

Petr Tichavský; Zbyněk Koldovský

Fast and accurate methods of independent component analysis: A survey

Kybernetika, Vol. 47 (2011), No. 3, 426--438

Persistent URL: <http://dml.cz/dmlcz/141594>

Terms of use:

© Institute of Information Theory and Automation AS CR, 2011

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

FAST AND ACCURATE METHODS OF INDEPENDENT COMPONENT ANALYSIS: A SURVEY

PETR TICHAŤAVSKÝ AND ZBYNĚK KOLDOVSKÝ

This paper presents a survey of recent successful algorithms for blind separation of determined instantaneous linear mixtures of independent sources such as natural speech or biomedical signals. These algorithms rely either on non-Gaussianity, nonstationarity, spectral diversity, or on a combination of them. Performance of the algorithms will be demonstrated on separation of a linear instantaneous mixture of audio signals (music, speech) and on artifact removal in electroencephalogram (EEG).

Keywords: Blind source separation, probability distribution, score function, autoregressive random processes, audio signal processing, electroencephalogram, artifact rejection

Classification: 94A12, 92-02, 92-04, 92-08

1. INTRODUCTION

Independent Component Analysis (ICA) and Blind Source Separation (BSS) represent a wide class of statistical models and algorithms that have one goal in common: to retrieve unknown statistically independent signals from their mixtures. In this paper, the classical real-valued square (invertible) instantaneous linear ICA model $\mathbf{X} = \mathbf{A}\mathbf{S}$ is addressed, where $\mathbf{S}, \mathbf{X} \in \mathbb{R}^{d \times N}$ contain d unknown independent source signals and their observed mixtures (respectively), each of length N , and $\mathbf{A} \in \mathbb{R}^{d \times d}$ is an unknown mixing matrix.

The goal is to estimate the mixing matrix \mathbf{A} or, equivalently, the de-mixing matrix $\mathbf{W} \triangleq \mathbf{A}^{-1}$ or, equivalently, the original source signals \mathbf{S} . Solution of this kind of problem is important, for example, in audio signal processing and in biomedical signal processing. The estimation problem is called “blind” if there is no prior information about the mixing system, represented by \mathbf{A} . Note that it is also possible to study the problem when the number of the original signals is greater than the number of the mixtures and vice versa. While the latter problem can be easily transformed to the square mixture using a dimensionality reduction through a principal component analysis, the former problem, called underdetermined, is more challenging. It exceeds the scope of this paper, however [10].

The nature of the original signals, represented by the rows of \mathbf{S} , is often hard to characterise. For example, if these signals are speech signals, it is possible to

model them in several ways; consequently, the signals may be separated by different methods depending on the selected statistical model.

There are three basic ICA approaches coming from different models of signals [25]. The first one assumes that a signal is a sequence of identically and independently distributed random variables. The condition of separability of such signals requires that only one signal is Gaussian at most, so the approach is said to be based on the *non-Gaussianity*. The second approach takes the *nonstationarity* of signals into account by modelling them as independently distributed Gaussian variables whose variance is changing in time. The third basic model considers weakly stationary Gaussian processes. These signals are separable if their spectra are distinct; therefore, the separation is based on the *spectral diversity*. All three features are present in speech signals, as is shown in Figure 1: Diagram (a) shows a speech signal in time domain, and diagrams (b), (c), (d) demonstrate its non-Gaussianity, nonstationarity and a non-uniform power spectrum. The normal probability plot in diagram (b) is defined as ordered response values versus normal order statistic medians [5]. If the points on this plot have formed a nearly linear pattern, then the normal distribution would be a good model for this data set. This is, however, not true in this case, because deviations of the points from a straight line are apparent.

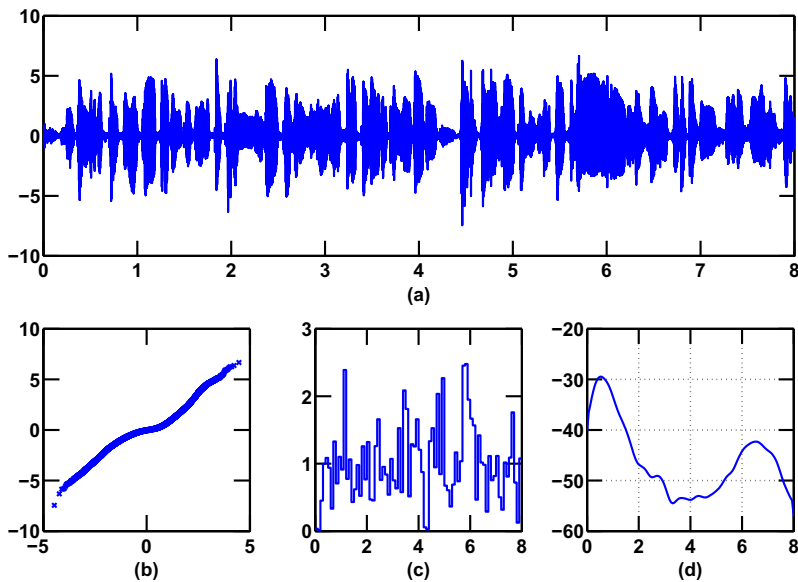


Fig. 1. (a) An 8 s long recording of a speech signal, sampled at 16 kHz. (b) Normal probability plot of the signal. (c) Variances of the signal in partitioning to 80 blocks of equal length. (d) Power spectrum density of the signal [dB/kHz].

The assumption common to all models in the ICA is the statistical independence of the original signals. Note that the observed signals, mixed through the matrix \mathbf{A} ,

are not mutually independent, in general. The solution of the ICA, whatever the model of \mathbf{S} is, consists in finding a matrix \mathbf{W} such that the signals (rows of) $\mathbf{W}\mathbf{X}$ are mutually statistically independent. This is fulfilled whenever $\mathbf{W} = \mathbf{\Lambda}\mathbf{P}\mathbf{A}^{-1}$ where $\mathbf{\Lambda}$ is a diagonal matrix with nonzero diagonal elements and \mathbf{P} is a permutation matrix. Hence, the ICA solution is ambiguous in the sense that the order, signs and scales of the original signals cannot be retrieved.

One possible extension of the ICA is the so-called Independent Subspace Analysis (ISA). ISA can be applied in situations where not all original signals can be separated from each other by a linear transformation. The goal here is to decompose the linear space spanned by rows of \mathbf{X} to a direct (orthogonal) sum of linear subspaces such that elements of each subspace are statistically independent of the others. The ISA problem can be approached by applying an ICA algorithm, which aims at separating each component from the others as much as possible. This step is followed by clustering of the obtained components according to their residual mutual dependence [14].

This paper provides a survey of several successful state-of-the-art methods that rely on the three principles mentioned above. The focus is on the methods whose performance approaches (or may approach) the best possible one given by the Cramér–Rao Lower Bound (CRLB) of the respective model. The following section presents three methods – EFICA, BGSEP and WASOBI – based on the basic models mentioned above. Section 3 describes MULTICOMBI, Block EFICA and BARBI assuming hybrid models, that is, models combining two of the three basic approaches. In both sections, each model is introduced with its necessary notation, and a survey of related methods and corresponding papers is given. Section 4 presents several experiments to compare the performance of the methods when working with different kinds of real-world signals. Section 5 concludes the paper.

2. BASIC ICA MODELS

2.1. Non-Gaussianity

The non-Gaussianity-based model assumes that each original signal is a sequence of i.i.d. random variables. This means that each sample of the i th original signal $s_i(n)$ has the probability density function (PDF) f_{s_i} . Since the signals are assumed to be independent, the joint density of $s_1(n), \dots, s_d(n)$ is equal to the product of the corresponding marginals, $f_{s_1, \dots, s_d} = \prod_{i=1}^d f_{s_i}$. Some ICA algorithms estimate the separating transformation by maximising the Kullback-Leibler divergence between the joint distribution of the separated signals and the product of the marginals. This is equal, by definition, to the mutual information (some authors call it *redundancy* or *multiinformation*) of the separated signals. This is equivalent to maximising neg-entropies of the separated signals, where the negentropy of a random variable is the Kullback-Leibler divergence between the variable and a Gaussian-distributed variable that has the same mean and variance. It can be shown that both of these approaches are equivalent to the maximum likelihood estimate which, however, requires simultaneous estimation of the PDFs of the separated signals [17]. The PDFs appear in the estimation in terms of so-called score functions – derivatives of log-

arithm of the PDF. These score functions can be estimated as non-parametric, see NPICA [2] or RADICAL [16]. Although these separation methods are usually accurate, they are computationally complex and cannot be used to separate more than few signals (less than 10), in practice.

Some other separation methods use a parametric modelling of the score functions of the separated signals. For instance, Pham et al. proposed mean square fitting of the score functions by linear combinations of given nonlinear functions in [18], to derive a blind separating algorithm.

A reasonably accurate and fast ICA algorithm can be obtained by minimising a contrast function, which can be quite an arbitrary nonlinear and non-quadratic statistic of the data, such as kurtosis. An example is the popular algorithm FastICA [9].

The finding of the k th row of \mathbf{W} in FastICA proceeds by optimising the contrast function

$$c(\mathbf{w}_k) = \hat{\mathbb{E}}[G(\mathbf{w}_k^T \mathbf{Z})], \quad (1)$$

where $\hat{\mathbb{E}}$ stands for the sample mean estimator, T denotes the matrix/vector transposition, and \mathbf{Z} is equal to the transformed matrix \mathbf{X} , such that rows of \mathbf{Z} are uncorrelated, $\mathbf{Z} = (\mathbf{X}\mathbf{X}^T/N)^{-1/2}\mathbf{X}$. G is a properly chosen nonlinear function whose derivative will be denoted by g . It can be shown that, ideally, g should be the score function of the separated signal, $\mathbf{w}_k^T \mathbf{Z}$ [3]. The original FastICA [9] utilises a fixed choice of G , e. g., such that $g(x) = x^3$ or $g(x) = \tanh(x)$. The estimation of the whole \mathbf{W} proceeds by finding all local extrema of $c(\mathbf{w}_k)$ on the unit sphere. The deflation approach estimates \mathbf{W} row by row so that each row must be orthogonal to the previous ones. Another approach, called *Symmetric FastICA*, orthogonalises all rows of \mathbf{W} after each iteration jointly, by means of the so-called symmetric orthogonalisation. Statistical properties of FastICA were analysed in [22].

The analysis of FastICA gave rise to a new algorithm, EFICA [12]. It is more sophisticated and its performance nearly achieves the corresponding Cramér–Rao bound, if the separated signals have a generalised Gaussian distribution.¹ Unlike FastICA and other algorithms, the outcome of EFICA does not produce strictly uncorrelated components; it was observed by several authors that the requirement that sample correlations of the separated signals must be zero may compromise the separation performance of algorithms [3].

EFICA is initialised by the outcome of Symmetric FastICA. Then, a special technique called a *test of saddle points* is applied to make sure that the global minimum of the contrast function has been found. The partly separated signals are used to form an adaptive contrast function, used in a fine tuning of the estimate of \mathbf{W} . EFICA does not differ much from FastICA in terms of computational complexity, so it retains its popular high speed property. Some further improvements of EFICA in terms of speed and accuracy were proposed in [24].

¹PDF of the generalised Gaussian distribution is proportional to $\exp(-|x|^\alpha/\beta)$, where α is a shape parameter and β controls the variance. The distribution includes a standard normal distribution for $\alpha = 2$, a Laplacean distribution for $\alpha = 1$ and a uniform distribution as a limit for $\alpha \rightarrow \infty$.

2.2. Nonstationarity

Let the original signals and the mixture be partitioned into M blocks of the same length $N_1 = N/M$, where N_1 is an integer,

$$\mathbf{S} = [\mathbf{S}^{(1)}, \dots, \mathbf{S}^{(M)}] \quad (2)$$

$$\mathbf{X} = [\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(M)}]. \quad (3)$$

Assume that each signal in each block $\mathbf{S}^{(\ell)}$ is Gaussian i.i.d., with zero mean and a variance $D_{k\ell}$, where $k = 1, \dots, d$ is the index of the signal and $\ell = 1, \dots, M$ is the index of the block. The signals are parameterised by a matrix \mathbf{D} with elements $D_{k\ell}$, which is unknown. The received mixture is parameterised by two unknown matrices, \mathbf{A} and \mathbf{D} .

The received data are by assumption, Gaussian distributed. The sufficient statistics for estimating \mathbf{A} and \mathbf{D} is the set of sample covariance matrices

$$\widehat{\mathbf{R}}_m = \frac{1}{N_1} \mathbf{X}^{(m)} (\mathbf{X}^{(m)})^T, \quad m = 1, \dots, M.$$

Theoretical covariance matrices obey the relation

$$\mathbf{R}_m = \mathbf{A} \mathbf{D}_m \mathbf{A}^T$$

where \mathbf{D}_m is a diagonal matrix containing the m th column of \mathbf{D} on its diagonal, and $\mathbf{D}_m = \mathbb{E}[\mathbf{S}^{(m)} (\mathbf{S}^{(m)})^T] / N_1$, $m = 1, \dots, M$.

We note that $\mathbf{W} = \mathbf{A}^{-1}$ can be found as a matrix that provides an approximate joint diagonalisation of the matrices $\{\widehat{\mathbf{R}}_k\}$, i. e., it has the property that all matrices $\{\mathbf{W} \widehat{\mathbf{R}}_k \mathbf{W}^T\}$ are approximately diagonal [21].

The approximate joint diagonalisation of a set of matrices can be performed in several ways, optimising several possible criteria, see, e. g., [26] for a survey of methods. In our case, it can be shown that the maximum likelihood (ML) estimator of the mixing/demixing matrices is achieved by minimising the criterion

$$C_{\text{LL}}(\mathbf{W}) = \sum_{m=1}^M \log \frac{\det \text{ddiag}(\mathbf{W} \widehat{\mathbf{R}}_m \mathbf{W}^T)}{\det(\mathbf{W} \widehat{\mathbf{R}}_m \mathbf{W}^T)} \quad (4)$$

where the operator “ddiag” applied to a square matrix \mathbf{M} nullifies the off-diagonal elements of the matrix. This criterion is meaningful only for positive definite target-matrices $\{\widehat{\mathbf{R}}_m\}$. An algorithm to minimise the criterion was proposed in [19].

In [26], a different method of joint diagonalisation of the matrices was proposed, which is asymptotically equivalent to the Pham’s estimator, but is more appealing computationally. It bears the name BGWEDGE (Block Gaussian Weighted Exhaustive Diagonalisation with Gauss itErations), and the corresponding separation algorithm is called BGSEP (Block Gaussian separation). Although theoretical computational complexity of Pham’s algorithm and BGWEDGE is the same, $O(d^2 M)$ operations per iteration, the latter algorithm is easier to parallelise. In matlab implementation, BGWEDGE is realised with fewer embedded “for” cycles and therefore it is faster in higher dimensions. Details of the BGWEDGE algorithm are rather technical and are omitted here to save some space.

2.3. Spectral Diversity

The third signal model assumes that the original signals may be stationary, but are distinguishable in the frequency domain. In particular, one may assume that the original signals are modelled as Gaussian autoregressive with a known order.

A sufficient statistic for joint estimation of the mixing/demixing matrix and autoregressive parameters of the separated sources is the set of the time-lagged estimated correlation matrices,

$$\hat{\mathbf{R}}_{\mathbf{x}}[\tau] = \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} \mathbf{x}[n]\mathbf{x}^T[n+\tau] \quad \tau = 0, \dots, M-1, \quad (5)$$

where $\mathbf{x}[n]$ denotes the n th column of \mathbf{X} and M is the order of the AR model.

Like in the previous subsection, the demixing matrix \mathbf{W} can be interpreted as a matrix that jointly diagonalises the matrices $\hat{\mathbf{R}}_{\mathbf{x}}[\tau]$, $\tau = 0, \dots, M-1$.

The first algorithm to realise the AJD was based on Jacobi rotations and is known under the acronym SOBI (Second Order Blind Identification)[1]. It has become quite popular in biomedical applications. SOBI, however, is not statistically efficient if the original signals obey the assumed AR model. Statistically efficient estimators of the mixing/demixing were independently proposed by Pham [20], Dégerine and Zaïdi [7], and Tichavský and Yeredor [26]. The latter algorithm is called WASOBI (weight adjusted SOBI). The weights in WASOBI are derived from AR modelling of partially separated signals. Unlike the other algorithms, WASOBI was shown to allow an approximately efficient separation even in high (100+) dimensional datasets.

3. HYBRID ICA MODELS

3.1. Block EFICA

Block EFICA is an ICA/BSS algorithm that relies both on non-Gaussianity and nonstationarity. Like the BGSEP algorithm, block EFICA assumes that the separated signal can be partitioned into a set of non-overlapping blocks so that the signals are stationary in each block. The signals may have different variances and even different distributions within distinct blocks.

The concept of block EFICA is very similar to that of EFICA. The main difference consists in that the optimal nonlinearities approximating score functions are estimated separately in each block of signals. Pham's parametric estimator from [18] is used for adaptive selection of the best linear combination of the functions from [24]. The second main difference is that the optimum weights for the refinement of the final estimate of \mathbf{W} are computed accordingly, respecting the piecewise stationary model.

Block EFICA asymptotically approaches CRLB under common assumptions when variance of the signals is constant. In cases where the variance of signals is changing, the algorithm is not optimal in theory, but its performance is close to the CRLB in practice. This was demonstrated by experiments with both synthetic and real-world signals [13].

3.2. BARBI

The abbreviation BARBI stands for the Block AutoRegressive Blind Identification. It is a separation method that relies on the signal nonstationarity and spectrum diversity. Like BGSEP and block EFICA, this method assumes that the mixture can be partitioned into L blocks, and in each of them, the separated signals are stationary and autoregressive of a given order. Therefore it can be viewed as an extension of BGSEP and WASOBI. The main idea consists in an approximate joint diagonalisation of the lagged covariance matrices like in (5), computed at each block separately. The number of these matrices is $L \times M$, where L is the number of blocks and M is the number of lags, i. e., the assumed AR order plus one. Unlike other ICA algorithms that are based on an AJD of some matrices, the AJD in BARBI incorporates a data-dependent weighting, which reflects the statistical model of the separated data. Therefore BARBI outperforms other separation methods in terms of accuracy if the assumed model is in accord with the reality.

BARBI can have two variants. The former variant, which is the only one so far to be programmed and tested, assumes that the AR parameters of each original signals may be completely different in each block. The total number of the estimated parameters is d^2 for all elements of the demixing matrix plus $d \times L \times M$ for AR parameters of all signals at all blocks separately. Such a large number of the estimated parameters has a negative impact on the separation performance if both L and M are large. This method will be called, for the sake of easy reference, BARBI-I.

In the second possible variant of BARBI, called BARBI-II, it is assumed that the AR coefficients of each original signal differ only by a multiplicative constant in different blocks. Again, the sufficient statistic is the same as in BARBI-I, i. e., the set of $L \times M$ time lagged covariance matrices, but the joint diagonalisation is constrained. For each original signal we would have only $L + M - 1$ parameters: L variances (one at each block) and $M - 1$ normalised AR coefficients.

3.3. MULTICOMBI

MULTICOMBI is an algorithm that combines EFICA and WASOBI to separate mixtures of signals that are either non-Gaussian or can be resolved in the spectral domain. It is based on the fact that these algorithms allow the estimation of not only the demixing matrix, but also the separation performance. The latter is measured in terms of the estimated interference-to-signal ratio (ISR) matrix, which predicts how much energy of the j th original signal is contained in the k th estimated signal. The ISR matrix is estimated by examining statistical properties of the separated signals. For instance, if some separated component is highly non-Gaussian, ISR of EFICA with respect to other components will be low, and vice versa: If there is a group of components that have nearly Gaussian distribution and cannot be well resolved from each other, the corresponding ISR submatrix will have large entries. Similarly, WASOBI produces an estimated ISR matrix which reveals structure of the mixture, i. e., components which have mutually similar spectra (and therefore they are hard to separate from one another) and vice versa.

MULTICOMBI applies both algorithms to the input data, which gives two dif-

ferent sets of independent components. In each set, the components are clustered according to their estimated ISR's. MULTICOMBI then accepts the clusters of the one algorithm that are separated from the other clusters better than all clusters of the other algorithm. The remaining (less well resolved) clusters of the winning algorithm are accepted as one merged cluster, unless it is empty. The procedure is applied recursively to each non-singleton cluster until all clusters are singletons, i. e., contain only one component and provide the output of MULTICOMBI.

In simulations, MULTICOMBI was shown to outperform other existing methods that rely on non-Gaussianity and spectral diversity, for instance ThinICA [6]. These methods are mostly based on a joint approximate diagonalisation of either cross-covariance, cumulant and cross-cumulant matrices. The (cross-)cumulants represent higher-order statistics taking the non-Gaussianity into account. Neither of these methods optimise the separation criterion to achieve the statistical efficiency given by the combined model.

4. SIMULATIONS

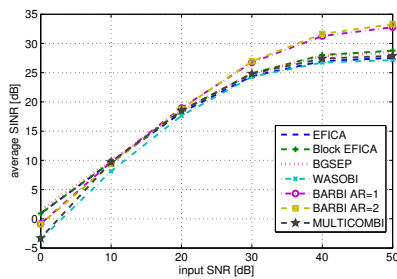
4.1. Separation of speech signals

This subsection presents a comparative study of performance of the above-mentioned algorithms in separation of a noisy linear instantaneous mixture of speech signals. Solution of this task might be a building block for separation of more challenging convolutive mixtures in the time domain [15].

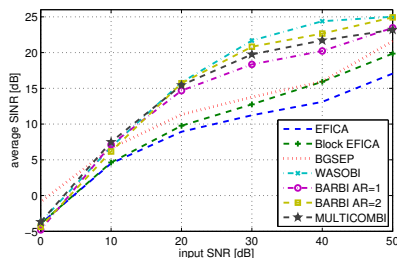
Twenty audio signals were considered for the experiment. Ten of them were speech signals and the other ten were pieces of music recordings. All signals were sampled at 8 kHz and normalised to have unit power (mean square). The recordings had the length of 5000 samples. The mixing matrix \mathbf{A} was chosen at random in each simulation trial, but it was normalised in the way that all rows of \mathbf{A}^{-1} had unit Euclidean norm. An independent Gaussian noise was added to the mixture to make the separation task more difficult and more realistic, symbolically $\mathbf{X} = \mathbf{AS} + \mathbf{N}$. The constraint on the norm of rows of \mathbf{A}^{-1} had the consequence that all signals in the mixture had the same signal-to-noise ratio (SNR).

The mixture has been processed by the seven ICA/BSS algorithms discussed in the paper. In BGSEP, block EFICA and in BARBI, the number of blocks was set to 10. In WASOBI, the AR order was set to 10. Two variants of BARBI were studied, with the AR order 1 and 2, respectively. The separated signals were sorted to best fit the original order of the signals. For each method and each signal, we have computed the resultant signal-to-interference-plus-noise ratio (SINR). The SINR values were averaged over 500 independent trials, for the speech signals and for the music signals separately. The results are shown in Figures 2(a) and 2(b), respectively.

Several conclusions can be drawn from the experiment. First, the music signals are harder to separate, in general, than the speech signals. They are less dynamical and more Gaussian. The best separation of the speech signals was obtained by BARBI with the AR order of 1 and 2. On the other hand, WASOBI has separated the music signals best. The other algorithms worked approximately equally well in separation of speech signals, but not so well in separating the music signals.



(a)



(b)

Fig. 2. Average SINR of speech signals (diagram (a)) and music signals (diagram (b)) obtained by 7 ICA/BSS algorithms from a noisy mixture versus varying SNR.

4.2. Artifact Elimination in Electroencephalogram

This subsection presents an example comparing performance of the above-mentioned algorithms in artifact elimination in Electroencephalogram (EEG). The EEG is a very complex multichannel biomedical signal, which is often corrupted by presence of some unwanted parasitic signals of various kinds. A typical example, which has been extensively studied in the literature, is an eye blinking. It has a typical U or V shape and can be observed in several channels simultaneously. For simplicity, these artifacts are considered in this paper as well.

Presence of artifacts makes automatic processing of EEG signals, which aims at diagnosis of brain diseases or in facilitating a human-computer interface, even more difficult than it already is without artifacts. For this reason there is a deep interest in designing automatic artifact removal procedures that would allow easier extraction of useful information. Methods of the ICA have proved to be very useful in this respect in the past [11].

Artifacts, including eye blinking, are assumed to be structurally simpler than the cerebral activity. Often, an artifact is like a short burst of some activity. When an ICA algorithm is applied to the EEG signal containing artifacts, a separation of the artifact from the rest of the data is possible if the artifact activity is concentrated

Table. Average Square Reconstruction Error.

EFICA	BGSEP	WASOBI	BEFICA	BARBI(1)	BARBI(2)	MULTICOMBI
0.142	0.116	3.69	0.177	5.06	1497	29.9

in a few (optimally in a single) “independent” components (IC) of the signal.

For purposes of this study, we skip the difficult question of how to recognise which IC represents the artifact. Our experiment assumes an EEG signal to be artifact-free, and an artifact of a known shape to be repeatedly added to the data at random time intervals. Of course, the shape of the artifact is not known to the separating algorithms. Note that typical eye blinking is most strongly present at the front electrodes on the scalp, FP1 and FP2. Since the artifact is known in advance, the artifact component is identified as the one which has the highest correlation with the true artifact. Once the artifact component is identified, it is replaced by zeros, and the reconstructed signal is computed by multiplying the matrix of the components by the estimated mixing matrix. The data, the independent components for one of the algorithms (BGSEP), and the reconstruction error are shown in Figures 3 through 5. In BGSEP, block EFICA and in BARBI, the number of blocks was set to 10. In WASOBI, the AR order was set to 10. Two variants of BARBI were studied, with the AR order 1 and 2, respectively.

A series of 70 similar experiments was conducted, with different noiseless data and with varying positions of the artifact. Resultant mean square reconstruction errors $\|\mathbf{Z} - \mathbf{X}\|_F / \|\mathbf{Y} - \mathbf{X}\|_F$ are summarised in Table. Here \mathbf{X} , \mathbf{Y} and \mathbf{Z} stand for the original (artifact-free) data, the data with the added artifact, and the reconstructed data, respectively, and $\|\cdot\|_F$ is the Frobenius norm.

The average errors are extremely large for some methods (WASOBI, BARBI and MULTICOMBI) due to cases of splitting the artifact into two or more components. It looks like these methods are completely wrong, but that is not the case: if *all* components that look like artifacts are deleted from the reconstruction, the error is not as huge. Also, the reconstruction error is not large if the single selected artifact component is subtracted not jointly, according to the estimated mixing matrix, but in each channel independently, minimising the norm of the reconstructed signal. Then, the latter reconstruction errors of all seven methods read 0.135, 0.115, 0.152, 0.171, 0.116, 0.175, and 0.135, respectively. In both methods of reconstruction, the best artifact rejection was achieved by BGSEP.

Note that another comparative study of performance of ICA methods in the context of the EEG signal processing was published in [8]. It includes more types of artifacts, but does not cover the most recent algorithms.

5. CONCLUSIONS

A survey of recent successful ICA algorithms is presented. The algorithms have been tested on separation of audio signals and on rejection of an eye blink artifact in a 19 channel EEG data. Although performance of the algorithms strongly depends on statistical properties of the separated signals, the results indicate that EFICA,

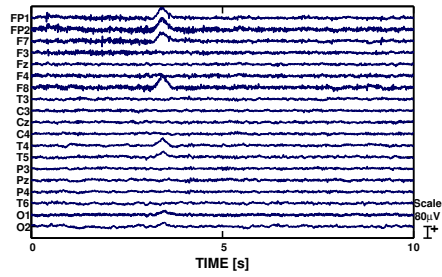


Fig. 3. A 19 channel EEG recording with one added artifact that mimics an eye blink.

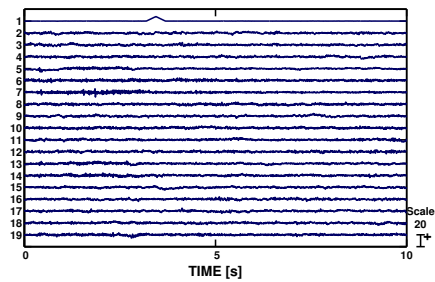


Fig. 4. Independent components of data in Figure 3 obtained by algorithm BGSEP.

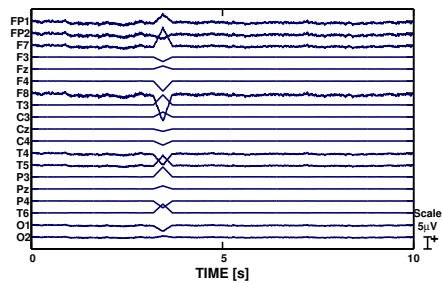


Fig. 5. Error in reconstruction of data in Figure 3 after excluding the first (artifact) component.

BGSEP and BARBI/WASOBI will be superior in audio signal processing, as well as in biomedical applications.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through Project 1M0572 and by the Grant Agency of the Czech Republic through Project 102/09/1278.

(Received July 1, 2010)

REFERENCES

- [1] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines: A blind source separation technique using second-order statistics. *IEEE Trans. Signal Processing* *45* (1997), 434–444.
- [2] R. Boscolo, H. Pan, and V. P. Roychowdhury: Independent component analysis based on nonparametric density estimation. *IEEE Trans. Neural Networks* *15* (2004), 55–65.
- [3] J.-F. Cardoso: Blind signal separation: statistical principles. *Proc. IEEE* *90* (1998), 2009–2026.
- [4] J.-F. Cardoso and D. T. Pham: Separation of non stationary sources. Algorithms and performance., In: *Independent Components Analysis: Principles and Practice* (S. J. Roberts and R. M. Everson, eds.), Cambridge University Press 2001, pp. 158–180.
- [5] J. Chambers, W. Cleveland, B. Kleiner, and P. Tukey: *Graphical Methods for Data Analysis*. Wadsworth, 1983.
- [6] S. Cruces, A. Cichocki, and L. De Lathauwer: Thin QR and SVD factorizations for simultaneous blind signal extraction. In: *Proc. European Signal Processing Conference (EUSIPCO)*, Vienna 2004, pp. 217–220.
- [7] S. Dégerine and A. Zaïdi: Separation of an instantaneous mixture of Gaussian autoregressive sources by the exact maximum likelihood approach. *IEEE Trans. Signal Processing* *52* (2004), 1492–1512.
- [8] A. Delorme, T. Sejnowski, S. Makeig: Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* *34* (2007) 1443–1449.
- [9] A. Hyvärinen and E. Oja: A fast fixed-point algorithm for independent component analysis. *Neural Computation* *9* (1997), 1483–1492.
- [10] A. Hyvärinen, J. Karhunen, and E. Oja: *Independent Component Analysis*. John Wiley & Sons, 2001.
- [11] C. J. James, C. W. Hesse: Independent component analysis for biomedical signals. *Physiol. Meas.* *26* (2005), R15–R39.
- [12] Z. Koldovský, P. Tichavský and E. Oja: Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér–Rao lower bound. *IEEE Trans. Neural Networks* *17* (2006), 1265–1277.
- [13] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, and S. Hosseini: Blind separation of piecewise stationary nonGaussian Sources. *Signal Process.* *89* (2009), 2570–2584.

- [14] Z. Koldovský and P. Tichavský: A comparison of independent component and independent subspace analysis algorithms. In: Proc. European Signal Processing Conference (EUSIPCO), Glasgow 2009, pp. 1447–1451.
- [15] Z. Koldovský and P. Tichavský: Time-domain blind separation of audio sources based on a complete ICA decomposition of an observation space. *IEEE Trans. Audio, Speech and Language Processing* *19* (2011), 406–416.
- [16] E. G. Learned-Miller and J. W. Fisher III: ICA using spacings estimates of entropy. *J. Machine Learning Research* *4* (2004), 1271–1295.
- [17] Te-Won Lee: Independent Component Analysis, Theory and Applications. Kluwer Academic Publishers, 1998.
- [18] D. T. Pham and P. Garat: Blind separation of mixture of independent sources through a quasi-maximum likelihood approach, *IEEE Trans. Signal Process.* *45* (1997), 1712–1725.
- [19] D.-T. Pham: Joint approximate diagonalization of positive definite Hermitian matrices. *SIAM J. Matrix Anal. Appl.* *22* (2001), 1136–1152.
- [20] D.-T. Pham: Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion. *Signal Process.* *81* (2001), 855–870.
- [21] D.-T. Pham and J.-F. Cardoso: Blind separation of instantaneous mixtures of non-stationary sources. *IEEE Trans. Signal Process.* *49* (2001), 1837–1848.
- [22] P. Tichavský, Z. Koldovský and E. Oja: Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis. *IEEE Trans. Signal Process.* *54* (2006), 1189–1203.
- [23] P. Tichavský, Z. Koldovský, and E. Oja: Corrections to “Performance analysis of the FastICA algorithm and Cramér–Rao Bounds for linear independent component analysis, TSP 04/06”. *IEEE Tran. Signal Process.* *56* (2008), 1715–1716.
- [24] P. Tichavský, Z. Koldovský, and E. Oja: Speed and accuracy enhancement of linear ICA techniques using rational nonlinear functions. *Lecture Notes in Comput. Sci.* *4666* (2007), 285–292.
- [25] P. Tichavský, Z. Koldovský, A. Yeredor, G. Gomez-Herrero, and E. Doron: A hybrid technique for blind non-Gaussian and time-correlated sources using a multicomponent approach. *IEEE Trans. Neural Networks* *19* (2008), 421–430.
- [26] P. Tichavský and A. Yeredor: Fast approximate joint diagonalization incorporating weight matrices. *IEEE Trans. Signal Process.* *57* (2009), 878–891.

*Petr Tichavský, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: tichavsk@utia.cas.cz*

*Zbyněk Koldovský, Faculty of Mechatronic and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec. Czech Republic.
e-mail: zbynek.koldovsky@tul.cz*