

BIOPHYSICS. MEDICAL PHYSICS. ENVIRONMENTAL PHYSICS

**A COMPARATIVE STUDY OF INDEPENDENT COMPONENT
ANALYSIS ALGORITHMS FOR ELECTROENCEPHALOGRAPHY**

RADU MUTIHAC¹ AND RADU CRISTIAN MUTIHAC²

¹ Department of Electricity and Biophysics, University of Bucharest, 077125 Bucharest, Romania

² DTNW and Duisburg-Essen University, 47798 Krefeld, Germany

(Received October 9, 2006)

Abstract. Application and quality assessment of four different algorithms in running temporal independent component analysis (ICA) decomposition of single-trial multichannel electroencephalographic (EEG) recordings onto temporally independent and spatially stationary source signals, as well as identification and possible removal of artifacts in EEG recordings were carried out. The differentiation between genuine cerebral and artifactual activities was performed on the assumption that they were anatomically and physiologically separate processes, which concluded in their temporal independence reflected in the higher-order statistics. The assumption of independence, even if weak, was proved reach in relevant inferences on brain activity. The ICA algorithms under study performed fairly similar in terms of source separation for both real and simulated EEG time series, though significant difference in computational demand was noted.

Key words: independent component analysis, blind source separation, electroencephalography, statistical analysis, Gaussian and non-Gaussian distributions, power spectrum density.

1. INTRODUCTION

The analysis of neuroimaging data is a rich source of both anatomical and functional information on the brain extracted from mixtures of unknown combinations of signals summing differently at different loci, and often contaminated with various artifacts and/or noise. In many cases, even the nature of the signal sources is a question of debate. The relative contribution and exact form of each of these components are largely unknown, suggesting analysis methods of the blind source separation-type (BSS) [1]. The current techniques in data analysis can be loosely dichotomized into (i) *hypothesis-driven* methods, like the general linear model (GLM) [2], and (ii) *data-driven* model-free methods, such as principal component analysis (PCA) and ICA [3]. These two approaches are complementary rather than competitive, and mirror the exploratory and confirmatory aspects of scientific investigation. The analytical hypothesis-driven techniques require *a priori* knowledge and/or specific assumptions about the time courses of processes

generating the measured signals. In contrast, no prior assumptions are necessary for the data-driven methods on: (i) time courses of activation of any components, (ii) whether a given component is due to specific psychophysiological activity or is related to machine noise or other artifacts. Imaging studies driven by hypotheses derived from cognitive psychology and related disciplines can, at best, support or refute currently formulated psychological models. Unanticipated time courses of activation of localized brain regions are less likely to be discovered with such analysis methods [4].

The main interest in functional brain studies lays in the electrical activity of firing neurons, which cannot be directly investigated by any magnetic resonance imaging (MRI) procedure. The electrical neural activity cannot entirely be inferred by analyzing the vascular process because: (i) the hemodynamic lag varies in a complex way from tissue to tissue, and (ii) no theory on the relationship between the electrical and hemodynamic processes is unanimously accepted. The analysis of brain electrical activity is an increasingly important area of research for both understanding and modeling the human brain, and for medical diagnosis and treatment as well, especially for developing automated patient monitoring and computer-aided diagnosis. Among various exploratory methods applied to electrical and hemodynamic recordings from the human brain, the ICA approach has been proved to reasonably fit the underlying assumptions in EEG, ERP, magnetoencephalography (MEG), and, more recently, in positron emission tomography (PET), single photon emission computed tomography (SPECT), and functional magnetic resonance imaging (fMRI). Extraction of relevant information on brain activity from measured electrical signals is affected by various artifacts due to volume conduction through cerebrospinal fluid, skull, and scalp, as well as generated by experimental imperfections. ICA can effectively remove artifacts and separate sources of the brain signals on the basis of minimal statistical suppositions on their underlying distributions.

The EEG and averaged event-related potentials (ERPs) are noninvasive measuring techniques for brain electrical activity recorded as variations of potential between points over the human scalp. The task in processing the EEG and ERP data is basically twofold: to perform *source identification* and *source localization*. Actually, the problem of determining the brain electrical sources from potential patterns recorded on the scalp surface is mathematically undetermined. The linear stationary noise-free ICA algorithms yield statistical estimations of the latent sources of activity from highly correlated EEG signals, irrespectively to their physical location and/or configuration. ICA is therefore a theoretically plausible approach to reliably solve the *source identification* problem in EEG data processing. Moreover, nonstationarities in EEG and behavioral states can also be tracked by ICA via changes in amount of residual correlation between ICA estimates. Since ICA may solve the identifiability of brain activity sources, it may be conceived as a preprocessing step prior to source localization [5].

ERPs are time series of voltages from the ongoing EEG, which are time-locked to a set of similar experimental events. ERP recordings are generally averaged prior to analyze aiming to increase their signal-to-noise ratio (S/N) relative to other non-time and phase-locked EEG activity, and non-neural artifacts. Initially, the related work carried out in EEG and ERP was distinct, in the sense that the EEG community primarily analyzed EEG data in the frequency domain using measures of power in standardized frequency bands, while the ERP community generally ignored the interaction between the ERPs and the ongoing EEG. Recent results [6] have shown by means of ICA that many features of an evoked response may actually be produced by event-related changes in the autocorrelation and cross-correlation structure of the ongoing EEG processes, each reflecting synchronous activity occurring continuously in some brain regions, or by small perturbations in their dynamics. It comes out that applying ICA to EEG/ERP data may constitute a potential source of information on mechanisms of neural synchronization as well.

The paper is structured as follows. In Section 2 the generative stationary linear noiseless ICA model is presented, the underlying assumptions and limitations are discussed, and the ICA estimation principles and the algorithm classes implementing ICA models are briefly overviewed. Section 3 summarizes the particularities of applying ICA to EEG/ERP data processing. Data acquisition and processing are described in Section 4. The results obtained by running four distinct ICA algorithms, including algorithm validation by artificially generated non-Gaussian signals are presented and discussed in Section 5, while conclusions are drawn and the perspectives of ICA are outlined in Section 6.

2. ICA MODEL

In data analysis, a frequent task is to find an adequate representation of multivariate data to facilitate subsequent processing and interpretation. Moreover, the transformed variables are expected to be the underlying *components* that optimally describe the intrinsic data structure and give deeper insight on the physical causes involved in the process of data generation. Linear transforms are often envisaged to accomplish this goal due to their computational and conceptual simplicity. Such a linear transform is ICA, which has emerged as a useful extension of the principal component analysis (PCA) and developed in context with BSS.

2.1. PCA AND ICA

PCA and ICA have major applications in exploratory data analysis (EDA), such as signal characterization, optimal feature extraction, and data compression, as well as providing bases of subspace classifiers in pattern recognition. Both PCA

and ICA are approaches to unsupervised learning (self-organization), shearing some common features. First, they are aiming at building generative models that are likely to have produced the observed data. In a neuromorphic approach, the model parameters are treated as the network weights trained in an unsupervised learning scheme. And secondly, they perform information preservation and redundancy reduction. Redundancy in the sensory input contains structural information about the environment. Completely non-redundant stimuli are indistinguishable from noise and the percept of structure is driven by the dependencies [7].

PCA is a way of encoding second-order dependencies in the data by rotating the orthogonal axes to correspond to the directions of maximum covariance (Fig. 1a,b). PCA decorrelates the input data, but does not address the high-order dependencies. *Decorrelation* simply means that variables cannot be predicted from each other using a *linear* predictor. There can still exist *nonlinear* dependencies between them. As for instance edges, defined by phase alignment at multiple spatial scales, are an example of high-order dependency in an image, like elements of shape and curvature are. Second-order statistics capture the amplitude spectrum of images but not the phase [8]. Coding mechanisms that are sensitive to phase are important for organizing a perceptual system [9].

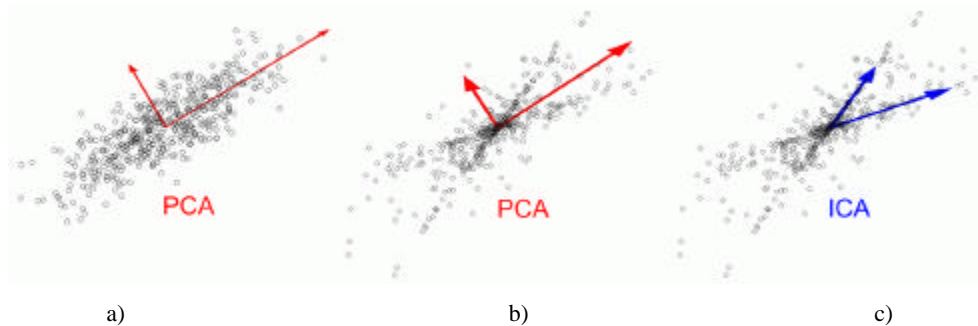


Fig. 1 – a) Gaussian distribution and PCA; b) non-Gaussian distribution and PCA; c) non-Gaussian distribution and ICA.

ICA is formulated as a generative linear latent variables model. ICA showed up as a generalization of PCA that separates the high-order dependencies in the input, in addition to second-order dependencies used up by PCA [10]. ICA does not constraint the axes to be orthogonal as PCA, rather it attempts to place them in the directions of maximal statistical dependencies in the data (Fig. 1c). Each estimate in ICA attempts to encode a portion of the dependencies in the input, so that the dependencies are removed from between the output components. The projection of the distributions on the ICA axes would have less overlap and the output distribution of the estimates would be kurtotic. So ICA, contrarily to other statistical methods searching for underlying factors in multivariate data, looks for components that are simultaneously *independent* and *non-Gaussian*.

In many cases the goal is to determine a subset of independent components (ICs) containing as much *interesting* structural information as possible. The non-Gaussianity of a distribution can be perceived as a measure of its *interestingness* [11–13]. Why should be a *sparse*, or, more general, a super-Gaussian distribution interesting? Friston [14] gives a simple and extremely compelling answer: because measurements of biological systems receive contributions from many sources (e.g., dipoles generated by neural activity), the observations usually represent a roughly linear mixture of interesting events. By the Central Limit Theorem in Statistics, this mixture conforms to a Gaussian distribution (Fig. 2). But the Gaussian is the most unstructured and non-committal distribution given its mean and variance. It follows that mixtures themselves are rather uninteresting and the only interesting things ought to be non-Gaussian (assuming that Gaussian distributions arise from mixing only). This perspective motivates why ICA is so pertinent to biological time series and compelling to data analysis in imaging neuroscience. Therefore, the ICA problem is solved on the basis of optimizing certain measures of “interestingness” (i.e., departure from Gaussianity), which leads to a numerical optimization problem [15, 16].

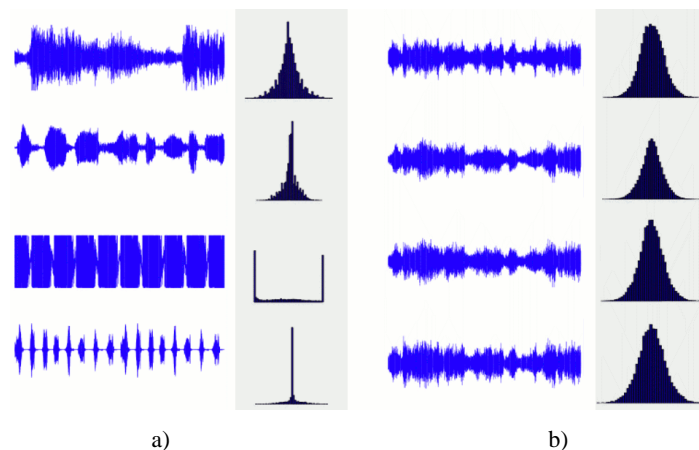


Fig. 2 – a) Super- and sub-Gaussian signal sources and their histograms;
b) their quasi-Gaussian linear mixtures and histograms.

ICA is also related to recent theories on the visual cortex, which assume that consecutive processing steps lead to a progressive reduction in the redundancy of the representation [17]. It is also related to work on sparse and low entropy coding [18]. When ICA is used to extract features, the principle of maximum non-Gaussianity reflects the connection to *sparse coding* that has been used in neuroscientific theories of feature extraction [19]. The idea in sparse coding is to represent data with a minimum of *active* components *active* at the same time. It turns out that this is equivalent, in some situations, to finding components that are

maximally non-Gaussian. The projection pursuit (PP) and sparse coding connections are related to a fundamental result stating that ICA gives a linear representation that is *as structured as possible*. This statement can be given a rigorous meaning by information-theoretic concepts [1] and shows that the ICs are in many ways easier to process than the original random variables (e.g., the ICs are easier to code or compress).

2.2. THE STANDARD ICA MODEL

Our linear stationary model considered hereafter assumes that $\mathbf{x}(t)$, $\mathbf{n}(t) \in \mathfrak{R}^N$ and $\mathbf{s}(t) \in \mathfrak{R}^M$ are three random (column) vectors for any sample index $t = 1, 2, \dots, T$ with zero mean and finite covariance, with the components of $\mathbf{s}(t)$ being statistically independent and at most one Gaussian, and \mathbf{A} a rectangular constant full column rank $N \times M$ matrix with at least as many rows as columns ($N \geq M$):

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) = \sum_{i=1}^M s_i(t) \mathbf{a}_i + \mathbf{n}(t). \quad (1)$$

Mixing is supposed to be instantaneous, so there is no time delay between the (latent) source variable $s_i(t)$ mixing into an observable (data) variable $x_j(t)$. The noise $\mathbf{n}(t)$ is assumed not to be correlated with the data. Then the ICA problem can be formulated as follows: given T realizations of $\mathbf{x}(t)$, estimate both the mixing matrix \mathbf{A} and the corresponding realizations of $\mathbf{s}(t)$ (Fig. 3). In BSS the task is somewhat relaxed, in the sense of searching to find the waveforms $\{s_i(t)\}$ of the sources knowing only the mixtures $\{x_j(t)\}$.

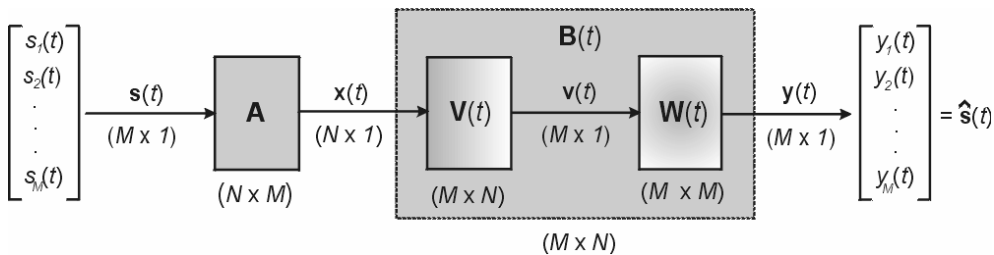


Fig. 3 – Overall processing of the (unknown) latent signal sources $\mathbf{s}(t)$ in the basic ICA model: linear mixing by the (unknown) constant mixing matrix \mathbf{A} , and separating by the demixing matrix $\mathbf{B}(t)$.

The model considered here is free of noise.

There are several limitations to solving the basic ICA model. If no suppositions are made about the noise (which is often the practical case), it cannot be introduced in the model but, eventually included in the signals, hence the noise-free ICA model may be expressed as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) = \sum_{i=1}^M \mathbf{a}_i s_i(t), \quad (2)$$

where \mathbf{a}_i , $i = 1, 2, \dots, M$ are the columns of \mathbf{A} . Next limitation refers to the size of the vectors involved. The size of $\mathbf{s}(t)$ (usually unknown) should not be greater than the size of data $\mathbf{x}(t)$, otherwise the problem becomes under-determined. If the size of $\mathbf{x}(t)$ is greater than the size of $\mathbf{s}(t)$ (e.g. there are more sensors than sources), the problem is over-determined and the extra data can be used for reducing the noise. This is accomplished by projecting the input data $\mathbf{x}(t)$ into its M -dimensional signal subspace using for example PCA whitening [20]. In any case, each source signal $s_i(t)$, $i = 1, 2, \dots, M$ is assumed here to be a stationary zero-mean stochastic process and only one of them is allowed to have a Gaussian distribution. Then the source separation consists in updating a $M \times N$ demixing matrix $\mathbf{B}(t)$, without resorting to any information about the spatial mixing matrix \mathbf{A} , so that the vector $\mathbf{y}(t) = \mathbf{B}(t)\mathbf{x}(t)$ becomes an estimate $\mathbf{y}(t) = \hat{\mathbf{s}}(t)$ of the original independent source signals $\mathbf{s}(t)$. It is expected that after a finite number of iterations, the separating matrix $\mathbf{B}(t)$ to converge to a fixed value $\mathbf{B}(t) \rightarrow \mathbf{B}$. Then the estimates of the ICA basis functions (vectors) $\{\mathbf{a}_i\}$ are the columns of the pseudo-inverse $\hat{\mathbf{A}} = \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1}$, which amounts to $\hat{\mathbf{A}} = \mathbf{B}^{-1}$ if \mathbf{B} is a square matrix (e.g. $M = N$) and the estimated independent components of the linear stationary ICA model are given by:

$$\hat{\mathbf{s}}(t) = \mathbf{y}(t) = \mathbf{B}\mathbf{x}(t). \quad (3)$$

Finally, ICA decomposition is unique up to sign, scaling and permutation. Various ranking procedures can be envisaged, though we consider two of them particularly significant in the spirit of the statistical significance of ICA. The first criterion is to rank the estimated ICs in decreasing order of their departure from Gaussianity (e.g., on the basis of kurtosis, negentropy, mutual information, ...). The second criterion used hereafter, which requires the computation of the (pseudo)inverse of the demixing matrix, \mathbf{B} , is ranking the estimated ICs in decreasing order of their percentage of variance accounted for in the data space (i.e., in their back-projections), which gives a feeling of the latent source energy.

In practice, applying the ICA model amounts to (i) determining an *estimation principle* for measuring the departure from Gaussianity of the estimated components and (ii) to selecting an adequate *algorithm* to run the computations

required by the estimation principle. The *statistical properties* (e.g. consistency, asymptotic variance, robustness) of the ICA method depend on the choice of the *objective* or *contrast* function (a real function of a probability distribution), whereas the *algorithmic properties* (e.g. convergence speed, memory requirements, numerical stability) of the ICA method depend on the optimization algorithm.

2.3. ESTIMATION PRINCIPLES FOR ICA

A self-consistent treatment of independence relies on information theory, which entails deriving the criterion for statistical independence from the statistical properties of data [21]. *Entropy* is such a criterion based on the amount of information contained in some occurrences of a random variable. In the case of a multidimensional *continuous* random variable \mathbf{x} distributed with the probability density function (PDF) $f_{\mathbf{x}}(\mathbf{x})$, the *differential entropy* is defined as:

$$H(\mathbf{x}) = -\int f_{\mathbf{x}}(\mathbf{u}) \log f_{\mathbf{x}}(\mathbf{u}) \, d\mathbf{u}. \quad (4)$$

If the random vector \mathbf{x} has d independent components $x_i, i = 1, 2, \dots, d$, then $f_{\mathbf{x}}(\mathbf{x})$ can be factorized such as:

$$f_{\mathbf{x}}(\mathbf{x}) = \prod_{i=1}^d f_i(u_i) \Leftrightarrow H(\mathbf{x}) = \sum_{i=1}^d H(x_i) \quad (5)$$

The components in the last summation, $H(x_i) = -\int f_{x_i}(u_i) \log f_{x_i}(u_i) \, du_i$, are the *marginal* or *pixel entropies*. In general, the entropy of a random vector \mathbf{x} is such as

$$H(\mathbf{x}) \leq \sum_{i=1}^d H(x_i).$$

Differential entropy is invariant to orthogonal transforms and it is upper bound, but is no longer invariant to invertible transforms as *entropy* is for discrete random processes. Therefore, two other concepts are employed as contrast functions for ICA that are endowed with the invariance property to invertible transforms, namely *negentropy* and *mutual information*.

For a multivariate continuous random variable \mathbf{x} with the density $f_{\mathbf{x}}(\mathbf{x})$, to which is associated a Gaussian variable \mathbf{x}_G with the same covariance matrix like \mathbf{x} , the *negentropy* is defined in terms of differential entropy:

$$J(\mathbf{x}) = H(\mathbf{x}_G) - H(\mathbf{x}) = \int f_{\mathbf{x}}(\mathbf{u}) \log \frac{f_{\mathbf{x}}(\mathbf{u})}{f_{\mathbf{x}_G}(\mathbf{u})} \, d\mathbf{u} = K(\mathbf{x} | \mathbf{x}_G). \quad (6)$$

It can be interpreted as a sort of *distance* from Gaussianity of the random variable \mathbf{x} , which is expressed by the Kullback-Leibler (KL) divergence. Though not really

a distance since it is not symmetric, the KL divergence behaves as a statistical measure of “distance” between two distributions. It is always nonnegative and takes the value 0 *iff* the distributions are identical. Hence negentropy is always non-negative, reaches its minimum for a Gaussian random variable, and it is invariant to linear invertible transforms.

The *mutual information* (MI) is also related with (differential) entropy. For the general case of N (scalar) random variables x_i , $i = 1, 2, \dots, N$ it is given by:

$$I(x_1, x_2, \dots, x_N) = \sum_i H(x_i) - H(\mathbf{x}) = \int f_{\mathbf{x}}(\mathbf{x}) \log \frac{f_{\mathbf{x}}(\mathbf{x})}{\prod_i f_{x_i}(x_i)} d\mathbf{x} = K\left(\mathbf{x} \mid \prod_i x_i\right) \quad (7)$$

Equation (6) shows that the MI is symmetric, zero *iff* the factorization of the joint density $f_{\mathbf{x}}(\mathbf{x})$ holds (e.g. the components are independent), and it is strictly positive otherwise. Comparing the form of negentropy in (5) with MI in (6), it comes out that if a Gaussian multivariate is a reasonable approximation to the product of the marginal densities, then negentropy is a means to estimate the MI and, implicitly, a measure of independence. Unlike the autocorrelation function, the MI also takes into account nonlinear correlations. An MI at zero means that the joint distribution of signals values holds the exact quantity of information as the signals considered separately, that is, the signals are independent.

2.4. ICA ALGORITHMS

Nonquadratic functions are generally involved by the estimation methods that require computationally demanding numerical algorithms. The current algorithms for ICA can loosely be classified in two categories [22]. One category contains *adaptive algorithms* generally based on stochastic gradient methods and implemented in neural networks [23, 24]. Adaptive algorithms may also be based either on optimization of cumulant-based contrast functions [25], or on “estimating equations” involving nonlinear distortions of the output $y(t)$ [26]. The neural adaptive algorithms exhibit slow convergence and their convergence heavily depends on the correct choice of the learning rate parameters. The second category relies on *batch computation* optimizing some relevant criterion functions. Generally, they imply complex matrix or tensorial operations. Neuromorphic block technique algorithms based on 2nd- and 4th-order cumulants [27], as well as (quasi)-likelihood approaches [28] were also proposed.

An alternative classification is based on the nature of the ICA implementations: *deterministic algorithms* like JADE [29], and *stochastic algorithms*, such as BS Infomax [1], FastICA [30], or EGLD [31]. We briefly present hereafter these algorithms employed throughout the experiments for both artificially generated and real EEG time series.

The basic idea of the *infomax* principle is to match the slope of the nonlinear transfer function of the elementary processing unit (e.g., neuron) in a network with the input PDF. The BS infomax nonlinear information maximization algorithm performs on-line stochastic gradient ascent in the MI between outputs and inputs of a neural-like network. Maximizing the information transfer in a nonlinear NN minimizes the MI among the outputs when optimization is done over both the synaptic weights *and* the nonlinear transfer function. By minimizing the MI between its outputs, the network factorizes the input into ICs.

EGLD algorithm for ICA is based on an ML approach in which the distributions of the source signals are iteratively estimated using the EGLD. The major benefit of the EGLD ML algorithm is that it also takes into account the skewness of the distributions. The score function of the EGLD is used as an ICA contrast function, which subsequently is subject to maximization.

The fixed-point algorithms are searching for the ICA solution by minimizing the MI among the estimated components. Since minimizing the MI is equivalent to negentropy maximization of individual outputs, the main task is to consider some valid approximations for negentropy [32] to make the computations tractable. Some remarkable features of the FastICA algorithm stem from its simplicity in implementation, high accuracy, fast convergence to local extrema of the contrast function, and, most important for the EEG time courses, simultaneously dealing with both sub-Gaussian and super-Gaussian sources.

The joint diagonalization of a set of square matrices consists in finding the orthonormal change of basis which makes the matrices as diagonal as possible. When all the matrices in the set commute, this can be achieved exactly. When this is not the case, it is always possible to optimize a joint diagonalization criterion. This defines an *approximate joint diagonalization*. When the matrices in the set are “almost exactly jointly diagonalizable”, this approach also defines something like the “average eigen-spaces” of the matrix set. The original implementation of the JADE algorithm was slightly modified in the sense of optimizing it to handle real-valued signals more efficiently. The additive noise term in the ICA model was deliberately skipped assuming a “relaxed” statistical independence of source signals. Consequently, the cumulant tensor could not be accurately retrieved by a set of eigen-matrices, which lead to jointly diagonalization of a set of cumulant matrices adequately selected. The set of cumulant matrices was further reduced by invoking the symmetry of cumulants in the case of real signals.

All four algorithms under test were designed to come up with the full separating matrix \mathbf{B} having the rows in such a way that the columns of its (pseudo)inverse were in decreasing order of the Euclidian norm. Consequently, the first recovered ICs corresponded to the most energetic source signals. The time courses of activation, which are the rows of the activation matrix $\mathbf{y}(t)$, $t = 1, 2, \dots, T$, were all set to be RMS-positive.

3. METHODS

3.1. SPECIFIC ASSUMPTIONS

Some supplementary assumptions to the basic ICA suppositions discussed in section 2.1 are necessary in order to formulate the ICA model for EEG/ERP time series. First, it is assumed that the brain subserves and optimizes *active* cognition. By all means, the interpretation of functional brain imaging data requires some suppositions on processing in the working brain that may not be entirely realistic and which preclude canonical methods of data analysis and experimental design. Next, the macroscopic brain signals (BOLD, EEG, MEG, ...) are considered to be largely produced by and reflect the dynamics of *top-down* brain processes rather than *bottom-up* sensory processes. If sources have no systematic overlap in time and/or space then they can be considered *independent* in time and/or space [33]. And finally, the noise is generally assumed *non-Gaussian* and shows up as one or more distinct ICs.

In the case of EEG/ERP time series, the rows of the input matrix, $\mathbf{x}(t)$, $t = 1, 2, \dots, T$, are the electromagnetic signals recorded at different electrodes and the columns are the measurements recorded at different time points t (frames). The temporal ICA method (tICA) estimates the demixing matrix \mathbf{B} that linearly decomposes the multichannel recordings into a sum of temporally independent and spatially fixed components $\mathbf{y}(t) = \mathbf{B} \mathbf{x}(t)$. The rows of the *activation* matrix, $\mathbf{y}(t)$, $t = 1, 2, \dots, T$, are the time courses of activation of the ICA components, and the columns of the pseudo-inverse matrix $\hat{\mathbf{A}} = \mathbf{B}^T (\mathbf{B} \mathbf{B}^T)^{-1}$, or $\hat{\mathbf{A}} = \mathbf{B}^{-1}$ for the square model, are the ICA basis vectors, also called *scalp maps*, which give the relative projection strengths of the respective components at each of the scalp sensors. These scalp weights indicate the scalp topography of each component and provide evidence for the physiological origins of each component. The projection of the k -th estimated component onto the original data $\mathbf{x}(t)$ is given by the outer product of the k -th row of the component activation matrix $\mathbf{y}(t)$, with the k -th column of the estimated mixing matrix $\hat{\mathbf{A}}$. In this framework, each estimated IC consists of a *time course* (activation matrix row time series) and a *scalp map* (pseudo-inverse demixing matrix columns).

We basically assumed that the multichannel EEG recordings are mixtures of underlying brain and artifactual signals. Source activations are temporally independent of one another across the input data as far as we have reasons to believe that the complexity of EEG dynamics can be modeled as a set of a

relatively small number of independent brain processes. Besides, the sources of eye and muscle activity, line noise, and cardiac signals are not generally time-locked to the sources of EEG activity, which is conceived to reflect synaptic activity of cortical neurons.

For EEG signals volume conduction is assumed to be instantaneous and summation of currents at the scalp sensors essentially linear; both assumptions are realistic for currents carried to the scalp electrodes at EEG frequencies [34]. Although the ICA model for EEG ignores the known variable synchronization of separate EEG generators by common subcortical or corticocortical influence [35], it is still adequate for identifying concurrent signal sources that are either situated too close together, or are too widely distributed such as to be separated by current localization techniques [36].

The common supposition that source distributions are super-Gaussian is compatible with the physiologically plausible assumption that an averaged ERP is composed of one or more overlapping series of relatively brief activations within spatially fixed brain regions performing separable stages of stimulus information processing. Nonetheless, sub-Gaussian ICs were detected in EEG data [37] including line noise, sensor noise and low frequency activity. In practice, however, sub-Gaussian components appear rarely in ERPs or in spontaneous EEG. The super-Gaussian statistics of ICs of ERP data may indicate that brain information processing is dominated by spatially sparse, intermittently synchronous brain structures [38].

It is also assumed that the sensors exceeds or at least equals the number of signal sources. In most ICA approaches, the number of sources is assumed equal to the number of sensors (i.e., square ICA). This is questionable since we have no prior information on the exact number of statistically independent signal sources contributing to the scalp EEG.

3.2. LIMITATIONS IN APPLYING ICA OF BIOMEDICAL TIME SERIES

ICA is currently used in two complementary ways to decompose an image sequence into a set of images and a corresponding set of time-varying image amplitudes: *spatial* ICA (sICA) [39], which finds a set of mutually independent component (IC) images and a corresponding dual set of unconstrained time courses, and *temporal* ICA (tICA) [1], which finds a set of IC time courses and a corresponding (dual) set of unconstrained images. ICA can find statistically independent signals by making use of the extra degrees of freedom implicit in the unconstrained dual signals, even if the underlying sources are not statistically independent. Therefore, ICA may yield independent signals, which are not the underlying sources.

The assumption of temporal independence used by tICA of EEG/ERP data cannot be satisfied when the training data is too small in the case of neural adaptive ICA algorithms, or when separate topographically discernable phenomena nearly co-occur in data. The separation reliability of ICA is largely influenced by the assumption that the sources of artifacts and cerebral activity remain spatially stationary in time, which is not always true all over the trials.

The inclusion of a noise model allows preventing data overfit in a systematic way. However, the square ICA does not include a noise model, instead the (latent) sources are assumed to be completely characterized by the data and the estimation of the mixing matrix. This precludes: (i) the assessment of statistical significance of the source estimates within the framework of a null-hypothesis testing, (ii) any threshold technique is superfluous being devoid of statistical meaning.

The number of brain sources of activity and the number of artifacts are generally not known, whereas the number of sensors is limited. It is likely that the number of small briefly active EEG sources and many small artifacts associated with a large number of electrodes to exceed the separation power of the ICA model. Model selection refers to determining the size as well as the complexity of the data model [40], which must be set outside ICA. Underestimation of the dimensionality discards valuable information and results in suboptimal signal extraction. Overfitting a noise-free generative model to noisy observations results in a large number of spurious components due to unconstrained estimation and factorization.

ICA does not discriminate between signals of interest (task-related, transiently task-related, and function-related) and signals not of interest (physiological rhythms, various artifacts, and noise). The estimated IC's are not endowed with any intrinsic meaning, which obscures the interpretation of the results rendering them domain specific. However, if some of the ICs are identified as artifactual activity, it is possible to "clean" the raw data having these components removed by zeroing the appropriate rows of the estimated activation matrix.

3.3. VALIDATION OF ICA

There are means to assess the separation quality performed by various ICA models. When reliable regression channels are available, such as electrooculographic (EOG) recordings, comparing artifact removal performed by ICA applied to EEG/ERP data with other techniques like PCA and multiple-lag regression can validate the ICA results [41]. Simultaneous recording and comparing different types of signals like concurrent EEG and fMRI, which respectively have good spatial (fMRI) and temporal (EEG) resolution, and checking the correlations can also validate ICA decompositions. However,

convergent physiological and behavioral evidence is needed since temporal and spatial independence of active brain regions is approximate only, so that the nature and functional significance of the estimated ICs must be alternatively assessed outside ICA. Another approach refers to simultaneously record and analyze the (presumable) correlation of concurrent types of signal, such as EEG and fMRI, which have good temporal (EEG) and spatial resolution (fMRI), respectively [42].

Artificially generated EEG-like data that mimic conditions under which ICA is likely to fail, such as simulated EEG recordings generated from a head model and dipole sources that include intrinsic noise and sensor noise, when the number of sources is larger than the number of sensors [43], can also be employed to estimate the separation power of ICA. Numerical simulations confirmed that the ICA method can accurately identify the time courses of activation and the scalp topographies of relatively large and temporally independent sources, even in the presence of a large number of low-level and temporally-independent source activities [44]. In the case of artificially generated signal sources, the accuracy of separating the independent components of an ICA algorithm can be estimated by means of some quantitative indexes as far as both the mixing and the estimated separation matrices are available. One index that we used is defined as *signal-to-interference ratio*:

$$SIR = -\frac{1}{N} \sum_{i=1}^N 10 \log_{10} \frac{\max(Q_i)^2}{Q_i^T Q_i - \max(Q_i)^2}, \quad (8)$$

where $\mathbf{Q} = \mathbf{BA}$ is the overall transforming matrix of the source components, Q_i is the i -th column of \mathbf{Q} , $\max(Q_i)$ is the maximum element of Q_i , and N is the number of both the signal sources and the recording channels. The higher SIR is, the better the separation performance of the algorithm. A second employed index, CTE , was the distance between the overall transforming matrix \mathbf{Q} and an ideal permutation matrix, which is interpreted as the *cross-talk error* [45]:

$$CTE = \sum_{i=1}^N \left(\sum_{j=1}^N \frac{|Q_{ij}|}{\max|Q_i|} - 1 \right) + \sum_{j=1}^N \left(\sum_{i=1}^N \frac{|Q_{ij}|}{\max|Q_j|} - 1 \right). \quad (9)$$

Above, Q_{ij} is the ij -th element of \mathbf{Q} , $\max|Q_i|$ is the maximum absolute valued element of the row i in \mathbf{Q} , and $\max|Q_j|$ is the maximum absolute valued element of the column j in \mathbf{Q} . A permutation matrix is defined so that on each of its rows and columns, only one of the elements equals to unity while all the other elements are zero. It means that CTE attains its minimum value zero for an exact permutation matrix (e.g. the decomposition is perfect) and goes positively higher

the more \mathbf{Q} deviates from a permutation matrix (e.g. a decomposition of lower accuracy).

Some statistical assessment of the discovered projections when applying exploratory methods like ICA can be performed by means of resampling-based techniques [46], such as on the basis of variance estimations which were found strongly correlated with the separation power [47]. Reliability estimation can be used to select an appropriate ICA-model, to enhance significantly the separation performance, and, most important, to discover the components that may have a physical interpretation. However, methods for rigorously testing the statistical reliability of ICA component time courses and areas of activation still need to be developed.

4. EXPERIMENTAL

Our working data due to K. Nayak [48] were multiple electrode time series EEG recordings collected from a study on temporal behavior aiming to differentiate healthy individuals from patients probably possessing Alzheimer or other type of dementia. The standard international 10-20 system (Fig. 4) was employed for EEG electrode placement on the scalp, which produced 19 simultaneous EEG measurements. The electric signals were collected at a sampling rate of 128 Hz during epochs of 8 seconds each, resulting in time series of 1024 data points. The measured analog signals were converted by an 8 bit ADC and band-pass filtered (0.1 to 30 Hz, 12 dB/octave roll-off). The EEG model consisted of a number of statistically independent but spatially fixed potential-generating systems, which may either be spatially restricted or widely distributed and eventually non-systematically overlapping.

We comparatively tested four different ICA algorithms, two of which were of on-line adaptive neural-type and two of batch-type. The neural algorithms were the Bell and Sejnowski (BS) infomax principle-based algorithm [1] with all subsequent refinements embedded in its extended version as in [49], and an ML approach involving the extended generalized lambda distribution (EGLD) as introduced by Eriksson and coworkers [31]. The algorithms of batch-type were the fixed-point fast ICA (FastICA) derived by Hyvärinen [50], and the JADE algorithm proposed by Cardoso and Souloumiac [29] on the basis of joint approximate diagonalization of eigen-matrices. The parameters of the algorithms and the nonlinearities involved to capture the higher-order dependencies were first tuned for optimum separation using artificially generated signals having both sub-Gaussian and super-Gaussian distributions [51]. Some codes in MATLAB [52] employed in topographic plotting and algorithm testing were adapted from the "ICA Toolbox for Psychophysiological Research" [53].

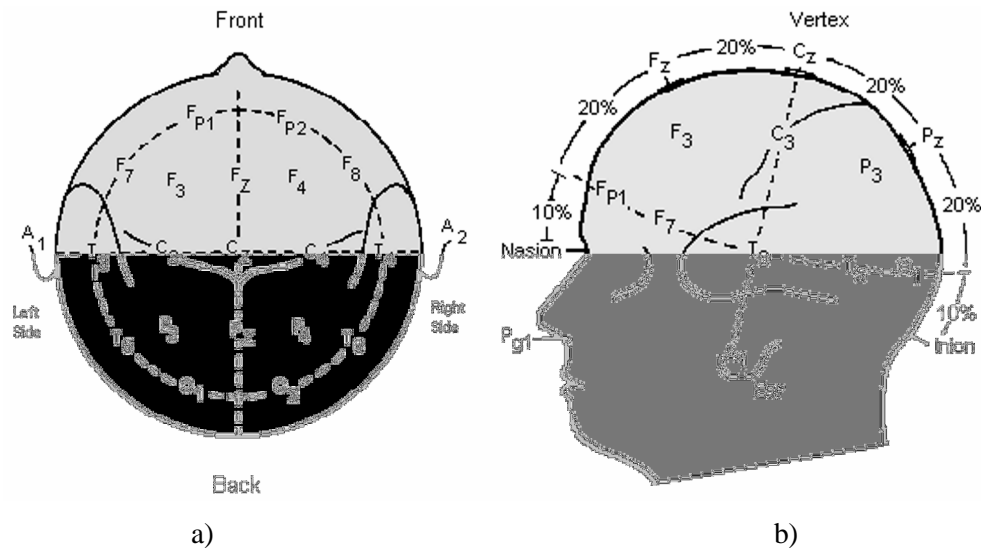


Fig. 4 – The standard international 10-20 system for 19 electrode placement:
 a) channel indexing – both hemispheres; b) odd indexes – left hemisphere.

5. RESULTS AND DISCUSSION

5.1. COMPARING ICA ALGORITHMS

As a rule of thumb, to get reliable decompositions from applying ICA it is advisable that the number of time points, T , to exceed the square of the number of available channels, M . We concatenated two single-trials of EEG signals corresponding to a healthy young patient with eyes opened and closed, respectively, in order to fully comply with this requirement and to ensure a relatively high Nyquist critical frequency for spectral analysis.

The estimated ICs (the time courses of activation *and* the scalp maps) were obtained using the BS infomax, EGLD ML, FastICA, and JADE algorithms, respectively. It was noticeable a good similarity of the first activation waveforms for all algorithms, yet a poor match for higher indexes (Fig. 5). The corresponding scalp maps of the first six estimated ICs by all four ICA algorithms are presented in Fig. 6, which support the relative high similarity displayed by the time courses (TCs) of activations of the first four ICA components.

The percentage of variance accounted for in the full set of the EEG recordings as in Fig. 7a by successive sums of ICA component activations in decreasing order of mean back-projected variance is presented in Fig. 7b. All algorithms display similar percentages of reconstruction, which suggest that at least the sequence of the most energetic components separated by any ICA algorithms

should be quite close. We limited our comparison to 12 estimated ICs since they capture more than 85% of variance in the raw data. Besides, it is very likely that the number of activity sources exceeds the number of channels, so that the least energetic components are expected to represent (linear) combinations of a certain number of independent components.

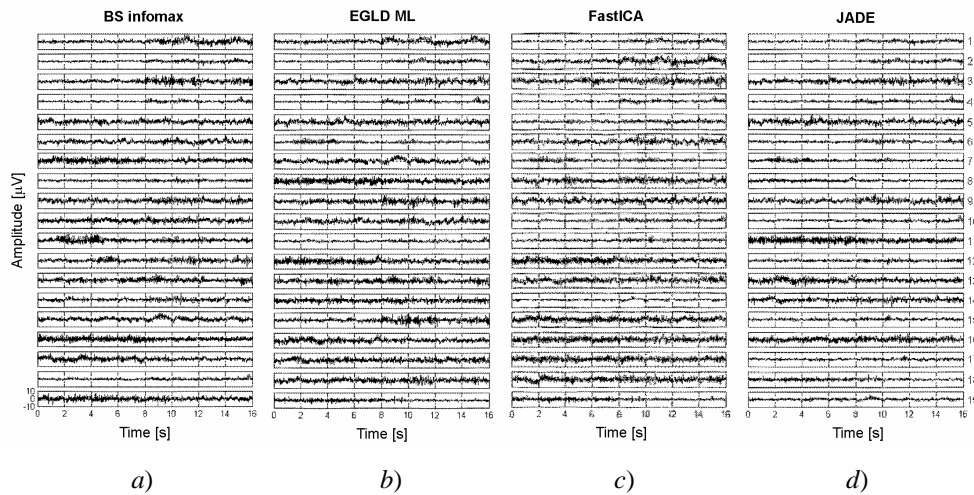


Fig. 5 – Time courses of activation separated by tICA algorithms for single-trial 2-epoch EEG data: a) BS extended infomax; b) EGLD ML; c) FastICA; d) JADE. The time courses of activations are displayed in decreasing order of their mean projected variance (e. g., most energetic components are the first).

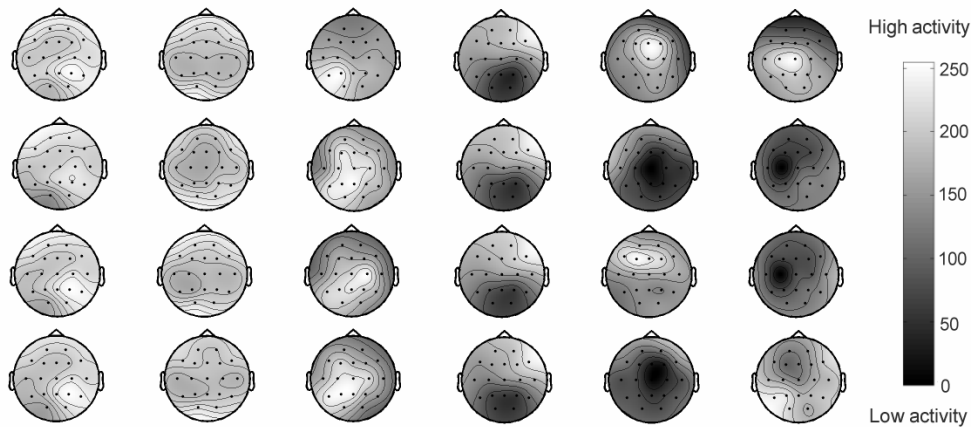


Fig. 6 – Topographic maps of the first most energetic six ICs separated by (from top to bottom): BS infomax, EGLD ML, FastICA, and JADE algorithm.

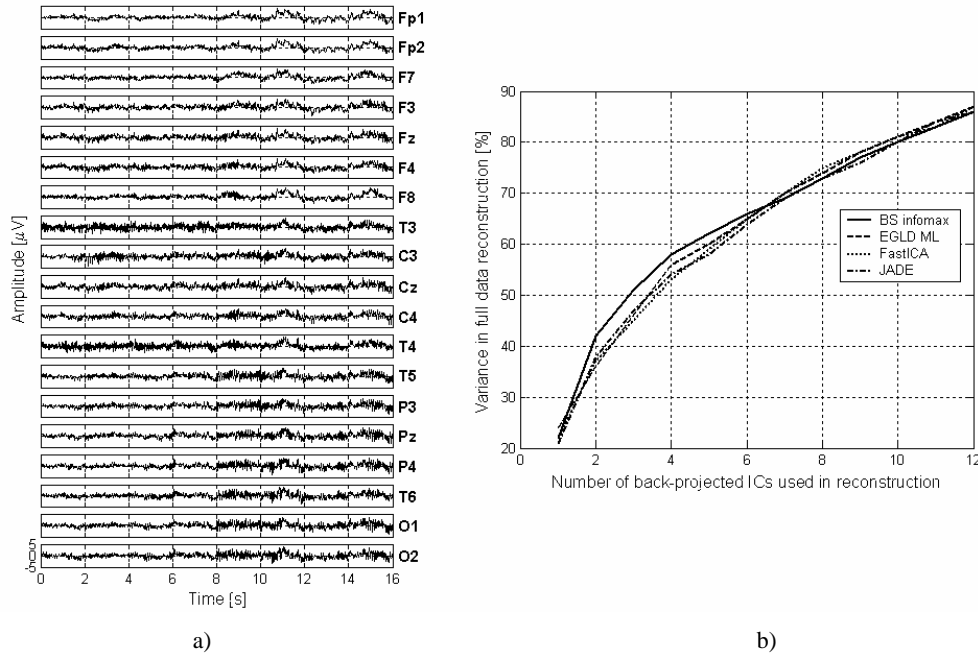


Fig. 7 – a) Full set of normalized raw EEG recordings; b) percentage of variance accounted for in the EEG recordings by the ICA component activations up to the sum of the first 12 most energetic components.

Complementarily, we analyzed also the percentage of variance accounted for in a *single* channel recording by successive sums of ICA component activations in decreasing order of mean projected variance. First, based on the similarity of the 4th waveform (Fig. 5) and scalp map (Fig.6) of the ICA decomposition by all four algorithms, we used it to investigate the percentage of variance accounted for in reconstruction of each channel (Fig. 8a). Then we selected the channel F8 on the basis of maximum percentage of variance accounted for by the 4th ICA component activation, which was the same for all algorithms, and used it in estimating the variance of channel reconstruction by successive sums of ICA component activations. The results show a relatively good match of one single channel reconstruction among all four algorithms (Fig. 8b), with a particular better fit between BS infomax and FastICA algorithm.

Traditionally, the EEG measurements were performed in the frequency domain, so that we compared the first 6 separated components by computing their power spectral density (PSD). The PSD estimators have a greater variety, though it is always agreed to relate the PSD normalization to a particular description of the function normalization. We considered appropriate for EEG signals to define the PSD for zero and discrete positive frequencies only, and its sum over these is the function mean squared amplitude. The PSD's were plotted in the frequency range up to 100 Hz because of two main reasons: (i) data available were already lowpass

filtered and (ii) it *never* makes sense to integrate the PSD of a sampled function outside of the Nyquist interval since, according to the sampling theorem, power there will have been aliased into the Nyquist interval. The `pwelch` function from the MATLAB signal processing toolbox was employed to plot the spectral densities. In Fig. 7a the PSD for all ICA component activations are presented, and in Fig. 7b the PSD's corresponding to the 4th waveform separated by each algorithm under study. The match is quite good among BS infomax, EGLD ML, and FastICA algorithm, yet JADE displays a clear overshoot around 10 Hz. The high frequency content in the range of 40 to 80 Hz is likely to reflect some intense activity of the scalp muscles.

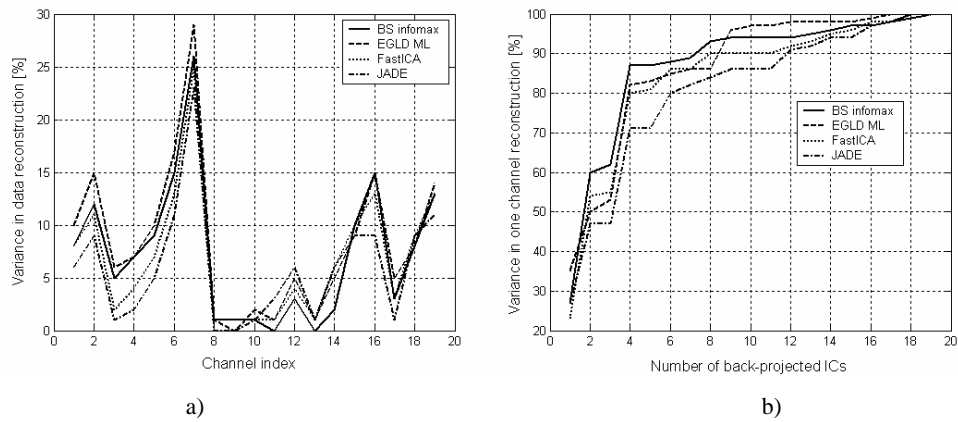


Fig. 8 – a) Percentage of variance accounted for in each recording channel by the 4-th ICA component activation; b) percentage of variance accounted for in one single channel (F8) reconstruction by successive sums of ICA time courses of activation.

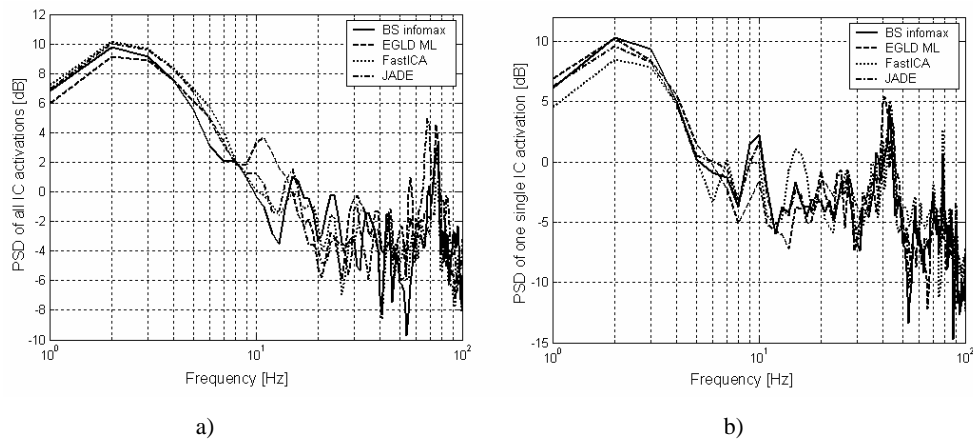


Fig. 9 – Power spectrum density: a) for all ICA component activations; b) for the 4-th ICA waveform corresponding to all algorithms under study.

5.2. REMOVING ARTIFACTS

Typical artifacts that affect to various extents any multichannel EEG recordings consist in eye movements, blinks and saccades, cardiac signals, muscle and line noise activity, as well as mechanical displacements of the measuring equipment. The currently widely spread method to discard artifacts is to reject the EEG segments with artifacts larger than an arbitrarily preset value. In certain cases,

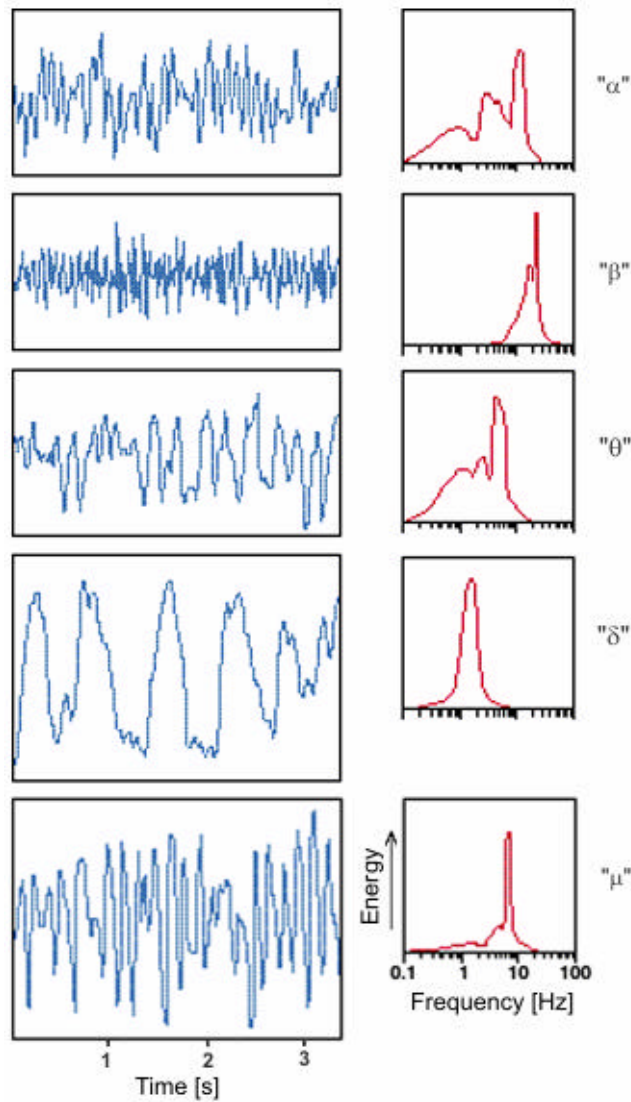


Fig. 10 – The brain waves with histograms.

particularly when limited data are available and/or many artifacts are present, removing a great deal of EEG recordings leads to an unacceptable loss of valuable data. Several methods were proposed for artifact rejection yet preserving the essential information in the EEG data. Frequently encountered methods include a spatio-temporal dipole model for eye artifact removal [54], regression in the time domain [55, 56] or frequency domain [57, 58] by performing simultaneous EEG and electrooculographic (EOG) recordings to derive parameters characterizing the appearance and spread of EOG artifacts in EEG channels. Nevertheless, simple regression in time domain for removing eye movements tends to overcompensate for blink artifacts and may additionally introduce artifacts into EEG data [59]. Though regression in the frequency domain can account for frequency-dependent transfer function differences from EOG to EEG, it is acausal and therefore unsuitable for real-time applications. The regression methods both in time and frequency are heavily depending on having a reliable regression channel and exhibit a common drawback that spread of excitation from eye movements and EEG signals is bidirectional [44]. Consequently, any regression method for artifact removal entails a certain loss of relevant EEG information contained in the EOG data. Besides, many noise sources (including muscle, electrode, and line noise) have no clear reference channels, so that regression methods cannot be used to remove them. Another approach to separating eye artifactual activity from brain signals consists in running PCA, but the separation is not complete, particularly at comparable amplitudes [41].

ICA separates the source components based on the higher-order statistics of their amplitude distributions over time, which implies the differentiation between strictly periodical signals, and regularly and irregularly occurring signals. Many artifacts belong to this last category [60]. The basic assumption in discriminating genuine EEG signals from artifacts is that the time courses of true EEG activity and artifacts are statistically independent, though in some ERPs studies (e.g., in the case of infrequent or painful stimuli) the genuine cerebral and ocular signals can be similarly time-locked to the stimulus. In many cases, the independence can be checked by the known differences in physiological origins of the signals. The independence between two signals is, anyway, a measure of the similarity between their joint amplitude distribution and the product of each (marginal) signal distribution computed throughout the *entire* signal; it seems, therefore, reasonable to assume that any of their strictly local relations during stimulation should not significantly affect their global statistical properties. It is also possible that the observed brain activity arises from more physically separable effective sources than the available number of EEG electrodes. In this case, if the dimensionality is high enough, it is reasonable to expect that the first separated independent components to display clearly the most strongly independent components, while the last components still consist of mixtures of the remaining signals.

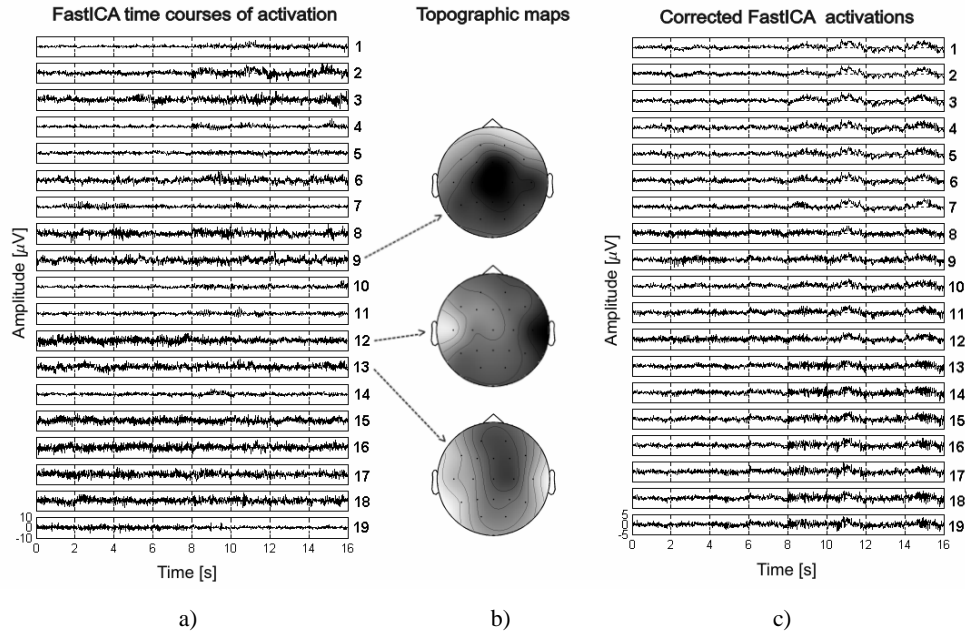


Fig. 11 – Filtering out some artifacts from: a) the raw ICA component activations; b) eye activity (TC 9) and temporal muscle activity (TCs 12 and 13); c) the corrected ICA time courses are decreased with a factor of 2 on average and recasted within the same range as the raw EEG recordings (Fig. 5a).

It seems plausible to conjecture that the artifacts, being clearly independent from the brain activity, to come out among the first estimated components. Once the independent time courses of artifactual sources are discarded, artifact-corrected EEG signals can be derived by projecting the sum of the remained non-artifactual ICA components back onto the scalp, such as $\mathbf{x}_0(t) = \hat{\mathbf{A}}\mathbf{y}_0(t)$, where $\mathbf{y}_0(t)$ is the matrix of the estimated components with the rows corresponding to artifactual activation(s) set to zero. The *back projection* of a component is in the original data space (e.g. recording channels Fz, Cz, etc.) and units (e. g., μV) as in the original data $\mathbf{x}(t)$. Since ICA is a linear decomposition, the projection of a sum of components is the same as the sum of the projections of the individual components, so that the artifact-free corrected patterns of EEG signals can be reliably retrieved.

In practice, the ability of ICA to remove various artifacts heavily relies on deciding which components are artifactual and which are not. It is a good practice to corroborate the information provided by the time courses of activation, such as their power spectral density, with the corresponding topographic scalp maps, which highlight the region(s) where the activity occurs. As for instance, intense prefrontal and frontal activity is likely to be the consequence of eye movements, blinks, or saccades, whereas lateral (temporal) activity is mainly due to muscles. A prominent

50 or 60 Hz component suggests the presence of line noise. It is a common practice to filter it out in the frequency domain, but this technique is undesirable if the research interest lays in high-frequency EEG phenomena. Traces with a broader irregular high frequency spectrum (50–100 Hz) are likely to be generated by high-frequency activity of the scalp muscles. Close to DC slowly varying waveforms are indicating slow head movements or slow mechanical drifts of the measuring equipment.

Signals having strong 7 or 10 Hz components may reflect the *theta* and *alpha* brain waves. Theta waves can accompany feelings of emotional stress and are characterized by moderately low frequencies, whereas alpha waves brought on by unfocusing one's attention have relatively large amplitude and moderate frequencies. Extremely low frequency oscillations (around 1 or 2 Hz) are identified with *delta* brain waves that occur during periods of deep sleep. An intense mental activity is reflected by *beta* waves, which are rapid oscillations with small amplitudes. Physical movements or the intention to move, are captured by the *mu* brain waves, which resemble croque wickets in shape. Based on their histograms, the brain rhythms (Fig. 10) can be wiped off from the original recordings, e.g., by correlation analysis.

The raw EEG data recordings (Fig. 5a) used in this study were filtered by removing some presumable artifacts and the “clean” time series are rendered in Fig. 11.

5.3. RELIABILITY OF ICA DECOMPOSITION

Alternatively, we generated 19 synthetic time series having the statistical properties as close as possible to the genuine EEG recordings and used them to assess all algorithms under study. All simulated signals were normalized to zero-mean and unit variance. Their simulated strengths ranged from 1.000 to 0.649, their absolute correlations were in the interval from 0.0001 to 0.0695, and their kurtosis encompasses values from 0.494 to 1.252. The same data set was run 120 times under identical conditions for all algorithms with the same randomly generated full column rank mixing matrix **A**.

Table 1

The main characteristics of ICA decomposition performed by the algorithms under study

<i>Figures of merit</i>	<i>BS infomax</i>	<i>EGLD ML</i>	<i>FastICA</i>	<i>JADE</i>
Absolute activation correlations	0.015 ± 0.011	0.000	0.000	0.000
Max absolute off-diagonal value in the matrix Q	1.093 ± 0.194	1.319 ± 0.721	1.014 ± 0.040	1.214 ± 0.000

Table 1 (continued)

Mean absolute off-diagonal value in the matrix \mathbf{Q}	0.110 ± 0.033	0.091 ± 0.042	0.071 ± 0.020	0.160 ± 0.000
Absolute correlations between source signals and corresponding ICA estimated activations	0.779 ± 0.082	0.791 ± 0.142	0.913 ± 0.027	0.632 ± 0.317
Signal-to-interference ratio (SIR) [dB]	8.17 ± 1.90	7.06 ± 1.71	6.50 ± 0.75	5.00 ± 0.00
Cross-talk error (CTE)	8.19 ± 2.41	4.59 ± 1.79	4.96 ± 0.72	3.50 ± 0.00
CPU processing time [s]	3.05 ± 0.05	110.1 ± 32.7	26.59 ± 2.83	28.46 ± 0.70

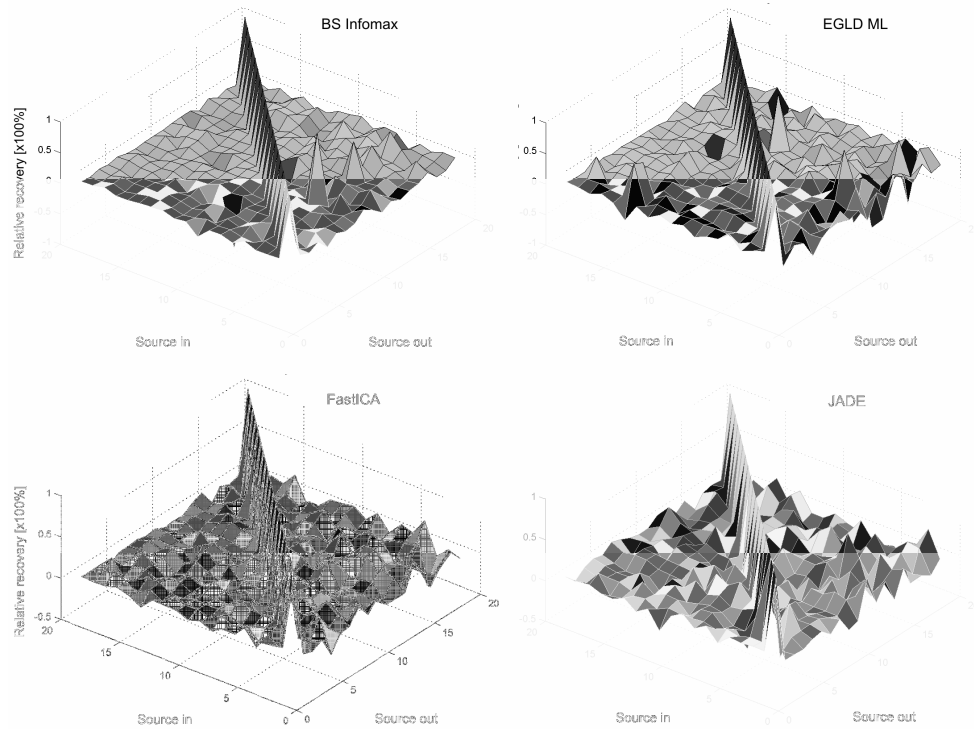


Fig. 12 – Relative recovery of artificially generated super-Gaussian sources by the ICA algorithms under study. Though the vertical axes were restricted from -1 to $+1$, some components may be inadvertently retrieved beyond these limits. A perfect recovery should come out with unit-height diagonal components only, while the rest of the landscape should be quite flat.

The relative recovery of the original artificially generated time series as output independent components yielded by ICA is showed for each algorithm in Fig. 12, where the plots were selected out of 120, such as to correspond

simultaneously as close as possible to the mean values of *SIR* and *CTE*. The items, which were subject of comparison for quality assessment of the ICA algorithms, are summarized in Table 1, where the values were averaged over 120 trials. The same data are plotted for convenience in Fig. 13.

Table 2

The score of the algorithms in performing ICA decomposition of artificially generated time series

<i>Figures of merit</i>	<i>BS infomax</i>	<i>EGLD ML</i>	<i>FastICA</i>	<i>JADE</i>
Absolute activation correlations	0	1	1	1
Max absolute off-diagonal value in the matrix Q	2	0	3	1
Mean absolute off-diagonal value in the matrix Q	1	2	3	0
Absolute correlations between source signals and corresponding ICA estimated activations	1	2	3	0
Signal-to-interference ratio (<i>SIR</i>) [dB]	3	2	1	0
Cross-talk error (<i>CTE</i>)	0	1	2	3
CPU processing time [s]	3	0	2	1
<i>Total points of relative merit</i>	10	8	15	6

The relevance of the data in Table 1 conclude in the ICA algorithm ranking by associating merit points to all performance items ranging from 0 to 3, in order of increasing performance (Table 2). Our simulations came out with the following sequence of ICA algorithms in order of decreasing performance of decomposition EEG-like time series from spatially-fixed linear mixtures: FastICA > BS infomax > EGLD ML > JADE. Nevertheless, it appears that in most respects, the algorithm characteristics are overlapping (Fig. 13) and one can not speak about “the best” ICA algorithm that outperforms all other approaches in *any* circumstances.

6. CONCLUSION

The flexibility of ICA approach by incorporating higher-order statistical information resides in transforming the PCA ill-posed problem associated with decorrelated decompositions into a well-posed problem of independent decompositions, that is, ICA avoids the non-uniqueness associated with PCA. As an exploratory data analysis, ICA is easy to perform and visualize. If sources can be shown to have distinct and consistent relationships to behavior or other

physiological signals, ICA filtering of brain recordings can reveal meaningful aspects of event-related brain dynamics associated with sensory and cognitive processing but hidden within correlated EEG responses at individual scalp sites [61].

The ICA approach applied to single-trial or averaged ERPs performs blind separation of multichannel complex EEG data into the time courses of activation and the corresponding scalp topographies of relatively large, temporally independent and spatially fixed sources, and further decomposes the remained mixed signals into subcomponents that may reflect the activity of functionally distinct generators of physiological activity. The simultaneous separation of true

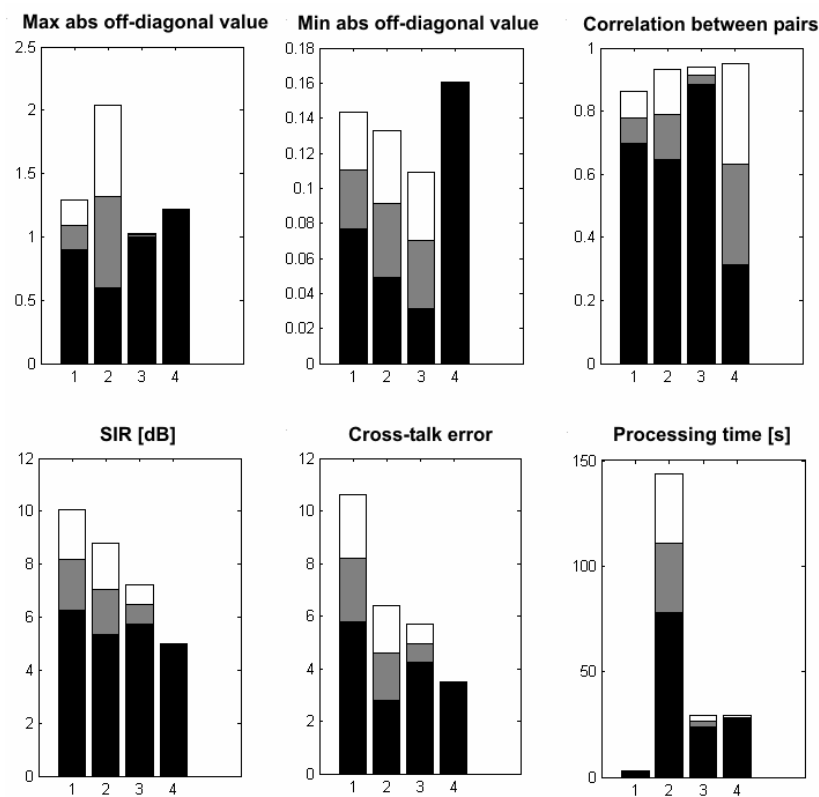


Fig. 13 – The data plotted are the following: a) maximum absolute off-diagonal values in \mathbf{Q} ; b) minimum absolute off-diagonal values in \mathbf{Q} ; c) correlations between source signals and their ICA estimated counterparts; d) signal to interference ratios; e) cross-talk errors; f) total processing time. The sequence of the algorithms is: (1) BS infomax, (2) EGLD ML, (3) FastICA, and (4) JADE.

EEG activity and its artifacts into distinct independent components based on their statistical properties and without need of "clean" reference channels circumvents the problem of mutual contamination of regressing and regressed channels and

prevents data loss. The decomposition of EEG data is carried out over the entire scalp and frequency band, which may display a variety of distinct relationships to task events, rather than focusing on activity at single frequencies in single scalp channels, or channel pairs. Pervasive artifacts of various types can be removed from EEG recordings making possible the analysis of highly contaminated EEG data. The ICA method allows identification and segregation of stimulus- and response-locked event-related activity in single trial EEG epochs. It makes also possible the investigation of the interaction between ERPs and the ongoing EEG, as well as monitoring the spatial structure of ongoing or averaged EEG activity in multiple brain areas, networks, or neural populations [41].

A general conclusion that has emerged from applying ICA to various brain recordings and which is fully supported by our research is the effectiveness of the independence assumption among artifacts and various forms of brain activity as a rich source of possible inferences. However, the underlying independence assumption in correction of EEG data by applying ICA is not always realistic: EEG activity may be correlated temporarily with some artifactual activity. ICA is computationally efficient; particularly the stabilized version of FastICA algorithm is attractive by its fast and reliable convergence, and the lack of parameters to be tuned, such as the learning rates in neuromorphic implementations. It works even when the assumption of independence is weak and it comes out with estimated components that are as independent as possible. This raises the question of the conceptual meaning of the ICA estimates. A possible answer may be given by the close connection between ICA and the algorithms adopted in theoretical neuroscience that emulate the way in which the brain optimally extracts information from sensory inputs [62]. The natural gradient incorporated in the BS extended infomax algorithm performs better than the original gradient ascent and is computationally less demanding. Though the BS algorithm is theoretically optimal in the sense of dealing with the MI, like all neural unsupervised algorithms, its performance heavily depends on the learning rates and its convergence is rather slow. The EGLD algorithm employing the ML principle separates skewed distributions, even for zero kurtosis. Therefore it is recommended in the cases where the source signals are skewed but distributed close to normality, when other approaches based on higher-order statistics are likely to fail. However, the source signals that significantly deviate from normality are not suited for modeling with extended generalized lambda or delta distributions. As both estimators for parameters and score functions are simple rational functions, the EGLD ML algorithm is expected to be computationally fast, though our simulations ranked it as the most computational demanding and, consequently, time consuming. In terms of computational time, the BS extended infomax algorithm was the fastest, FastICA more faithfully retrieved the sources among all algorithms under test, while the JADE algorithm came out with a full transform matrix \mathbf{Q} that is the closest to unity. If properly tuned, any ICA algorithms, either based on infomax

principle, maximum likelihood, approximate diagonalization of eigen-matrices, or nonlinear approximations of negentropy, perform quite similar, apart from the required computational time and number of parameters to set, and yield consistent outcomes given the same non-Gaussian distributions.

REFERENCES

1. Bell A.J. and Sejnowski T.J., *An information maximization-approach to blind separation and blind deconvolution*, Neural Comput., **7**, 1129–1159 (1995).
2. Friston K.J., Holmes A.P., Worsley K.J., Poline J.-P., Frith C.D., Frackowiak R.S.J., *Statistical parametric maps in functional imaging: A general linear approach*, Hum. Brain Mapp., **2**, 189–210 (1995).
3. Mutihac R., *Exploratory versus confirmatory analysis of fMRI data*, ESMRMB 2005, Functional MRI, Basel, Switzerland, Sept. 15–18, 2005, S203-204.
4. McKeown M.J., Makeig S., Brown G.G., Jung T.-P., Kindermann S.S., Bell A.J., Sejnowski T.J., *Response from Martin McKeown, Makeig, Brown, Jung, Kindermann, Bell and Sejnowski*, Trends in Cognitive Sciences, **2**, 10, 375 (1998a).
5. Zhukov L., Weinstein D., and Johnson C., *Independent component analysis for EEG source localization*, IEEE Eng. Med. Biol., **19**, 87–96 (March 2000).
6. McKeown M.J., Humphries C., Iragui V. and Sejnowski T.J., *Spatially fixed patterns account for the spike and wave features in absence seizures*, Brain Topogr., **12**, 107–116 (1999).
7. Barlow H.B., *Possible principles underlying the transformations of sensory messages*, in *Sensory Communications* (WA Rosenblith, Ed.), MIT Press, Cambridge MA, 1961, pp. 217–234.
8. Field D.J., *What is the goal of sensory coding?*, Neural Comput., **6**, 559–601 (1994).
9. Atick J.J., *Could information theory provide an ecological theory of sensory processing?*, Network, **3**, 213-251 (1992).
10. Mutihac R. and Van Hulle M.M., *Comparison of PCA and ICA for BSS*, Romanian Reports in Physics, **56**, 1, 25–37 (2004).
11. Huber P.J., *Projection pursuit*, The Annals of Statistics, **13**, 2, 435–475 (1985).
12. Friedman J.H., *Exploratory projection pursuit*, J. of the American Statistical Association, **82**, (397), 249–266, 1987.
13. Jones M.C. and Sibson R., *What is projection pursuit?*, J. of the Royal Statistical Society, ser. A, **150**, 1–36 (1987).
14. Friston K.J., *Modes or models: A critique on independent component analysis for fMRI*, Trends in Cognitive Sciences, **2**(10), 373–375 (1998).
15. Comon P., *Independent component analysis, A new concept?*, Signal Proces., **36**, 287–314 (1994).
16. Cardoso J.-F., *Blind signal separation: Statistical principles*, Proc. IEEE, **9**, 2009–2025 (Oct. 1998).
17. Simoncelli E.P. and Olshausen B.A., *Natural image statistics and neural representation*, Annu. Rev. Neurosci., **24**, 193-216 (2001).
18. Olshausen B.A. and Field D.J., *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*, Nature, **381**, 607–609 (1996).
19. Van Hateren J.H. and Van der Schaaf A., *Independent component filters of natural images compared with simple cells in primary visual cortex*, Proc. Royal. Society, ser. B, **265**, 359–366, 1998.
20. Karhunen J., Cichocki A., Kasprzak W. and Pajunen P., *On neural blind separation with noise suppression and redundancy reduction*, Int. Journal of Neural Systems, **8**, 219–237 (1997).
21. Stetter M., *Advanced Statistics*, Presentation at the *EU Advanced Course in Computational Neuroscience*, The Abdus Salam ICTP, Trieste, Italy, 2001.

22. Mutihac R. and Mutihac R.C., *Independent component analysis of fMRI data – Algorithms, model selection, and performance*, *ESMRMB 2006.*, fMRI Methods, Warsaw, Poland, Sept. 21–23, 2006, p. 129.
23. Amari S., Cichocki A. and Yang H.H., *A new learning algorithm for blind signal separation*, *Advances in Neural Information Processing*, MIT Press, Cambridge, MA, **8**, 757–763 (1995).
24. Mutihac R. and Van Hulle M.M., *A comparative survey on adaptive neural network algorithms for ICA*, *Romanian Reports in Physics*, **55**, 1, 49–73 (2003).
25. Moreau O. and Macchi O., *New self-adaptive algorithms for source separation based on contrast functions*, *Proc. IEEE Signal Processing Workshop on Higher Order Statistics*, 1993, pp. 215–219.
26. Cardoso J.-F., Belouchrani A. and Laheld B., *A new composite criterion for adaptive and iterative blind source separation*, *Proc. ICASSP*, 1994, pp. 273–276.
27. Cardoso J.-F. and Souloumiac A., *Blind beamforming for nongaussian signals*, *IEE Proc.-F.*, **140**, 362–370 (June 1993).
28. Pham D.-T., Garrat P. and Jutten C., *Separation of a mixture of independent sources through a maximum likelihood approach*, *Proc. EUSIPCO*, 1992, pp. 771–774.
29. Cardoso J-F and Souloumiac A, *Jacobi angles for simultaneous diagonalization*, *J. Mat. Anal. Appl.*, **17**, 161–164 (1996).
30. Hyvärinen A. and Oja E., *A fast fixed-point algorithm for independent component analysis*, *Neural Comput.*, **9**, 1483–1492 (1997).
31. Eriksson J., Karvanen J. and Koivunen V., “*Source distribution adaptive maximum likelihood estimation of ICA model*”, in *Proc. 2nd Int’l. Workshop on ICA and BSS*, 2000, pp. 227–232.
32. Hyvärinen A., *New approximations of differential entropy for ICA and projection pursuit*, in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, **10**, 273–279 (1998).
33. Calhoun V.D., Adali T., Hansen L.K., Larsen J. and Pekar J.J., *ICA of functional MRI data: An overview*, *ICA 2003*, Nara, Japan, April 2003.
34. Nunez P., *Electric Fields of the Brain*, Oxford University Press, Oxford, 1981.
35. Galambos R. and Makeig S., *Dynamic changes in steady-state potentials*, in *Dynamics of Sensory and Cognitive Processing of the Brain* (E. Basar, Ed.), Springer-Verlag, Berlin, 1989, pp. 102–122.
36. Makeig S., Bell A.J., Jung T.-P. and Sejnowski T.J., *Independent component analysis of electroencephalographic data*, in *Advances in Neural Information Processing Systems* (D. Touretzky, M. Mozer, and M. Hasselmo, Eds.), MIT Press, Cambridge MA, **8**, 145–151 (1996).
37. Jung T.-P., Humphries C., Lee T.-W., Makeig S., McKeown M.J., Iragui V. and Sejnowski T.J., *Extended ICA removes artifacts from electroencephalographic recordings*, in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, **10**, 894–900 (1998).
38. McKeown M.J., *Detection of consistently task-related activations in fMRI data with hybrid independent component analysis*, *Neuroimage*, **11**, 24–35 (2000).
39. McKeown M.J., Makeig S., Brown G.G., Jung T.-P., Kindermann S.S., Bell A.J. and Sejnowski T.J., *Analysis of fMRI data by blind separation into independent spatial components*, *Hum. Brain Map.*, **6**, 160–188 (1998).
40. Mutihac R., *On model complexity estimation in non-square ICA of fMRI data*, *ISMRM 2005*, Miami Beach, Florida, May 7–13, 2005, p. 1594.
41. Jung T.-P., Humphries C., Lee T.-W., McKeown M.J., Iragui V., Makeig S. and Sejnowski T.J., *Removing electroencephalographic artifacts by blind source separation*, *Psychophysiol.*, **37**, 163–178 (Feb. 2000).
42. Jung T.-P., Makeig S., Townsend J., Westerfield M., Hicks B., Courchesne E., and Sejnowski T.J., *Single-trial ERPs during continuous fMRI scanning*, *Society for Neuroscience Abstracts*, **25**, 1389 (1999).
43. Makeig, S., Enghoff S., Jung T.-P. and Sejnowski T.J., *A natural basis for efficient brain-actuated control*, *IEEE Trans. Rehab. Eng.*, **8**, 208–211 (2000).

44. Jung T.-P., Makeig S., Bell A.J. and Sejnowski T.J., *Independent component analysis of electroencephalographic and event-related data*, in Processing and Neural Modeling (P. Poon and J. Brugge, Eds.), pp. 189–197, Plenum Press, New York, 1998.
45. Yang H.H. and Amari S., *Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information*, Neural Comput., **9**, 1457–1482 (1997).
46. Friman O. and Westin C.-F., *Resampling fMRI time series*, NeuroImage, **25**, 859–867 (2005).
47. Meinecke F., Ziehe A., Motoaki Kawanabe M. and Müller K.-R., *Assessing reliability of ICA projections – A resampling approach*, Oberwolfach Seminar, Germany, Sept. 2000.
48. Nayak K., SCRI, Florida State University, EEG data at <http://www.scri.fsu.edu/~nayak/postit.html>.
49. Lee T.-W., Girolami M., Bell A.J. and Sejnowski T.J., *A unifying information-theoretic framework for independent component analysis*, Computers and Mathematics with Applications, **39**, 1–21 (2000).
50. Hyvärinen A., *Fast and robust fixed-point algorithms for ICA*, IEEE Trans. on Neural Networks, **10**, 626–634 (March 1999).
51. Mutihac R. and Van Hulle M.M., *Neural network implementations of independent component analysis*, IEEE Workshop on Neural Networks for Signal Processing, Martigny, Switzerland, Sept. 4–6, 2002, pp. 505–515.
52. *** Matlab R13, The MathWorks Inc., <http://www.mathworks.com>.
53. Makeig *et al.*, ICA Toolbox for Psychophysiological Research. Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, available at <http://www.cnl.salk.edu/~ica.htm>.
54. Berg P. and Scherg M., *Dipole model for eye movements and blinks*, Electroencephalog. clin. Neurophysiol., **79**, 36–44 (1991).
55. Hillyard S.A. and Galambos R., *Eye-movement artifact in the CNV*, Electroencephalog. clin. Neurophysiol., **28**, 173–182 (1970).
56. Verleger R., Gasser T. and Mocks J., *Correction of EOG artifacts in event-related potentials of EEG: Aspects of reliability and validity*, Psychoph., **19**, 472–480 (April 1982).
57. Whitton J.L., Lue F. and Moldofsky H., *A spectral method for removing eye-movements artifacts from the EEG*, Electroencephalog. clin. Neurophysiol., **44**, 735–741 (1978).
58. Woestenburg J.C., Verbaten M.N. and Slangen J.L., *The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain*, Biol. Psychol., **16**, 127–147 (1983).
59. Weerts T.C. and Lang P.J., *The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV)*, Biol. Psychol., **1**, 1–19 (Jan. 1973).
60. Vigário R., Sarela J., Jousmäki V., Hämäläinen M., and Oja E., *Independent component approaches to the analysis of EEG and MEG recordings*, IEEE Trans. Biomed. Eng., **47**, 589–593 (2000).
61. Bartlett M.S., Makeig S., Bell A.J., Jung T.-P. and Sejnowski T.J., Society for Neuroscience Abstracts, **21**, 437, 1995.
62. McKeown M.J., Jung T.P., Makeig S., Brown G.G., Kindermann S.S., Lee T.-W. and Sejnowski T.J., *Spatially independent activity patterns in fMRI data during the Stroop color-naming task*, Proc. Natl. Acad. Sci. USA, **95**, 803–810 (Feb. 1998).