [18] and wireless communications [20]. This article focuses on the use of ICA in BCI systems. Several studies dealing with ICA-based BCI systems have been reported during the last decade [31]. Nevertheless, most of these studies have only explored a limited number of ICA methods, and mainly FastICA [25] and INFOMAX [30]. In addition, the performance of ICA algorithms for arbitrary electrophysiological sources is still almost unknown. This prevents us from choosing the best method for a given application, and may limit the role of these methods in BCI systems. To overcome these limitations, the purpose of our study is i) to show the interest of ICA in BCI, ii) to identify ICA techniques that are appropriate to BCI, iii) to present a comparative performance analysis of six algorithms in BCI operational context, and iv) to build a reference for BCI community researchers, especially for those who are not experts of ICA techniques.

## TWO DECADES OF ICA

Hérault and Jutten seem to be the first (around 1983) to informally use the concept of ICA, especially in order to solve the blind source separation (BSS) problem [4]. A few years later, Comon presents a mathematical formulation of ICA and shows how higher order (HO) cumulants can be used to solve the problem of ICA: the HO contrast-based method COM2 arises from this work (see [14] and references therein). In parallel, Cardoso and Souloumiac develop the JADE algorithm [10], based on a joint approximate diagonalization (JAD). While these two approaches use both second order (SO) and fourth order

(FO) statistics, other approaches attempt to exploit SO statistics only. This is made possible thanks to the color of the sources, assumed unknown but different. Fety is the first to exploit covariance matrices at two different delay lags [21]; the complete theoretical background is given only a few years later by Comon et al. [15]. The same kind of approach is developed independently several years later by Tong [40], Belouchrani et al. [7] and Ziehe and Müller [49], who give rise to the so-called AMUSE, SOBI, and TDSEP methods, respectively. In 1999, Müller et al. propose a modified version of JADE, which uses the color of sources through both SO and FO statistics. More recently Albera et al. present an extension of SOBI to FO statistics [20], called FOBIUM, dealing with ICA especially in underdetermined contexts (more components than observations). Authors also propose an algebraic method [1], named ICAR, using the matrix redundancies of the FO covariance matrix, well-known as the quadricovariance matrix. As pointed out by Parra and Sajda in [36], under the assumption of non-Gaussian, nonwhite or nonstationary sources, ICA can be easily reformulated as a generalized eigenvalue problem.

Whereas the previous methods identify simultaneously the independent components, Delfosse and Loubaton [16] propose to extract one component at a time, which is now referred to as deflation procedures. A few years later, Hyvarinen et al. propose the FastICA method, which iteratively maximizes a FO contrast. While the first version of this algorithm is of deflation type, as that of Delfosse and Loubaton [16], Hyvarinen et al. [25] propose later a "simultaneous" version of FastICA whose joint orthonormalization step is similar to the one presented by Moreau [34]. Instead of exploiting, explicitly or implicitly, the SO and the FO statistics to solve the problem of ICA, some approaches use directly the independence assumption. In fact, Lee et al. present an information maximization approach [30] based on parameterized probability distributions that have sub- and super-gaussian regimes to derive a general learning rule, which is optimized using a natural gradient algorithm proposed by Amari et al. [3]. Pham proposes to use non-parametric estimates of the likelihood or the mutual information [37]. Another algorithm, based on a minimization of a non-parametric estimator of Renyi's mutual information as a criterion for ICA is introduced by Erdogmus et al. [17]. Note that Renyi's entropy is not yet proved to be better than Shannon's to address the BSS problem [41].

The previous list of ICA methods is not exhaustive, which shows that a great deal has been written on the subject. However, most of the ICA-based BCI systems presently use only FastICA [25] or INFOMAX [30]. So, after a brief survey of the concept of ICA, we propose hereafter to help the BCI scientists to choose the most appropriate ICA method among a class of six algorithms, namely INFOMAX [30], FastICA [25], COM2 [14], JADE [10], SOBI [7], and ICAR [1].

### THE CONCEPT OF ICA
ICA is very useful in the case of noninvasive BCI systems. Such BCI systems generally exploit EEG which has a high time resolution (below 100 ms). This temporal precision allows to explore the timing of basic neural processes at the level of cell assemblies. More particularly, EEG consists of measurements of a set of $N$ electric potential differences between pairs of scalp electrodes. The sensors may be either directly glued to the skin at selected locations right above cortical regions of interest, as the motor area for instance, or fitted in an elastic cap for rapid attachment with near uniform coverage of the entire scalp. Research protocols can use up to 256 electrodes. Then the $N$-dimensional set of recorded signals can be viewed as one realization of a random vector process $\{x[m]\}_{m \in \mathbb{N}}$. The ICA of $\{x[m]\}_{m \in \mathbb{N}}$ consists in looking for an overdetermined ($N \times P$) mixing matrix $A$ (i.e. $P$ is smaller than or equal to $N$) and a P-dimensional source vector process $\{s[m]\}_{m \in \mathbb{N}}$ whose components are the most statistically independent as possible so that the linear observation model below holds:

$$\forall m \in \mathbb{N} \quad x[m] = A s[m] + v[m], \tag{1}$$

where $\{v[m]\}_{m \in \mathbb{N}}$ is an $N$-dimensional noise vector process independent from the source process; bold faced lowercases denote vectors, whereas bold uppercases denote matrices. In other words, ICA consists of searching for a ($N \times P$) separator matrix $W_o$ such that $y_o[m] = W_o^\top x[m]$ is an estimate of the source vector $s[m]$. It is worth noting that once $y_o[m]$ is computed, only its components of interest for the considered BCI application have to be selected. Now how can one justify model (1) in practice?

Let's consider the instance of subjects who would learn how to control the amplitude of Mu waves by visualizing motor activities, such as smiling, chewing, or swallowing. When people move their hands a brain wave called the Mu wave gets blocked and disappears completely. Such a suppression also occurs when a person watches someone else waving his hand, but not if he or she watches a similar movement of an inanimate object. These people could thus learn how to drive a cursor up or down on a computer screen by controlling the amplitude of Mu waves. In such an example, when noninvasive measurements as EEG are used, the surface sensors record the result of the Mu wave diffusion from the motor cortex towards the scalp, corrupted by artifacts such as eye movements. The diffusion of electromagnetic waves in the head is now well-known by the biomedical community and can be modeled as a linear static transformation [2]. As far as the artifacts are concerned, they can be considered as additive perturbations whose weightings depend on the physiological nature of the artifacts. Eventually for this instance, only the statistical independence between the Mu wave and the other sources is crucial since only the Mu wave is of interest for this BCI application. In fact, it can be justified by the physiological independence between the Mu wave and the other sources such as ocular and cardiac activities.

### A CLASS OF STATISTICAL TOOLS TO PERFORM ICA
One of the fundamental questions one should ask oneself in order to choose the most appropriate ICA method in a BCI context is how to characterize the statistical independence of a set

of $P$ random signals $\{y_p[m]\}_{m \in \mathbb{N}}$ when one realization of each signal is available.

## ENTROPY AND INFORMATION

First, recall that a random vector $y = [y_1, \cdots, y_P]^\top$ has mutually independent components if and only if its probability density function (PDF) $p_y$ can be decomposed as the product of the $P$ marginal PDFs, $p_{y_p}$, where $p_{y_p}$ denotes the PDF of the $p$th component $y_p$ of $y$. Then a natural way of checking whether $y$ has independent components is to measure a pseudo-distance between $p_y$ and $\prod_p p_{y_p}$. Such a measure can be chosen among the large class of $f$-divergences. If the Kullback divergence is used, we get the mutual information (MI) of $y$:

$$\mathrm{MI}(y) = \int_{\mathbb{R}^P} p_y(u) \log \left( \frac{p_y(u)}{\prod_{p=1}^P p_{y_p}(u_p)} \right) du. \qquad (2)$$

It can be shown that the MI vanishes if and only if the $P$ components of $y$ are mutually independent, and is strictly positive otherwise.

Another measure based on the PDF of $y$ is the differential entropy (DE) of $y$:

$$S(y) = - \int_{\mathbb{R}^P} p_y(u) \log(p_y(u)) du = -\mathrm{E}[\log(p_y(y))] \qquad (3)$$

sometimes referred to as Shannon's joint entropy, where $\mathrm{E}[\cdot]$ denotes the mathematical expectation. This entropy is not invariant by an invertible change of coordinates, but only by orthogonal transforms. A fundamental result in information theory is that the DE can be used as a measure of non-gaussianity. Indeed, among the random vectors having an invertible covariance matrix, the Gaussian vector is the one that has the largest entropy. Then, to obtain a measure of non-Gaussianity of $y$ that is i) zero only for a Gaussian vector, ii) always positive, and iii) invariant by any linear invertible transformation, one often uses a normalized version of the DE, called negentropy, and given by:

$$J(y) = S(z) - S(y), \qquad (4)$$

where $z$ stands for the Gaussian vector with the same mean and covariance matrix as $y$. Since MI and negentropy are simply related to each other [14], estimating the negentropy allows to estimate the MI. However, even if we have at hand consistent estimators of PDFs (as Parzen estimators [47]), the integral computation (3) is time consuming.

The INFOMAX and FastICA methods succeed in avoiding this exact computation. On one hand, INFOMAX solves the ICA problem by maximizing the DE of the output of an invertible nonlinear transform of $y[m] = W^\top x[m]$ with respect to $W$ using the natural gradient algorithm [3]. In practice, nonlinearities whose derivative are sub-Gaussian (resp. super-Gaussian) PDFs are sufficient for sub-Gaussian (resp. super-Gaussian) sources [30]. On the other hand, in its deflationary implementation, FastICA extracts the $p$th ($1 \le p \le P$) source by maximiz-

ing an approximation of the negentropy $J(w_p^\top x[m])$ with respect to the ($N \times 1$) vector $w_p$. This maximization is achieved using an approximate Newton iteration. To prevent all vectors $w_p$ from converging to the same maximum (which would yield several times the same source), the $p$th output has to be decorrelated from the previously estimated sources after every iteration. A simple way to do this is a deflation scheme based on a Gram-Schmidt orthogonalization.

Another way to avoid the exact computation of the negentropy consists in using another measure of statistical independence less natural but easier to compute. The contrast function [14, definition 5] built from the data cumulants satisfies this condition. Let's recall the definition of cumulants and let's show why they are so attractive tools in the ICA framework.

## CUMULANTS

Let $\Phi_x(u) = \mathrm{E}[\exp(iu^\top x)]$ be the first characteristic function of a random vector $x$. Since $\Phi_x(0) = 1$ and $\Phi_x$ is continuous, then there exists an open neighborhood of the origin, in which $\Psi_x(u) = \log(\Phi_x(u))$ can be defined. Remind that the $r$th order moments are the coefficients of the Taylor expansion of $\Phi_x$ about the origin; similarly cumulants, denoted by $\mathcal{C}_{i,j,\cdots,\ell,x}$, are the coefficients of the second characteristic function, $\Psi_x$. For a $N$-dimensional random vector $x$, SO cumulants can be arranged in a ($N \times N$) matrix, which is the well-known covariance matrix, denoted by $R_x$. In the same way, it is possible to store the FO cumulants of $x$ in a ($N^2 \times N^2$) matrix, $Q_x$, called the quadricovariance matrix.

But why are cumulants useful to build a good optimization criterion dedicated to the extraction of independent components? Why are they more appropriate than moments? This comes essentially from two important properties: i) if at least two components or groups of components of $x$ are statistically independent, then all cumulants involving these components are null. For instance, if all components of $x$ are mutually independent, then $\mathcal{C}_{i,j,\cdots,\ell,x} = \delta[i, j, \cdots, \ell] \mathcal{C}_{i,i,\cdots,i,x}$, where the Kronecker $\delta[i, j, \cdots, \ell]$ equals 1 when all its arguments are equal and is null otherwise . And ii) if $x[m]$ is Gaussian, then all its HO cumulants are null. So HO cumulants may be seen as a distance to normality. Note that moments do not enjoy these two key properties. On the other hand, moments and cumulants share two other useful properties. iii) They are both symmetric arrays, since the value of their entries does not change by permutation of their indices. Consequently, $R_x$ and $Q_x$ are necessarily symmetric matrices. Lastly, iv) moments and cumulants satisfy the multi-linearity property [32]. To illustrate this, let $x$ be a random vector satisfying $x = As$, where $A$ is a ($N \times P$) matrix and $s$ a random vector with statistically independent components. We know that the ($P \times P$) covariance matrix of $s$, $R_s$, is diagonal, and that the covariance matrix of $x$ may be written as:

$$R_x = A R_s A^\top. \qquad (5)$$

Actually, this is nothing else but the expression of the multi-linearity property at order 2. Similarly at order 4, one can define a

$(P \times P)$ diagonal matrix $\zeta_s$ containing marginal source cumulants, $\mathcal{C}_{p,p,p,s}$. Then, from properties i), iii), and iv), one can deduce that $Q_x$ has the following algebraic structure:

**[TABLE 1] NUMERICAL COMPLEXITY OF THE SIX ANALYZED ICA METHODS.**

| ALGORITHMS | FLOPS |
|---|---|
| SOBI | $TMN^2/2 + 4N^3/3 + (T-1)N^3/2 + ItP^2[4P(T-1) + 17(T-1) + 4P + 75]/2$ |
| COM2 | $B + \min\{12If_4(P)P^2 + 2ItP^3 + 3Mf_4(P) + MP^2, 13ItMP^2/2\} + ItP^2Q/2$ |
| JADE | $B + \min\{4P^6/3, 8P^3(P^2+3)\} + 3Mf_4(P) + ItP^2(75 + 21P + 4P^2)/2 + MP^2$ |
| ICAR | $3Mf_4(P) + 2N^6/3 + P^2(3N^2 - P)/3 + N^2P + P^2N^3 + 7P^2N^2 + ItP^2(4N^4 - 8N^3 + 25N^2)/2$ |
| FastICA | $B + J[2P(P+M) + 5MP^2/2]$ |
| INFOMAX | $B + J[P^3 + P^2 + P(5M+4)]$ |

$$Q_x = (A \oslash A)\, \zeta_s\, (A \oslash A)^\top, \qquad (6)$$

where $\oslash$ denotes the column-wise Kronecker product [1].

Now, $r$th order cumulants can be related to moments of order smaller than or equal to $r$ using the Leonov-Shiryaev formula [32]. For instance, for any zero-mean random vector $x$ symmetrically distributed, we have:

$$
\begin{aligned}
\mathcal{C}_{i,j,x} &= \mathrm{E}[x_i x_j] \\
\mathcal{C}_{i,j,k,\ell,x} &= \mathrm{E}[x_i x_j x_k x_\ell] - \mathrm{E}[x_i x_j]\mathrm{E}[x_k x_\ell] \\
&\quad - \mathrm{E}[x_i x_k]\mathrm{E}[x_j x_\ell] - \mathrm{E}[x_i x_\ell]\mathrm{E}[x_j x_k].
\end{aligned} \qquad (7)
$$

Consequently, cumulants of a stationary-ergodic process can be easily estimated using both the Leonov-Shiryaev formula and sample statistics [32]. The SOBI, COM2, JADE, and ICAR methods perform ICA from the cumulants of the data. More precisely, SOBI uses the SO cumulants while COM2 and JADE use both the SO and FO cumulants. As far as ICAR is concerned, it uses only the FO cumulants of the data. Next, SOBI, JADE, and ICAR take advantage of the algebraic structure of the covariance (5) and/or quadricovariance matrices (6); they consider the problem of ICA as a generalized eigenvalue problem [36], while COM2 explicitly maximizes a contrast function based on the FO cumulants of the data by rooting successive polynomials. Finally, SOBI, JADE, and ICAR use the JAD method to extract the independent components.

To conclude this section, let us compare the six previously seen ICA methods to each other. First, the four cumulant-based algorithms constitute a semi-algebraic solution to the ICA problem, in the sense that they terminate within a finite number of iterations. On the contrary, INFOMAX and FastICA are iterative methods, whose convergence to local optima is possible. Moreover, all these methods except ICAR can extract components whose FO marginal cumulants have different signs. The latter scenario may occur in biomedical contexts. Another difference is the need for a spatial whitening (also called standardization) [14]. This preprocessing, based on SO cumulants, is mandatory for FastICA, COM2, JADE, and SOBI. It is not necessary but recommended in INFOMAX, in order to improve its speed of convergence [25, Chapter 9]. Regarding ICAR, it uses only FO cumulants without any standardization. Consequently, it is asymptotically insensitive to the presence of a Gaussian noise with a nondiagonal covariance matrix. In addition, contrary to the five other methods, SOBI needs that all sources are not temporally white, which is generally satisfied by BCI sys-

tems. Eventually, these six methods, namely INFOMAX, FastICA, COM2, JADE, SOBI, and ICAR, require the stationarity-ergodicity assumption to ensure an asymptotical mean square convergence. However, such an assumption is very rarely fulfilled in biomedical contexts and a consistence analysis is difficult in the presence of such complex biomedical signals. But the good behavior of FastICA, COM2, and JADE given by recent computer results [29] shows that the stationnarity-ergodicity assumption is not absolutely necessary. As far as COM2 and JADE are concerned, even if the sample statistics do not estimate accurately the cumulants of the data, they still satisfy reasonably well the above mentioned basic properties i) to iv).

### NUMERICAL COMPLEXITY

This section aims at giving some insights into the numerical complexity of the six ICA algorithms studied in this article, for given values of $N$, $P$ and data length $M$. The numerical complexity of methods is calculated in terms of number of floating point operations (flops). A flop is defined as a multiplication followed by an addition; according to the usual practice, only multiplies are counted, which does not affect the order of magnitude of the computational complexity. Now denote by $f_4(P) = P(P+1)(P+2)(P+3)/24$ the number of free entries in a fourth order cumulant tensor of dimension $P$ enjoying all symmetries, $It$ the number of sweeps required by a joint diagonalization process (SOBI, JADE, ICAR) or by contrast function optimization algorithms (COM2), $T$ the number of delay lags used in SOBI, $J$ the maximal number of iterations considered in iterative algorithms (FastICA, INFOMAX), $Q$ the complexity required to compute the roots of a real 4th degree polynomial by Ferrari's technique (we may take $Q \approx 30$ flops), and $B = \min\{MN^2/2 + 4N^3/3 + PNM, 2MN^2\}$ the number of flops required to perform spatial whitening. Then for given values of $N$, $P$, $M$, $It$, $T$, $J$, and $B$, the computational complexities are given in Table 1. It is difficult to compare computational complexities because the input parameters are different. But one can say that for a comparable performance, SOBI requires a smaller amount of calculations (when the requested assumptions are satisfied), the iterative algorithms INFOMAX and FastICA require generally a larger amount of calculations, whereas COM2, ICAR, and JADE appear close to each other in the picture. See [13], [48] for more details. The number of electrodes used in BCI systems can vary for example from one electrode [27] to 41 electrodes [24]. However, in the perspective of using BCI systems in ambulatory conditions, a reduction of the number of sensors $N$ is necessary. Regarding the number of samples $M$, we can say that $M$ is

generally low (between 300 snapshots and 5,000 snapshots). Most BCI systems exploit four types of neurophysiological signals, namely P-300 ERP, SSVEP , and EEG rhythm. These signals have generally a short time support (a few seconds) with a sampling rate that generally varies between 100 Hz and 1,000 Hz. As an example, the numerical complexity of the six studied algorithms is calculated in our experiment (described in the section "Comparative Perfomance Analysis in BCI Context"), where the ICA technique is used to extract and denoise the Mu activity. The input parameters of Table 1 are $N = 6$, $P = 3$, $M = 5$, 120 (trial duration equal to $20s$ with sampling rate of 256 Hz), $T = 5$, $Q = 30$, $It = 6$ for SOBI, $It = 4$ for COM2, $It = 7$ for JADE, $It = 11$ for ICAR, $J = 60$ forr FastICA, $J = 31$ for INFOMAX, $f_4(P) = 15$ and $B = 184608$. These parameters have been chosen so as to allow each algorithm to perform reasonably well. Thus the numerical complexity of SOBI, COM2, JADE, ICAR, FastICA, and INFOMAX is about $5.10^5$, $5.10^5$, $5.10^5$, $5.10^5$, $3.10^6$ and $3.10^6$, respectively. Clearly, for this specific example, SOBI, COM2, JADE, and ICAR require a smaller amount of calculations, whereas INFOMAX and FastICA need a larger amount of calculations.

## WHY ICA-BASED BCI SYSTEMS: A BIBLIOGRAPHICAL SURVEY

Promising results have been reported in biomedical signal processing using ICA techniques. They include fetal ECG extraction, evoked potentials (EP) enhancement, categorized brain signals detection, spindles detection and estimation, and EEG/MEG artifacts reduction. Therefore, it appears natural to consider ICA techniques as a potential tool for building BCI systems. Four types of neurophysiological signals have been mainly investigated in this context: the P-300 ERP, the SSVEP, the EEG oscillation rhythms and the ERS/ERD phenomenon. The aim of this section is to provide an overview of ICA-based BCI systems, to show how the ICA technique can be applied and how the informative independent components can be automatically chosen.

### *P-300 EVOKED POTENTIALS*

P-300 is a positive ERP, which occurs over the parietal cortex with a latency of about 300 ms after rare or task-relevant stimuli. The P-300 can be obtained in all stimulus modalities (auditory as well as visual and somatosensory modalities) and can even be produced by the omission of a stimulus in a regular train of stimuli [35]. The P-300 has useful properties, which make it very interesting for BCI applications. For example: i) The P-300 is parietally maximal, with amplitude inversely related to the relative probability of the evoking stimulus, and directly related to its task relevance. ii) The latency of the P-300 correlates to some extent with categorization or evaluation of the stimulus and consequently is related to the task-difficulty [35]. Due to the poor SNR as well as to the presence of artifacts (such as ocular, muscular, and cardiac activities), the P-300 wave can be buried in the signal collection. Hence, the main objective when applying ICA to P-300-based BCI systems is two-fold: to denoise the EEG signal in order to enhance the SNR of the P-300, and to separate ERP responses to target and nontarget ones.

The first point was considered by Bayliss and al. in [5]. Authors described an experiment demonstrating the existence of a P-300 wave when facing red stoplights and the absence of this signal when facing yellow stoplights in a virtual driving environment. The P-300 wave occurs if the red stoplights are less frequent than the yellow stoplights and the subject is asked to stop his virtual car at the red light. The data were recorded from eight electrodes (Fz, Cz, CPz, Pz, P3, P4 as well as two vertical EOG channels). Bayliss et al. pointed out that most of artifacts were due to eye movements. They showed that an ICA technique was able to separate the background EEG signal and eye movements from the P-300 signal. Indeed, after training the mixing matrix $A$ of (1) on the first seven trials, matrix $A$ was then used to find sources of the red or yellow lights. Note that the decision whether the obtained source represents a P-300 ERP or not was set by correlation with the red and yellow light signal references obtained, from prior controlled experiment, by averaging the EEG recorded in red light and yellow light trials, respectively (when the subject ran a red light, the trial was canceled). Authors showed that the ICA performance was similar to that obtained by a robust Kalman filter.

In [45], Xu et al. dealt with the second point and proposed to enhance the P-300 wave detection in the P-300 speller paradigm (described in [19]) used to record the database IIb of BCI Competition 2003 [9]. Briefly, the experiment consisted in displaying a $(6 \times 6)$ grid containing letters on a computer monitor. The user was asked to select a letter in an alphabet and to count the number of times that a row or column containing the letter flashes. Flashes occurred at about 10 Hz and the desired letter flashed twice in every set of twelve flashes. As classical methods for enhancing the detection of P-300 components are time consuming, authors proposed to use an ICA technique in the training phase. The 64-raw EEG channels were first filtered through a 2–8 Hz bandpass filter, the data dimension was reduced from 64 to 22 by means of PCA, then ICA was applied on the dimension reduced data to derive a $(22 \times 22)$ mixing matrix, $A$. Thus the trained matrix $A$ was used in the testing phase in order to identify the target P-300 responses. The key issue of this study is the selection of the meaningful independent components (ICs). Indeed, according to the a priori physiological knowledge, authors proposed two additional post-processing steps, namely the temporal manipulation of ICs and the spatial manipulation of ICs (see [45] for details). They showed that the proposed algorithm for P-300 detection based on ICA provided a perfect accuracy (100%) in the competition.

### *AUDITORY EVENT-RELATED POTENTIALS*

The auditory event-related potential (AERP) is the brain response time-locked to an auditory stimulus. AERPs are very small electrical potentials [35] (2–10 $\mu$v for cortical AERPs to much less than 1 $\mu$v to the deeper structures). Their low voltage combined with relatively high background EEG activity and other artifacts requires the use of highly sensitive amplifiers and robust signal processing techniques for denoising and for extracting significant features. The use of AERPs in BCI systems

is motivated by some particular problems encountered in communication with patients suffering from severe motor disabilities. In some severe cases, the eyes are completely immobile and the pyramidal cells of the motor cortex are degenerated. Thus, the BCI paradigms using the visual modality or the imagined-movements signals become too limited. Hill et al. presented in [24] an experiment on healthy subjects, based on AERPs, to develop BCI systems. In their study, two auditory stimuli were presented to 15 healthy untrained subjects: i) the right-ear sequence consisted of eight beeps (pure frequencies) at 1,500 Hz (nontarget) and 1,650 Hz (target) and ii) the left-ear sequence consisted of seven beeps at 800 Hz (nontarget) and 880 Hz (target), starting 70 ms after the onset of the right-ear sequence.

EEG signals were recorded using 39 electrodes and eye movements were recorded by means of an electrode positioned on the side and slightly below the left eye. All trials were first visually inspected to reject those corrupted by large artifacts. The retained trials were then classified using a support vector machine (SVM) technique. To evaluate the performance of the classifier, the retained trials from a single subject were split up into ten nonoverlapping partitions of equal size: each partition was used in turn as a test set for evaluating the performance of the classifier trained on the other 90% of the trials. Before starting the SVM classifier, ICA was applied to the training set and a $(40 \times 40)$ mixing matrix $A$ was obtained. This matrix was used to extract ICs of both the training set and the testing set. The SVM classifier was then trained and tested. Authors compared the performance of the classifier with and without ICA. They reported that ICA generally improves the result up to 14%.

### STEADY-STATE VISUAL EVOKED POTENTIALS
The visual system can be studied noninvasively by recording scalp EEG overlying the visual cortex. These potentials constitute the visual evoked potentials (VEPs) and reflect the output features of the entire visual pathway [35]. The SSVEP is the response of a continuous rapid visual stimulus [35]. Typically, the visual stimulus is generated using a frequency of 9–17 Hz (Fs) flashing lights. This high simulation rate allows saving time and thereby makes SSVEP attractive for BCI applications. The Fourier analysis is one of the conventional methods used for studying the SSVEP. Indeed, a typical power spectrum of SSVEP wave, introduced by Fs Hz stimulation, presents fundamental harmonics and second harmonics at Fs and two Fs, respectively. However, due to the small amplitude of SSVEPs these harmonics can be drowned in background EEG rhythm. To accurately detect the frequency of occurrence of SSVEP, an appropriate bipolar derivation should be selected. Wang et al. proposed, using ICA, a method for selecting both the signal channel and the reference [43] which enhances the SNR. In their experiment, a blinking light-emitting diode (LED) modulated by a square wave was used as the stimulator with stimulation frequencies rising from 9 Hz to 17 Hz. One-min-long data were recorded, in each test with different simulation frequencies, from 13 channels located between Pz and Oz (electrodes positioned over visual cortex). ICA was then applied to EEG signals and 13 ICs were derived, and the associated mixing matrix

was then estimated. The power spectrum of each IC was analyzed and the four most significant powers at stimulation frequency were supposed to be related to SSVEP signals and the remaining powers were considered as the contribution of background noise. Two groups of EEG signals were then reconstructed from the source mixture and these two kinds of sources, namely the SSVEP and background activity groups, respectively. Thus, one bipolar derivation with higher correlation of background activity and lower correlation of SSVEP signal was selected as the optimal derivation. Authors also showed that their method had been successfully applied to a BCI-based environmental controller presented in [22].

### MU RHYTHM AND OTHER ACTIVITIES FROM SENSORIMOTOR CORTEX
EEG contains a fairly wide frequency spectrum. Nevertheless, the relevant frequency range from the psychophysiological viewpoint lies between 0.1 Hz and 100 Hz [35]. Six important waves are categorized by their frequency band or their location: Alpha (8–13 Hz), Beta (14–30 Hz), Theta (4–7.5 Hz), Delta (0.1–3.5 Hz), Gamma (above 30 Hz), and Mu, which is in the same frequency band as Alpha (the most common frequency of the Mu rhythm is 10 Hz), but Mu wave has a spacial distribution essentially confined to the precentral-postcentral region (activity focused over motor cortex). Several factors suggest that the EEG rhythms could be good signal features for BCI systems. Indeed, all waves are usually associated with specific activity. For example, Beta rhythm is associated with active thinking and attention, the Theta rhythm is associated with emotional processes, the posterior Alpha rhythm is temporally blocked by an influx of light (opening the eye) and other mental activities [35], and Mu is affected by movements or movement imagery. The reactivity of EEG by short-term attenuation and intensification of power in particular frequency bands is labeled ERD and ERS, respectively. Several studies using ICA to extract one specific feature or feature subsets, have been conducted. We decided to present in detail two studies.

The first one [39] is a pilot study aimed at classifying motor imagery for BCI systems, using ICA as a spatio-temporal filter. More precisely, the study was focused on the Mu rhythm which decreases or desynchronizes with movement or movement imagery (an ERD appears in contralateral brain region). Authors conduct an experiment where the subject was asked to imagine right or left-hand movement according to a timetable. EEG were recorded using 59 electrodes but only those located around sensorimotor cortex were used in the study. After a preprocessing step, ICA was used to extract ICs related to the left and the right motor imagery task. The estimated mixing matrix was then sorted based on the norm of columns in ascending order. Authors applied two source analysis methods to reconstruct equivalent neural sources corresponding to the motor imagery (see [39] for details). Note that only the first IC, corresponding to the first column of the ordered matrix, was used in the source analysis step. Authors showed that mixing ICA plays an important role in extracting a useful feature that identifies

the imagined hand movement. A promising classification rate (about 80%) of left- or right-hand movement imagery was obtained on human subject studies, based only on a single trial and without any training procedure.

The second study [27] is interesting because ICA was applied to only one recording channel. The 2003 international BCI competition database III [9] was exploited by authors to show that a rhythmic activity between 8–12 Hz can be extracted in the case of hand imagined movements. The goal of the study was to control a feedback bar by means of imagined left or right-hand movements. ICA was then applied, as above-mentioned, on each EEG signal recorded from three electrodes C3, Cz, and C4. Authors calculated the Fourier transform (FT) of each column of the obtained mixing matrix and extracted all components corresponding to columns exhibiting a peak in the 8–12 Hz frequency band. The chosen ICs were then projected to the measurement space and summed. Authors showed, by applying this technique on the EEG recorded from C3, Cz, and C4 electrodes, that an ERD of activity in the 8–12 Hz band can be clearly identified.

### HOW TO SELECT THE INFORMATIVE INDEPENDENT COMPONENTS?

One of the challenging tasks in BCI is to reliably detect, enhance, and localize very weak brain activities corrupted by noise and various interfering artifacts of physiological and non-physiological origin. The applications presented before show that ICA is a promising approach for improving BCI systems. In fact, for these applications, ICA has been successfully applied for providing feature subsets with high classification accuracy. However, one important problem that arises when ICA is used in practical BCI systems, is to automatically select and classify independent sources of interest. Several experimental advances in BCI select manually the relevant ICs or estimate the cross-correlation between ICs and the reference signals corresponding to the specific features. To transform these experimental advances into viable BCI systems, an automatic on-line selection of the relevant ICs is necessary. The solution to these problems can be decomposed into two stages. In a first step, the recorded brain signals are decomposed into useful signal and noise subspaces using standard techniques like PCA, factor analysis (FA), singular value decomposition (SVD), nonlinear adaptive filtering, etc. Since the methods used in the first step always contain the risk of including ICs that do not represent the components of interest (other physiological artifacts, such as eye movements, cardiac activities, …), a second step allowing to automatically select only the component of interest is needed. For this purpose, Wang et al. [42] developed two algorithms. In the first one, in order to assist the user in choosing the related P300 components, they propose: i) to increase the target and nontarget ratio in the training data, ii) to create a template as a topography which presents maximal signal intensity at the vertex region, and iii) to use the correlation between the template and each column of the inverse unmixing matrix.

In the second algorithm, authors show that if a priori knowl-edge about the spatial projections of some sources is available, then it can be incorporated into the ICA model by means of constraints on some columns of the mixing matrix; this provides an assistance to the component selection (see [26] for details).

Another method allowing to select the components of interest consists in exploiting the spectral information of particular sources. As in the case of BCI systems using the EEG rhythm, we have some ideas about the spectral information on the relevant components we wish to extract. For example, in [27], authors select the source of interest by using a mere FT of each column of the obtained mixing matrix. All components corresponding to columns with a peak in the 8–12 Hz frequency band are selected and projected to the measurement space.

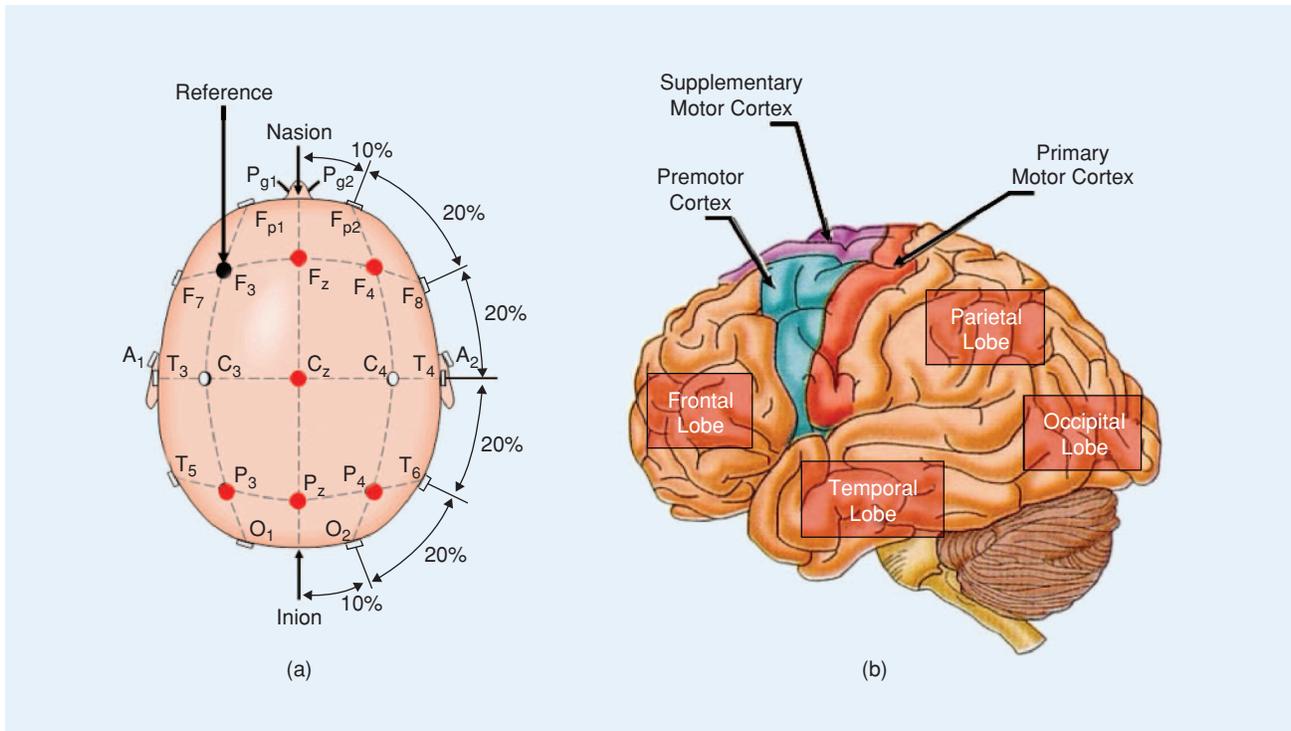## COMPARATIVE PERFORMANCE ANALYSIS IN BCI CONTEXT

### DATA GENERATION

The main goal of this subsection is to explain how to obtain realistic data for comparing ICA methods in the context of BCI systems based on the Mu rhythm when seven surface electrodes are used to achieve EEG recordings [see Figure 2(a)]. In such a context, as depicted in the section on "Two Decades of ICA," the surface observations can be considered as a noisy mixture of one source of interest, namely the Mu wave, and artifact sources such as the ocular and cardiac activities. The intracerebral Mu wave, located in the motor cortex [see Figure 2(b)], is simulated using the parametric model of Jansen [28] whose parameters are selected to derive a Mu-like activity. The ocular and cardiac signals are issued from our polysomnographic database [38]. Concerning the additive noise, it is modeled as the sum of the instrumental and physiological noises. A Gaussian vector process is used to simulate the instrumental noise while a brain volume conduction of 200 independent EEG sources, generated using the Jansen model [28], is simulated in order to produce a surface background EEG activity. Finally, the mixing matrix is defined as the concatenation of a two column matrix, modeling the head volume conduction [2] of the Mu and ocular sources, and a unit weighting vector associated with the cardiac source, which implies that the surface electrodes are uniformly corrupted by the cardiac activity.

### PERFORMANCE CRITERION

Two separators, $W_o^{(1)}$ and $W_o^{(2)}$ can be compared with the help of the criterion introduced by Chevalier [12]. The quality of the extracted component is directly related to its signal to interference-plus-noise ratio (SINR). More precisely, the SINR of the $p$th source at the $i$th output of the separator $W_o = [w_o^{(I)}, \ldots, w_o^{(P)}]$ is defined by:

$$\text{SINR}_p\left[w_o^{(i)}\right] = \pi_p \frac{|w_o^{(i)\top} a_p|^2}{w_o^{(i)\top} R_{\boldsymbol{v}_p} w_o^{(i),}} \qquad (8)$$

where $\pi_p$ represents the power of the $p$th source, $w_o^{(i)}$ the $i$th column of the separator $W_o$ and $R_{vp}$ is the total noise covari-

[FIG2] (a) The international 10–20 electrodes placement system and (b) surface of left cerebral hemisphere, viewed from the side.

ance matrix for the $p$th source, corresponding to the estimated data covariance matrix $R_x$ in the absence of the component $p$. On the basis of these definitions, the restitution quality of the $p$th source at the output of the separator $W_o$ is evaluated by computing the maximum of $\text{SINR}_p[w_o^{(i)}]$ with respect to $i$ where $i$ $(1 \le i \le P)$. This quantity is denoted by $\text{SINRM}_p$. The performance of a source separator $W_o$ is then defined by the following line vector $\text{SINRM}(W_o)$:

$$\text{SINRM}(W_o) = (\text{SINRM}_1[W_o], \dots, \text{SINRM}_P[W_o]). \quad (9)$$

In a given context, a separator $W_o^{(1)}$ is better than another one $W_o^{(2)}$ for retrieving the source $p$, provided that $\text{SINRM}_p[W_o^{(1)}] > \text{SINRM}_p[W_o^{(2)}]$. The criterion given by (9) allows for a quantification of the component analysis performed by ICA algorithms. However, the use of this criterion requires to know its upper bound, which is achieved by the optimal source separator, in order to completely evaluate the performance of a given ICA method. It is shown in [12] that the optimal source separator corresponds to the separator $W_o(\text{SMF})$ whose columns are the spatial matched filters (SMF) associated with the different sources. It is defined to within a diagonal matrix and a permutation by $W_o(\text{SMF}) = R_x^{-1}A$ where $A$ is the the true mixture.

### COMPUTER RESULTS

To conduct a comparative performance study of the six ICA algorithms presented in the section "A Class of Statistical Tools to Perform ICA," two experiments are envisaged from the data described in the section "Data Generation." We just focus on the ability of the six algorithms to extract the Mu source which is
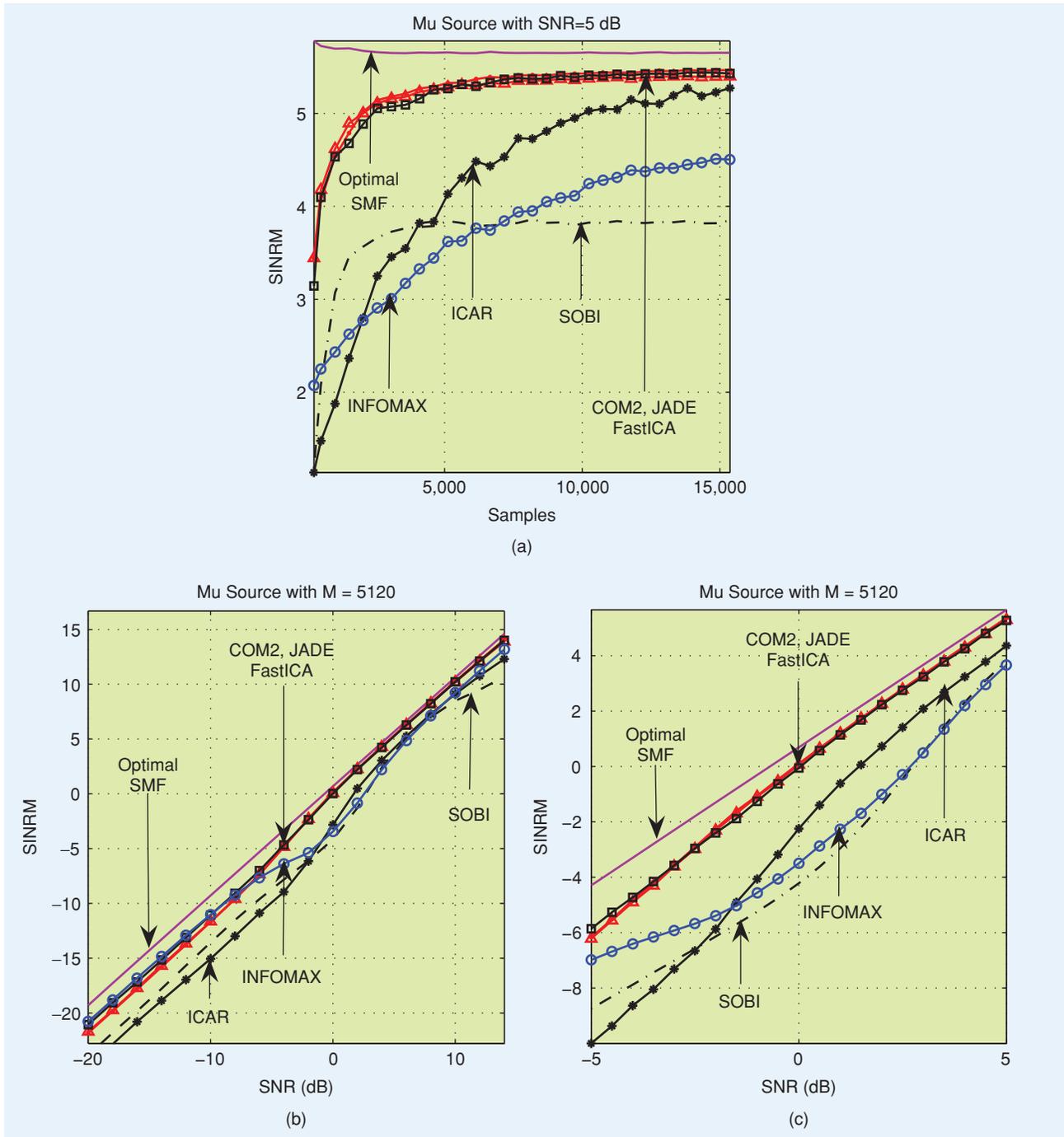
considered as the informative component in the case of the Mu rhythm based BCI application. All results reported hereafter are obtained by averaging over 200 realizations. Moreover, INFO-MAX is implemented with a prior spatial whitening step.

### INFLUENCE OF THE DATA LENGTH FOR A FIXED SNR

In this experiment, we set the SNR to 5 dB for each source. The $\text{SINRM}_p$ $(1 \le p \le 3)$ at the output of the six ICA methods, namely SOBI, COM2, JADE, ICAR, FastICA, and INFOMAX, is computed as a function of the number of samples (with a sampling rate of 256 Hz). Figure 3(a) displays the obtained results corresponding to the Mu source. We observe the good behavior of COM2, JADE, and FastICA. ICAR is slightly less effective to reconstruct the Mu source and needs more samples to converge to the optimal solution. One of the reasons could be the fact that ICAR require that all sources are non-Gaussian. But the estimated PDF of the simulated Mu source is close to the Gaussian PDF, especially when the number of snapshots is low. Another reason is the fact that the sign of the FO marginal cumulant of the Mu source is different from the sign of the two other FO marginal source cumulants, which is not supported a priori by ICAR. Finally, whereas INFOMAX seems to require a very large amont of samples to converge, the separator estimated by SOBI appears to be biased for this simulation scenario.

### INFLUENCE OF THE SNR

In this section we study the behavior of the six ICA methods as a function of the SNR, which is assumed to be the same for each source. The number of samples is set to 5120, which correlates to 20 s. Figure 3(b) illustrates that, when varying the

[FIG3] Evolution (in dB) of SINRM of sources as: (a) a function of data length with a SNR equal to 5 dB (b) as a function of SNR. (c) Represents the zoom in (b) for SNR between −5 dB and 5 dB.

SNR from −20 to 14 dB, COM2, JADE, and FastICA exhibit quasi-optimal performance. Regarding SOBI and ICAR algorithms, they are less effective in comparison to the three previous methods for all values of SNR. Moreover, ICAR seems to be a little effective than SOBI beyond −2 dB and the inverse appears to be true below −2 dB [see Figure 3(c) which is the zoom in of the Figure 3(b)]. The INFOMAX algorithm behaves like the COM2, JADE, and FastICA for SNR values lower than

−6 dB but, as depicted in Figure 3(c), its performance is similar with SOBI beyond −6 dB.

**CONCLUSIONS**
BCI allows a person to control external devices using electrophysiological phenomena such as evoked or spontaneous EEG features (P-300, EEG rhythm, SSVEP, etc.). EEG signals generally contain an unknown mixture of components represent-

ing brain processes. They are often contaminated by artifacts either physiological, such as ECG, EMG, or nonphysiological (power line supply, electrode displacement, etc.). Thus, reliable signal processing tools for both preprocessing and extracting significant features are crucial in BCI systems. Previous works on biomedical signals showed that ICA holds great promises for blindly separating artifacts from relevant signals and for decomposing mixtures into subcomponents that may index electrophysiological activities of interest. In addition, all studies emphasize the fact that ICA provides feature subsets with higher classification accuracy than original EEG signals.

Nevertheless, the application of ICA in the field of BCI systems is at its beginning. Indeed, the majority of studies reported in the literature are conducted on healthy subjects and the fact that ICA-based BCI systems provide good results with healthy subjects does not necessary guarantee a success with target user groups. In addition, the automatic selection of the informative component would allow an online ICA-based BCI, but it is not an easy task. Some recent studies detailed in this article show that ICA can be used in an automated fashion. But up to now, most BCI systems using an automated selection of the informative independent components have been tested only in laboratories with well defined protocols.

Another limitation of the studies dealing with ICA-based BCI seems to be the fact that most of them have only explored a limited number of ICA algorithms, namely FastICA and INFOMAX. This last point is investigated in the present article and six classical ICA methods are presented. Then a numerical complexity study, a comparative performance analysis on electrophysiological data, reproducing real scalp EEG recordings, is conducted. The obtained results show that the numerical complexity can largely vary, depending on the ICA algorithm that has been used whereas the algorithm behavior depends both on the number of snapshots and on the value of the SNR. More precisely, SOBI, COM2, JADE, and ICAR provide a smaller amount of calculations in comparison to FastICA and INFOMAX. Regarding performance, COM2, JADE, and FastICA lead to the best results. Thus, taking into account both numerical complexity and performance, COM2 and JADE seem to be the best methods in our BCI simulation.

Another result, already obtained in one of our previous works [29], shows that the performance of ICA algorithms also depends on the physiological nature of the source that is extracted. All these results strengthen us in our conviction that the choice of an appropriate ICA method can significantly improve the performance of ICA-based BCI systems. With this goal, it is important to collect as much statistical/physiological prior information as possible on sources, such as their temporal color, their non-Gaussianity, the sign of their marginal cumulants, or the nature of the additive noise.

## ACKNOWLEDGMENT

## AUTHORS
*Amar Kachenoura* (amar.kachenoura@univ-rennes1.fr) received the engineer degree in electronic engineering from the University of Mouloud Mammeri de Tizi-Ouzou, Tizi-Ouzou, Algeria, in 1998 and the DEA degree from École Central de Nantes (ECN), Nantes, France, in 2002. He received his Ph.D. in signal processing from the University of Rennes 1, Rennes, France, in 2006. He is post-doctorant in Laboratoire Traitement du Signal et de l'Image (LTSI), INSERM U642-University of Rennes 1, Rennes, France. His research interests focus mainly on biomedical and statistical signal processing, BSS, ICA, and BCI.

*Laurent Albera* (laurent.albera@univ-rennes1.fr) received the DEA degree in Automatic and Signal Processing in 2001 from the University of Science of Orsay in Paris. From 2001–2004, he was committed for a study within the framework of a research contract convention industrielle de formation par la recherche (CIFRE) between Thales Communications (formerly Thomson-CSF), the French National Centre for Scientific Research CNRS, and the University of Nice Sophia-Antipolis. In 2003, he received his Ph.D. degree in Sciences from the University of Nice Sophia-Antipolis, France. He is now an assistant professor at the University of Rennes 1 and is affiliated with the research group LTSI (Laboratoire Traitement du Signal et de l'Image). His research interests include multilinear algebra and high-order statistics in order to perform ICA and source localization.

*Lotfi Senhadji* (lotfi.senhadji@univ-rennes1.fr) received his Ph.D. degree from the University of Rennes 1, France, in signal processing and telecommunications in 1993. He is a professor at the University of Rennes 1 in the Department of Electrical Engineering, and he is the head of the INSERM research laboratory, LTSI. His main research efforts are focused on nonstationary signal processing with particular emphasis on wavelet transforms and time-frequency representations for detection, classification, and interpretation of biosignals. He has been investigator of several grants in the field of health technology with public institutions or companies. He has published more than 70 research papers in journals and conferences, and he made contributions to five handbooks. He is a member of the council of the French Society of Biomedical Engineering (SFGBM), the French Alliance for Biomedical Engineering (AGBM), and senior member of IEEE Engineering in Medicine and Biology Society and IEEE Signal Processing Society.

*Pierre Comon* (p.comon@ieee.org) received the Doctorate degree in 1985 from the University of Grenoble, France. He later received the habilitation to lead researches (HLR) in 1995 from the University of Nice, France. Since 1998, he has been research director with CNRS at Laboratory I3S, Sophia-Antipolis, France. His research interests include high-order statistics, tensors, blind deconvolution and equalization, and statistical signal and array processing for digital communications or biomedical applications. He was Associate Editor of the *IEEE Transactions on Signal Processing* from 1995–1998, and a member of the French National Committee of Scientific

Research from 1995–2000. From 2001–2004 he acted as launching Associate Editor with the *IEEE Transactions on Circuits and Systems I,* in the area of blind techniques. He was also IEEE Distinguished Lecturer from 2002–2003. He is a member of the editorial board of *Signal Processing*, and member of the SAM and SPTM technical committees in the Signal Processing Society. He received the Monpetit Prize from the French Academy of Sciences in 2005, and the Individual Technical Achievement Award from EURASIP in 2006. He is a member of IEEE and SIAM, and is an emeritus member of the SEE.

## REFERENCES

[1] L. Albera, A. Ferreol, P. Chevalier, and P. Comon, "ICAR, a tool for blind source separation using fourth order statistics only," *IEEE Trans. Signal Processing*, vol. 53, no. 10, pp. 3633–3643, Oct. 2005.

[2] L. Albera, A. Ferreol, D. Cosandier-Rimele, I. Merlet, and F. Wendling, "Brain source localization using a fourth order deflation scheme," *IEEE Trans. Biomed. Eng.*, submitted for publication.

[3] S. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.

[4] B. Ans, J. Herautl, and C. Jutten, "Adaptative neural architectures: Detection of primitives," in *COGNITIVA'85*, Paris, France, June 1985, pp. 593–597.

[5] J.D. Bayliss and D.H. Ballard, "Single trial p300 recognition in a virtual environment," in *CIMA'99. Proceedings of Soft Computing in Biomedicine*, Rochester, New York, USA, June 1999.

[6] A.J. Bell and T.J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[7] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[8] N. Birbaumer and L.G. Cohen, "Brain-computer interfaces: Communication and restoration of movement in paralysis," *J. Physiol.*, vol. 579, no. 3, pp. 621–636, 2007.

[9] B. Blankertz, K.R. Muller, G. Curio, T.M. Vaughan, G. Schalk, J.R. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroder, and N. Birbaumer, "The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1044–10 511, Mar. 2004.

[10] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc.-Part-F, Special issue on application of Hight-Order Statistics*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[11] J.F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 1, pp. 161–164, 1996.

[12] P. Chevalier, "Optimal separation of independent narrow-band sources: Concept and performances," *Signal Proces., Elsevier,* vol. 73, no. 1, pp. 27–47, 1999.

[13] P. Chevalier, L. Albera, P. Comon, and A. Ferreol, "Comparative performance analysis of eight blind source separation methods on radiocommunications signals," in *IJCNN 04, Proc. IEEE Int. Joint Conf. Neural Networks*, Budapest, Hungary, July 2004, pp. 278-283.

[14] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[15] P. Comon and J.L. Lacoume, "Statistiques d'ordres supérieurs pour le traitement du signal," Ecole Predoctorale de Physique, P. Flandrin and J. L. Lacoume, Eds. Les Houches summer school, 30 aout—10 septembre 1993.

[16] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Process.*, vol. 45, no. 1, pp. 59–83, 1995.

[17] D. Erdogmus, K. II, and J.C. Principe, "Blind source separation using Renyi's alpha-marginal entropies," *Special Issue on Blind Source Separation and Independent Component Analysis, Neurocomputing,* vol. 49, no. 1, pp. 25–38, Dec. 2002.

[18] F. Esposito, T. Scarabino, A. Hyvärinen, J. Himberg, E. Formisano, S. Comani, G. Tedeschi, R. Goebel, E. Seifritz, and F.D. Salle, "Independent component analysis of fMRI group studies by self-organizing clustering," *NeuroImage*, vol. 25, no. 1, pp. 193–205, 2005.

[19] L.A. Farwell and E. Donchin, "Talking off the top of your head: Toward the mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.

[20] A. Ferreol, L. Albera, and P. Chevalier, "Fourth order blind identification of underdetermined mixtures of sources (FOBIUM)," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1640–1653, May 2005.

[21] L. Fety, "Méthodes de traitement d'antenne adpatées aux radiocommunications," Ph.D. dissertation, Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, 1988.

[22] X. Gao, D. Xu, M. Cheng, and S. Gao, "A BCI-based environmental controller for the motion-disabled," *IEEE Trans. Rehab. Eng.*, vol. 11, no. 2, pp. 137–140, June 2003.

[23] S. Haykin, *Unsupervised Adaptive Filtering, Vol. I, Blind Source Separation*. New York: Wiley Interscience, 2000.

[24] N.J. Hill, T.N. Lal, K. Bierig, N. Birbaumer, and B. Scholkopf, "Attentional modulation of auditory event-related potentials in a brain-computer interface," in *Proc. IEEE Int. Workshop Biomedical Circuits and Systems*, Singapore, 2004, pp. 17–19.

[25] A. Hyvarinen, J. Karhunen, and P. Oja, *Independent Component Analysis, Ser.* New York: Wiley Interscience, 2001.

[26] C.J. James and C.W. Hesse, "Independent component analysis for biomedical signals," *Physiolog. Meas.*, vol. 26, no. 1, pp. R15–R39, 2005.

[27] C.J. James and S. Wang, "Blind source separation in single-channel EEG analysis: An application to BCI," in *Proc. 28th Annu. Int.Conf. IEEE Enginering in Medicine and Biology Society*, New York, USA, Aug. 2006, pp. 6544-6547.

[28] B.H. Jansen and V.G. Rit, "Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns?" *Biol. Cybern.*, vol. 73, no. 4, pp. 357–366, 1995.

[29] A. Kachenoura, L. Albera, and L. Senhadji, "Blind source separation methods applied to synthesized polysomnographic recordings: A comparative study" in *Proc. EMBS 07, 29th Ann. Int. Conf. IEEE Engineering in Medicine and Biology Society*, Lyon, France, Aug. 2007, pp. 3868–3871.

[30] T.W. Lee, M. Girolami, and T.J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources," *Neural Comput.*, vol. 11, no. 2, pp. 417–441, 1999.

[31] S.G. Mason, A. Bashashati, M. Fatourechi, K.F. Navarro, and G.E. Birch, "A comprehensive survey of brain interface technology designs," *Ann. Biomed. Eng.*, vol. 35, no. 2, pp. 137–169, 2007.

[32] P. McCullagh, *Tensor Methods in Statistics*. London: Chapman & Hall, 1987.

[33] K.R. Müller, R. Vigario, F. Meinecke, and A. Ziehe, "Blind source separation techniques for decomposing event-related brain signal," *Int. J. Bifurcation and Chaos*, vol. 14, no. 2, pp. 773–791, 2004.

[34] E. Moreau, "Criteria for complex sources separation," in *EUSIPCO'96, Proc. European Signal Processing Conf.*, Trieste, Italy, Sept. 1996, pp. 931–934.

[35] E. Niedermeyer and F.L. Dasilva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields,* 4th ed. Baltimore, MD: Williams & Wilkins, 1999.

[36] L. Parra and P. Sajda, "Blind source separation via generalized eigenvalue decomposition," *Special Issue on Independent Components Analysis, J. Mach. Learn. Res.*, vol. 4, pp. 1261–1269, 2003.

[37] D.T. Pham, "Mutual information approach to blind source separation of stationary sources," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1935–1946, July 2002.

[38] F. Poree, A. Kachenoura, H. Gauvrit, C. Morvan, G. Carrault, and L. Senhadji, "Blind source separation for ambulatory sleep recording," *IEEE Trans. Inform. Technol. Biomed.*, vol. 10, no. 2, pp. 293–301, 2006.

[39] L. Qin, L. Ding, and B. He, "Motor imagery classification by means of source analysis for brain-computer interface applications," *J. Neural Eng.*, vol. 1, no. 3, pp. 135–141, 2004.

[40] L. Tong, V. Soon, R. Liu, and Y. Huang, "AMUSE: A new blind identification algorithm," in *Proc. IEEE Int. Symp. Circuits and Systems*, New Orleans, USA, 1990, pp. 1784-1787.

[41] F. Vrins, D.-T. Pham and M. Verleysen "Is the general form of Renyi's entropy a contrast for source separation?" in *Proc. ICA 2007, 7th Int. Conf. Independent Component Analysis and Signal Separation*, London, UK, Sept. 2007, pp. 9–12.

[42] W. Wang and C. James, "Enhancing evoked responses for BCI through advanced ICA techniques," in *Proc. MEDSIP 06, 3rd Int. Conf. Advances in Medical, Signal, and Information Processing*, Glasgow, UK, July 2006, pp. 38–41.

[43] W. Wang, Z. Zhang, X. Gao, and S. Gao, "Lead selection for SSVEP-based brain-computer interface," in *Proc. EMBS 04, 26th Int. Conf. Engineering in Medicine and Biology Society*, San Fransisco, USA, Sept. 2004, pp. 4507–4510.

[44] J. Wolpaw, N. Birbaumer, D.J. Mcfarland, G. Pürtsceller, and T.M. Vaughan, "Brain-computer interfaces for communication and control," *Electroencephalogr. Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[45] N. Xu, X. Gao, X. Hong, X. Miao, S. Gao, and F. Yang, "BCI competition 2003-data set iib: Enhancing P-300 wave detection using ICA-based subspace projections for BCI applications," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1067–1072, 2004.

[46] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1545–1553, July 2002.

[47] T.Y. Young and T.W. Calvert, *Classification, Estimation and Pattern Recognition*. New York: American Elsevier, 1974.

[48] V. Zarzoso and P. Comon, "How fast is FastICA?" in *Proc. European Signal Processing Conf.*, Florence, Italy, Sept. 2006.

[49] A. Ziehe and K.R. Müller, "TDSEP- An efficient algorithm for blind separation using time structure," in *Proc. 8th Int. Conf. Artificial Neural Networks*, Berlin, 1998, pp. 675–680.

SP