BLIND SEPARATION OF NONSTATIONARY AND TEMPORALLY CORRELATED SOURCES FROM NOISY MIXTURES

Seungjin CHOI § and Andrzej CICHOCKI †

§ Department of Electrical Engineering
Chungbuk National University, KOREA
Email: schoi@lisa.chungbuk.ac.kr

† Laboratory for Advanced Brain Signal Processing
Brain-Style Information Systems Research Group
Brain Science Institute, RIKEN, JAPAN
and Warsaw University of Technology, POLAND
Email: cia@brain.riken.go.jp

Appeared in Proc. IEEE NNSP, pp. 405-414, Sydney, Austrailia, Dec. 11-13, 2000.

Abstract. In this paper we present a new method of blind source separation that is robust with respect to additive white noise. Our method exploits the nonstationarity and temporal structure of sources. The method needs only multiple time-delayed correlation matrices of the observation data at several different time-windowed frames to estimate the mixing matrix. We present an implementation based on the joint diagonalization (JD). Extensive simulations verify the high performance of the proposed method, especially in low SNR environment.

INTRODUCTION

Blind source separation (BSS) is a fundamental problem that is encountered in many practical applications such as telecommunications, array signal processing, speech processing (cocktail party problem), biomedical signal analysis where multiple sensors are involved. In the context of BSS, the m-dimensional observation vector $\boldsymbol{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ is assumed to be generated by

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{v}(t), \qquad (1)$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the mixing matrix, $\boldsymbol{s}(t)$ is the n-dimensional source vector ($n \leq m$), and $\boldsymbol{v}(t)$ is additive white noise that is statistically independent of $\boldsymbol{s}(t)$.

Throughout this paper, the following assumptions are made:

- The mixing matrix $\boldsymbol{A}$ is of full column rank.
- Sources are mutually uncorrelated but are temporally correlated stochastic signals with zero mean. In other words, sources are not i.i.d. in time domain, i.e.,

$$E\{s_i(t)s_i(t-\tau)\} \neq 0, \tag{2}$$

  for $\forall \tau \neq 0$ or at least for some $\tau \neq 0$. The statistical expectation operator is denoted by $E\{\cdot\}$.
- Sources are nonstationary signals in the sense that their variances are time varying.

The task of BSS is to recover sources from their linear instantaneous mixtures without resorting to any prior knowledge. Sources can be recovered blindly by either estimating the mixing matrix $\boldsymbol{A}$ or its pseudo-inverse $\boldsymbol{W} = \boldsymbol{A}^{\#}$ (corresponding to the demixing system).

Two indeterminacies cannot be resolved in BSS without some prior knowledge. They include scaling and permutation ambiguities. Thus if the estimate of the mixing matrix, $\hat{\boldsymbol{A}}$ satisfies $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A} = \hat{\boldsymbol{A}}^{\#}\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}$ where $\boldsymbol{G}$ is the global transformation which combines the mixing and demixing system, $\boldsymbol{P}$ is some permutation matrix, $\boldsymbol{\Lambda}$ is some nonsingular diagonal matrix, and the superscipt # denotes the pseudo-inverse, then $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{s}})$ and $(\boldsymbol{A}, \boldsymbol{s})$ are said to be related by a waveform-preserving relation.

Most methods of BSS have focused on statistically independent stationary non-Gaussian sources, so higher-order statistics was necessary. In such a case, BSS can be formulated as an ICA problem [9]. When sources are spatially uncorrelated but temporally correlated, second-order statistics is sufficient for successful separation. Along this line, several methods have been developed [13, 11, 3, 2, 1, 8]. When sources are nonstationary, it was shown in [10] that spatial decorrelation (second-order statistics) could successfully separate sources. A recurrent network and an associated gradient-based algorithm were proposed in [10], but only noise-free mixtures were considered.

This paper has two contributions. We first consider the case of noise-free mixture and show that we can separate the mixtures of nonstationary sources using the joint diagonalization method. Since the joint diagonalization method [5, 6] is numerically robust and fast, the resulting method is also very fast, compared to gradient-based adaptive BSS algorithms. Next we extend this method to the case of noisy mixtures. By exploiting both nonstationarity and temporal information of sources, we develop a robust BSS algorithm that is robust to additive white noise.

NOISE-FREE MIXTURES

Let us consider the linear data model (1) in the zero noise limit. For the sake of simplicity, we assume $m = n$. Then the observation vector $\boldsymbol{x}(t)$ is simply

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t). \tag{3}$$

The output of the demixing system $\boldsymbol{W}$, $\boldsymbol{y}(t)$ is

$$\boldsymbol{y}(t) = \boldsymbol{W}\boldsymbol{x}(t) = \widehat{\boldsymbol{A}}^{-1}\boldsymbol{x}(t). \tag{4}$$

Methods of BSS based on second-order statistics exploit either temporal correlations or nonstationarity of sources. In the case where sources are temporally correlated stationary stochastic processes, the estimate of the mixing matrix, $\widehat{\boldsymbol{A}}$ is a linear transformation that jointly diagonalizes two or more time-delayed correlation matrices of the observation vector $\boldsymbol{x}(t)$. Along this line, various methods have been developed. They rely on the generalized eigen-decomposition [13], recurrent network [11], and the joint diagonalization [3].

In the case where the variances of sources are time varying, spatial decorrelation (using equal-time statistics) is sufficient for the task of BSS. Matsuoka *et al.* [10] proposed the following cost function $\mathcal{J}(\boldsymbol{W})$

$$\mathcal{J}(\boldsymbol{W}) = \frac{1}{2}\left\{\sum_{i=1}^{n} \log E\left\{y_i^2(t)\right\} - \log\det\left(E\left\{\boldsymbol{y}(t)\boldsymbol{y}^T(t)\right\}\right)\right\}, \tag{5}$$

where $\det(\cdot)$ denotes the determinant of a matrix. It was shown in [10] that the cost function (5) is a nonnegative function that takes the minimum iff $E\{y_i(t)y_j(t)\} = 0$ for $\forall i, j = 1, \dots, n,\ i \neq j$. A linear feedback network with an associated learning algorithm was proposed in [10].

In principle, two different statistics is sufficient for source separation. This is explained by the following theorem that was already exploited in several literature [4, 13].

Theorem 1 *Let $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$, $\boldsymbol{D}_1$, $\boldsymbol{D}_2 \in \mathbb{R}^{n\times n}$ be diagonal matrices with nonzero diagonal entries. Suppose that $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ satisfies the following decompositions:*

$$\begin{aligned} \boldsymbol{D}_1 &= \boldsymbol{G}\boldsymbol{\Lambda}_1\boldsymbol{G}^T, & (6)\\ \boldsymbol{D}_2 &= \boldsymbol{G}\boldsymbol{\Lambda}_2\boldsymbol{G}^T. & (7)\end{aligned}$$

Then the matrix $\boldsymbol{G}$ is the generalized permutation matrix, i.e., $\boldsymbol{G} = \boldsymbol{P}\boldsymbol{\Lambda}$ if $\boldsymbol{D}_1^{-1}\boldsymbol{D}_2$ and $\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Lambda}_2$ have distinct diagonal entries.

Proof: From (6), there exists an orthogonal matrix $\boldsymbol{Q}$ such that

$$\left(\boldsymbol{G}\boldsymbol{\Lambda}_1^{\frac{1}{2}}\right) = \left(\boldsymbol{D}_1^{\frac{1}{2}}\right)\boldsymbol{Q}. \tag{8}$$

Hence,

$$\boldsymbol{G} = \boldsymbol{D}_1^{\frac{1}{2}} \boldsymbol{Q} \boldsymbol{\Lambda}_1^{-\frac{1}{2}}. \tag{9}$$

Substitute (9) into (7) to obtain

$$\boldsymbol{D}_1^{-1} \boldsymbol{D}_2 = \boldsymbol{Q} \boldsymbol{\Lambda}_1^{-1} \boldsymbol{\Lambda}_2 \boldsymbol{Q}^T. \tag{10}$$

Sine the right-hand side of (10) is the eigen-decomposition of the left-hand side of (10), the diagonal elements of $\boldsymbol{D}_1^{-1}\boldsymbol{D}_2$ and $\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Lambda}_2$ are the same. From the assumption that the diagonal elements of $\boldsymbol{D}_1^{-1}\boldsymbol{D}_2$ and $\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Lambda}_2$ are distinct, the orthogonal matrix $\boldsymbol{Q}$ must have the form $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is an diagonal matrix whose diagonal elements are either $+1$ or -1. Hence, we have

$$\begin{aligned} \boldsymbol{G} &= \boldsymbol{D}_1^{\frac{1}{2}} \boldsymbol{P} \boldsymbol{\Psi} \boldsymbol{\Lambda}_1^{-\frac{1}{2}} \\ &= \boldsymbol{P} \boldsymbol{P}^T \boldsymbol{D}_1^{\frac{1}{2}} \boldsymbol{P} \boldsymbol{\Psi} \boldsymbol{\Lambda}_1^{-\frac{1}{2}} \\ &= \boldsymbol{P} \boldsymbol{\Lambda}, \end{aligned} \tag{11}$$

where

$$\boldsymbol{\Lambda} = \boldsymbol{P}^T \boldsymbol{D}_1^{\frac{1}{2}} \boldsymbol{P} \boldsymbol{\Psi} \boldsymbol{\Lambda}_1^{-\frac{1}{2}} \qquad \text{Q.E.D.} \tag{12}$$

Now we describe our method of BSS based on the joint diagonalization when sources are nonstationary. Let us define a symmetric matrix $\boldsymbol{M}_y(t_k, \tau)$ by

$$\boldsymbol{M}_y(t_k, \tau) = \frac{1}{2} \left\{ \boldsymbol{R}_y(t_k, \tau) + \boldsymbol{R}_y^T(t_k, \tau) \right\}, \tag{13}$$

where $\boldsymbol{R}_y(t_k, \tau) = E\{\boldsymbol{y}(t_k)\boldsymbol{y}^T(t_k - \tau)\}$ is the correlation matrix of the demixing system output $\boldsymbol{y}(t)$. In practice $\boldsymbol{R}_y(t_k, \tau)$ is evaluated using the samples in the kth time-windowed data frame. We have to point out that $\{\boldsymbol{M}_y(t_k, \tau)\}$ carry different statistics for nonstationary signals, whereas they give the same statistics regardless of t_k for stationary stochastic signals. In similar manner, we define $\boldsymbol{M}_s(t_k, \tau)$ and $\boldsymbol{M}_x(t_k, \tau)$ for the source vector $\boldsymbol{s}(t)$ and for the observation vector $\boldsymbol{x}(t)$, respectively. Note that $\boldsymbol{M}_s(t_k, \tau)$ is a nonsingular diagonal matrix from the assumptions. For the moment, we consider only equal-time correlation matrix ($\tau = 0$).

With these definitions, we have the following relation:

$$\boldsymbol{M}_y(t_k, 0) = (\boldsymbol{W}\boldsymbol{A})\, \boldsymbol{M}_s(t_k, 0)\, (\boldsymbol{W}\boldsymbol{A})^T, \quad \text{for } k = 1, \ldots, K, \tag{14}$$

where K is the number of data frames available. If the demixing system outputs $\{\boldsymbol{y}(t)\}$ are spatially uncorrelated, then $\{\boldsymbol{M}_y(t_k, 0)\}$, $k = 1, \ldots, K$ are diagonal matrices. Under this condition, invoking Theorem 1, one can easily see that $\boldsymbol{W}\boldsymbol{A}$ is a generalized permutation matrix if there exist i and j ($i \neq j$) such that $\boldsymbol{M}_s^{-1}(t_i, 0)\boldsymbol{M}_s(t_j, 0)$ had distinct diagonal elements.

The simultaneous diagonalization of $\boldsymbol{M}_x(t_j,0)$ and $\boldsymbol{M}_x(t_j,0)$ requires

$$\begin{aligned} \boldsymbol{W}\boldsymbol{M}_x(t_i,0)\boldsymbol{W}^T &= \boldsymbol{\Lambda}_i, \\ \boldsymbol{W}\boldsymbol{M}_x(t_j,0)\boldsymbol{W}^T &= \boldsymbol{\Lambda}_j, \end{aligned} \tag{15}$$

where $\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Lambda}_j$ are diagonal matrices. The demixing system $\boldsymbol{W}$ corresponds to the generalized eigenvector matrix that satisfies

$$\boldsymbol{M}_x^{-1}(t_i,0)\boldsymbol{M}_x(t_j,0)\boldsymbol{W}^T = \boldsymbol{W}^T\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Lambda}_j. \tag{16}$$

For successful separation, it is required to select two data frames (t_i and t_j) such that $\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Lambda}_j$ has distinct diagonal elements. The performance relies on the choice of two different data frames that are used to compute the correlation matrices. However it is not guaranteed that this condition is satisfied for data frames that is chosen in advance.

In order to overcome this drawback we employ the joint diagonalization method [6] which finds a unitary transform that jointly diagonalizes several matrices. Here we refer to the simultaneous diagonalization (SD) as the method of diagonalizing two different matrices and the joint diagonalization (JD) as the method of diagonalizing more than two matrices. Two methods are summarized below.

Nonstationary source separation: SD (NSS-SD)

(1) We divide the data into two non-overlapping frames and calculate $\boldsymbol{M}_x(t_1,0)$ and $\boldsymbol{M}_x(t_2,0)$.

(2) We find the generalized eigenvector matrix $\boldsymbol{U}$ that satisfies

$$\boldsymbol{M}_x(t_2,0)\boldsymbol{U} = \boldsymbol{M}_x(t_1,0)\boldsymbol{U}\boldsymbol{\Lambda}. \tag{17}$$

Then the demixing system $\boldsymbol{W}$ is given by $\boldsymbol{W} = \boldsymbol{U}^T$ or the estimate the mixing matrix is $\widehat{\boldsymbol{A}} = \boldsymbol{U}^{-T}$.

Nonstationary source separation: JD (NSS-JD)

(1) We whiten the whole data by a linear transformation. Let us denote the whitened data by $\boldsymbol{z}(t) = \boldsymbol{Q}\boldsymbol{x}(t)$ where $\boldsymbol{Q}$ is a whitening transformation matrix.

(2) We calculate $\boldsymbol{M}_z(t_k,0)$ for $k = 1,\ldots,K$, where K denotes the number of non-overlapped data frames.

(3) We find a joint diagonalizer of $\{\boldsymbol{M}_z(t_k,0)\}$ which satisfies

$$\boldsymbol{V}^T\boldsymbol{M}_z(t_k,0)\boldsymbol{V} = \boldsymbol{\Lambda}_k, \tag{18}$$

where $\{\boldsymbol{\Lambda}_k\}$ is a set of diagonal matrices.

(4) Then the demixing system $\boldsymbol{W}$ is given by is $\boldsymbol{W} = \boldsymbol{V}^T\boldsymbol{Q}$ or the estimate of the mixing matrix is $\widehat{\boldsymbol{A}} = \boldsymbol{Q}^{-1}\boldsymbol{V}$.

NOISY MIXTURES

Here we show how the method can be extended to handle the noisy mixtures. In the presence of additive noise, we have

$$\boldsymbol{M}_x(t_k, 0) = \boldsymbol{A}\boldsymbol{M}_s(t_k, 0)\boldsymbol{A}^T + \boldsymbol{A}\boldsymbol{M}_v(t_k, 0)\boldsymbol{A}^T. \tag{19}$$

Thus the joint diagonalization is not directly applicable to estimate the mixing matrix. However, the time-delayed correlation matrices of the observation vector satisfy

$$\boldsymbol{M}_x(t_k, \tau_k) = \boldsymbol{A}\boldsymbol{M}_s(t_k, \tau_k)\boldsymbol{A}^T, \tag{20}$$

for $\tau_k \neq 0$ because $\boldsymbol{M}_v(t_k, \tau_k) = 0$, $\forall \tau_k \neq 0$ due to the assumption of white noise. If sources are nonstationary and temporally correlated, then $\{\boldsymbol{M}_s(t_k, \tau_k)\}$ is nonsingular diagonal matrices. Either simultaneous diagonalization or joint diagonalization method can be applied just like the case of noise-free mixtures.

In the case of noisy mixtures, methods NSS-SD and NSS-JD are modified slightly. Equal-time correlation matrices are replaced by time-delayed correlation matrices so that the effect of noise is negligible. The pre-whitening procedure in NSS-JD is performed using the time-delayed correlation matrix.

Algorithm Outline: NSS-TD-JD

(1) The pre-whitening is performed based on a time-delayed correlation matrix $\boldsymbol{M}_x(\tau) = \boldsymbol{U}_x \boldsymbol{\Lambda}_x \boldsymbol{U}_x^T$ for some $\tau \neq 0$ in order to reduce the white noise effect, where $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are the eigenvector and eigenvalue matrices of $\boldsymbol{M}_x(\tau)$, respectively. The whitened data $\boldsymbol{z}(t)$ is $\boldsymbol{z}(t) = \boldsymbol{\Lambda}_x^{-\frac{1}{2}} \boldsymbol{U}_x^T \boldsymbol{x}(t)$.

(2) We divide the data into K non-overlapping blocks and calculate $\boldsymbol{M}_z(t_k, \tau_j)$ for $k = 1, \ldots, K$ and $j = 1, \ldots, J$. In other words, At each time-windowed data frame, we compute J different time-delayed correlation matrices.

(3) We find a joint diagonalizer $\boldsymbol{V}$ of $\{\boldsymbol{M}_z(t_k, \tau_j)\}$ that satisfies

$$\boldsymbol{V}^T \boldsymbol{M}_z(t_k, \tau_j)\boldsymbol{V} = \boldsymbol{\Lambda}_{k,j}, \tag{21}$$

where $\{\boldsymbol{\Lambda}_{k,j}\}$ is a set of diagonal matrices.

(4) Then the demixing system $\boldsymbol{W}$ is given by $\boldsymbol{W} = \boldsymbol{V}^T \boldsymbol{\Lambda}_x^{-\frac{1}{2}} \boldsymbol{U}_x^T$ or the estimate of the mixing matrix is $\widehat{\boldsymbol{A}} = \boldsymbol{U}_x \boldsymbol{\Lambda}_x^{\frac{1}{2}} \boldsymbol{V}$.

Remarks

- Our methods exploits both nonstationarity and temporal information of sources since they rely on multiple time-delayed correlation matrices calculated using the samples at several different data frames.

- The key step in the case of noisy mixtures lies in the pre-whitening procedure. Conventionally the whitening was done in such a way that the equal-time correlation matrix of the data is diagonalized. When the number of sources and sensors are equal, this whitening procedure might enhance the noise. Our whitening procedure aims at diagonalizing a time-delayed correlation matrix of the input data. Due to the whiteness of noise, this whitening process is insensitive to the noise. Similar idea was also reported in [12].
- For stationary sources, $M_z(t_k, \tau_j) = M_z(\tau_j)$. In such a case, the method NSS-TD-JD becomes identical to the SOBI method [3] except for the different pre-whitening step. Thus our method also works for even stationary sources provided that sources are temporally correlated.

NUMERICAL EXPERIMENTS

Two numerical experiments are presented here. The first experiment was designed to evaluate the effectiveness of the proposed method in the presence of several Gaussian signals. In this experiment, we used three speeches that are sampled at 8 kHz and two Gaussian signals whose variances are slowly varying. With these sources we generated five mixtures using a randomly chosen mixing matrix. The second experiment was designed to show the robustness of the proposed method in the presence of additive white noise. We used 3 digitized voice signals and 2 music signals, all of which were sampled at 8 kHz and generated five mixtures using randomly chosen mixing matrix.

In order to evaluate the performance of algorithms in both experiments, we calculated the performance index (PI) defined by

$$\mathrm{PI} = \frac{1}{2(n-1)} \sum_{i=1}^{n} \left\{ \left(\sum_{k=1}^{n} \frac{|g_{ik}|^2}{\max_j |g_{ij}|^2} - 1 \right) + \left(\sum_{k=1}^{n} \frac{|g_{ki}|^2}{\max_j |g_{ji}|^2} - 1 \right) \right\}, \quad (22)$$

where g_{ij} is the (i,j)-element of the global system matrix $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A}$ and $\max_j g_{ij}$ represents the maximum value among the elements in the ith row vector of $\boldsymbol{G}$, $\max_j g_{ji}$ does the maximum value among the elements in the ith column vector of $\boldsymbol{G}$. The performance index defined in (22) tells us how far the global system matrix $\boldsymbol{G}$ is from a generalized permutation matrix.

Experiment 1

We tested our two methods (NSS-SD and NSS-JD), the JADE algorithm [5], and the flexible ICA algorithm (FLEXICA, one exemplary adaptive ICA algorithm) [7]. Methods based on higher-order statistics such as JADE and FLEXICA are expected to give degraded performance since mixtures contain two Gaussian sources. The observation vector $\boldsymbol{x}$ has 10000 data points.

In NSS-SD, only two frames of data (with non-overlapping) were used to produce two different equal-time correlation matrices. Simultaneous diagonalization (based on generalized eigen-decompositon) was employed to estimate the mixing matrix. In NSS-JD, we used 100 different frames of data (with non-overlapping) to calculate 100 different equal-time correlation matrices of the observation vector. The joint diagonalization method was used to estimate the mixing matrix. In each method, the global system matrix $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A} = \hat{\boldsymbol{A}}^{-1}\boldsymbol{A}$ was computed (see Figure 1). In Hinton diagram, each square's area represents the magnitude of element of the matrix and each square's color represent the sign of the element (red for negative value, green for positive value). The NSS-JD showed the best performance, although NSS-SD was also successful for the separation of 5 different sources. The JADE and FLEXICA had difficulty in the separation of two Gaussian sources, as expected, although they were successful in the separation of speech signals.

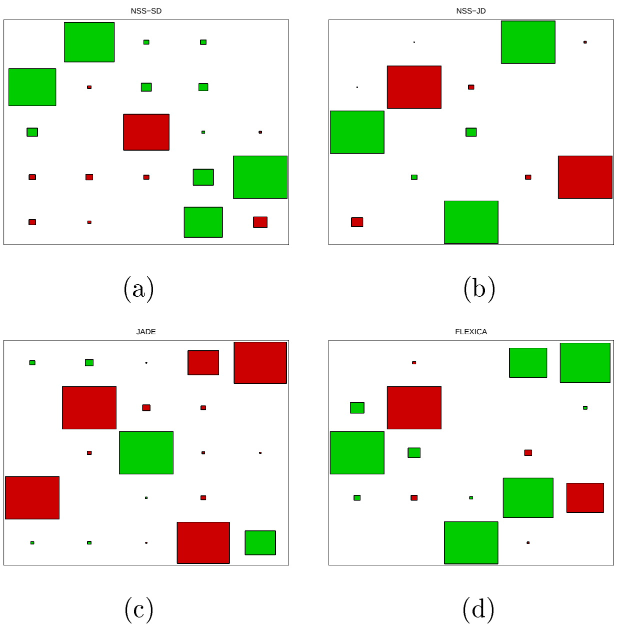

Figure 1: Hinton diagrams of global system matrices: (a) NSS-SD; (b) NSS-JD; (c) JADE; (d) FLEXICA.

Experiment 2

We have evaluated the performance of five different methods, given noisy mixtures. Here we exploited both nonstationarity and temporal information of sources. In NSS-SD, we computed two different time-delayed correlation matrices (with the time lag being 1 for both matrices) of the observation vector and employed the generalized eigen-decomposition to estimate the mixing matrix. In NSS-JD, we used 50 different frames of data (with non-overlapping) to calculate 50 different time-delayed correlation matrices (with time lag being 1 for all of them) of the observation vector. Then, the joint daignalization method was employed to estimate the mixing matrix. In NSS-TD-JD, we also used 50 different non-overlapped frames of data to compute the time-delayed correlation matrices. Unlike NSS-JD, for each data frame we computed five different time-delayed correlation matrices and then employed the joint diagonalization method to estimate the demixing/mixing matrix. The performance index of the methods are shown in Figure 2. At each SNR, we performed 10 different runs and averaged the value of performance index. The NSS-TD-JD and NSS-JD showed the best performance in this experiment. The NSS-SD also performs quite well in low SNR (7dB-20dB), however, its performance was worse than both NSS-TD-JD and NSS-JD because only two different statistics was used. Other methods such as JADE, FLEXICA, SOBI showed good performance in high SNR (above 25dB only).

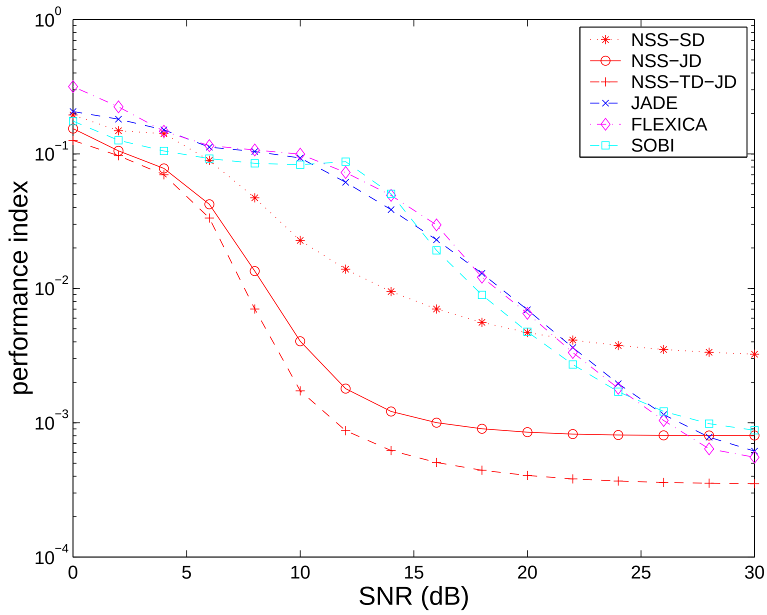

Figure 2: The comparison of performance of various BSS algorithms. Our methods (NSS-TD-JD, NSS-JD, NSS-SD) show improved performance, compared to JADE [5], SOBI [3], and FLEXICA [7], especially in low SNR.

CONCLUSIONS

We presented new methods of BSS when sources are nonstationary and temporally correlation in the presence of additive noise. The proposed methods required only multiple time-delayed correlation matrices to estimate the mixing matrix. The efficient implementation based on the joint diagonalization was explained. The robustness of the proposed methods (NSS-JD, NSS-TD-

JD) in low SNR was verified by numerical experiments.

ACKNOWLEDGMENT

This work was supported in part by Korean Ministry of Science and Technology under Brain Science and Engineering Research Program and by by Brain Science Institute, RIKEN in JAPAN.

REFERENCES

[1] S. Amari, "Estimating Functions of Independent Component Analysis for Temporally Correlated Signals," **Neural Computation**, vol. 12, 2000, in press.

[2] H. Attias and C. E. Schreiner, "Blind Source Separation and Deconvolution: The Dynamic Component Analysis Algorithms," **Neural Computation**, vol. 10, pp. 1373–1424, 1998.

[3] A. Belouchrani, K. Abed-Merain, J. F. Cardoso and E. Moulines, "A Blind Source Separation Technique Using Second Order Statistics," **IEEE Trans. Signal Processing**, vol. 45, pp. 434–444, Feb. 1997.

[4] J. F. Cardoso, "Source Separation Using Higher-order Moments," in **Proc. ICASSP**, 1989.

[5] J. F. Cardoso and A. Souloumiac, "Blind Beamforming for Non Gaussian Signals," **IEE Proceedings-F**, vol. 140, no. 6, pp. 362–370, 1993.

[6] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," **SIAM J. Mat. Anal. Appl.**, vol. 17, no. 1, pp. 161–164, Jan. 1996.

[7] S. Choi, A. Cichocki and S. Amari, "Flexible Independent Component Analysis," in T. Constantinides, S. Y. Kung, M. Niranjan and E. Wilson (eds.), **Neural Networks for Signal Processing VIII**, 1998, pp. 83–92.

[8] A. Cichocki and R. Thawonmos, "On-line Algorithm for Blind Signal Extraction of Arbitrarily Distributed but Temporally Correlated Sources Using Second-order Statistics," **Neural Processing Letters**, vol. 12, no. 1, pp. 91–98, 2000.

[9] P. Comon, "Independent Component Analysis, A New Concept?" **Signal Processing**, vol. 36, no. 3, pp. 287–314, 1994.

[10] K. Matsuoka, M. Ohya and M. Kawamoto, "A Neural Net for Blind Separation of Nonstationary Signals," **Neural Networks**, vol. 8, no. 3, pp. 411–419, 1995.

[11] L. Molgedey and H. G. Schuster, "Separation of a Mixture of Independent Signals Using Time Delayed Correlations," **Physical Review Letters**, pp. 3634–3637, 1994.

[12] K. R. Müller, P. Philips and A. Ziehe, "JADE$_{TD}$: Combining Higher-order Statistics and Temporal Information for Blind Source Separation (with Noise)," in **Proc. ICA'99**, Aussois, France, 1999, pp. 87–92.

[13] L. Tong, V. Soon, Y. F. Huang and R. Liu, "Indeterminacy and Identifiability of Blind Identification," **IEEE Trans. Circuits and Systems**, vol. 38, pp. 499–509, May 1991.