A BLIND EXTRACTION OF TEMPORALLY CORRELATED BUT STATISTICALLY DEPENDENT ACOUSTIC SIGNALS

Andrzej Cichocki, Tomasz Rutkowski,
Allan Kardec Barros, Sang-Hoon Oh.
Brain Science Institute, RIKEN,
2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
e-mail: cia@brain.riken.go.jp.
http://www.bsp.brain.riken.go.jp

Abstract. **In this paper we propose a batch learning algorithm for sequential blind extraction of arbitrary distributed but generally not i.i.d. (independent identically distributed) temporally correlated sources, possibly dependent speech signals from from linear mixture of them. The proposed algorithm is computationally very simple and efficient, it is based only on the second order statistics and in contrast to the most known algorithms developed for the sequential blind extraction and independent component analysis, do not assume statistical independence of source signals neither non-zero kurtosis for the sources, thus statistical dependent signals including sources with extremely low or even zero kurtosis (colored Gaussian with different spectra) can be also successfully extracted. Extensive computer simulation confirm the validity and high performance of the proposed algorithm.**

INTRODUCTION

Recently there was an increase of interest in blind signal processing, especially, instantaneous blind source separation and related problem independent component analysis (ICA) due to potential applications in many areas, including speech enhancement, rotating machine vibration analysis and diagnosis, geophysical data processing, data mining, wireless communication and biomedical signal processing [1-18]. Most the known and efficient techniques assume that primary sources are statistically independent, however in many applications the sources are not completely independent and developed ICA algorithms may fail to extract signals of interest. In this paper we assume that sources can be statistically dependent or highly correlated but they have various temporal structures.

For our purposes, the problem of blind source separation (BSS) and blind source extraction (BSE) can be formulated as follows: we observe sensor signals $\mathbf{x}(k) = [x_1(k), \ldots, x_m(k)]^T$ described by matrix equation:

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{n}(k),\tag{1}$$

where $\mathbf{A}$ is $m \times n$ full-rank unknown mixing matrix, $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ ($n \leq m$) is a vector of unknown temporally correlated sources (with different autocorrelation) and $\mathbf{n}(k)$ is a vector of additive white (i.i.d.) noise. The objective is to estimate all primary source signals (BBS problem) or extract some of them sequentially one-by-one (BSE problem) assuming that they are non i.i.d. temporally correlated with different autocorrelation functions.

In this paper we assume that 'interesting' source signals, (which should be estimated) have specific temporal structures and can be modeled by unknown stable autoregressive models [1, 13, 3] (see Fig.1) as:

$$s_i(k) = \tilde{s}_i(k) - \sum_{p=1}^{M} a_{ip} s_i(k-p) = \tilde{s}_i(k) - A_i(z)s_i(k),\tag{2}$$

where $\tilde{s}_i(k)$ is i.i.d. (white) innovative process and $s_i(k)$ are signals to be estimated. To such category belong natural speech signals and some biomedical signals. Temporal structures of sources can be modeled by more general means, e.g. using ARMA (autoregressive moving average) process or HMM (Hidden Markov Model) which would be able to represent high-order temporal statistics and facilitates EM learning [1, 2].

There are in general two approaches for estimating the source signals: the simultaneous blind source separation approach and sequential (one by one) blind extraction. In the blind source separation (BSS) approach a neural network and associated learning algorithm performs separation of all sources simultaneously, while in the blind source extraction (BSE) the sources are extracted on-by-one by eliminating the already extracted sources from their mixture by using deflation techniques. Blind signal extraction approach has several advantages over simultaneous blind source separation: like global convergence (avoiding spurious equilibrium points), relative low complexity of learning algorithms, especially when the number of sources is large, extraction only of signals of interest with stochastic specific features [6, 10, 8, 11].

The main objective of this paper is to present a simple linear processing element and associated new and efficient batch learning algorithm for BSE of temporally correlated, arbitrary distributed but generally non i.i.d. (not independent identically distributed), i.e. $E\{s_i(k-p)s_i(k)\} \neq 0$ for some p $(p = 1, 2, \ldots, L)$.

Most of the known and efficient algorithms for blind source extraction assume that sources are statistically independent, they are non-Gaussian and have non-zero kurtosis. Their performance and convergence for low value of kurtosis may be often poor [6, 7, 10, 8]. Furthermore, those algorithms use typically higher order statistics and some nonlinear activation functions

whose optimal choice depends on the unknown distribution of sources. Especially, some acoustic signals and biomedical source signals can be dependent and they can be characterized by extremely low value of normalized kurtosis and due to nonstationarities their distribution may change in time. In this paper we propose a simple and efficient algorithm which avoids these problems in the sense that it employs only second order statistics and enables to successfully extract non-stationary statistically dependent temporally correlated sources and also sources with extremely low or even zero value of normalized kurtosis (colored Gaussian signals with different spectra).

It should be mentioned that recently several researchers have developed a number of efficient algorithms for sequential blind source extraction, especially works of Delfosse and Loubaton [7], Hyvärinen and Oja [10], Malouche and Macchi [11], Douglas and Kung [9] and Tugnait [15]. Moreover, the use of second statistics for BSS problem and problem of separation of temporally correlated sources has been discussed by Amari [1], Molgdey and Schuster [12], Pham and Garat [14], Belouchrani et al. [4], Pearlmutter and Parra [13], Choi and Cichocki [5]. However, our approach is different and it has some advantages that may not be found in others at the same time. It is computationally simple; it provides relative fast convergence; it extracts the components sequentially so that the user can decide in advance how many sources with specific stochastic properties should be extracted; does not assume non-zero kurtosis for the sources; does not need the sources to be stationary. Furthermore, the proposed algorithm is robust in respect to white additive noise what often leads to smaller errors (cross-talking between estimated sources).

BLIND EXTRACTION OF A SINGLE SOURCE

Let us assume for simplicity that we want to extract only one single source signal, say $s_j(k)$, from the available sensor vector $\mathbf{x}(k)$. For this purpose we design a single processing unit described as (see Figs. 1 and 2):

$$y_1(k) = \mathbf{w}_1^T \mathbf{x}(k) = \sum_{j=1}^{m} w_{1j} x_j(k), \tag{3}$$

$$\varepsilon_1(k) = y_1(k) - \sum_{p=1}^{L} b_{1p} y_1(k-p)$$

$$= \mathbf{w}_1^T \mathbf{x}(k) - \mathbf{b}_1^T \widetilde{\mathbf{y}}_1, \tag{4}$$

where $\mathbf{w}_1 = [w_{11}, w_{12}, \ldots, w_{1m}]^T$,
$\tilde{\mathbf{y}}_1 = [y_1(k-1), y_1(k-2), \ldots, y_1(k-L)]^T$, $\mathbf{b}_1 = [b_{11}, b_{12}, \ldots, b_{1L}]^T$, and $B_1(z) = \sum_{p=1}^{L} b_{1p} z^{-p}$ is a transfer function of the corresponding FIR filter. It should be noted that the FIR filter can have in a special case a sparse
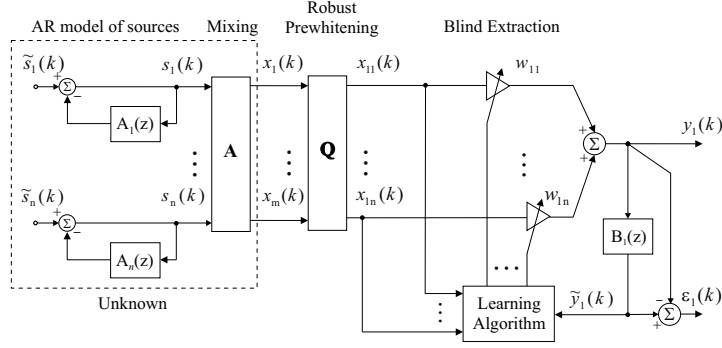
Figure 1: Block diagram illustrating implementation of learning algorithm for blind extraction of a single temporally correlated source modeled by AR model.

representation, i.e. only one single processing unit, say with delay p and $b_{1p} \neq 0$ can be used instead of L delays.

The processing unit has two outputs: $y_1(k)$ which estimates the extracted source signals, and $\varepsilon_1(k)$, which represents an error, after passing the output signal $y_1(k)$ by the FIR filter (see Fig. 2).

Our objective is to estimate the optimal values of vectors $\mathbf{w}_1$ and $\mathbf{b}_1$, in such way that the processing unit extracts successfully one of the sources. This is achieved if the global vector defined as $\mathbf{g}_1 = \mathbf{A}^T \mathbf{w}_1 = \left(\mathbf{w}_1^T \mathbf{A}\right)^T = c_j \mathbf{e}_j$ contains only one nonzero element, say in the j-th row, then $y_1(k) = c_j s_j$, where c_j is an arbitrary nonzero scaling factor. For this purpose we reformulate the problem as minimization of a cost function $J(\mathbf{w}_1, \mathbf{b}_1) = E\left\{\varepsilon_1^2\right\}$. The main justification of applying such cost function is an assumption that primary sources signals (signals of interest) have a specific temporal structure and can be modeled by the autoregressive model (2) and we want to estimate unknown innovative processes (and simultaneously identify the temporal structure and estimate the primarily sources $s_i(t)$ [1, 13, 3] (see Fig.1).

Derivation of Basic Learning Algorithm

The cost function can be evaluated as follows:

$$
\begin{aligned}
J(\mathbf{w}_1, \mathbf{b}_1) &= E\left\{\varepsilon_1^2\right\} \\
&= \mathbf{w}_1^T \hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{w}_1 - 2\mathbf{w}_1^T \hat{\mathbf{R}}_{\mathbf{x}\bar{\mathbf{y}}_1} \mathbf{b}_1 + \mathbf{b}_1^T \hat{\mathbf{R}}_{\bar{\mathbf{y}}_1 \bar{\mathbf{y}}_1} \mathbf{b}_1,
\end{aligned}
\tag{5}
$$

where $\hat{\mathbf{R}}_{\mathbf{xx}} \simeq E\{\mathbf{xx}^T\}$, $\hat{\mathbf{R}}_{\mathbf{x}\bar{\mathbf{y}}_1} \simeq E\{\mathbf{x}\tilde{\mathbf{y}}_1^T\}$ and $\hat{\mathbf{R}}_{\bar{\mathbf{y}}_1 \bar{\mathbf{y}}_1} \simeq E\{\tilde{\mathbf{y}}_1 \tilde{\mathbf{y}}_1^T\}$, mean estimation of true values of covariance and cross-correlation matrices: $\mathbf{R}_{\mathbf{xx}}, \mathbf{R}_{\mathbf{x}\bar{\mathbf{y}}_1}$, $\mathbf{R}_{\bar{\mathbf{y}}_1 \bar{\mathbf{y}}_1}$, respectively. In order to estimate vectors $\mathbf{w}_1$ and $\mathbf{b}_1$ we evaluate the gradients of the cost function and equalize them to zero as follows:

$$
\frac{\partial J_1(\mathbf{w}_1, \mathbf{b}_1)}{\partial \mathbf{w}_1} = 2\hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{w}_1 - 2\hat{\mathbf{R}}_{\mathbf{x}\bar{\mathbf{y}}_1} \mathbf{b}_1 = \mathbf{0},
\tag{6}
$$

$$\frac{\partial J_1\left(\mathbf{w}_1, \mathbf{b}_1\right)}{\partial \mathbf{b}_1} = 2\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1\tilde{\mathbf{y}}_1}\mathbf{b}_1 - 2\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1\mathbf{x}}\mathbf{w}_1 = \mathbf{0}. \tag{7}$$

Solving the above matrix equations we obtain an iterative algorithm

$$\mathbf{w}_1 = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{x}\tilde{\mathbf{y}}_1}\mathbf{b}_1, \tag{8}$$

$$\mathbf{b}_1 = \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1\tilde{\mathbf{y}}_1}^{-1}\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1\mathbf{x}}\mathbf{w}_1 = \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1\tilde{\mathbf{y}}_1}^{-1}\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1 y_1}, \tag{9}$$

where the matrices $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1\tilde{\mathbf{y}}_1}$ and $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1 y_1}$ are estimated on basis of parameters $\mathbf{w}_1$ obtained in previous iteration step.

Remark: It should be emphasized here that in our derivation we have assumed that $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1\tilde{\mathbf{y}}_1}$ and $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1 y_1}$ are independent of actual evaluated vector $\mathbf{w}_1$, i.e. they are estimated on value of $\mathbf{w}_1(k-1)$ in the previous iteration step. This two phase procedure is similar to EM schemes: (i) freeze the covariance and cross-correlation matrices and learn the parameters of the processing unit $(\mathbf{w}_1, \mathbf{b}_1)$; (ii) freeze $\mathbf{w}_1$ and $\mathbf{b}_1$ and learn new statistics (i.e. matrices $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_1 y_1}$ and $\mathbf{R}_{\tilde{\mathbf{y}}_1\tilde{\mathbf{y}}_1}$) of estimated source signal, then go back to (i) and repeat. Hence, in phase (i) our algorithm extract a source signal, whereas in phase (ii) it learns the statistics of the source.

In order to avoid the trivial solution $\mathbf{w}_1 = \mathbf{0}$ we can perform the normalization of the vector $\mathbf{w}_1$ to unit length in each iteration step as $\mathbf{w}_{1*} = \mathbf{w}_1 / \|\mathbf{w}_1\|$ (what ensures that $E\{y_1^2\} = 1$).

The above algorithm can be considerably simplified. It should be noted that in order to avoid the inversion of autocorrelation matrix $\mathbf{R}_{\mathbf{xx}}$ in each iteration step we can apply as preprocessing the standard pre-whitening or standard PCA (principal component analysis) and next normalization of sensor signals to unit variance. In such case $\hat{\mathbf{R}}_{\mathbf{xx}} \approx \mathbf{I}_n$ and the algorithm simplifies to

$$\mathbf{w}_1 = \hat{\mathbf{R}}_{\mathbf{x}\tilde{\mathbf{y}}_1}\mathbf{b}_1 = \hat{\mathbf{R}}_{\mathbf{x}\overline{\mathbf{y}}_1}, \tag{10}$$

where $\hat{\mathbf{R}}_{\mathbf{x}\overline{y}} \cong \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}\left(k\right)\overline{y}_1\left(k\right)$ and $\overline{y}_1 = \mathbf{b}_1^T\tilde{\mathbf{y}}_1 = \sum_{p=0}^{L}b_{1p}y_1\left(k-p\right)$.

Remark: It should be noted that in order to reduce influence of additive white noise we can perform whitening by diagonalizing instead of standard covariance matrix $\hat{\mathbf{R}}_{xx} = \left\langle \mathbf{x}(k)\mathbf{x}^T(k)\right\rangle$ the following covariance matrix

$$\tilde{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{2}\left[\left\langle \mathbf{x}(k)\mathbf{x}^T(k-\Delta)\right\rangle + \left\langle \mathbf{x}(k-\Delta)\mathbf{x}^T(k)\right\rangle\right] = \mathbf{V}\Lambda\mathbf{V}^T \tag{11}$$

with linear transformation $\mathbf{x}_1 = \mathbf{Q}\mathbf{x} = \mathbf{V}^T\Lambda^{-1/2}\mathbf{x}$. Furthermore, it is worth also to note that the learning rule (9) can be simplified as

$$b_{1p} = \frac{E\{y_1(k-p)y_1(k)\}}{E\{y_1^2(k-p)\}}, \quad b_{iq} = 0 \text{ for } q \neq p \tag{12}$$

if we use single delay units with time delay $\tau = p$ with $1 \leq p \leq L$ instead of FIR filter with lengths L. The above formulas (10) and (9) or (12) builds up our basic simplified learning algorithm.
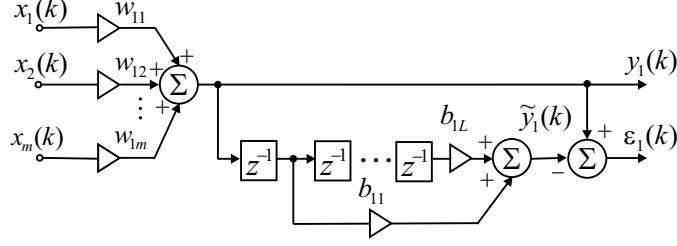
Figure 2: The neural network structure for single processing unit.

A length of FIR filter should be chosen sufficiently large but a value of $L \approx 10$ was enough in our experiments. However, as shown by our extensive computer simulations, in practice it is sufficient to use only a single time delay unit with suitably chosen delay $\tau = p$ if some a priori information about temporal structure of source signals is available. The suitable choice of single time delay p can be determined form the characteristic pics in autocorrelation function of sensor signals [3]. Using different time delays $\tau = p$ we able to extract different sources. However, we have not guarantee that we extract various signals for different time delays, thus to avoid this problem in general case we can apply deflation approach described in the next section.

DEFLATION AND BLIND IDENTIFICATION OF MIXING MATRIX ALGORITHM

We now describe a simple and efficient deflation procedure. After the successful extraction of the first source signal $y_1(k) \sim s_j(k)$, we can apply the deflation procedure which removes the previously extracted signals from the mixtures. This procedure may be recursively applied to extract sequentially the rest of the mixing source signals. This means that we require an on-line linear transformation [6] given by,

$$\mathbf{x}_{i+1}(k) = \mathbf{x}_i(k) - \widetilde{\mathbf{w}}_i y_i(k) \qquad i = 1, 2, \dots, \tag{13}$$

where $\mathbf{x}_1(k) = \mathbf{x}(k) = [x_1(k), \dots, x_m(k)]^T$, which ensures minimization of the mean squares cost (loss) function

$$\tilde{J}_i(\widetilde{\mathbf{w}}_i) = E\{\mathbf{x}_{i+1}^T \mathbf{x}_{i+i}\} =$$
$$E\{\mathbf{x}_i^T \mathbf{x}_i\} - 2\widetilde{\mathbf{w}}_i^T E\{\mathbf{x}_i y_i\} + \widetilde{\mathbf{w}}_i^T \widetilde{\mathbf{w}}_i E\{y_i^2\}. \tag{14}$$

Minimization of this cost function with respect to $\widetilde{\mathbf{w}}_i$ leads to a simple updating equation,

$$\widetilde{\mathbf{w}}_i = \hat{\mathbf{a}}_i = \frac{E\{\mathbf{x}_i y_i\}}{E\{y_i^2\}} = \frac{E\{\mathbf{x}_i \mathbf{x}_i^T\} \mathbf{w}_i}{E\{y_i^2\}}, \tag{15}$$

where $\hat{\mathbf{a}}_i$ is, in fact, an estimated column of the mixing matrix $\mathbf{A}$ neglecting arbitrary scaling and permutation of columns ambiguities. It is worth to
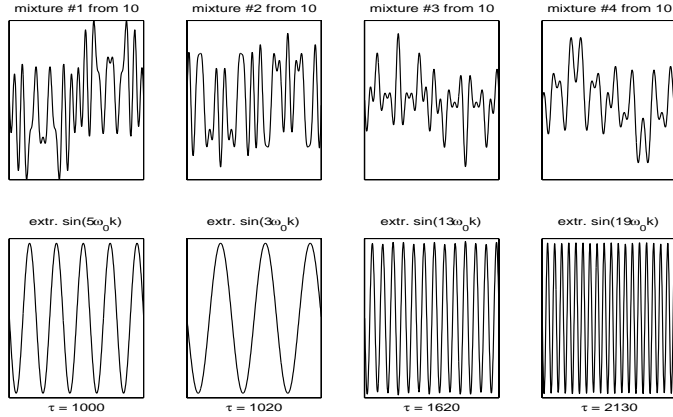
Figure 3: Exemplary blind extraction of four sinusoidal signals for different time delays.

noting that in special case when sensor signals are pre-whitened the above formula simplifies to $\widetilde{\mathbf{w}}_i = \hat{\mathbf{a}}_i = \mathbf{w}_1$ since in such case the mixing and demixing matrix are orthogonal, i.e. $\mathbf{W} = \mathbf{A}^T$. The procedure can be continued until all the all source signals are recovered, that is, until the amplitude of each signal $\mathbf{x}_{i+1}$ is below a preassigned threshold. It can be proved that this deflation optimization procedure has no spurious (undesired) minima and after the convergence the algorithm (15) estimates one column of the mixing matrix $\mathbf{A}$ with scaling and permutation indeterminacy.

SIMULATION RESULTS

We have extensively carried out many simulations where we synthetically mixed source signals and extracted them by the proposed algorithm. Due to limit of space here we show only two illustrative examples.

EXAMPLE 1: In the first example we used ten sinusoidal signals $s_i(k) = \sin\left[(2i+1)\omega_0 k\right]$ where $\omega_0 = 2\pi 100$ and $i = 0, 1, \ldots, 9$, that have been mixed by randomly generated mixed nonsingular matrix $\mathbf{A}$. It should be noted the that sources are dependent. The autocorrelation functions of sensor signals have been plotted and suitable time delays $\tau_i = p_i = \{270, 280, 360, 420, 430, 1000, 1020, 1620, 2130\}$ has been estimated on basis of picks of autocorrelation functions. The exemplary four mixed and extracted signals are depicted on Fig. 3. Since source signals are harmonic, there is also possibility to apply FFT to sensor signals to estimate time delays, but that approach is sensitive to phase shifts of source sinusoidal signals.

EXAMPLE 2: In the second example three natural speech signals of 3 different persons pronouncing simultaneously the same sentence (hello goodmorning -in Polish language) have been mixed using randomly chosen nonsingular mixing matrix. The source signals are highly correlated and standard ICA algorithms provide rather poor performance. On basis of common picks
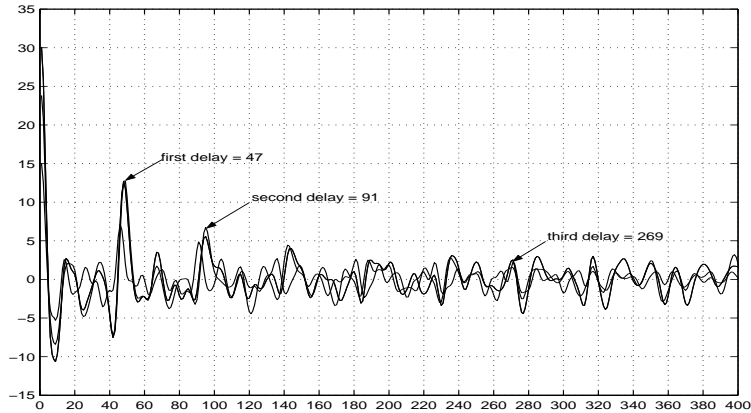
Figure 4: The autocorrelation functions of 3 mixed speech signals.

of all autocorrelation functions we estimated the desired time delays (see Fig.4). Simulation results are shown in Fig.5. The original speech signals have been reconstructed very well.

In fact, we found that the most of known ICA algorithms for the sequential blind signal extraction fail to separate successfully all such source signals because some of source signals are statistically dependent. However, second order statistics algorithms can do also such separation (see e.g. [5]).

DISCUSSIONS AND CONCLUSION

We have developed a batch efficient algorithm for sequential (one by one) extraction of sources with temporal structures. The algorithm is computationally simple; it provides relative fast convergence; it extracts the components sequentially so that the user can decide in advance how many sources with specific stochastic properties should be extracted; does not assume non-zero kurtosis for the sources; does not need the sources to be stationary neither statistically independent. Furthermore, the proposed algorithm is robust in respect to white additive noise what leads to smaller errors (cross-talking between estimated sources). We have confirmed by extensive computer simulations that the derived algorithm performs well when the source signals are temporally correlated even if they have very low value of the normalized kurtosis or even they statistically dependent. In more general, case, when some of 'interesting' source signals are i.i.d. signals we can switch between the fast and very efficient fixed point algorithm [10] of Hyvärinen and Oja and algorithm proposed in this paper in order to separate successfully and efficiently all desired sources.

REFERENCES

[1] S. Amari, "ICA of temporally correlated signals - learning algorithm," in **Proceedings of ICA'99: International workshop on blind signal separation and independent component analysis**, Aussois, France, Jan. 1999, pp. 13–18.

[2] H. Attias, **Independent factor analysis with temporally structured factors**, MIT Press, Cambridge, MA, chap. 12: Advances in Neural Information Processing Systems, p. (in print), edited by: T. Leen et al.

[3] A. K. Barros and A. Cichocki, "RICA - Reliable and robust program for Independent Component Analysis," Report and matlab program of riken, **Brain Science Institute RIKEN**, 2-1 Hirosawa, Wako-shi, Saitama, 351-0198 JAPAN, http://www.riken.nagoya.jp/sensor/allan/RICA or http://go.to/RICA.

[4] A. Belouchrani, K. Meraim and J.-F. Cardoso, "A blind source separation technique using second order statistics," **IEEE Transactions on Signal Processing**, vol. 45, pp. 434–444, February 1997.

[5] S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures," **Electronic Letters**, vol. 36, pp. 848–849, April 2000.

[6] A. Cichocki, R. Thawonmas and S. Amari, "Sequential blind signal extraction in order specified by stochastic properties," **Electronics Letters**, vol. 33, no. 1, pp. 64–65, January 1997.

[7] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach," **Signal Processing**, vol. 45, pp. 59 – 83, 1995.

[8] S. Douglas and S.-Y. Kung, "KuicNet algorithm for blind deconvolution," in **Proceedings of the 1998 IEEE Workshop on Neural Networks for Signal Processing**, New York, 1998, pp. 3–12.

[9] S. Douglas and S.-Y. Kung, "An Ordered-rotation KuicNet algorithm for separating arbitrarily-distributed sources," in **Proceedings of ICA'99: International workshop on blind signal separation and independent component analysis**, Aussois, France, January 1999, pp. 419–425.

[10] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," **Neural Computation**, vol. 9, pp. 1483 – 1492, 1997.

[11] Z. Malouche and O. Macchi, "Adaptive un-supervised extraction of one component of linear mixture with a single neuron," **IEEE Transactions Neural Networks**, vol. 9, no. 1, pp. 123–138, 1998.

[12] L. Molgedey and H. Schuster, "Separation of a mixture of independent signals using time-delayed correlations," **Physical Review Letters**, vol. 72, no. 23, pp. 3634–3637, 1994.

[13] B. A. Pearlmutter and L. C. Parra, "Maximum Likelihood blind source separation: A context-sensitive generalization of ICA," in **Proceedings of NIPS'96**, 1997, vol. 9, pp. 613–619.

[14] D. Pham and P. Garat, "Blind separation of mixtures of independent sources through a quasi-maximum likelihood approach," **IEEE Transactions on Signal Processing**, vol. 45, no. 7, pp. 1712–1725, July 1997.

[15] J. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," **IEEE Transactions on Signal Processing**, vol. 45, pp. 268–271, January 1997.
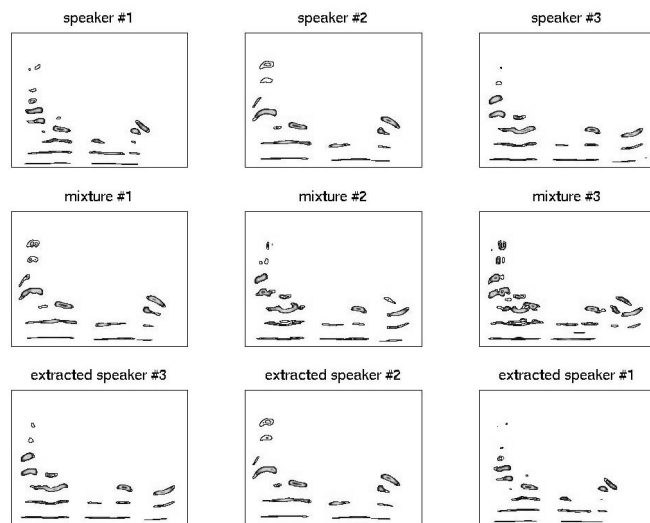
(a)
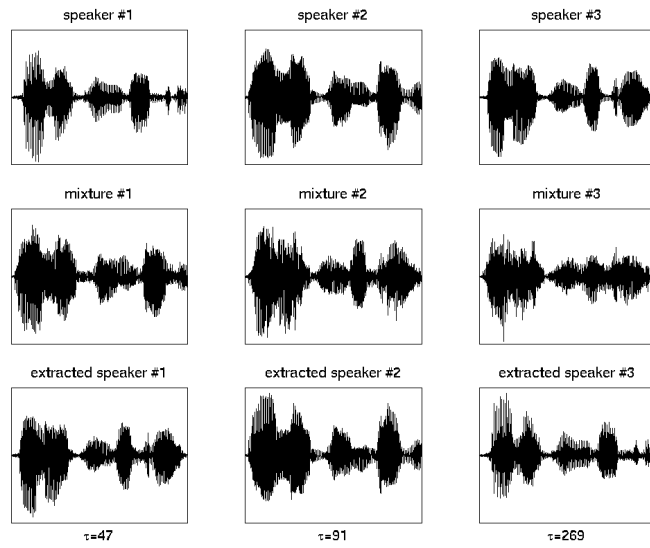
(b)

Figure 5: Example 2 Blind extraction of dependent speech signals for different time delays: (a) in time domain, (b) in time-frequency domain.