Proceedings of the 5th WSEAS Int. Conf. on SIGNAL, SPEECH and IMAGE PROCESSING, Corfu, Greece, August 17-19, 2005 (pp62-67)

A Reference Suite Design for Blind Signal Separation

MARKUS BORSCHBACH
Dept. of Mathematics and Natural Science, Institute for Computer Science,
University of Münster, Einsteinstr. 62, D-48149 Münster, GERMANY
http://cs.uni-muenster.de/u/flyer/

Abstract: - According to the common underlying mathematical model for Independent Component Analysis (ICA), the fulfillment of a BSS for either a linear and scalar type of composition, a convolutive and linear type or a nonlinear type has different conditions. So far, several approaches have been developed in the last decades for stationary and non-stationary data. To identify key research priorities, the different origins of neural network approaches for BSS are briefly reviewed and divided by classes of specific theoretical and application features. A principal guideline for the design of reference data sets for the comparison of all the existing ICA methods by its individual strengths and weaknesses for performing BSS is developed.

Key-Words: Statistical Models of ICA, Blind Deconvolution (BD), Multi-channel-BD, Reference Data Suite.

1 Introduction

In the last decade, the interest in Blind Signal Separation for the application in various fields, like processing of biomedical or geophysical data, speech or image processing and wireless communication code recognition is growing steadily (refer to section 4.5 and [2], [3] for an introduction). Starting with an example of blind acoustic source separation, the main underlying task is described in mathematical terms briefly in section 2. The estimation of Independent Components for a given blind observation of a predetermined time interval is the goal of any ICA-algorithm. First, two general cases are distinguished by stationary signals and time varying, e.g. by moving of the sources. Further on, the main approaches for stationary ICA are reviewed in section 3. Despite this brain storming facts of an ICA-overview, section 4 reveals key questions for a comparison of the ICA approaches and features which have to be fulfilled by the design of reference data sets for different application areas of ICA, again starting by the cocktail party problem for stationary acoustic signals. The proposed concept is completed by a list of existing databases for signals and algorithms. In the conclusion, the necessity of standard test reference data sets and more over, its'importance or senselessness for the development of new approaches and algorithms are discussed. Moreover, the similarities and difference of the presented concept to the most closely related work is summarized.

2 Blind Acoustic Source Separation

The initial situation of ICA and BSS is as follows. Given are linear noisy mixture vectors $\vec{u} \in R^m$,

$$\vec{u}(t) = M \cdot \vec{s}(t) + \vec{n}(t), \tag{1}$$

(see figure 1for a sample consisting of four mixtures) which are assumed to represent a linear transformation of the unknown source signal vector $\vec{s} = (s_1, ..., s_n)^T$ (each component of $\vec{s}$ is a single source) and the noise vector $\vec{n} \in R^n$ by the unknown but constant $m \times n$ mixing matrix M with full column rank. The source signals $s_i(t), i = 1, \cdots, n$ have to meet the following conditions: (a) they are statistically independent, (b) $E(s_i) = 0, Var(s_i) = 1$. The goal of BSS is to find a $n \times m$ ***separating matrix*** B such that the components y_i of the output vectors $\vec{y}(t) = B \cdot \vec{u}(t)$ have a maximal degree of statistical independence (which can be measured by a so called contrast function [5]).

By now, two underlying choices are made. First, the amplitude spectrum in the time domain of acoustic signals is used, which can be expressed by the electric field of a loudspeaker which originates one of the mixtures over a given time interval. Second, the mixtures are assumed as linear, but also non-linear dynamic systems exist. If the nonlinear characteristic is known or can be estimated and the inverse exists, in some cases the system can be simplified to a linear model [2], [3].

As an alternative representation of the noisy mixtures, the frequency domain requires for each frequency a single BSS-task. More over, each separated source appears in different order at each frequency. Therefore, a permutation problem arises. In signal processing, the frequency domain representation has advantages for convolutive mixtures in some cases [3] (see page 383 and chapter 19).

The different neural networks proposed for performing BSS consist of an input and an output layer, each consisting of n units. Both layers are fully connected to each other. The connection weights are the n-dimensional vectors $\vec{w}_i, i = 1, \cdots, n$ which form the columns of the $n \times n$ matrix W. The output of the i-th output unit $y_i, i = 1, \ldots, n$ is

$$y_i = \vec{w}_i^T \vec{x} \quad \text{or} \quad \vec{y} = W^T \vec{x} \tag{2}$$

where $\vec{x} = (x_1, \cdots, x_n)^T$ is the input vector and $\vec{y} = (y_1, \cdots, y_n)^T$ is the output vector. The neural approaches

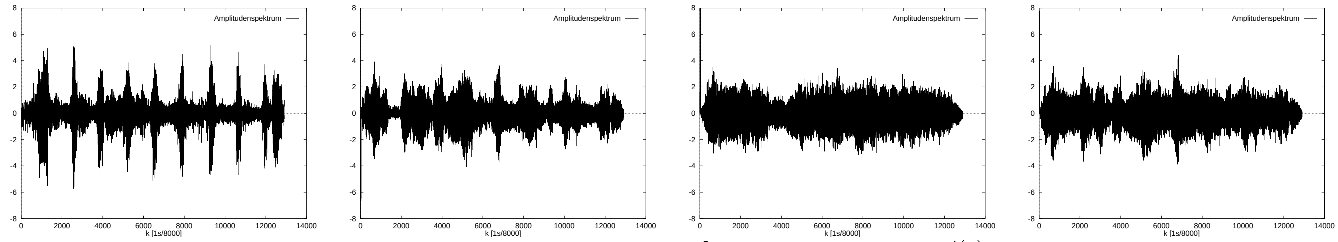

Fig. 1: Component $u_1 \ldots u_4$ of noisy mixture vector $\vec{u}(t)$

for BSS differ in the kind of learning iteration and whether a preliminary transformation of the observed noisy mixture vectors $\vec{u} \in R^m$ into the input vector $\vec{x}$, e.g. whitening [13], is necessary (transformation $\vec{x} = T(\vec{u})$ for each input vector $\vec{x}$) or not ($\vec{x} = \vec{u}$). In figure 2, the components of the separated signals are illustrated. Refer to [14], for full details of the learning process, e.g. the remaining error during the process, experimental environment, statistical features of the data and the illustration of all ICA steps. Moreover, estimating the underlying structure of the mixture, the so called ICA matrix H of the estimated sources, whose columns form the ICA basis, is the final task of ICA. In general, it can be determined directly from the separation matrix [12]. As a final step, the estimated observed mixture vector $\vec{\hat{u}} = \vec{H} \cdot \vec{y}$ can be determined. A linear mixture of stationary sources, is the simplified application case of ICA. The problem of non-linear mixtures will be revisited in the context of convolutive mixtures and non-stationary data in section 4.

3 Essential ICA Features

3.1 Existing approaches

In the case of a linear mixture of stationary sources (1) and the presence of additiv noise ($n(t)$), at least nine different approaches for robust BSS exist [2] (see page 321 for an overview). The existing approaches are divided by different criteria. As the most obvious assumption, appears the fulfillment of the independence criteria for identifying the sources in the observed mixtures. Therefore, the approaches in [3] are denoted here in an order of the following classes: class one is defined by ICA-tasks deriving the maximum likelihood estimation, e.g. the Bell Sejnowski algorithm [4], a natural gradient algorithm, a fast fixed point algorithm and the infomax principle ([3] see page 214 for references of different forms and versions). For this class of approaches, the maximization is realized by different numerical maximizations of the likelihood.

Class two is defined by ICA-tasks fulfilling the maximization of non-gaussianity, derived by a contrast function for non-gaussainity either by extremes of kurtosis, or by approximating negentropy. For both contrast functions, the maximization is realized either by a gradient algorithm or a fixed-point algorithm. For estimating more than one independent component, different vectors have to be uncorrelated by a form of orthogonalization ([3], see page 192).

Class three is defined by ICA-tasks derived by the the minimization of mutual information, e.g. approximation based on cumulants [5]. In a theoretical sense, besides the approximation by negentropy, this principle is equivalent to the one of class one and class two algorithms. For some different forms and versions of this class of algorithms refer to [3] (see page 226).

Class four is defined by ICA-tasks deriving non-linear decorrelation, e.g. Jutten-Herault-algorithm [6], Cichocki-Unbehauen-algorithm [7], estimation function approach from Amari-Cardoso [8] for neural networks [9], the relative gradient for equivariant adaptive separation via independence (EASIE) by Cardoso-Laheld [10] and learning rules for non-linear PCA [11]. For a collection of different forms and versions of this class of algorithms refer to [3] (see page 261).

3.2 Implications for application

In the theoretical sense, all the different contrast and objective functions perform an almost equal task. But more important, are practical limitation and possible advantages of each approach. Examples for different requirements of specific approaches are summarized in the following.

1. Requirements for the observed signals
   e.g. limitation on the number of different sources in the mixtures; limitations on the mixing matrix like required full column rank or being regular with or without special requirements on the diagonal elements; necessity of standardization of the observed mixture vectors; stationary or non-stationary sources in the observed mixtures; linear or non-linear composition of the sources in the observed mixtures (see [12] for a review of fourteen approaches regarding the features listed so far (implication for convolutive mixtures and non-linear instead of linear are revisited later)); necessity of a preprocessing step (e.g., by principal component analysis [13]); acceptance of noise besides an assumed structure (1) without any noise.

2. Assumption on the underlying sources
   e.g. selection of the activation function for the assumed statistical property of all underlying sources (either super- or subgaussian) [14] [13]; acceptance

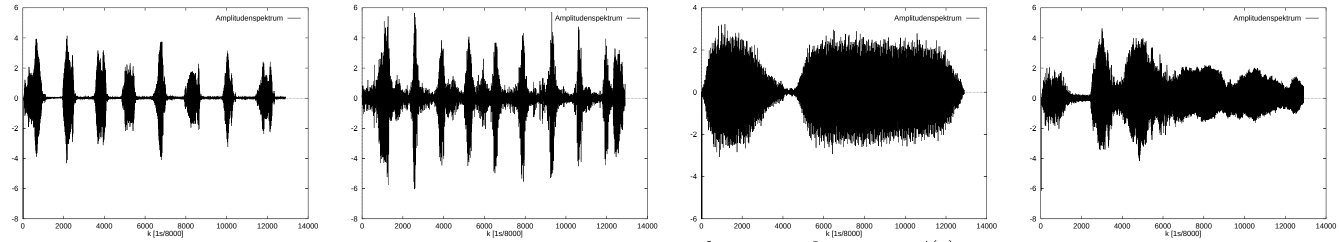

Fig. 2: Component $y_1 \cdots y_4$ of estimated sources $\vec{u}(t)$

of mixed sources of both (non-gaussian) without adjustment of a proper activation function [15].

3. Real world data
e.g. some algorithms require a square composition matrix with no noise as the underlying structure of the observed mixtures, even from the theoretical view others don't; preprocessing by PCA reduces the number of observed mixtures to the necessary amount (the estimated mixing matrix is square); adaptive or on-line algorithms for a permanently changing mixing matrix in the underlying structure of the observed signals; batch algorithms for a stationary underlying structure (called batch, block or off-line estimation, depending on the context, see [3] page 271). Noise at different stages: at each assumed source, noise as a single or a composition of a higher number of assumed sources, overall noise at each observed mixture, noise as a result of the demixing, noise as artifact reduction [2] (e.g. for medical data processing [16],[17]).

4 Test Set Priorities

For application of ICA-algorithms, all approaches capable of separating a non-trivial number of sources in the observed mixtures (either simultaneous or sequential) with robust behavior in the presence of noise are considered, e.g. see [18], [2], [19], [20]. Besides the existing empirical comparison for specific applications based on sample data sequences [17], [19], overall reference data sets for each domain are missing. Starting from the linear case for the underlying ICA-structure, different degrees for the design of the reference test sets arise from the different theoretical background, e.g. acceptance of underlying non-gaussian sources, or limitation to either sub- or super-gaussian sources. Moreover, for the deconvolution of convolutive mixtures a similar structure for the design of reference data sets is necessary. In the nonlinear case, the highest degree of assumptions for the design of reference data sets arises. For comparison of each test data set domain, a specific set of error functions is defined. In the following, the concept for reference data sets in the linear case, the convolutive case and the nonlinear case is universal framework for both, for each existing application domain and for every new application domain.

4.1 Reference sets in the linear case

The concept for the design of reference test data sets is divided into the scalable complexity of synthetic constructed mixtures as follows.

1. Determine a range of the number of observed mixtures $\{s(i) \mid i \in [1 \ldots n]\}$ which is useful according to the application domain, or a special task
e.g. in the idealized form of the four Chinese cocktail party problem and for a multichannel recording device for MEG-data the number of observed compositions are predetermined by the system. Besides such problems of fixed size, reference data sets summarize all reasonable cases, like a range from 64 to 168 for all recently existing MEG recording devices. This is done by constructing single data sets of observed data sets either in a step size according to every existing systems or by a reasonable approximated size of steps.

2. Determine a range of the underlying composition by different features of the underlying composition matrix.
The influence of the regularity of the underlying composition matrix for the BSS task is represented by a reasonable step size.

3. According to the application case, each test set is divided into permutation of different possible combinations of either super- or sub-gaussian sources, and the number of occurring sources with an almost gaussian distribution depending on the application.

4. Selected cases of the test suite build upon step 1.- to 3., are tested upon different amount of noise added at first in a singular case to the stage of composing to a single source, to a single observed mixture and in further cases to a rising number of sources and observed mixtures for each group, again in a reasonable step size. The kind of noise, is inspired by the kind occurring most likely in the application. Otherwise, the choice is a standard distributions of noise [21].

To limit the overall sum of occurring data sets in the test suite, at first only a coarse amount of each listed feature (step one to four) is used, e.g. the maximum and minimum number of observed mixtures. In a further step, the coarse setup is driven into a more granulate one, e.g. by the importance of an application case or the failure of single algorithm on a specific step. For instant, if an ICA-algorithm fails for a data set with the maximum number of observed mixture, a set of lower but close numbers of observed signals should be used, to compare the results of the algorithm result to other approaches.

4.2 Non-stationary data and composition

Depending on the application, the assumed statistical feature of the estimated sources can vary between stationary and non-stationary. Moreover, the underlying composition of the sources can vary from time invariant to instantaneous. For both cases, the test suite is adapted according to the following steps.

5. The batch size for selected cases of the test suite build upon step 1.- to 3., are reduced for the observed signals from an arbitrary high amount to a small one with a coarse step size. Starting from a small amount of the batch size or estimated according to the application domain or task, the speed of the moving components determines the necessary batch size for a successful separation of the estimated sources, e.g. for a practical constellation of the cocktail party problem based on stationary sources and moving microphones or sources. In the case of an underlying time invariant (stationary) composition structure (mixing matrix) and non-stationary sources, the necessary batch size is adapted to meet the estimated statistical feature of the stationary process for the sources [22]. According to this constellation, in [9] (see page 46) the feature of a wide sense stationary process can be found or the context of a quasi-stationary signal in [22]. If the estimated underlying structure and the sources are non-stationary, the batch size has to be small enough to fulfill both conditions for a successful separation.

4.3 Reference sets for convolutive mixtures

For a far amount of applications on real world data, the assumed underlying time invariant and linear model of a scalar composition of unknown sources receives a significant good result, even if the underlying model is not of scalar type or slightly time variant. Driven by the domain of application, it can be useful to extend the assumed underlying model, e.g. in the case of the cocktail party problem for acoustic speech recognition, each single speaker induce multiple superpositions at each microphone, due to reverberation and echo effects (multi-path fading) of the physical location, e.g. reflection on walls or objects. Moreover, the recordings at each microphone have different time delays due to the propagation speed of voices in the air. The constellation of a convolute of a single source at several observation is identified as the principle of SIMO, Single Input Multiple Output. Moreover, the separation is called blind deconvolution or blind equalization. Therefore, the linear and scalar type of the assumed underlying structure (1) is extended to a linear and convolutive one (3).

$$\vec{u}(t) = \sum_{l:=-\delta}^{+\delta} \vec{m}(l) \cdot s(t-l) + \vec{n}(t), \tag{3}$$

In comparison to the linear model of scalar type, parameter δ denotes the overall number of superpositions of a source s(t) at each observed signal (components of vector $\vec{u}(t)$) and the time shift of each superposition in respect to the source signal $s(t)$. Moreover, the vector $\vec{m}$ scales the amplitude of each superposition, e.g. depending on the exact condition of reflection.

6. To design synthetic reference data sets for blind deconvolution appropriate ranges of parameters for the mixing system (3) are necessary. For instance, applications areas are digital communication, cabel HDTV, global positioning systems, biomedical processing (Elektroencephalography (EEG)) [2](see page 336 for some references) and processing of seismographic signals. In the case of blind deconvolution, the number of estimated source signals in the preceding steps for constructing test data sets is reduced to one. Either extended or different approaches [2] (see page 336) [3] (see page 357-361) are necessary to identify a mapping of the separation system from a multiple deconvolutive input to a single output. Again, the underlying model can be assumed time invariant or time variant (recall step five). Therefore, it is necessary to determine deconvolution parameters $\vec{m}$ (3) either once or for any quasi-stationary blocksize.

A SIMO system is extended to a multichannel convolutive model by adaption of model parameter m as a vector in (3) to a matrix in (4).

$$\vec{u}(t) = \sum_{l:=-\delta}^{+\delta} M(l) \cdot \vec{s}(t-l) + \vec{n}(t), \tag{4}$$

In the context of the cocktail party problem, each observed mixture contents multiple superpositions of each speaker.

7. For synthetic reference data sets for multichannel blind deconvolution (MBD) the mixing system is extended to multiple sources (4). The useful range of parameters sets is predetermined by the application domain (refer to step 6). Like in the preceding step, approaches for BBS are extended for MBD [2] (see page 354 for some references) and moreover, the frequency domain representation of the signals is necessary for performing MBD in many applications. As in the SIMO case before, the underlying model can be assumed time invariant or time variant (recall step five and six). Therefore, it is necessary to determine deconvolution parameters of matrix M (4) either once or for any quasi-stationary blocksize.

The necessity for the convolutive model is depending strongly on the application domain. For instance, it appears in EEG recordings due to distortion by the sculp and the slow propagation velocity, but it did not appear in MEG recordings [3] (see page 446). Moreover, the constellation of the microphone and speaker position and the special features of the surrounding are determining the necessary complexity of the assumed underlying model, see [2] for example (page 35).

4.4 Reference sets for the nonlinear case

In general, the non-linear case of the underlying model is not solvable by ICA [23]. Therefore, some explicit models for non-linearity are invertible by ICA approaches. For example, the post non-linear model (5, PNL) with a vector of non-linear functions $\vec{f}$.

$$\vec{u}(t) = \vec{f}(M \cdot \vec{s}(t)), \tag{5}$$

8. Synthetic reference data sets for an underlying nonlinear ICA are restricted to all specific invertible models, like PNL, Wiener model, Hammerstein model to list only a few [2] (see page 443 et seq.) [3] (see page 315 et seq.). A comparison of the existing approaches by reference data sets seems one of the most challenging problems. In respect to the guideline steps listed so far, the comparison of the capability due to the different nonlinear mixing seems to be the most promising.

4.5 Reference data and source collections

Meanwhile, a number of web-spaces collect different ICA-approaches and data sources. For further developments, all available sources with a proper documentation should be collected as a part of a global reference, for instance called ICA-test suite reference center. As a starting point, a few selected resources known to the author are listed. The date of existing public access has been March 11th, 2005.

- ICA-Central:
  http://www.tsi.enst.fr/icacentral/icalistArchive/
  ICA-central mailing list, data sets, algorithms, submitting interface for data and code.

- ICALAB-toolboxes:
  www.bsp.brain.riken.jp/ICALAB/
  ICA, BSS, PCA, BSE matlab-packages for signal and image processing package. At least 18 different approaches so far, sample files as benchmarks [2].

- ICA-CNL (Computational Neurobiology Laboratory at the Salk Institute):
  http://www.cnl.salk.edu/ tewon/ica_cnl.html
  Some algorithm code, matlab ICA toolbox for EEG analysis, animated examples and demos, links.

- ICA-link collection
  http://www.cis.hut.fi/projects/ica/book/
  Algorithm and data sets, links to the International workshop on ICA from 2000-2004, with all papers online besides the proceedings of 2004 [1].

Although enormous amounts of uncorrelated data sets can be found, a systematic reference data suite for each specific application domain is still missing.

5 Conclusion

Like all ideas of major influence on the development of science, the idea of a rigorous test suite is common sense in many different areas. Even for BSS, any comparison and even each approach tested on test data sets requires a fundamental understanding on the statistical variety of data occuring for application cases. The closest approach to the one presented in this paper can be found in [18]. Despite the consequent divided cases for each robust underlying statistical model in the test suite concept here, all major dependencies are identified in a consequently manner. Despite this analysis, the contribution is more in a practical sense by four sample routines for evaluation software for BSS and a first attempt of a small collection of sample test sets with online access.
A major conclusion for comparison is as follows. The concept presented here, fills the gap between the theoretical similarities of ICA approaches and the necessity for significant empirical verification. To overcome the limitation of only limited simulation on sample data sets, a universal concept for the design of consequent test suites for any application area of BSS, BD, MBD and the nonlinear are concerned. The sharp difference between the three main assumed underlying compositions is important.

References

[1] C.G. Puntonet and A. Prieto (Ed.), Independent Component Analysis and Blind Signal Separation, *Proc. of Fifth Int. Conference, ICA 2004*, LNCS 3195.

[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, Wiley 2002.

[3] E. Oja, A. Hyvärinen and J. Karhunen, *Independent Component Analysis*. (NY: Wiley, 2003).

[4] A. Bell and T. Sejnowski, An Information-Maximisation Approach to Blind Signal Separation and Blind Deconvolution, *J. of Neural Computation*, Vol.7(1), 1995, pp. 1129–1159.

[5] P. Comon, Independent Component Analysis – A New Concept, *IEEE Trans. on Signal Processing*, Vol. 36, 1994, pp. 287-314.

[6] C. Jutten and J. Herault, Blind Separation of Sources, Part I : An Adaptive Algorithm Based on Neuromimetic Architecture, *IEEE Trans. on Signal Processing*, Vol. 24(1), 1991, pp.1–10.

[7] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, (NY: John Wiley, 1993).

[8] S.-I. Amari and J-F. Cardoso, Blind Source Separation-semiparametric statistical approach, *IEEE Transaction on Signal Processing*, Vol. 45(11), 1997, pp. 2692-2700.

[9] E. Oja, From neural learning to independent components, *Neurocomputing*, Vol. 22, 1998, pp. 187–199.

[10] B. Laheld and J.F. Cardoso, Adaptive Source Separation with Uniform Performance, *EU-SIPCO*, Vol. 2, 1994, pp. 183–186.

[11] M. Girolami, *Self-Organizing Neural Networks- Independent Component Analysis and Blind Source Separation*, (Springer Verlag, 1999).

[12] C. Hagen, *Neural Networks and Multivariate Statistical Data Analysis* (in German), (University of Darmstadt: dissertation, 1997).

[13] M. Borschbach, B. Freisleben and C. Hagen, Blind Separation of Acoustic Signals Using a Neural Network, *Proc. of the 15th Int. Conference on Applied Informatics* , 1997, pp. 96–99.

[14] M. Borschbach, B. Freisleben and C. Hagen, A Neural Network for Blind Acoustic Signal Separation, *Proc. of the 1996 Int. Conference on Robotics, Vision and Parallel Processing for Industrial Automation*, Vol. 1, 1996, pp. 67–72.

[15] M. Borschbach, B. Freisleben and C. Hagen, A Neural Network for the Blind Separation of Non-Gaussian Sources, *Proc. of IEEE World Congress, Int. Join Conference on Neural Networks IJCNN'1998*, pp. 837-842.

[16] M. Borschbach, Analysing MEG-data by a combination of different Neural Networks, *Proc. of First Asia-Pacific Bioinformatics Conference*, Volume 19, 2003, pages 181-187.

[17] M. Borschbach and M. Schulte, Performance Analysis of Learning Rules for the Blind Separation of Magnetencephalography Signals, *Proc. of ICA'99, First Int. Workshop on Independent Component Analysis and Signal Separation*, 1999, pp. 341-346.

[18] D. Schobben, K. Torkkola and P. Smaragdis, Evaluation Blind Signal Separation Methods, *Proc. of ICA'99, First Int. Workshop on Independent Component Analysis and Signal Separation*, 1999, pp. 261-266.

[19] X. Giannakopoulos, J. Karhunen and E. Oja, An experimental comparison of neural algorithms for independent component analysis and blind separation, *Int. J. of Neural Systems*, 1999, Vol. 9, No. 2.

[20] N.N. Schraudolph and X. Giannakopoulos, Online Independent Component Analysis With Local Learning Rate Adaptation, In *Advances in Neural Information Processing Systems 12 (NIPS'12)*, 2000.

[21] M. Borschbach, M. Schulte and A. Knief, Aspects of Neuronal Learning Rules for the Blind Separation of Independent Brain Activities, *Proc. of ICS'99, IASTED Int. Conference on Intelligent Systems and Control*, 1999, pp. 143-148.

[22] A. Ahmed. *Signal Separation*. (University of Cambridge: PhD-thesis, 2000).

[23] A. Taleb and C. Jutten. Source separation in postlinear mixtures. *IEEE Trans. on Signal Processing*, Vol. 47(10), 1999, pp. 2807-2820.