REAL WORLD BLIND SEPARATION OF CONVOLVED NON-STATIONARY SIGNALS

M. Kawamoto[1], A. K. Barros[1], A. Mansour[1], K. Matsuoka[2], and N. Ohnishi[3]

1.Bio-mimetic Control Research Center, RIKEN,
2271-130 Anagahora, Shimoshidami, Moriyama-ku, Nagoya, 463-0003, Japan
E-mail:kawa@nagoya.riken.go.jp
2. Department of Control Engineering, Kyushu Institute of Technology (Japan),
3. Graduate School of Engineering, Nagoya University (Japan)

ABSTRACT

In this paper, a method of blind separation for convolved non-stationary signals (e.g., speech signals and music) is presented. Our method achieves blind separation by forcing mixed signals to uncorrelate with each other. The validity of the proposed method has been confirmed by a computer simulation and an experiment in an anechoic room [7]. In this paper, we apply our method to an experiment which extracts two source signals from their mixtures observed in a normal room. The experiment is implemented in a noisy environment. Moreover, we test our algorithm using the data obtained from Computational Neurobiology Lab.'s Blind Source Separation Web Page.

1. INTRODUCTION

We present a method of blind separation for a convolutive mixture:

$$\boldsymbol{x}(t) = \overline{\boldsymbol{A}}(z)\boldsymbol{s}(t), \tag{1}$$

where $\boldsymbol{x}(t) = [x_1(t), ..., x_N(t)]^T$ and $\boldsymbol{s}(t) = [s_1(t), \ldots, s_N(t)]^T$. $\overline{\boldsymbol{A}}(z)$ is a matrix which has elements $\overline{a}_{ij}(z)$ $(i, j = 1, ..., N)$:

$$\overline{a}_{ij}(z) = \sum_{k=-\infty}^{\infty} a_{ij}(k) z^{-k} \qquad (i, j = 1, ..., N), \tag{2}$$

where z^{-k} is a delay operator, i.e., $s_i(t)z^{-k} = s_i(t-k)$. In this paper, the sources $\boldsymbol{s}(t)$ are assumed to be nonstationary signals (e.g., speech signals, music), and the source signals are separated from their mixtures $\boldsymbol{x}(t)$ (observed signals) by using **the nonstationarity properties** of the sources. Nonstationarity of the sources implies that the auto-correlations of the sources change with time t. Our method **does not require any additional information about whether the sources are super-Gaussian or sub-Gaussian**. We only make use of **the second-order moments** of the observed signals. Methods using second-order moments for separating the sources $\boldsymbol{s}(t)$ from the observed signals $\boldsymbol{x}(t)$ have been proposed by Chan et al. [?], Ehlers et al. [3], Gerven et al. [4,5], and Lindgren et al. [13]. An attractive feature of our method, differently from those, is that only one set of cross-correlation data is used and non-minimum phase systems can be treated.

Our method separates the sources from the observed signals by modifying the parameters of an adaptive filter such that a cost function takes the minimum (zero) at any time. The validity of the proposed method is confirmed by an experiment that extracts two source signals from their mixtures observed in a normal room.

2. SOURCE SIGNALS

Source signals $s_i(t)$ $(i = 1, ..., N)$ are assumed to be mutually independent with zero mean. From this property of source signals, the auto-correlation matrix $\boldsymbol{R}(t, \tau)$ of $\boldsymbol{s}(t)$ becomes a diagonal matrix:

$$\begin{aligned} \boldsymbol{R}(t,\tau) &= E[\boldsymbol{s}(t)\boldsymbol{s}(t-\tau)^T] \\ &= diag\,\{E[s_1(t)s_1(t-\tau)], ..., E[s_N(t)s_N(t-\tau)]\} \\ &\equiv diag\,\{r_1(t,\tau), ..., r_N(t,\tau)\}, \end{aligned} \tag{3}$$

where $diag\{...\}$ represents a diagonal matrix with the diagonal element $\{...\}$, and $E[x]$ is the ensemble average of x.

Our aim is to extract source signals from the observed signals $x_i(t)$ $(i = 1, ..., N)$. To this end, we make the following assumptions.

Assumption 1 $\overline{\boldsymbol{A}}(z)$ does not have poles or zeros on the unit circle $|z| = 1$.

Assumption 2 $s_i(t)$ $(i = 1, ..., N)$ are nonstationary signals whose auto-correlations $r_i(t, \tau)$ $(i = 1, ..., N; \forall\tau)$ change independently with time t.

3. SEPARATION PROCESS

An adaptive feedforward network (see Figure 1) is used to separate source signals from the observed signals $x_i(t)$ $(i = 1, ..., N)$. The network outputs can be written as:

$$\begin{aligned} y_i(t) &= x_i(t-L) + \sum_{\substack{j=1 \\ j\neq i}}^{N} \sum_{k=0}^{M} b_{ij}(k) x_j(t-k) \\ &\qquad (i = 1, ..., N; 0 \le L < M) \qquad (4) \\ &= \sum_{j=1}^{N} \bar{b}_{ij}(z) x_j(t), \qquad (5) \end{aligned}$$

where $\bar{b}_{ij}(z) = \sum_{k=0}^{M} b_{ij}(k) z^{-k}$ $(i, j = 1, ..., N; i \neq j)$ represent the transfer function between the j-th input signal and the i-th output signal, and $\bar{b}_{ii}(z) = z^{-L}$ $(i = 1, ..., N)$ represent delay time L between the i-th input and the i-th output. Eqn (5) can be rewritten in vector notation as

$$\boldsymbol{y}(t) = \overline{\boldsymbol{B}}(z)\boldsymbol{x}(t), \qquad (6)$$

where $\boldsymbol{y}(t) = [y_1(t), ..., y_N(t)]^T$, $\overline{\boldsymbol{B}}(z) = [\bar{b}_{ij}(z)] = \sum_{k=0}^{M} \boldsymbol{B}(k) z^{-k}$, $\boldsymbol{B}(k) = [b_{ij}(k)]$.

Substituting eqn (1) into eqn (6), we have
$\boldsymbol{y}(t) = \overline{\boldsymbol{B}}(z)\overline{\boldsymbol{A}}(z)\boldsymbol{s}(t) \equiv \boldsymbol{C}(z)\boldsymbol{s}(t)$,
where $\boldsymbol{C}(z) \equiv \overline{\boldsymbol{B}}(z)\overline{\boldsymbol{A}}(z)$. If $\overline{\boldsymbol{B}}_0(z)\overline{\boldsymbol{A}}(z) = \boldsymbol{D}(z)\boldsymbol{P}$, the outputs of the network become the filtered and permuted source signals, i.e., $\bar{\boldsymbol{s}}(t) = [\bar{s}_1(t), ..., \bar{s}_N(t)]^T = \boldsymbol{D}(z)\boldsymbol{P}\boldsymbol{s}(t)$. Here, $\boldsymbol{P}$ is an arbitrary permutation matrix, and $\boldsymbol{D}(z)$ is a diagonal matrix expressed as

$$\boldsymbol{D}(z) = diag\Big\{ \sum_{k=-\infty}^{\infty} d_1(k) z^{-k}, ..., \sum_{k=-\infty}^{\infty} d_N(k) z^{-k} \Big\}.$$

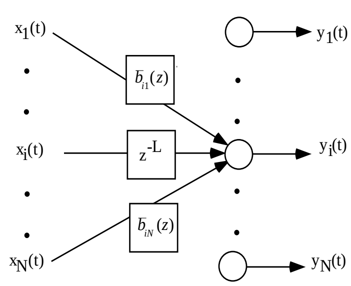

Figure 1: Signal separation network

$\bar{s}_i(t)(i = 1, ..., N)$ can also be regarded as source signals, because $\bar{s}_i(t)(i = 1, ..., N)$ are mutually independent signals. Therefore, our goal is now to find the matrix $\overline{\boldsymbol{B}}_0(z)$ satisfying $\boldsymbol{C}(z) = \boldsymbol{D}(z)\boldsymbol{P}$.

4. SEPARATION METHOD

In order to find the matrix $\overline{\boldsymbol{B}}_0(z)$ satisfying $\boldsymbol{C}(z) = \boldsymbol{D}(z)\boldsymbol{P}$, we use the following function:

$$\begin{aligned} Q(t, \overline{\boldsymbol{B}}(z)) &= \frac{1}{2}\Big\{ \sum_{i=1}^{N} log E[y_i(t-L)^2] \\ &\qquad - log det E[\boldsymbol{y}(t-L)\boldsymbol{y}(t-L)^T] \Big\}. \qquad (7) \end{aligned}$$

Note that the parameter L of eqn (7) represents the same delay as the one in eqn (4). In our method, time $t - L$ is regarded as $t = 0$. Therefore, our algorithm has access to both future and past values of the observed signals, that is, $\{\boldsymbol{x}(t), ..., \boldsymbol{x}(t-L+1)\}$ and $\{\boldsymbol{x}(t-L-1), ..., \boldsymbol{x}(t-M)\}$, respectively. Owing to this, our proposed algorithm can be applied to non-minimum phase systems. The function given by eqn (7) evaluates only one set of cross-correlations, $E[y_i(t-L)y_j(t-L)](i, j = 1, ..., N; i \neq j)$, and data outside that set, for example, $E[y_i(t)y_j(t-\tau)](i, j = 1, ..., N; i \neq j; \forall\tau)$ are not taken into account.

Matrix $\overline{\boldsymbol{B}}_0(z)$ (satisfying $\boldsymbol{C}(z) = \boldsymbol{D}(z)\boldsymbol{P}$) is found by minimizing the function $Q(t, \overline{\boldsymbol{B}}(z))$. In order to minimize the cost function (7) the steepest descent method is used:

$$\begin{aligned} \Delta\boldsymbol{B}(k) &\doteq -\alpha \frac{\partial Q(t, \overline{\boldsymbol{B}}(z))}{\partial \boldsymbol{B}(k)} = -\alpha \left[\frac{\partial Q(t, \overline{\boldsymbol{B}}(z))}{\partial b_{ij}(k)} \right] \\ &\qquad (k = 0, ..., M), \qquad (8) \end{aligned}$$

where α is a small positive constant. The symbol $\doteq$ in eqn (8) indicates that only the non-diagonal elements on the left-hand side of eqn (8) are equivalent to those on right-hand side.

Calculating the right-hand side of eqn (8), we have

$$\begin{aligned} \Delta\boldsymbol{B}(k) &\doteq \alpha z^{-k} \Big\{ \boldsymbol{I} - (diag E[\boldsymbol{y}(t-L)\boldsymbol{y}(t-L)^T])^{-1} \\ &\qquad \times E[\boldsymbol{y}(t-L)\boldsymbol{y}(t-L)^T] \Big\} \overline{\boldsymbol{B}}(z)^{-T} \\ &\qquad (k = 0, ..., M), \qquad (9) \end{aligned}$$

where $diag\mathbf{X}$ represents a diagonal matrix with the diagonal elements of matrix $\mathbf{X}$.

In practice, $E[\boldsymbol{y}(t-L)\boldsymbol{y}(t-L)^T]$ is replaced by its instantaneous value $\boldsymbol{y}(t-L)\boldsymbol{y}(t-L)^T$. To estimate $diag E[\boldsymbol{y}(t-L)\boldsymbol{y}(t-L)^T]$, we use the following moving average:

$$\begin{aligned} \phi_i(t) &= \beta\phi_i(t-1) + (1-\beta) y_i(t-L)^2 \\ &\qquad (i = 1, ..., N; 0 < \beta < 1). \qquad (10) \end{aligned}$$

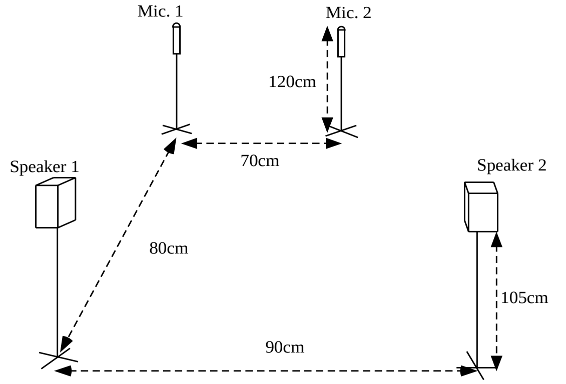

Figure 2: The configuration of two speakers and two microphones

Then, eqn (9) becomes

$$\Delta \boldsymbol{B}(k) \doteq \alpha z^{-k}\Big\{\boldsymbol{I} - \boldsymbol{\Phi}(t)^{-1} \times \boldsymbol{y}(t-L)\boldsymbol{y}(t-L)^T\Big\}\overline{\boldsymbol{B}}(z)^{-T} \quad (k = 0, ..., M), \tag{11}$$

where $\boldsymbol{\Phi}(t) = diag\{\phi_1(t), ..., \phi_N(t)\}$. Eqns (10) and (11) are used to update $\boldsymbol{B}(k)$ $(k = 0, ..., M)$.

5. EXPERIMENTS: N=2

The validity of the proposed method has been confirmed by computer simulation and an experiment in an anechoic room [7]. In this experiment, we applied the proposed method to extract source signals from their mixtures observed in a normal room. We consider the case in which the number of sources and observed signals is two, i.e., N = 2. In this case, eqn (11) becomes

$$\begin{aligned}\Delta b_{12}(k) &= -\frac{\alpha y_1(t-L)y_2(t-k)}{(1 - z^{2L}b_{12}(z)b_{21}(z))\phi_1(t)} \\ \Delta b_{21}(k) &= -\frac{\alpha y_2(t-L)y_1(t-k)}{(1 - z^{2L}b_{12}(z)b_{21}(z))\phi_2(t)} \\ &\qquad (k = 0, ..., M).\end{aligned} \tag{12}$$

In this section, we use the simplified algorithm obtained by omitting the common term $1/(1 - z^{2L}b_{12}(z)b_{21}(z))$ of eqn (12):

$$\Delta b_{ij}(k) = -\alpha y_i(t-L)y_j(t-k)/\phi_i(t) \quad (i, j = 1, 2; i \neq j). \tag{13}$$

Several experiments have been performed to demonstrate the validity of our method. This section describes three of them.

Example 1: The source signals $s_1(t)$ and $s_2(t)$ were parts of a speech given from one male person. And they were input at the same time to two speaker devices. The observed signals $x_1(t)$ and $x_2(t)$ were detected by two microphones (nondirectional microphones). This experiment was implemented in a normal room with air conditioning and computer noises. The configuration of the two speakers and two microphones is shown in Fig. 2. Parameters M and L of eqn (4) were set to 800 and 100, respectively. The parameters of the learning algorithm were chosen as $\alpha =$ 0.00001 (see eqn (13)) and $\beta = 0.9$ (see eqn (10)). The initial values of $b_{ij}(k)$ $(k = 0, ..., 799; i, j = 1, 2; i \neq j)$ and $\phi_i(t)$ were set to 0 and 1, respectively.

Fig. 3 shows the plots of $s_i(t)$, $x_i(t)$, and $y_i(t)$ (i = 1, 2). It can be seen that the output signals $y_1(t)$ and $y_2(t)$ are close to the original speech signals $s_1(t)$ and $s_2(t)$, respectively. Therefore, one can see that our method could separate the source signals from their mixtures observed in a normal room.

Example 2: In this example, source signals $s_1(t)$ and $s_2(t)$ were music and a male voice, respectively. The configuration of two speakers and two microphones is the same as the case of example 1. We used the same parameters (M, L, α, β) and the same initial values of $b_{12}(k)$ and $b_{21}(k)$ as in example 1.

Fig. 4 shows the plots of $s_i(t)$, $x_i(t)$, and $y_i(t)$ (i = 1, 2). It can be seen that the output signals $y_1(t)$ and $y_2(t)$ are close to the original signals $s_1(t)$ and $s_2(t)$, respectively.

Example 3: In this example, observed signals $x_1(t)$ and $x_2(t)$ are the data which were obtained from Computational Neurobiology Lab.'s Blind Source Separation Web Page (http://www.cnl.salk.edu / tewon/Blind/ blind.html, Audio Examples page, 2. Speech-Speech Separation). [On source signals and the configuration of two speakers and two microphones, see his home page.] Parameters M and L of eqn (4) were set to 200 and 50, respectively. The parameters of the learning algorithm were chosen as $\alpha = 0.00001$, $\beta = 0.9$, and the initial values of $b_{12}(k)$, $b_{21}(k)$ $(k = 0, ..., 199)$, and $\phi_i(t)$ were set to 0, 0, and 1, respectively.

Fig. 5 shows the plots of $x_i(t)$, our result $y_i(t)$, and Te-Won's result $h_i(t)$ (i = 1, 2). We confirmed that our proposed algorithm can completely separate original signals from their mixtures $x_i(t)$ (i = 1, 2).

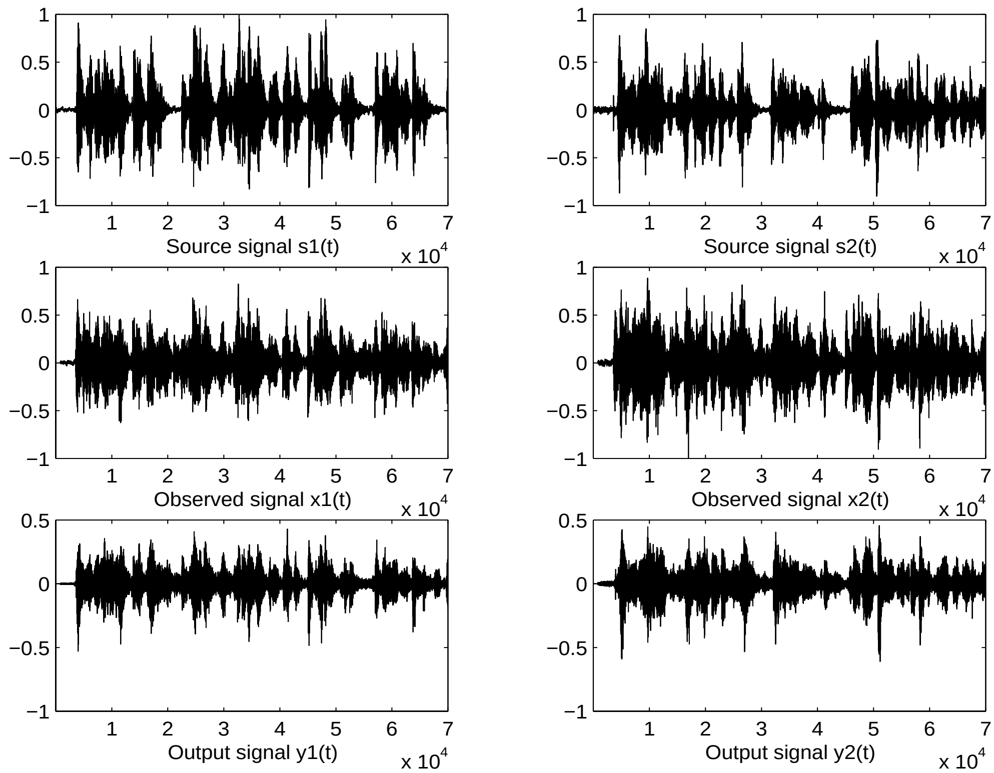

Figure 3: The plots of $s_i(t)$,$x_i(t)$,$y_i(t)$ $(i = 1, 2)$

6. CONCULSION

We have presented a method of blind separation for convolved nonstationary signals.

We have shown the results of an experiment of real world blind separation for convolved nonstationary signals. The experiment was implemented in a normal room. The room had air conditioning and computer noises. It has been shown that our method can separate two original signals from their mixtures observed in an ordinary room. In example 3 of section 5, we used the data which were obtained from Computational Neurobiology Lab.'s Blind Source Separation Web Page. We have confirmed in this example that our proposed method can separate two speech signals from their mixtures observed in a normal office.

7. REFERENCES

[1] D. C. B. Chan, S. J. Godsill, and P. J. W. Rayner, "Blind signal separation by output decorrelation," NIPS '96 Workshop on Blind Signal Processing and Their Applications.

[2] A. Cichoki, S. Amari, and J. Cao, "Blind separation of delayed and convolutive signals with self-adaptive learning rate," Proceedings of 1996 International Symposium on Nonlinear Theory and its Applications, pp. 229-232, 1996.

[3] F. Ehlers and H. G. Schuster, "Blind Separation of Convolutive Mixtures and an Application in Automatic Speech Recognition in a Noisy Environment," IEEE Trans. Signal Processing, Vol. 45, No. 10, pp.2608-2612, 1997.

[4] S. Van Gerven, D. Van Compernolle, H. L. Nguyen Thi, and C. Jutten, "Blind separation of sources: a comparative study of a 2-nd and a 4-th order solution," Proceedings of Signal Processing VII, Theories and Applications, pp. 1153-1156, 1994.

[5] S. Van Gerven and D. Van Compernolle, "Signal separation by symmetric adaptive decorrelation; Stability, convergence, and uniqueness," IEEE Trans. on SP, 43(7); pp. 1602-1612, July 1995.

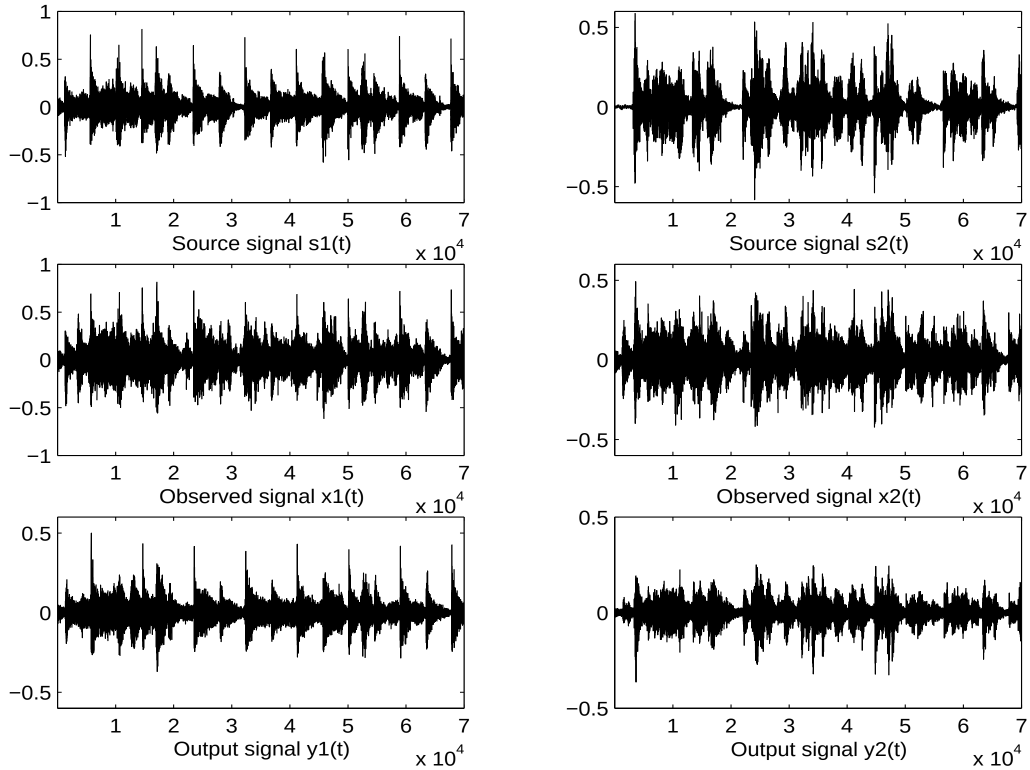

Figure 4: The plots of $s_i(t)$,$x_i(t)$,$y_i(t)$ $(i = 1, 2)$

[6] M. Kawamoto, K. Matsuoka, and M. Oya, "Blind separation of sources using temporal correlation of the observed signals," IEICE Trans. Fundamentals, Vol. E80-A, No. 4, April 1997.

[7] M. Kawamoto, A. K. Barros, and N. Ohnishi, "A neural network for blind separation of convolved non-stationary signals," Proc. International ICSC Workshop on Independent and Artificial Neural Networks 98, Tenerife, Spain, pp.1374-1379, 1998.

[8] M. Kawamoto, K. Matsuoka, and N. Ohnishi, "Blind signal separation of convolved non-stationary signals," Proc. International Symposium on Nonlinear Theory and its Applications, Hawaii, pp.1001-1004, 1997.

[9] T-W. Lee, A. J. Bell, and R. H. Lambert, "Blind separation of delayed and convolved sources, " Advances in Neural Information Processing Systems 9, pp. 758-764, 1997.

[10] T-W. Lee, A.J. Bell and R. Orglmeister, "Blind Source Separation of Real World Signals," Proceedings of IEEE International Conference Neural Networks, June 97, Houston, pp 2129-2135.

[11] 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.

[12] K. Torkkola, "Blind separation of convolutive sources based on information maximization," Neural Networks for Signal Processing VI, pp. 423-432, 1996.

[13] U. A. Lindgren and H. Broman, "Source separation using a criterion based on second-order statistics," IEEE Trans. SP, Vol. 46, No. 7, July 1998.

[14] J. Xi and J. P. Reilly, "Blind separation of signals in a convolutive environment," Proc. ICASSP, pp. 1327-1331, 1997.

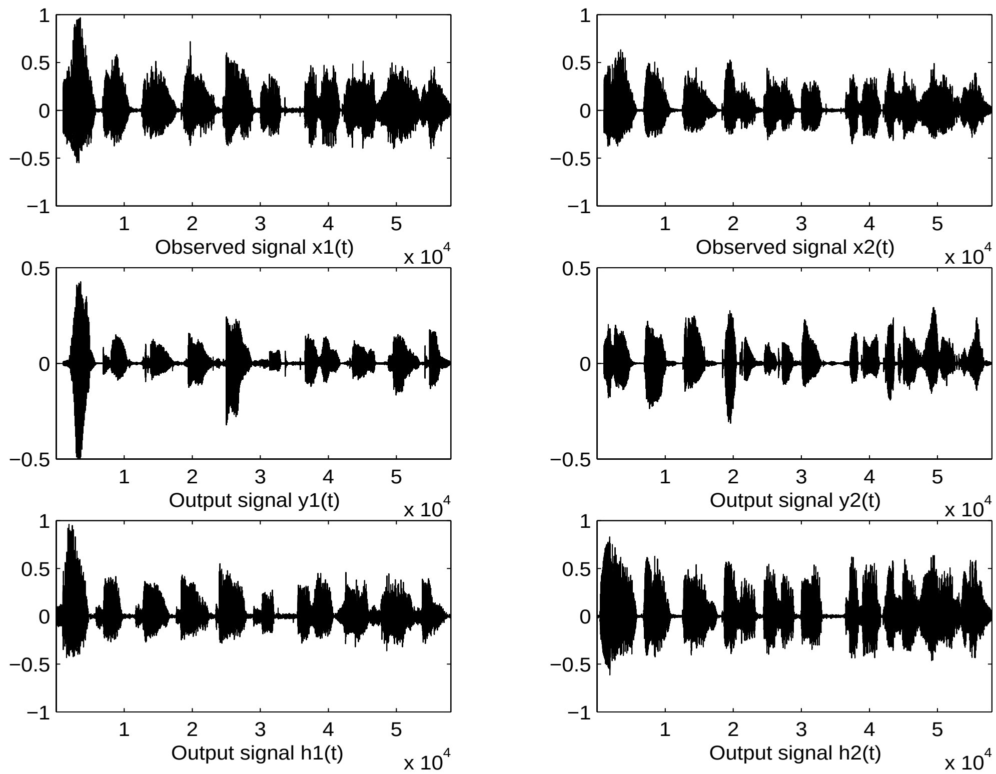

Figure 5: The plots of $x_i(t)$,$y_i(t)$,$h_i(t)$ $(i = 1, 2)$