HIGH-FIDELITY BLIND SEPARATION OF ACOUSTIC SIGNALS USING SIMO-MODEL-BASED ICA WITH INFORMATION-GEOMETRIC LEARNING

Tomoya TAKATANI, Tsuyoki NISHIKAWA, Hiroshi SARUWATARI, and Kiyohiro SHIKANO

Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192, JAPAN

E-mail: {tomoya-t, tsuyo-ni, sawatari, shikano}@is.aist-nara.ac.jp

ABSTRACT

We propose a new Single-Input Multiple-Output (SIMO)-model-based ICA with information-geometric learning algorithm for high-fidelity blind source separation. The SIMO-ICA consists of multiple ICAs and a fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system. The 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. Thus, the separated signals of SIMO-ICA can maintain the spatial qualities of each sound source. In order to evaluate its effectiveness, separation experiments are carried out under a reverberant condition. The experimental results reveal that the signal separation performance of the proposed SIMO-ICA is the same as that of the conventional ICA-based method, and that The sound quality of the separated sound in SIMO-ICA is superior to that of the conventional method.

1. INTRODUCTION

Blind source separation (BSS) is the approach taken to estimate original source signals using only the information of the mixed signals observed in each input channel. This technique is applicable to high-quality hands-free telecommunication systems. In recent works on BSS based on independent component analysis (ICA) [1], various methods have been proposed to deal with a means of separation of acoustical sounds which corresponds to the convolutive mixture case [2, 3]. However, the conventional ICA-based BSS approaches are basically means of extracting each of the independent sound sources as a *monaural* signal, and consequently they have a serious drawback in that the separated sounds cannot maintain information about the directivity, localization, or spatial qualities of each sound source. This prevents any BSS methods from being applied to binaural signal processing [4] or high-fidelity sound reproduction systems [5].

In this paper, we propose a new blind separation framework for Single-Input Multiple-Output (SIMO)-model-based acoustic signals using the extended ICA algorithm, SIMO-ICA. The SIMO-ICA consists of multiple ICA parts and a fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system. The fidelity controller as well as each of the ICA parts is designed on the basis of information-geometric theory proposed by, e.g., Amari et al. [6]. Namely, all of the procedures for optimization of the separation filters are conducted by the information-geometric learning algorithm. In the SIMO-ICA scenario, unknown multiple source signals which are mixed through unknown acoustical transmission channels are detected at the microphones, and these signals can be separated, not into monaural

source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Thus, the separated signals of the SIMO-ICA can maintain the spatial qualities of each sound source.

In order to evaluate its effectiveness, separation experiments are carried out under a reverberant condition. The experimental results reveal that the signal separation performance of the proposed SIMO-ICA is the same as that of the conventional ICA, and the sound quality of the separated signals in SIMO-ICA is remarkably superior to that in the conventional ICA.

2. MIXING PROCESS AND CONVENTIONAL BSS

2.1. Mixing process

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

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

where $\boldsymbol{s}(t) = [s_1(t), \cdots, s_L(t)]^{\mathrm{T}}$ is the source signal vector, $\boldsymbol{x}(t) = [x_1(t), \cdots, x_K(t)]^{\mathrm{T}}$ is the observed signal vector. Also, $\boldsymbol{a}(n) = [a_{kl}(n)]_{kl}$ is the mixing filter matrix with the length of N and $\boldsymbol{A}(z) = [A_{kl}(z)]_{kl} = [\sum_{n=0}^{N-1} a_{kl}(n)z^{-n}]$ is the z-transform of $\boldsymbol{a}(n)$, 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.

2.2. Conventional ICA-based BSS method

In the BSS method, we consider the time-domain ICA (TDICA), in which each element of the separation matrix is represented as an FIR filter. The separated signal $\boldsymbol{y}(t) = [y_1(t), \cdots, y_L(t)]^{\mathrm{T}}$ is expressed as

$$\begin{aligned} \boldsymbol{y}(t) &= \sum_{n=0}^{D-1} \boldsymbol{w}(n)\boldsymbol{x}(t-n) = \boldsymbol{W}(z)\boldsymbol{x}(t) \\ &= \boldsymbol{W}(z)\boldsymbol{A}(z)\boldsymbol{s}(t), \qquad (2) \end{aligned}$$

where $\boldsymbol{w}(n) = [w_{ij}(n)]_{ij}$ is the separation filter matrix, $\boldsymbol{W}(z)$ is the z-transform of $\boldsymbol{w}(n)$, and D is the filter length of $\boldsymbol{w}(n)$. In our study, the separation filter matrix is optimized by minimizing the Kullback-Leibler divergence (KLD) between the joint probability density function (PDF) of $\boldsymbol{y}(t)$ and the product of marginal PDFs of $y_l(t)$. The iterative learning rule is given by [7]
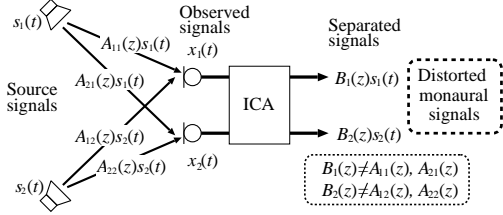
Fig. 1. Input and output relations in conventional ICA.

$$
\boldsymbol{w}^{[j+1]}(n) = \boldsymbol{w}^{[j]}(n) + \eta \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}^{[j]}(t)) \right. \right.
$$
$$
\left. \left. \cdot \boldsymbol{y}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \cdot \boldsymbol{w}^{[j]}(d), \qquad (3)
$$

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 time-averaging operator, and off-diag$\boldsymbol{W}(z)$ is the operation for setting every diagonal element of the matrix $\boldsymbol{W}(z)$ to be zero. Also, $\boldsymbol{\varphi}(\cdot)$ is the nonlinear vector function, e.g., the l-th element is set to be $\tanh(y_l(t))$.

2.3. Problems in conventional ICA

The conventional ICA is basically a means of extracting each of the independent sound sources as a monaural signal (see Fig. 1). In addition, the quality of the separated sound cannot be guaranteed, i.e., the separated signals can still include spectral distortions possibly because the modified separated signals which convolved with arbitrary linear filters are also mutually independent. As shown in Fig. 1, $y_l(t) = B_l(z)s_l(t)$, where $B_l(z)$ ($\neq A_{kl}(z)$) is an arbitrary filter, is a possible solution obtained from the conventional ICA using Eq. (3). Therefore, the conventional ICA has a serious drawback in that the separated sounds cannot maintain information about the directivity, localization, or spatial qualities of each sound source. In order to resolve the problem, particularly for the sound quality, Matsuoka et al. have proposed a modified ICA based on the Minimal Distortion Principle [8]. Since Matsuoka's method should simultaneously minimize the two different kinds of cost functions, namely, the KLD and the Euclidean distance between the separated signal vector and the observed signal vector, the method has the problem that the additional parameter for balancing the cost functions should be required, but is excessively sensitive for the convergence in the iterative learning of the separation filter matrix. The theoretical indication in selecting the parameter has not been presented. Also, this method is valid only for monaural outputs, and the fidelity of the output signals as SIMO-model-based signals cannot be guaranteed.

3. PROPOSED ALGORITHM: SIMO-ICA

In order to resolve the above-mentioned fundamental problems, we propose a new blind separation method for SIMO-model-based acoustic signals using SIMO-ICA. SIMO-ICA consists of $(L-1)$ ICA parts and a *fidelity controller*, and each ICA runs in parallel under the fidelity control of the entire separation system (see Fig. 2). The separated signals of the l-th ICA ($l = 1, \cdots, L-1$) in SIMO-ICA are defined by

$$
\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = [y_k^{(\mathrm{ICA}l)}(t)]_{k1} = \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(n)\boldsymbol{x}(t-n)
$$
$$
= \boldsymbol{W}_{(\mathrm{ICA}l)}(z)\boldsymbol{x}(t), \qquad (4)
$$

where $\boldsymbol{w}_{(\mathrm{ICA}l)}(n) = [w_{ij}^{(\mathrm{ICA}l)}(n)]_{ij}$ is the separation filter matrix in the l-th ICA. Regarding the fidelity controller, we calculate the following signal vector, in which the all elements are to be mutually independent,

$$
\boldsymbol{y}_{(\mathrm{ICA}L)}(t) = \boldsymbol{x}(t-D/2) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}(t). \qquad (5)
$$

Hereafter, we regard $\boldsymbol{y}_{(\mathrm{ICA}L)}(t)$ as an output of a *virtual* "L-th" ICA, and define its *virtual* separation filter matrix as

$$
\boldsymbol{w}_{(\mathrm{ICA}L)}(n) = \boldsymbol{I}\delta(n - \frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(n), \qquad (6)
$$
$$
\boldsymbol{W}_{(\mathrm{ICA}L)}(z) = \boldsymbol{I}z^{-\frac{D}{2}} - \sum_{l=1}^{L-1} \boldsymbol{W}_{(\mathrm{ICA}l)}(z). \qquad (7)
$$

From Eq. (6) and Eq. (7), we can rewrite Eq. (5) as

$$
\boldsymbol{y}_{(\mathrm{ICA}L)}(t) = \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA}L)}(n) \cdot \boldsymbol{x}(t-n)
$$
$$
= \boldsymbol{W}_{(\mathrm{ICA}L)}(z) \cdot \boldsymbol{x}(t). \qquad (8)
$$

The reason why we use the word "*virtual* " here is that the L-th ICA does not have own separation filters differing from the other ICAs, and $\boldsymbol{w}_{(\mathrm{ICA}L)}(n)$ and $\boldsymbol{W}_{(\mathrm{ICA}L)}(z)$ are subject to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ and $\boldsymbol{W}_{(\mathrm{ICA}l)}(z)$ ($l = 1, \cdots, L-1$).

To explicitly show the meaning of the fidelity controller, we rewrite Eq. (5) as

$$
\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-D/2) = [0]_{k1}. \qquad (9)
$$

Equation (9) means a constraint to force the sum of all ICAs' output vectors $\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t)$ to be the sum of all SIMO components $[\sum_{l=1}^{L} A_{kl}(z)s_l(t-D/2)]_{k1} (= \boldsymbol{x}(t-D/2))$. Here the delay of $D/2$ is used as to deal with nonminimum phase systems. Using Eq. (4) and Eq. (5), we can obtain the appropriate separated signals and maintain their spatial qualities as follows.

Theorem: If the independent sound sources are separated by Eq. (4), and simultaneously the signals obtained by Eq. (5) are also mutually independent, then the output signals converge on unique solutions, up to the permutation, as

$$
\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \text{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}} \right] \boldsymbol{P}_l \boldsymbol{s}(t-D/2), \qquad (10)
$$

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

$$
\sum_{l=1}^{L} \boldsymbol{P}_l = [1]_{ij}. \qquad (11)
$$

Proof of Theorem: The necessity is obvious. The sufficiency is shown below. Let $\boldsymbol{D}_l(z)$ ($l = 1, \cdots, L$) be arbitrary diagonal polynomial matrices and $\boldsymbol{Q}_l$ be arbitrary permutation matrices. The general expression of the l-th ICA's output is given by

$$
\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \boldsymbol{D}_l(z)\boldsymbol{Q}_l \boldsymbol{s}(t-D/2). \qquad (12)
$$

If $\boldsymbol{Q}_l$ are not exclusively-selected matrices, i.e., $\sum_{l=1}^{L} \boldsymbol{Q}_l \neq [1]_{ij}$, then there exists at least one element of $\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t)$ which does not include all components of $s_l(t-D/2)$ ($l = 1, \cdots, L$). This obviously makes the left-hand side of Eq. (9) nonzero because the observed signal vector $\boldsymbol{x}(t-D/2)$ includes all components
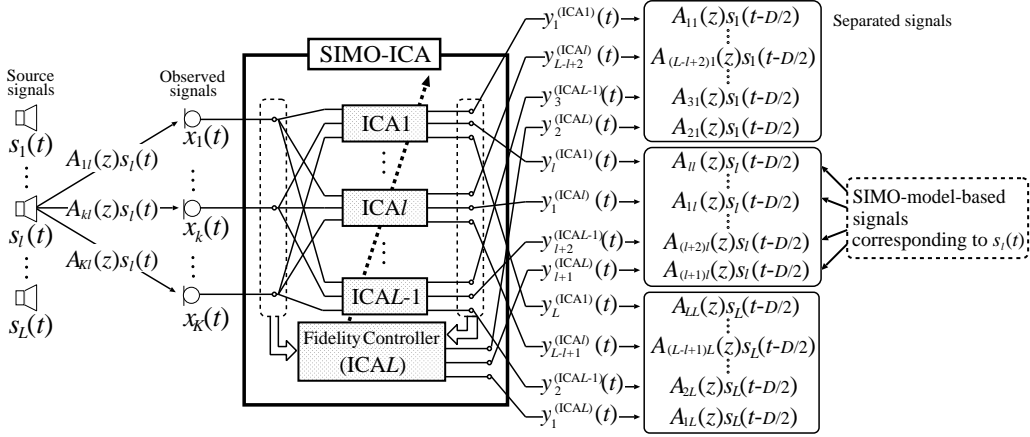
Fig. 2. Example of input and output relations in proposed SIMO-ICA, where permutation matrices $\boldsymbol{P}_l$ is given by Eq. (16).

of $s_l(t - D/2)$ in each element. Accordingly, $\boldsymbol{Q}_l$ should be $\boldsymbol{P}_l$ specified by Eq. (11), and we obtain

$$\boldsymbol{y}_{(\text{ICA}l)}(t) = \boldsymbol{D}_l(z)\boldsymbol{P}_l\boldsymbol{s}(t - D/2). \tag{13}$$

In Eq. (13) under Eq. (11), the arbitrary diagonal matrices $\boldsymbol{D}_l(z)$ can be substituted by $\text{diag}[\boldsymbol{B}(z)\boldsymbol{P}_l^{\text{T}}]$, where $\boldsymbol{B}(z) = [B_{ij}(z)]_{ij}$ is a single arbitrary matrix, because all diagonal entries of $\text{diag}[\boldsymbol{B}(z)\boldsymbol{P}_l^{\text{T}}]$ for all l are also exclusive. Thus,

$$\boldsymbol{y}_{(\text{ICA}l)}(t) = \text{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_l^{\text{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2). \tag{14}$$

Substitution of Eq. (14) in Eq. (9) leads to the following equation

$$\sum_{l=1}^{L} \text{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_l^{\text{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2)$$
$$- \left[\sum_{l=1}^{L} A_{kl}(z)s_l(t - D/2)\right]_{k1}$$
$$= \left[\sum_{l=1}^{L}\{B_{kl}(z) - A_{kl}(z)s_l(t - D/2)\}\right]_{kl} = [0]_{k1}. \tag{15}$$

Equation (15) is satisfied if and only if $B_{kl}(z) = A_{kl}(z)$ for all k and l. Thus, Eq. (14) results in Eq. (10). This completes the proof of Theorem.

Obviously the solutions given by Eq. (10) provide necessary and sufficient SIMO components, $A_{kl}(z)S_l(t - D/2)$, for each l-th source. However, the condition Eq. (11) allows multiple possibilities for the combination of $\boldsymbol{P}_l$. For example, one possibility is shown in Fig. 2 and this corresponds to

$$\boldsymbol{P}_l = [\delta_{im(k,l)}]_{ki}, \tag{16}$$

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

$$m(k,l) = \begin{cases} k + l - 1 & (k + l - 1 \leq L) \\ k + l - 1 - L & (k + l - 1 > L) \end{cases}. \tag{17}$$

In this case, Eq. (10) yields

$$\boldsymbol{y}_{(\text{ICA}l)}(t) = [A_{km}s_m(t - D/2)]_{k1} \quad (l \leq L). \tag{18}$$

In order to obtain Eq. (10), the natural gradient [6] of KLD of Eq. (8) with respect to $\boldsymbol{w}_{(\text{ICA}l)}(n)$ should be added to the iterative learning rule of the separation filter in l-th ICA ($l = 1, \cdots, L -$

1). The new iterative algorithm of l-th ICA part ($l = 1, \cdots, L-1$) in SIMO-ICA is given as

$$\boldsymbol{w}_{(\text{ICA}l)}^{[j+1]}(n)$$
$$= \boldsymbol{w}_{(\text{ICA}l)}^{[j]}(n) - \alpha \sum_{d=0}^{D-1}\left[\left\{\text{off-diag}\left\langle\boldsymbol{\varphi}\big(\boldsymbol{y}_{(\text{ICA}l)}^{[j]}(t)\big)\right.\right.\right.$$
$$\left.\left.\cdot\,\boldsymbol{y}_{(\text{ICA}l)}^{[j]}(t - n + d)^{\text{T}}\right\rangle_t\right\}\cdot\boldsymbol{w}_{(\text{ICA}l)}^{[j]}(d)$$
$$- \left\{\text{off-diag}\left\langle\boldsymbol{\varphi}\big(\boldsymbol{x}(t - \frac{D}{2}) - \sum_{l=1}^{L-1}\boldsymbol{y}_{(\text{ICA}l)}^{[j]}(t)\big)\right.\right.$$
$$\left.\left.\cdot\left(\boldsymbol{x}(t - n + d - \frac{D}{2}) - \sum_{l=1}^{L-1}\boldsymbol{y}_{(\text{ICA}l)}^{[j]}(t - n + d)^{\text{T}}\right)\right\rangle_t\right\}$$
$$\cdot\left(\boldsymbol{I}\delta(d - \frac{D}{2}) - \sum_{l=1}^{L-1}\boldsymbol{w}_{(\text{ICA}l)}^{[j]}(d)\right)\right], \tag{19}$$

where α is the step-size parameter. In Eq. (19), the updating $\boldsymbol{w}_{(\text{ICA}l)}(n)$ for all l should be simultaneously performed in parallel because each iterative equation is associated with the others via $\sum_{l=1}^{L-1}\boldsymbol{y}_{(\text{ICA}l)}^{[j]}(t)$. Also, the initial values of $\boldsymbol{w}_{(\text{ICA}l)}(n)$ for all l should be different.

After the iterations, the separated signals should be classified into SIMO components of each source because the permutation arises. This can be easily achieved by using a cross correlation between time-shifted separated signals, $\max_n\langle y_k^{(\text{ICA}l)}(t)y_{k'}^{(\text{ICA}l')}(t - n)\rangle_t$, where $l \neq l'$. The large value of the correlation indicates that $y_k^{(\text{ICA}l)}(t)$ and $y_{k'}^{(\text{ICA}l')}(t)$ are SIMO components of the same sources.

4. EXPERIMENT AND RESULTS

4.1. Conditions for experiment

As a preliminary study on the proposed SIMO-ICA, we carried out the source separation experiment using a simple microphone array, neglecting the effect of the head-related transfer function (HRTF) [4]. A two-element array with an interelement spacing of 4 cm is assumed. The speech signals are assumed to arrive from two directions, $-30°$ and $40°$. The distance between the microphone
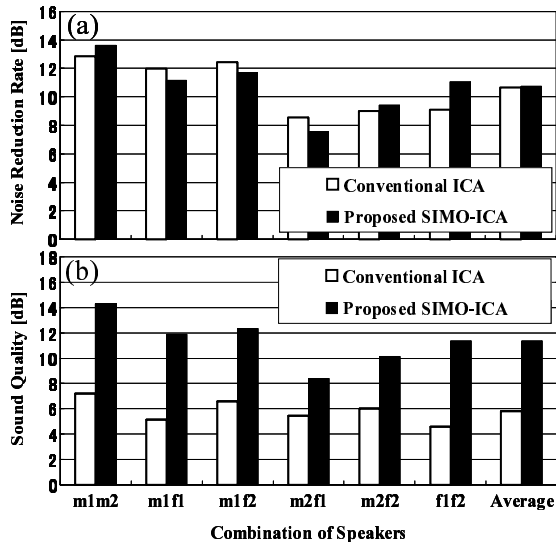
Fig. 3. Results of (a) NRR and (b) SQ in the reverberation time of 150 ms.

array and the loudspeakers is 1.15 m. Two kinds of sentences, spoken by two male and two female speakers, are used as the original speech samples. Using these sentences, we obtain 6 combinations. The sampling frequency is 8 kHz and the length of speech is limited to 3 seconds. The source signals are the original speech convolved with impulse responses specified by the reverberation times (RT) of 150 ms. The length of $w(n)$ is 512, and the initial value is Null-Beamformer [3] whose directional null is steered to $\pm 60°$. The number of iterations in ICA is 5000. Regarding the conventional ICA given for comparison, we used the conventional nonholonomic ICA [7].

In this experiment, two objective evaluation scores [9] are defined as described below. First, *noise reduction rate* (NRR), 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 signals. Secondly, *sound quality* (SQ) indicates the quality of the each separated signal. This evaluation score does not take into account the separation performance of the output signals.

4.2. Results and discussion

The step-size parameter α is changed from 5.0×10^{-8} to 5.0×10^{-6} in order to find the optima which give the best performance. Figure 3 (a) shows the results of NRR for different speaker combinations. The bars on the right of this figure correspond to the averaged results of each combination. In the averaged scores, the improvement of NRR in SIMO-ICA is 0.1 dB compared with that in the conventional ICA. From these results, it is evident the signal separation performance of the proposed SIMO-ICA is almost the same as that of the conventional ICA-based method.

Figure 3 (b) shows the result of SQ for different speaker combinations. The bars on the right of each figure correspond to the averaged results of each combination. In the averaged scores, compared with the conventional ICA, the improvement of SQ is 5.5 dB. From these results, it is evident that the sound quality of the separated signals in SIMO-ICA is obviously superior to that of the separated signals in the conventional ICA-based method.

Overall, the results indicate the following points. In SIMO-

ICA, the addition of a fidelity controller is effective in compensating for the spatial qualities of the separated SIMO-model-based signals. There is no deterioration in the separation performance (NRR) even with the additional compensation of sound quality in SIMO-ICA. Therefore, we can conclude that the proposed SIMO-ICA is applicable to binaural signal processing and high-fidelity sound reproduction systems.

5. CONCLUSION

We propose a new blind separation framework for SIMO-model-based acoustic signals using the extended ICA algorithm, SIMO-ICA. SIMO-ICA is an algorithm for separating the mixed signals, not into monaural source signals but into SIMO-model-based signals of independent sources without loss of their spatial qualities. In order to evaluate its effectiveness, separation experiments are carried out using 2 microphones and 2 sources under the condition that the RT is set to be 150 ms. The experimental results reveal that the signal separation performance of the proposed SIMO-ICA is the same as that of the conventional ICA-based method, and the spatial qualities of the separated sound in SIMO-ICA are superior to that in the conventional ICA-based method. Therefore, we can conclude that the proposed SIMO-ICA is applicable to binaural signal processing and high-fidelity sound reproduction systems.

7. REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Processing,* vol.36, pp.287–314, 1994.

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing,* vol.22, pp.21–34, 1998.

[3] H. Saruwatari, T. Kawamura, and K. Shikano, "Blind source separation for speech based on fast-convergence algorithm with ICA and beamforming," *Proc. Eurospeech2001,* pp.2603–2606, Sept. 2001.

[4] J. Blauert, *Spatial Hearing (revised edition).* Cambridge, MA: The MIT Press, 1997.

[5] Y. Tatekura, H. Saruwatari, and K. Shikano, "Sound reproduction system including adaptive compensation of temperature fluctuation effect for broad-band sound control," *IEICE Trans. Fundamentals,* vol.E85–A, no.8, pp.1851–1860, Aug. 2002.

[6] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications,* John Wiley & sons, Ltd, West Sussex, 2002.

[7] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," *Proc. International Workshop on ICA and BSS (ICA'99),* pp.371–376, 1999.

[8] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. International Conference on ICA and BSS,* pp.722–727, Dec. 2001.

[9] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "SIMO-Model-Based independent component analysis for high-fidelity blind separation of acoustic signals," *Proc. International Symposium on ICA and BSS,* pp.993–998, Apr. 2003.