Blind Separation For Instantaneous Mixture of Speech Signals: Algorithms and Performances.

Ali MANSOUR, Mitsuru KAWAMOTO and Noburo OHNISHI.

Bio-Mimetic Control Research Center (BMC - RIKEN),

2271-130, Anagahora, Shimoshidami, Moriyama-ku, Nagoya 463 (JAPAN)

email: mansour@ieee.org, kawa@nagoya.riken.go.jp, and ohnishi@sheraton.ohnishi.nuie.nagoya-u.ac.jp

Tel/Fax: +81 - 52 - 736 - 5867 / 5868

http://www.bmc.riken.go.jp

Abstract

Because it can be found in many applications, the Blind Separation of Sources (BSS) problem has raised an increasing interest. According to the BSS, one should estimate some unknown signals (named sources) using multisensor output signals (i.e. observed or mixing signals). For the Blind Separation of Sources (BSS) problem, many algorithms have been proposed in the last decade. Most of these algorithms are based on High Order Statistics (HOS) criteria.
In this paper, we focus on the blind separation of non-stationary signals (music, speech signal, etc) from their linear mixtures. At first, we present briefly the idea behind the separation of non-stationary sources using Second Order Statistics (SOS). After that, we introduce and compare three possible separating algorithms.

Keywords: Decorrelation, Second order Statistics, Whiteness, Blind Separation of Sources, Natural Gradient, Kullback-Leibler Divergence, Hadamard Inequality, Jacobi Diagonalization, and Joint Diagonalization.

1 Introduction

This problem was initially proposed by Hérault *et al.* to study some biological phenomena [1]. Actually, the BSS model can be found in different situations [?]: radio-communication (in mobile-phone as SDMA (Spatial Division Multiple Access) and free-hand phone), speech enhancement [2], separation of seismic signals [3], sources separation method applied to nuclear reactor monitoring [4], airport surveillance [5], noise removal from biomedical signals [6], *etc.*

In our laboratory (BMC), we are involving in the application of signal processing and BSS [7, 8] in robotics and artificial life as in the following scenario: In our environment, there are many kinds of sound sources, human voices, phone bell, fan noise, radio and so on. We, human, can discriminate each of sounds overlapping each other and recognize what sound exits at which direction. Thus we can understand our environment by sense of audition. This is called auditory scene analysis. Our goal is the realization of a new generation of smart robots. These robots, using sound discrimination along with sound separation among other capabilities, should imitated the behavior of human been.

In this paper, we briefly show that the second order statistics is enough to separate the instantaneous mixture of independent non-stationary signals. In addition, We also discuss and compare the behavior of three different algorithms for BSS of non-stationary signals.

- The first one the algorithm of Matsuoka *et al.* [9, 10] is based on on the minimization of Hadamard's inequality. This algorithm use indirectly the time correlation information of the sources to achieve the separation [10].

- The second algorithm use directly that information in the sense that it minimize the correlation matrix of the estimated sources. In another word that algorithm use the decorrelation (or whiteness) process of the estimated signals at different times.

- Finally, an algorithm based on a modified Jacobi diagonalization approach is discussed.

2 Transmission Model

Let us denote by $X = (x_i)$ the $p \times 1$ unknown source vector, $Y = (y_i)$ the p observation signals and by $S = (s_i)$ the p estimated sources (see Fig. 1). Let $\mathbf{M} = (m_{ij})$ denotes the channel effect or the unknown full-rank mixing matrix and $\mathbf{W} = (w_{ij})$ is the weight matrix. The relationships between the
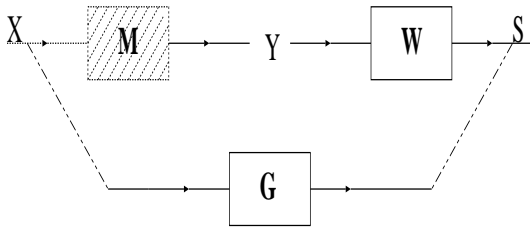
Figure 1: Channel Model

different vectors are given in the following:

$$
\begin{aligned}
Y &= \mathbf{M}X, &&(1)\\
S &= \mathbf{W}Y = \mathbf{W}\mathbf{M}X = \mathbf{G}X, &&(2)
\end{aligned}
$$

here $\mathbf{G}$ stands for the Global matrix. It is widely known that in the context of blind separation of instantaneous mixtures, one can only separate the source up to a scale factor and a permutation order [11]. In other words, the separation is considered achieved when the global matrix $\mathbf{G}$ becomes:

$$
\mathbf{G} = \mathbf{P}\boldsymbol{\Delta} \qquad (3)
$$

where $\mathbf{P}$ is any full-rank permutation matrix and Δ is any full-rank diagonal matrix.

3   Separation Approach

At first, Matsuoka et al. [9, 10] have showed that blind separation for non-stationary signals can be achieved by making the mixed signals uncorrelate with each other, if the variances of the source signals fluctuate independently of each other.

Independently from the previous approach and for two signals, it has been shown that the decorrelation of the output signals makes the weight matrix coefficients belong to a set of hyperbolas. And these hyperbolas have two intersection points which correspond to the blind separation solutions of non-stationary signals [12].

In general and for two or more sources, it was proved [12, 13] that the decorrelation of the output signals at any time means the separation of the non-stationary statistically independent sources. In other words, for the case of independent non-stationary (up to second-order statistics) sources such speech signals where the power of the signals can be considered as time variant, we proved, using geometrical information, that the decorrelation of the output signals at any time leads to the separation of the independent sources. In other words, for these kinds of sources, any algorithm can separate the sources if at

the convergence of this algorithm the covariance matrix of the output signals becomes a diagonal matrix at any time.

4   Algorithms & Experimental Results

In this section we discuss the ideas of three different algorithms and some experimental results are presented.

4.1   Minimization of Hadamard's inequality

Given that Hadamard's inequality [14] of an arbitrary positive semidefinite matrix $\mathbf{R} = (r_{ij})$ is defined by

$$
\prod_{i=1}^{p} r_{ii} \geq \det\{\mathbf{R}\}, \qquad (4)
$$

where the equality holds if and only if the matrix $\mathbf{R}$ is a diagonal matrix.

Matsuoka et al. [9, 10, 15] suggest the separation of non-stationary signals by minimizing, with respect to the weight matrix $\mathbf{W}$, a modified version of Hadamard's inequality (4) of the estimated source's covariance matrix $\mathbf{R} = \mathrm{E}\{S(n)\ \ S(n)^T\}$. Their practical method uses a nonnegative function $Q(\mathbf{W}, t)$ which takes the minimum (zero) only when the mixed signals are uncorrelated with each other, and achieves blind separation by modifying the parameters of the network such that the cost function takes the minimum:

$$
\min_{\mathbf{W}} \sum_{i=1}^{p} \log \mathrm{E}\{s_i^2(n)\} - \log \det\{\mathrm{E}\{S(n)\ \ S^T(n)\}\}.
$$
$$(5)$$

The last function is a nonnegative function that takes the minimum (zero) only when the output signals are uncorrelated with each other. The separating matrix $\mathbf{W}$ is obtained by minimizing the function (5) with steepest decent method as:

$$
\Delta w_{ij}^{\star} = \alpha \frac{y_i(t)y_j(t)}{\Phi_i}. \qquad (6)
$$

here $\mathbf{W}^{\star} = (w_{ij}^{\star}) = \mathbf{W}^{-1}$, $\Delta w_{ij}^{\star} = w_{ij}^{\star}(t) - w_{ij}^{\star}(t-1)$, $\Phi_i(t)$ denotes the moving average of $E[y_i^2(t)]$ given by $\Phi_i(t) = \beta\Phi_i(t) + (1-\beta)y_i^2(t)$, and α and β were set to 10^{-4} and 0.9 respectively. The separated signals are found after 15000 iterations using that algorithms. The performances of this algorithm is shown in Fig. 2.
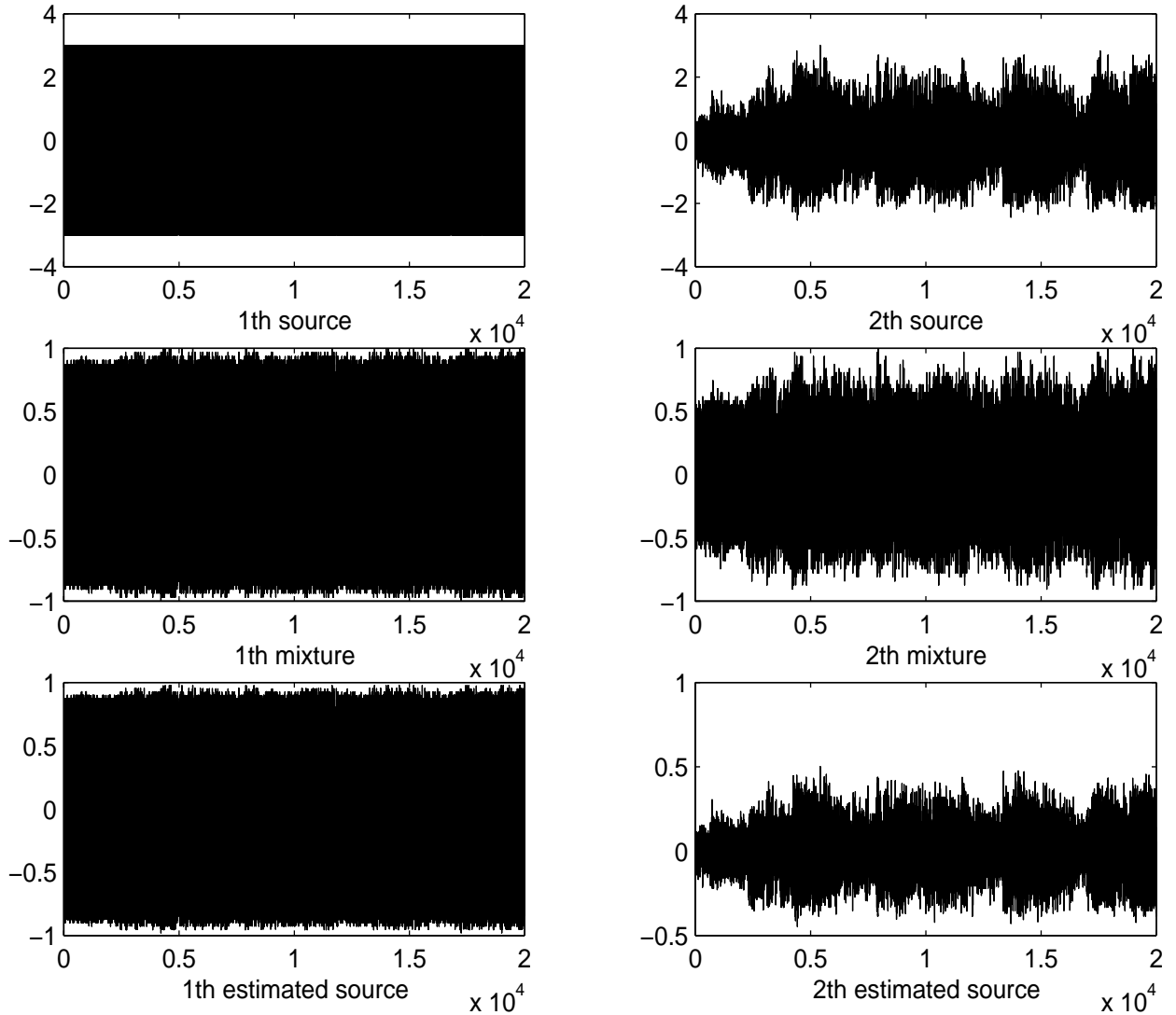
27

Figure 2: Hadamard's inequality: First column contains the signals of the first channel (i.e., first source, first mixture signal and the first estimated source), the second column contains the signals of the second channel.

4.2 Minimization of a Kullback divergence

The kull-back divergence between two random zero mean Gaussian vectors V_1 and V_2, with respectively two covariance matrix $\mathbf{I}$ and $\mathbf{R}$, is given by:

$$\delta(\mathbf{R}, \ \mathbf{I}) = \frac{1}{2}(\mathrm{Trace}\{\mathbf{R}\} - \log\det(\mathbf{R})) \geq 0, \qquad (7)$$

where $\mathbf{I}$ it can be considered as the $p \times p$ identity matrix, and $\mathbf{R} = \mathrm{E}\{S(n) \ \ S(n)^T\}$ is the $p \times p$ covariance matrix of the estimated sources $S(n)$. Thus the minimization of divergence (7) makes the matrix $\mathbf{R}$ close to an identity matrix (i.e., a diagonal matrix) and induces the separation of the sources, as we ex-

plained in the previous section.

The minimization of divergence (7) [13] is achieved according to the natural gradient [16, 17]. The advantage of this approach is that the algorithm and the updating rules are simple. However the convergence point of this criterion (7) is a $\mathbf{W}^*$ that makes the matrix $\mathbf{R}$ close to an identity matrix (i.e., a special diagonal matrix). It is obvious that this condition is more restrictive than the initial condition described in the previous section where $\mathbf{R}$ must simply be a diagonal matrix. In another hand and in many cases, the convergence has been observed over 3000 to 20000 iterations, depending on the sources (music, speech, with or without silent effect, high power, etc) and the mixing matrix.
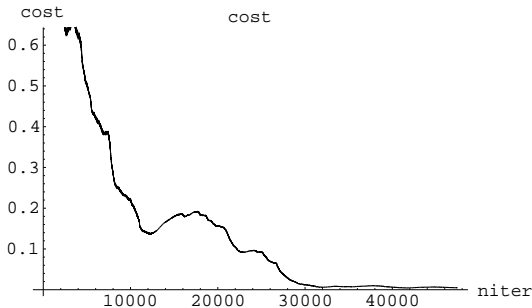
28

Figure 3: Evaluation of the Kullback Divergence with respect to the iteration number.

We conducted many experiments and found that the crosstalk was between -15 dB and -23 dB. The evaluation of the cost function with respect to the iteration number is shown in Fig. 3. Fig. 4 shows the experimental results of the separation of two speech sources.

4.3 Jacobi Diagonalization Method

It has been shown that the joint diagonalization [18] (i.e. a generalization of the Cyclic Jacobi diagonalization method [19]) of the fourth order cumulants matrices can separate blindly the independent sources [20]. In addition, Belouchrani *et al.* [21], using the joint diagonalization (i.e. JADE algorithm), have derived a second order statistics criterion to separate correlated stationary signals.

According to the previous study [12], one can separate non-stationary sources (speech or music) from an instantaneous mixture by looking for a weight matrix **W** that can diagonalize the covariance matrix of the output signals. Unfortunately, the Cyclic Jacobi method can not directly be used to achieve our goal because the sources are assumed to be a second order non-stationary signals, therefore the covariance matrix of such signals are time variant. Using the joint diagonalization algorithm proposed by cardoso and soulamic [18], one can jointly diagonalize a set of q covariance matrix $\mathbf{R}_i = \mathrm{E}\{S(n)S(n)^T\}$, here $1 \leq i \leq q$. The joint diagonalization algorithm is a modified version of the cyclic Jacobi method that minimize the following function with respect to a matrix **V**:

$$\mathrm{JOff}(\mathbf{R}_1, \cdots, \mathbf{R}_q) = \sum_i \mathrm{Off}(\mathbf{V}^T \mathbf{R}_i \mathbf{V}) \qquad (8)$$

here the function Off of a matrix $\mathbf{R} = (r_{ij})$ is defined by: $\mathrm{Off}(\mathbf{R}) = \sum_{i \neq j} r_{ij}^2$. It is obvious that $JOff(\mathbf{R}_1, \cdots, \mathbf{R}_q) = 0$ when $\mathbf{V}^T \mathbf{R}_i \mathbf{V}$ is a diagonal matrix for every i. Because the estimation error and

the noise, one can not minimize $JOff(\mathbf{R}_1, \cdots, \mathbf{R}_q)$ to the lower limit (i.e. 0). In our experimental study, the number q of the covariance matrices $\mathbf{R}_i$ has been chosen between 10 and 25. The covariance matrices $\mathbf{R}_i$ have been estimated according to the adaptive estimator of [22] over some slipping windows of 500 to 800 samples and shifted 100 to 200 samples for each $\mathbf{R}_i$. All the previous limits have been determined by an experimental study using our data base signals. In addition, we should mention that we used a threshold to reduce the silence effect: When ever the observation signals at time n_0 is less than the predefined threshold ϵ, it will not be considered as input signals.
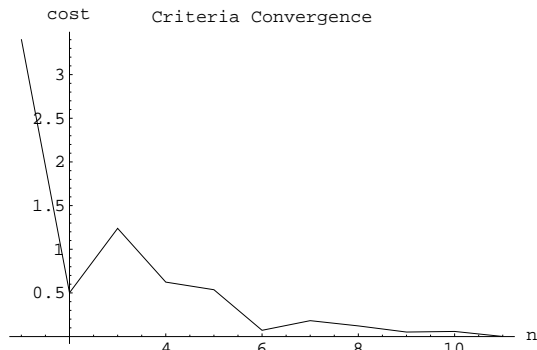

Figure 5: Evaluation of the Jacobi Diagonalization with respect to the iteration number.

We conducted many experiments and found that the crosstalk was between -17 dB and -25 dB. Fig. 5 shows the evaluation of the cost function with respect to the iteration number. The experimental study shows that the convergence of this algorithm are obtained in few iterations. Fig. 6 shows the experimental results of the separation of two speech sources.

5 Conclusion

In this paper, the separation of non-stationary sources (up to second order statistics, as music or speech signals) is investigated. The idea of three different approaches is discussed and the experimental results of three algorithms have been shown.

In some experiments the second criterion of the subsection 4.2 shows better results than the other algorithms but its performances and convergences depends more on the type of the signals and the mixing matrix than the other algorithms. The first algorithm 4.1 shows, in general, better performances than the others. We should also mention that modified versions of that algorithm were proposed to separate signals in real world applications and for con-
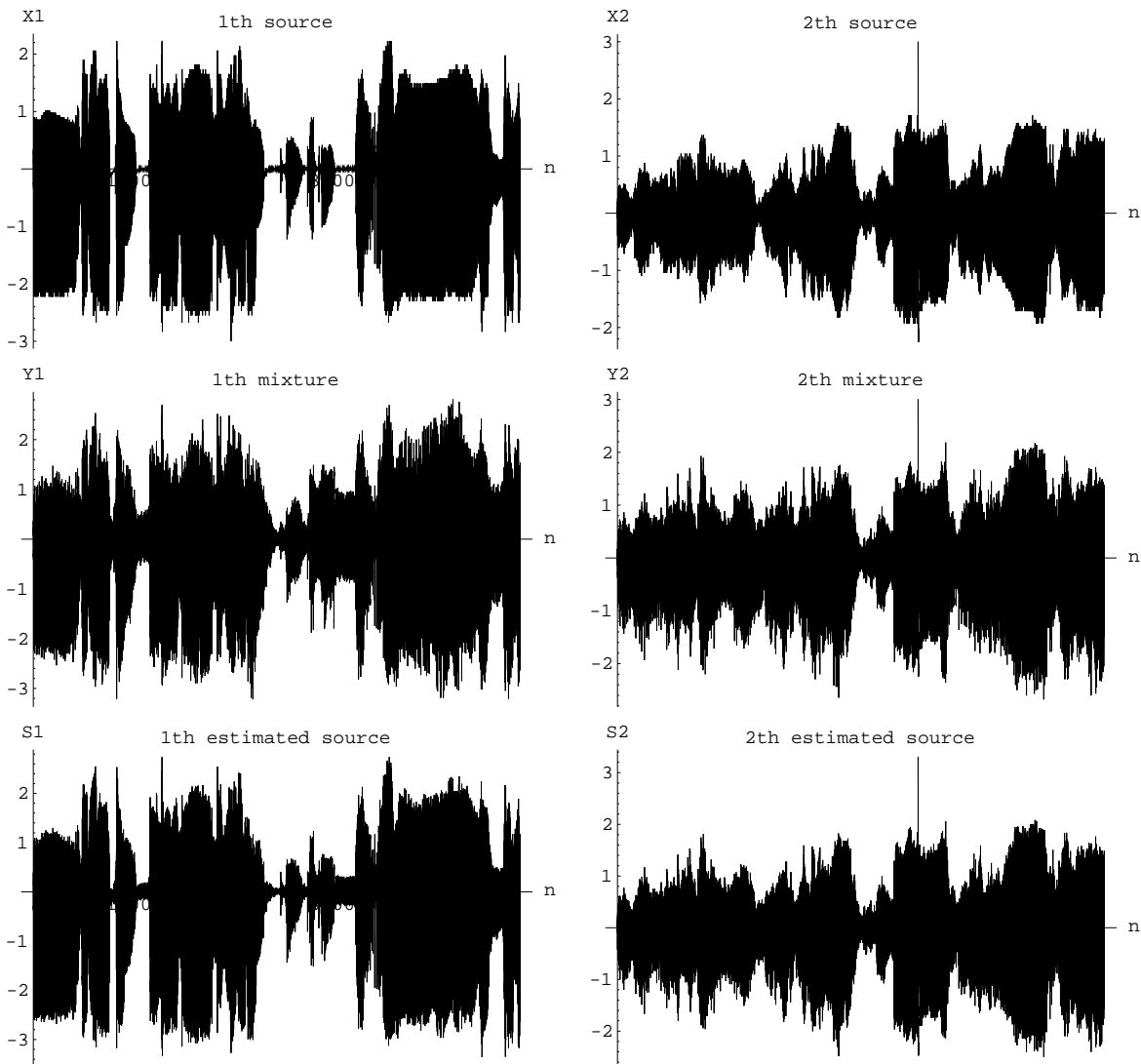
29

Figure 4: Kullback Divergence: First column contains the signals of the first channel (i.e., first source, first mixture signal and the first estimated source), the second column contains the signals of the second channel.

volutive mixtures (i.e. channel with memory effect) [23, 24, 25]. But in another hand, its performances depends on the algorithm parameters. Finally, the convergence of the third one (subsection 4.3) depends less on the type of the sources. But depending on the sources and the channel, his performances results at the convergence can not be satisfactory enough.

To conclude our paper, one should mention that the separation of non-stationary signals in real world is far to be considered as completely achieved. In another hand, the performances of the algorithms can change depending on the channel (anechoic chamber, normal, room, echo chamber), the types of the sources (sampling rates, speech, music or mixed signals) and on the algorithms parameters. These reasons make the classification and the comparison among the different criteria and algorithms, are very difficult.

References

[1] J. Hérault and B. Ans, "Réseaux de neurones à synapses modifiables: Décodage de messages sensoriels composites par une apprentissage non supervisé et permanent," *C. R. Acad. Sci. Paris*, vol. série III, pp. 525–528, 1984.

[2] L. Nguyen Thi, *Séparation aveugle de sources à large bande dans un mélange convolutif*, Ph.D. thesis, INP Grenoble, January 1993.

[3] N. Thirion, J. MARS, and J. L. BOELLE, "Separation of seismic signals: A new concept based on a blind algorithm," in *Signal Processing*
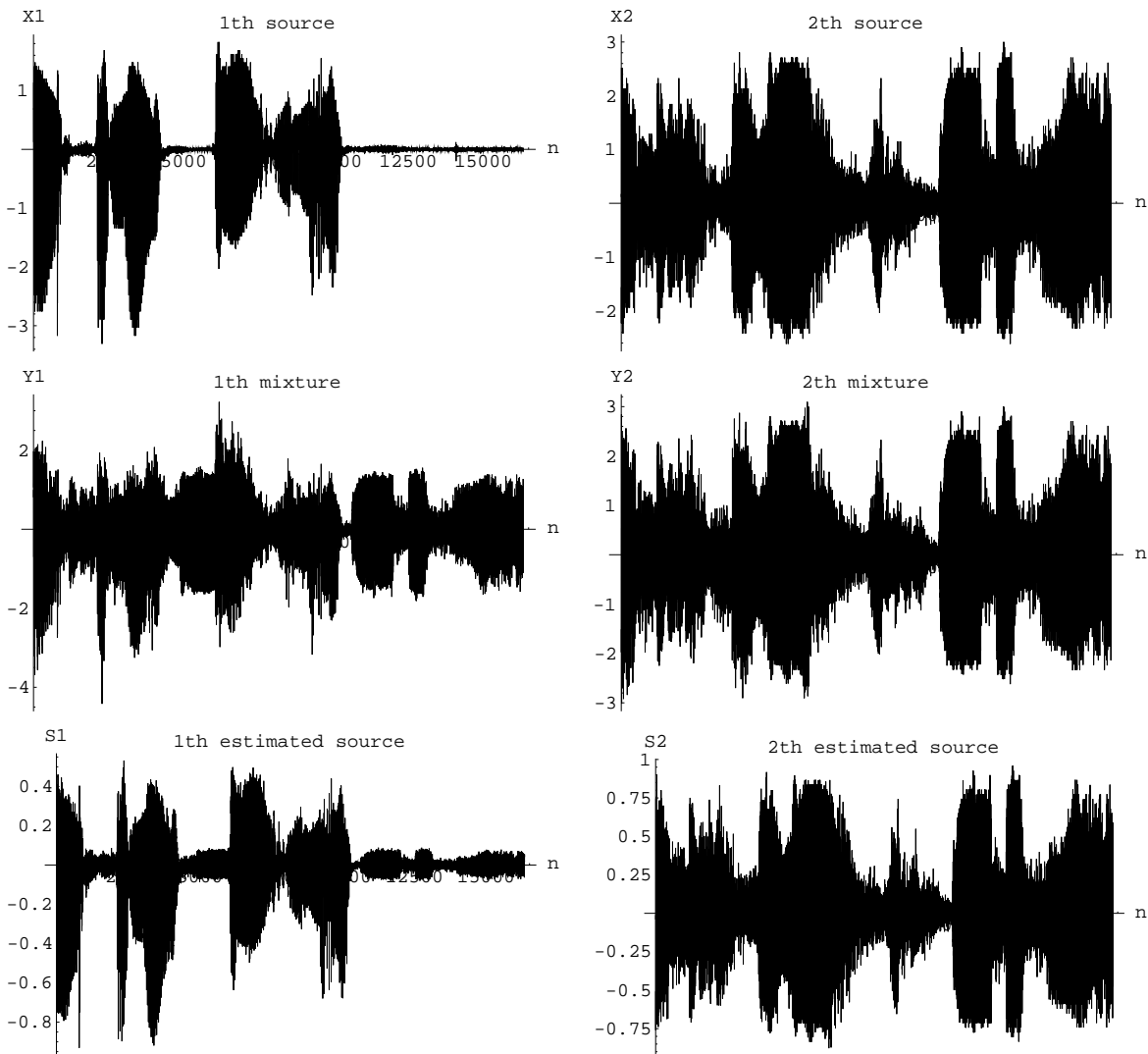
Figure 6: Jacobi Diagonalization: First column contains the signals of the first channel (i.e., first source, first mixture signal and the first estimated source), the second column contains the signals of the second channel.

VIII, Theories and Applications, Triest, Italy, September 1996, pp. 85–88, Elsevier.

[4] G. D'urso and L. Cai, "Sources separation method applied to reactor monitoring," in *Proc. Workshop Athos working group*, Girona, Spain, June 1995.

[5] E. Chaumette, P. Common, and D. Muller, "Application of ica to airport surveillance," in *HOS 93*, South Lake Tahoe-California, 7-9 June 1993, pp. 210–214.

[6] A. Kardec Barros, A. Mansour, and N. Ohnishi, "Removing artifacts from ecg signals using independent components analysis," *NeuroComputing*, vol. 22, pp. 173–186, 1999.

[7] A. Mansour and N. Ohnishi, "Multichannel blind separation of sources algorithm based on cross-cumulant and the levenberg-marquardt method.," *IEEE Trans. on Signal Processing*, vol. 47, no. 11, pp. 3172–3175, November 1999.

[8] A. Mansour, C. Jutten, and P. Loubaton, "An adaptive subspace algorithm for blind separation of independent sources in convolutive mixture," *IEEE Trans. on Signal Processing*, vol. 48, no. 2, pp. 583–586, February 2000.

[9] K. Matsuoka, M. Oya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411–419, 1995.

[10] M. Kawamoto, K. Matsuoka, and M. Oya, "Blind separation of sources using temporal cor-

relation of the observed signals," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E80-A, no. 4, pp. 111–116, April 1997.

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

[12] A. Mansour, "The blind separation of non stationary signals by only using the second order statistics.," in *Fifth International Symposium on Signal Processing and its Applications (ISSPA'99)*, Brisbane, Australia, August 22-25 1999, pp. 235–238.

[13] A. Mansour, A. Kardec Barros, and N. Ohnishi, "Blind separation of sources: Methods, assumptions and applications.," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E83-A, no. 8, pp. 1498–1512, 2000, Special Section on Digital Signal Processing in IEICE EA.

[14] B. Noble and J. W. Daniel, *Applied linear algebra*, Prentice-Hall, 1988.

[15] H. C. Wu and J. C. Principe, "Simultaneous diagonalization in the frequency domain (sdif) for source separation," in *First International Workshop on Independent Component Analysis and signal Separation (ICA99)*, Aussois, France, 11-15 January 1999, pp. 245–250.

[16] S. I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 4, pp. 251–276, 1998.

[17] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Signal Processing*, vol. 44, no. 12, December 1996.

[18] J. F. Cardoso and A. Soulamic, "Jacobi angles for simultaneous diagonalization," *SIAM*, vol. 17, no. 1, pp. 161–164, 1996.

[19] G. H. Golub and C. F. Van Loan, *Matrix computations*, The johns hopkins press- London, 1984.

[20] J. F. Cardoso and A. Soulamic, "Blind beamforming for non-gaussian signals," December 1993.

[21] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "Second-order blind separation of correlated sources," in *Int. Conf. on Digital Sig.*, Nicosia, Cyprus, july 1993, pp. 346–351.

[22] A. Mansour, A. Kardec Barros, and N. Ohnishi, "Comparison among three estimators for high order statistics.," in *Fifth International Conference on Neural Information Processing (ICONIP'98)*, Kitakyushu, Japan, 21-23 October 1998, pp. 899–902.

[23] M. Kawamoto, A. Kardec Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Real world blind separation of convolved non-stationary signals.," in *First International Workshop on Independent Component Analysis and signal Separation (ICA99)*, Aussois, France, 11-15 January 1999, pp. 347–352.

[24] M. Kawamoto, A. Kardec Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Blind signal separation for convolved non-stationary signals," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. J82-A, no. 8, pp. 1320–1328, August 1999, Japanese paper.

[25] M. Kawamoto, A. Kardec Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Blind signal separation for convolved non-stationary signals," *To appear in the Electronics and Communications in Japan Part 3*, 2000, Published by John Wiley & Sons, Inc.