

EVALUATION OF BLIND SEPARATION AND DECONVOLUTION FOR CONVOLUTIVE SPEECH MIXTURE USING SIMO-MODEL-BASED ICA

Hiroaki Yamajo, Hiroshi Saruwatari, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano

Graduate School of Information Science, Nara Institute of Science and Technology, 630-0192 Japan

ABSTRACT

We propose a new two-stage blind separation and deconvolution (BSD) algorithm for a convolutive mixture of speech, in which a new Single-Input Multiple-Output (SIMO)-model-based ICA (SIMO-ICA) and blind multichannel inverse filtering are combined. SIMO-ICA can separate the mixed signals, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. After SIMO-ICA, a simple blind deconvolution technique for the SIMO model can be applied even when each source signal is temporally correlated. The simulation results reveal that the proposed method can successfully achieve the separation and deconvolution for a convolutive mixture of speech.

1. INTRODUCTION

Blind separation and deconvolution (BSD) of sources is an approach taken to estimate original source signals using only the information of mixed signals observed in each input channel. In the BSD framework, not only the source separation but also the deconvolution of the transmission channel characteristics are considered. For the BSD based on independent component analysis (ICA), various methods have been proposed to deal with the separation and deconvolution for the convolutive mixture of independently, identically distributed (i.i.d.) source signals [1, 2]. These BSD methods require the specific assumptions that the source signals are mutually independent and each source signal is also temporally independent. However, the latter assumption does not hold in many practical acoustic mixtures of sound signals such as speech. The application of the conventional ICA-based BSD to speech often yields the negative results, e.g., the separated speech is adversely decorrelated and whitened. In order to solve the problem, we have proposed a novel BSD approach that combines information-geometry theory and multichannel signal processing [3]. In this approach, the BSD problem is resolved into two stages: new blind separation technique using a Single-Input Multiple-Output (SIMO)-model-based ICA (SIMO-ICA) and the deconvolution in the SIMO-model framework.

In the previous report[3], we dealt with real-world data, but it is hard to say that we could make clear whether the proposed BSD can obtain exact source signals or not. With real-world data, it is difficult to evaluate the performance of the system accurately due to background noise, too long reverberation, and so on. In this paper, we give the objective indication of the performance in the first stage, and properly evaluate the performance of the proposed method using the artificial transmission channels. In addition, we show that the proposed method can be regarded as a square FIR-type filter matrix, and we discuss the channel identifiability of such a system. The simulation results reveal that the pro-

posed method can achieve the separation and deconvolution for a convolutive mixture of speech when we set the SIMO-ICA's filter length sufficiently long.

2. MIXING PROCESS AND CONVENTIONAL BSD

In this study, the number of microphones is K and the number of multiple sound sources is L . The observed signals in which multiple source signals are mixed linearly are expressed as

$$\mathbf{x}(t) = \sum_{n=0}^{N-1} \mathbf{a}(n)\mathbf{s}(t-n) = \mathbf{A}(z)\mathbf{s}(t), \quad (1)$$

where $\mathbf{s}(t) = [s_1(t), \dots, s_L(t)]^T$ is the source signal vector, and $\mathbf{x}(t) = [x_1(t), \dots, x_K(t)]^T$ is the observed signal vector. Also, $\mathbf{a}(n)$ is the mixing filter matrix with the length of N , and $\mathbf{A}(z)$ is the z-transform of $\mathbf{a}(n)$; these are given as

$$\begin{aligned} \mathbf{a}(n) &= [a_{kl}(n)]_{kl}, \\ \mathbf{A}(z) &= [A_{kl}(z)]_{kl} = \left[\sum_{n=0}^{N-1} a_{kl}(n)z^{-n} \right]_{kl}, \end{aligned} \quad (2)$$

where z^{-1} is used as the unit-delay operator, i.e., $z^{-n} \cdot x(t) = x(t-n)$, $a_{kl}(n)$ is the impulse response between the k -th microphone and the l -th sound source, and $[X]_{ij}$ denotes the matrix which includes the element X in the i -th row and the j -th column. Hereafter, we only deal with the case of $K = L$ in this paper.

In the time-domain ICA (TDICA), the separated signal $\mathbf{y}(t) = [y_1(t), \dots, y_L(t)]^T$ is expressed as

$$\mathbf{y}(t) = \sum_{n=0}^{D-1} \mathbf{w}(n)\mathbf{x}(t-n) = \mathbf{W}(z)\mathbf{x}(t), \quad (4)$$

where $\mathbf{w}(n)$ is the separation filter matrix, $\mathbf{W}(z)$ is the z-transform of $\mathbf{w}(n)$, and D is the filter length of $\mathbf{w}(n)$. In the ICA-based BSD framework assuming i.i.d. sources, Amari [1] proposed the holonomic TDICA algorithm which optimizes the separation filter by minimizing the Kullback-Leibler divergence between the joint probability density function (PDF) of $\mathbf{y}(t)$ and the product of marginal PDFs of $y_l(t)$. The iterative learning rule is given by

$$\begin{aligned} \mathbf{w}^{[j+1]}(n) &= \mathbf{w}^{[j]}(n) + \eta \sum_{d=0}^{D-1} \left\{ \mathbf{I}\delta(n-d) \right. \\ &\quad \left. - \left\langle \varphi(\mathbf{y}^{[j]}(t))\mathbf{y}^{[j]}(t-n+d)^T \right\rangle_t \right\} \cdot \mathbf{w}^{[j]}(d) \end{aligned} \quad (5)$$

where η is the step-size parameter, the superscript $[j]$ is used to express the value of the j -th step in the iterations, $\langle \cdot \rangle_t$ denotes the

This work was partly supported by Core Research for Evolutional Science and Technology (CREST) in Japan.

time-averaging operator, and \mathbf{I} is the identity matrix. $\delta(n)$ is a delta function, where $\delta(0) = 1$ and $\delta(n) = 0$ ($n \neq 0$). $\varphi(\cdot)$ is the nonlinear vector function.

3. PROPOSED TWO-STAGE BSD ALGORITHM

In this section, we propose a new two-stage BSD algorithm combining SIMO-ICA and blind multichannel inverse filtering. In the proposed method, the separation/deconvolution problems can be solved efficiently using the following reasonable assumptions. **(A1)** The assumption of the mutual independence among the acoustic sound sources usually holds, and consequently, this can be used in the SIMO-ICA-based separation. **(A2)** The temporal-correlation property of the source signals and the nonminimum phase property of the mixing system can be taken into account in the blind multichannel inverse filtering for the SIMO model. The detailed process using the proposed algorithm is as follows.

3.1. First stage: SIMO-ICA for source separation

In this stage, a new blind separation method using SIMO-ICA is conducted. SIMO-ICA consists of multiple ICA parts and a *fidelity controller*, and each ICA runs in parallel under fidelity control of the entire separation system. The separated signals of the l -th ICA in SIMO-ICA are defined by

$$\begin{aligned} \mathbf{y}_{\text{ICA}l}(t) &= [y_k^{(l)}(t)]_{k1} = \sum_{n=0}^{D-1} \mathbf{w}_{\text{ICA}l}(n) \mathbf{x}(t-n) \\ &= \mathbf{W}_{\text{ICA}l}(z) \mathbf{x}(t), \end{aligned} \quad (6)$$

where $\mathbf{w}_{\text{ICA}l}(n)$ is the separation filter matrix in the l -th ICA, and $\mathbf{W}_{\text{ICA}l}(z)$ is the z -transform of $\mathbf{w}_{\text{ICA}l}(n)$. Regarding the fidelity controller, we introduce the following new cost function to be minimized,

$$\left\langle \left\| \sum_{l=1}^L \mathbf{y}_{\text{ICA}l}(t) - \mathbf{x}(t - D/2) \right\|^2 \right\rangle_t, \quad (7)$$

where $\|\mathbf{x}\|$ is the Euclidean norm of vector \mathbf{x} . Using (6) and (7), we can obtain the appropriate separated signals and maintain their spatial qualities as follows.

Theorem: If the independent sound sources are separated by (6), and simultaneously (7) is minimized to be zero, then the output signals converge on unique solutions, up to the permutation, as

$$\mathbf{y}_{\text{ICA}l}(t) = \text{diag} \left[\mathbf{A}(z) \mathbf{P}_l^T \right] \mathbf{P}_l \mathbf{s}(t - D/2), \quad (8)$$

where \mathbf{P}_l ($l = 1, \dots, L$) are exclusively-selected permutation matrices which satisfy

$$\sum_{l=1}^L \mathbf{P}_l = [\mathbf{1}]_{ij}. \quad (9)$$

As for the proof of theorem, we have given in [3].

Obviously the solutions given by (8) provide necessary and sufficient SIMO components, $A_{ki}(z) s_l(t - D/2)$, for each l -th source. There, however, is an arbitrariness in a selection of \mathbf{P}_l . For example, one possible selection is set permutation matrices, \mathbf{P}_l to following equation,

$$\mathbf{P}_l = [\delta_{im(k,l)}]_{ki}, \quad (10)$$

where δ_{ij} is Kronecker's delta function, and

$$m(k,l) = \begin{cases} k+l-1 & (k+l-1 \leq L) \\ k+l-1-L & (k+l-1 > L) \end{cases}. \quad (11)$$

In this case, (8) yields

$$\mathbf{y}_{\text{ICA}l}(t) = [A_{km(k,l)}(z) s_{m(k,l)}(t - D/2)]_{k1}. \quad (12)$$

In order to obtain (8), the gradient of (7) with respect to $\mathbf{w}_{\text{ICA}l}(n)$ should be added to the iterative learning rule of the separation filter. The natural gradient [1] of (7) is given as

$$\begin{aligned} & \left\{ \frac{\partial}{\partial \mathbf{w}_{\text{ICA}l}(n)} \left\langle \left\| \sum_{l=1}^L \mathbf{y}_{\text{ICA}l}(t) - \mathbf{x}(t - D/2) \right\|^2 \right\rangle_t \right\} \\ & \quad \cdot \mathbf{W}_{\text{ICA}l}(z^{-1})^T \mathbf{W}_{\text{ICA}l}(z) \\ & = 2 \sum_{d=0}^{D-1} \left\langle \left(\sum_{l=1}^L \mathbf{y}_{\text{ICA}l}(t) - \mathbf{x}(t - D/2) \right) \right. \\ & \quad \left. \cdot \mathbf{y}_{\text{ICA}l}(t - n + d)^T \right\rangle_t \cdot \mathbf{w}_{\text{ICA}l}(d). \end{aligned} \quad (13)$$

By combining (13) with the nonholonomic TDICA [4], we can obtain a new iterative algorithm in the l -th ICA of SIMO-ICA as

$$\begin{aligned} & \mathbf{w}_{\text{ICA}l}^{[j+1]}(n) \\ & = \mathbf{w}_{\text{ICA}l}^{[j]}(n) \\ & \quad - \alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \varphi(\mathbf{y}_{\text{ICA}l}^{[j]}(t)) \mathbf{y}_{\text{ICA}l}^{[j]}(t - n + d)^T \right\rangle_t \right. \\ & \quad \left. + \beta \left\langle \left(\sum_{l=1}^L \mathbf{y}_{\text{ICA}l}^{[j]}(t) - \mathbf{x}(t - D/2) \right) \right. \right. \\ & \quad \left. \left. \cdot \mathbf{y}_{\text{ICA}l}^{[j]}(t - n + d)^T \right\rangle_t \right\} \cdot \mathbf{w}_{\text{ICA}l}^{[j]}(d), \end{aligned} \quad (14)$$

where α and β are the step-size parameters; α is for the control of the total update quantity and β is for fidelity control. In (14), updating of $\mathbf{w}_{\text{ICA}l}(n)$ for all l should be simultaneously performed in parallel in terms of l because each iterative equation is associated with the others via $\sum_{l=1}^L \mathbf{y}_{\text{ICA}l}^{[j]} = \sum_{l=1}^L \mathbf{W}_{\text{ICA}l}^{[j]}(z) \mathbf{x}(t)$. Also, the initial values of $\mathbf{w}_{\text{ICA}l}(n)$ for all l should be different. If not, each ICA has the same set of inputs and will produce the same outputs. This results in an undesired solution. However, if we use different initial values, then the convergence on the appropriate SIMO solution is guaranteed by the simultaneous minimization of (6) and (7).

3.2. Second stage: Blind multichannel inverse filtering for deconvolution

In this stage, first, consider the blind channel identification corresponding to the first sound source $s_1(t)$, where we deal with the case of $K = L = 2$. Note that this can be easily extended to the general case ($K > 2$) by picking up the arbitrary two SIMO components from SIMO-ICA's outputs. In this process, the room transfer functions, $A_{11}(z)$ and $A_{21}(z)$, can be estimated by a subchannel matching approach [5, 6, 7] in an SIMO framework because we have already resolved the mixing process of the sources into a simple SIMO model through SIMO-ICA in the previous stage. The subchannel matching approach can work even for the temporally correlated signal. Regarding the blind channel identification corresponding to another sound source $s_2(t)$, we can estimate $A_{12}(z)$ and $A_{22}(z)$ using the same approach.

Finally, we can estimate the multichannel inverse filters, $G_{11}(z)$ and $G_{21}(z)$ for $\hat{A}_{11}(z)$ and $\hat{A}_{21}(z)$, and $G_{12}(z)$ and $G_{22}(z)$ for

$\hat{A}_{12}(z)$ and $\hat{A}_{22}(z)$, based on the multiple-input/output inverse theorem (MINT) [8]. In the MINT method, the exact inverse of the room acoustics can be uniquely determined, even when $\hat{A}_{kl}(z)$ has the nonminimum phase properties, if $\hat{A}_{kl}(z)$ does not have any common zeros in the z -plane. For example, the recovered signals $\hat{s}_l(t)$ under (12) are given as

$$\hat{s}_1(t) = G_{11}(z)y_1^{(1)}(t) + G_{21}(z)y_2^{(2)}(t), \quad (15)$$

$$\hat{s}_2(t) = G_{12}(z)y_1^{(2)}(t) + G_{22}(z)y_2^{(1)}(t). \quad (16)$$

The accurate estimation of the filter length N of the room impulse responses is indispensable for improving the system identification performance. There are various methods for filter-length estimation and we use the Furuya's method [7] in this work.

3.3. Discussion on identifiability

In this section, we first derive the entire filter used in the proposed method. Secondly we discuss the channel identifiability of the proposed BSD.

Using (6) and (7), we can express the recovered source signals (15) and (16) as

$$\hat{s}_1(t) = [G_{11}(z), G_{21}(z)] \begin{bmatrix} W_{11}^{(\text{ICA1})}(z) & W_{12}^{(\text{ICA1})}(z) \\ W_{21}^{(\text{ICA2})}(z) & W_{22}^{(\text{ICA2})}(z) \end{bmatrix} \cdot \mathbf{x}(t), \quad (17)$$

$$\hat{s}_2(t) = [G_{12}(z), G_{22}(z)] \begin{bmatrix} W_{11}^{(\text{ICA2})}(z) & W_{12}^{(\text{ICA2})}(z) \\ W_{21}^{(\text{ICA1})}(z) & W_{22}^{(\text{ICA1})}(z) \end{bmatrix} \cdot \mathbf{x}(t). \quad (18)$$

Thus, we obtain the entire input-output relation,

$$[\hat{s}_1(t), \hat{s}_2(t)]^T = \bar{\mathbf{W}}(z)\mathbf{x}(t), \quad (19)$$

where

$$\bar{\mathbf{W}}(z) = \begin{bmatrix} G_{11}(z)W_{11}^{(\text{ICA1})}(z) + G_{21}(z)W_{21}^{(\text{ICA2})}(z), & G_{11}(z)W_{12}^{(\text{ICA1})}(z) + G_{21}(z)W_{22}^{(\text{ICA2})}(z) \\ G_{12}(z)W_{11}^{(\text{ICA2})}(z) + G_{22}(z)W_{21}^{(\text{ICA1})}(z), & G_{12}(z)W_{12}^{(\text{ICA2})}(z) + G_{22}(z)W_{22}^{(\text{ICA1})}(z) \end{bmatrix}. \quad (20)$$

$\bar{\mathbf{W}}(z)$ is the resultant separation filter matrix, and is represented as a square (2×2) polynomial matrix with a finite order of less than $D + N - 1$. Here, N corresponds to the length of the multichannel inverse filter $G_{ij}(z)$, and is automatically determined in accordance with the length of $\mathbf{A}(z)$. On the other hand, D , the length of the separation filter $W_{ij}^{(\text{ICA}l)}(z)$ in SIMO-ICA, can be arbitrarily set by the user.

Previous studies [9, 10, 11] have indicated that the channel identification cannot be realized in the case of $K = L$ without special assumptions. Therefore the proposed BSD cannot obtain the exact source signals in theory because the entire filter is a square polynomial matrix. Since the deconvolution in the second stage can be performed exactly, it is considered that the separation to the SIMO model in the first stage includes a few residuals. In practice, however, we can reduce the residuals by setting filter length D in the first stage to be sufficiently long; this can be shown in the next simulation. Thus, the SIMO-model-based signals are approximately reproduced in this case. Overall, the identifiability almost holds under the assumption that we are allowed to use the long FIR filters in SIMO-ICA as well as (A1) and (A2).

4. SIMULATIONS

4.1. Conditions for experiment

The mixing filter matrix $\mathbf{A}(x)$ is taken to be $A_{11}(z) = 1 - 0.7z^{-1} - 0.3z^{-2}$, $A_{21}(z) = z^{-1} + 0.7z^{-2} + 0.4z^{-3}$, $A_{12}(z) = z^{-1} + 0.7z^{-2} + 0.4z^{-3}$, and $A_{22}(z) = 1 - 0.7z^{-1} - 0.3z^{-2}$. Two sentences spoken by two male speakers are used as the original speech samples $\mathbf{s}(t)$. The sampling frequency is 8 kHz and the length of speech is limited to 7 seconds. The number of iterations in ICA is 15000.

We carry out the following two experiments.

(Experiment 1) We evaluate SIMO-ICA while the length of the separation filter, D , is varied from 4 to 128 taps. We change the step-size parameter α among $1 \times 10^{-6} \sim 2 \times 10^{-6}$, and set β to be 6×10^{-4} , and we find optima which give the best performances.

(Experiment 2) We compare three methods as follows: (a) conventional **holonomic ICA** (ICA-based BSD) [1] given by (5), (b) conventional **nonholonomic ICA** [4] given by (14) with setting $\beta=0$, and (c) **proposed two-stage BSD**. In SIMO-ICA, the step-size parameter α is 2×10^{-6} and β is 6×10^{-4} . Also, η is 1×10^{-6} in the holonomic ICA, and α is 1×10^{-6} in the nonholonomic ICA; these are optima which provide the best performances. The length of the separation filter is set to be 64 taps.

In these experiments, three objective evaluation scores are defined as described as follows. First, *SIMO-model accuracy* (SA) is defined as

$$SA_l = \frac{1}{K} \sum_k 10 \log_{10} \left\{ \frac{\langle \|ref_{kl}(t)\|^2 \rangle_t}{\langle \|y_k^{(l)}(t) - ref_{kl}(t)\|^2 \rangle_t} \right\}, \quad (21)$$

where $ref_{kl}(t) = A_{kl}(z)s_l(t)$. The SA is used as to indicate a degree of similarity between the SIMO-ICA's outputs $y_k^{(l)}(t)$ and SIMO-model-based signals $ref_{kl}(t)$. Secondly, *noise reduction rate* (NRR) [12], defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB, is used as the objective indication of separation performance, where we do not take into account the distortion of the separated signal. The SNRs are calculated under the assumption that the speech signal of the undesired speaker is regarded as noise. Thirdly, *mel cepstral distortion* (melCD) is used as the indication of deconvolution performance. In this study, we defined the melCD as the distance between the spectral envelope of the original source signal $s_l(t - D/2)$ and that of the separated output. The 40th-order Mel-scaled cepstrum based on the smoothed FFT spectrum is used. The melCD will be decreased to zero if the separation-deconvolution processing is performed perfectly.

4.2. Results and discussion

Figure 1 shows the results of SA, where the SA increases as the length of the separation filter, D , is increased to more than the length of the mixing system. In particular, the SA of about 30 dB, which is sufficiently accurate for the following deconvolution process, is achieved when the filter length is set to 64 taps. Thus, the SIMO-ICA can reproduce the SIMO-model-based signals using the sufficiently long filter. This result supports the discussion on the identifiability of the proposed method as described in Sect. 3.3.

When the channel identification was performed in the second stage, the proposed method could blindly estimate the length of $\mathbf{a}(n)$ at four taps successfully by using an existing Furuya's method [7] for SIMO model.

Figure 2 shows the results of NRR and melCD for different methods. From the results of NRR, it is evident that the separation performance of the holonomic ICA is too poor, but those of the

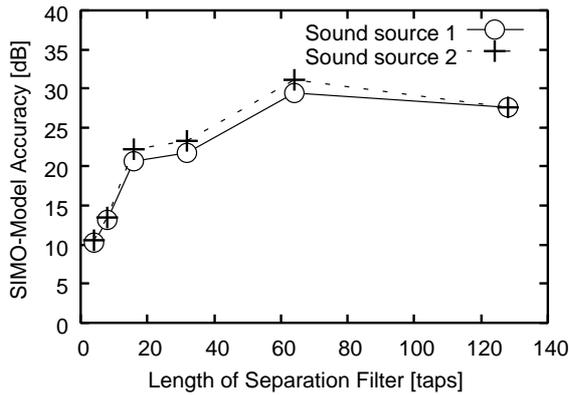


Fig. 1. SIMO-model accuracy of SIMO-ICA with different filter length.

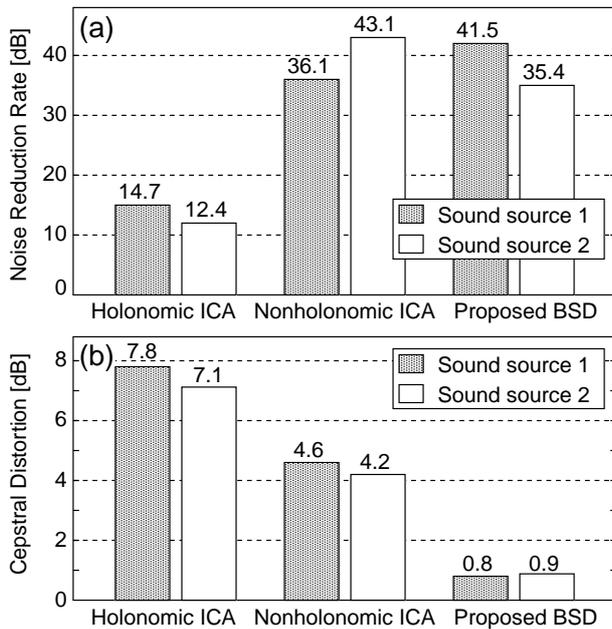


Fig. 2. Simulation Results for different methods with regard to (a) noise reduction rate, and (b) cepstral distortion.

proposed method and the nonholonomic ICA are high and comparable as far as the only separation performance is concerned. As for the distortion of the separated speech, which is an important issue from the practical viewpoint, there is a considerable difference between these methods, and this will be discussed in the next. From the results of melCD, first, it is evident that the melCD of the holonomic ICA is obviously high, i.e., the resultant speech is whitened by the decorrelation in the conventional method. Next, the result of the nonholonomic ICA shows that there are still some distortions in the separated signals. Finally, regarding the results of the proposed method, there is a considerable reduction of melCD. These results indicate that the proposed BSD algorithm can successfully achieve the separation and deconvolution for a convolutive mixture of temporally correlated signals using the sufficiently long separation filter in SIMO-ICA.

5. CONCLUSION

We proposed a new BSD framework in which SIMO-ICA and blind multichannel inverse filtering are efficiently combined. In

order to evaluate its effectiveness, a separation-deconvolution experiment was carried out assuming 2 microphones and 2 speech sources. The simulation results revealed that the conventional ICA-based method includes adverse spectral distortion due to the inherent whitening effect, and the spectral distortion can be considerably reduced by using the proposed two-stage BSD algorithm.

6. REFERENCES

- [1] S. Amari, S. Douglas, A. Cichocki, and H. H. Yang, "Multi-channel blind deconvolution and equalization using the natural gradient", *Proc. IEEE Int. Workshop on Wireless Communication*, pp.101–104, April 1997.
- [2] S. Haykin (ed.), *Unsupervised Adaptive Filtering*, John Wiley & Sons, Ltd., New York, 2000.
- [3] H. Saruwatari, T. Takatani, H. Yamajo, T. Nishikawa, K. Shikano, "Blind separation and deconvolution for real convolutive mixture of temporally correlated acoustic signals using SIMO-model-based ICA", *Proc. 4th Int. conf. on ICA and BSS (ICA2003)*, pp.549–554, Apr. 2003.
- [4] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," *Proc. Int. Workshop on ICA and BSS (ICA'99)*, pp.371–376, 1999.
- [5] H. Xu and L. Tong, "A deterministic approach to blind identification of multi-channel FIR systems," *Proc. ICASSP94*, pp.581–584, 1994.
- [6] Z. Ding and Y. Li, *Blind Equalization and Identification*, Marcel Dekker, Inc., New York, 2001.
- [7] K. Furuya and Y. Kaneda, "Two-channel blind deconvolution of nonminimum phase FIR system," *IEICE Trans. Fundamentals*, vol.E80-A, no.5, pp.804–808, 1997.
- [8] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol.36, no.2, pp.145–152, Feb. 1988.
- [9] K. Diamantaras, P. Petropulu, and B. Chen, "Blind two-input-two-output FIR channel identification based on frequency domain second-order statistics," *IEEE Trans. Signal Processing*, vol.48, no.2, pp.534–542, Feb. 2000.
- [10] Y. Inouye and K. Hirano, "Cumulant-based blind identification of linear multi-input-multi-output systems driven by colored inputs," *IEEE Trans. Signal Processing*, vol.45, no.6, pp.1543–1552, June 2000.
- [11] Y. Hua and J. Tugnait, "Blind identifiability of FIR-MIMO systems with colored input using second order statistics," *IEEE Signal Processing Letters*, vol.7, no.12, pp.348–350, Dec. 2000.
- [12] S. Araki, S. Makino, R. Mukai, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolved mixture of speech," *Proc. 3rd Int. conf. on ICA and BSS (ICA2001)*, pp.132–137, Dec. 2001.