SIMO-MODEL-BASED INDEPENDENT COMPONENT ANALYSIS FOR HIGH-FIDELITY BLIND SEPARATION OF ACOUSTIC SIGNALS

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 newly propose a novel 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 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 both nonreverberant and reverberant conditions. The experimental results reveal that (1) the signal separation performance of the proposed SIMO-ICA is the same as that of the conventional ICA-based method, and that (2) the spatial quality of the separated sound in SIMO-ICA is remarkably superior to that of the conventional method, particularly for the fidelity of the sound reproduction.

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, 4]. 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[5] or high-fidelity sound reproduction systems[6].

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. 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 SIMO-ICA can maintain the spatial qualities of each sound source.

In order to evaluate its effectiveness, separation experiments are carried out under nonreverberant and reverberant conditions. The experimental results reveal that (1) the signal separation performance of the proposed SIMO-ICA is the same as that of the conventional ICA, and (2) the sound quality of the separated signals in SIMO-ICA is remarkably superior to that in the conventional ICA, particularly for the spatial quality and the fidelity of the sound reproduction.

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

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

where $\boldsymbol{s}(t)$ is the source signal vector, $\boldsymbol{x}(t)$ is the observed signal vector, $\boldsymbol{a}(n)$ is the mixing filter matrix with the length of N, and $\boldsymbol{A}(z)$ is the z-transform of $\boldsymbol{a}(n)$; these are given as

$$\boldsymbol{s}(t) = [s_1(t), \cdots, s_L(t)]^{\mathrm{T}}, \tag{2}$$

$$\boldsymbol{x}(t) = [x_1(t), \cdots, x_K(t)]^{\mathrm{T}}, \tag{3}$$

$$\boldsymbol{a}(n) = \begin{bmatrix} a_{11}(n) & \cdots & a_{1L}(n) \\ \vdots & \ddots & \vdots \\ a_{K1}(n) & \cdots & a_{KL}(n) \end{bmatrix}, \tag{4}$$

$$\boldsymbol{A}(z) = [A_{ij}(z)]_{ij} = \left[\sum_{n=0}^{N-1} a_{ij}(n)z^{-n}\right]_{ij}, \tag{5}$$

where z^{-1} is used as the unit-delay operator, i.e., $z^{-n} \cdot x(t) = x(t-n)$, a_{kl} 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 a FIR filter. In the TDICA, we optimize the separation matrix by using only the fullband observed signals without subband processing

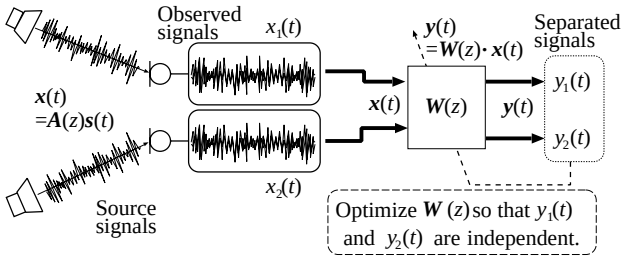

Fig. 1. Configuration of conventional TDICA.

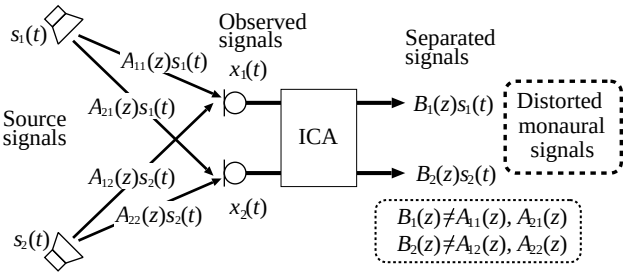

Fig. 2. Input and output relations in conventional ICA.

(see Fig. 1). The separated signal $\boldsymbol{y}(t) = [y_1(t), \cdots, y_L(t)]^{\mathrm{T}}$ is expressed as

$$\begin{aligned} \boldsymbol{y}(t) &= \sum_{n=0}^{D-1} \boldsymbol{w}(n)\boldsymbol{x}(t-n) = \boldsymbol{W}(z)\boldsymbol{x}(t) \\ &= \boldsymbol{W}(z)\boldsymbol{A}(z)\boldsymbol{s}(t), \end{aligned} \tag{6}$$

where $\boldsymbol{w}(n)$ 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 Kullbuck-Leibler divergence 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]

$$\begin{aligned} &\boldsymbol{w}^{[j+1]}(n) \\ &= \boldsymbol{w}^{[j]}(n) \\ &\quad - \alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}^{[j]}(t))\boldsymbol{y}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \\ &\quad \cdot \boldsymbol{w}^{[j]}(d), \end{aligned} \tag{7}$$

where α 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, we define the nonlinear vector function $\boldsymbol{\varphi}(\cdot)$ as

$$\boldsymbol{\varphi}(\boldsymbol{y}(t)) = [\tanh(y_1(t)), \cdots, \tanh(y_L(t))]^{\mathrm{T}}. \tag{8}$$

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. In addition, the quality of the separated sound cannot be guaranteed, i.e., the separated signals can possibly include spectral distortions because the modified separated signals which convolved with arbitrary linear filters are still mutually independent (see Fig. 2). 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]. However, this method is valid for only 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. 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 (see Fig. 3). The separated signals of the l-th ICA in SIMO-ICA are defined by

$$\begin{aligned} \boldsymbol{y}_{\mathrm{ICA}l}(t) &= [y_k^{(l)}(t)]_{k1} \\ &= \sum_{n=0}^{D-1} \boldsymbol{w}_{\mathrm{ICA}l}(n)\boldsymbol{x}(t-n) \\ &= \boldsymbol{W}_{\mathrm{ICA}l}(z)\boldsymbol{x}(t), \end{aligned} \tag{9}$$

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

$$\begin{aligned} &C(\boldsymbol{w}_{\mathrm{ICA}1}(n), ..., \boldsymbol{w}_{\mathrm{ICA}L}(n)) \\ &\quad \equiv \left\langle \| \sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t - D/2) \|^2 \right\rangle_t, \end{aligned} \tag{10}$$

where $\| \boldsymbol{x} \|$ is the Euclidean norm of vector $\boldsymbol{x}$. Using Eq. (9) and Eq. (10), we can obtain the appropriate separated signals and maintain their spatial qualities as follows.

Theorem: If the independent sound sources are separated by Eq. (9), and simultaneously Eq. (10) is minimized to be zero, then the output signals converge on unique solutions, up to the permutation, as

$$\boldsymbol{y}_{\mathrm{ICA}l}(t) = \mathrm{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2), \tag{11}$$

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

$$\sum_{l=1}^{L} \boldsymbol{P}_l = [1]_{ij}. \tag{12}$$

Proof of Theorem: The necessity is obvious. The sufficiency is shown below. Let $\boldsymbol{D}_l(z)$ $(l = 1, ..., L)$ 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

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

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

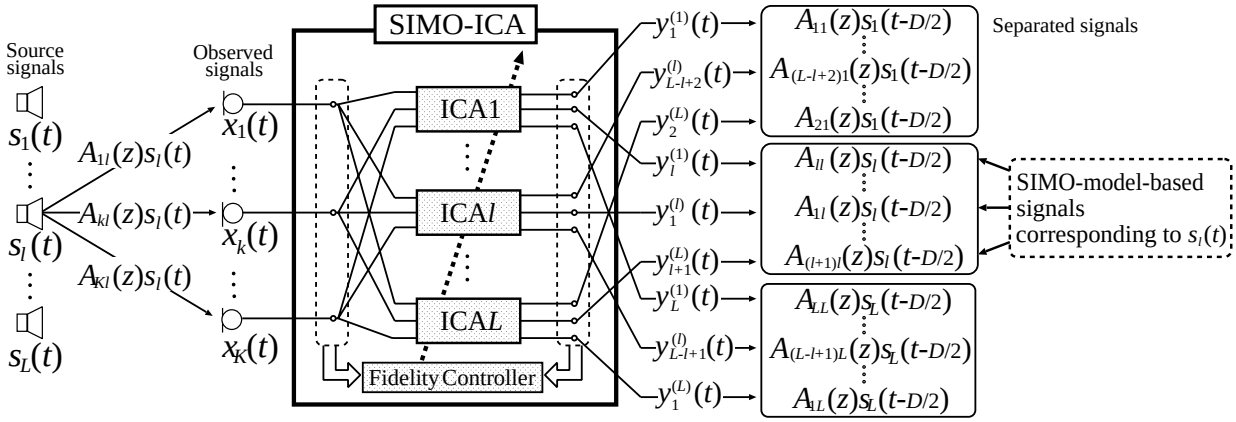

Fig. 3. Example of input and output relations in proposed SIMO-ICA, where exclusively-selected permutation matrices $\boldsymbol{P}_l$ are given by Eq. (18).

not include all components of $s_l(t - D/2)$ $(l = 1, ..., L)$. This obviously makes the cost function Eq. (10) be nonzero because the observed signal vector $\boldsymbol{x}(t - D/2)$ includes all components of $s_l(t - D/2)$ in each element. Accordingly, $\boldsymbol{Q}_l$ should be $\boldsymbol{P}_l$ specified by Eq. (12), and we obtain

$$\boldsymbol{y}_{\mathrm{ICA}l}(t) \quad = \quad \boldsymbol{D}_l(z)\boldsymbol{P}_l\boldsymbol{s}(t - D/2). \tag{14}$$

In Eq. (14) under Eq. (12), the arbitrary diagonal matrices $\boldsymbol{D}_l(z)$ can be substituted by $\mathrm{diag}[\boldsymbol{B}(z)\boldsymbol{P}_l^{\mathrm{T}}]$, where $\boldsymbol{B}(z) = [B_{ij}(z)]_{ij}$ is a single arbitrary matrix, because all diagonal entries of $\mathrm{diag}[\boldsymbol{B}(z)\boldsymbol{P}_l^{\mathrm{T}}]$ for all l are also exclusive. Thus,

$$\boldsymbol{y}_{\mathrm{ICA}l}(t) \quad = \quad \mathrm{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2), \tag{15}$$

and consequently

$$\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t) \quad = \quad \left[\sum_{l=1}^{L}B_{kl}(z)s_l(t - D/2)\right]_{k1}. \tag{16}$$

Substitution of Eq. (16) in Eq. (10) leads to the following equation.

$$\begin{aligned} &C(\boldsymbol{w}_{\mathrm{ICA}1}(n), ..., \boldsymbol{w}_{\mathrm{ICA}L}(n)) \\ &= \left\langle \left\| \left[\sum_{l=1}^{L}B_{kl}(z)s_l(t - D/2)\right]_{k1} \right.\right. \\ &\qquad \left.\left. - \left[\sum_{l=1}^{L}A_{kl}(z)s_l(t - D/2)\right]_{k1} \right\|^2 \right\rangle_t \\ &= \sum_{l=1}^{L}\sum_{k=1}^{K}(B_{kl}(z) - A_{kl}(z))^2 \cdot \left\langle s_l(t - D/2)^2 \right\rangle_t, \end{aligned} \tag{17}$$

where we used the relation, $\langle s_l(t - D/2)s_{l'}(t - D/2)\rangle_t = 0$ $(l \neq l')$. Since $\langle s_l(t - D/2)^2\rangle_t$ are positive, the cost function given by Eq. (17) becomes zero if and only if $B_{kl}(z) = A_{kl}(z)$ for all k and l. Thus, Eq. (15) results in Eq. (11). This completes the proof of Theorem.

Obviously the solutions given by Eq. (11) provide necessary and sufficient SIMO components, $A_{kl}(z)S_l(t - D/2)$, for each l-th source. There, however, is an arbitrariness in a selection of $\boldsymbol{P}_l$. For example, one possible selection is shown in Fig. 3 and this corresponds to

$$\boldsymbol{P}_l = [\delta_{im(k,l)}]_{ki}, \tag{18}$$

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

$$m(k,l) = \begin{cases} k + l - 1 & (k + l - 1 \leq L) \\ k + l - 1 - L & (k + l - 1 > L) \end{cases} \tag{19}$$

In this case, Eq. (11) yields

$$\boldsymbol{y}_{\mathrm{ICA}l}(t) = [A_{km}s_m(t - D/2)]_{k1}, \tag{20}$$

In order to obtain Eq. (11), the gradient of Eq. (10) with respect to $\boldsymbol{w}_{\mathrm{ICA}l}(n)$ should be added to the iterative learning rule of the separation filter given by Eq. (7).

The (standard) gradient of Eq. (10) is given as

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{w}_{\mathrm{ICA}l}(n)}\left\langle \| \sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t - \frac{D}{2}) \|^2 \right\rangle_t \\ &\quad = \quad 2\left\langle \left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t - \frac{D}{2})\right) \cdot \boldsymbol{x}(t - n)^{\mathrm{T}} \right\rangle_t. \end{aligned} \tag{21}$$

Here $\boldsymbol{x}(t-n)^{\mathrm{T}}$ is expressed as the following equation from Eq. (9):

$$\boldsymbol{x}(t - n)^{\mathrm{T}} = \boldsymbol{y}_{\mathrm{ICA}l}(t - n)^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ICA}l}(z)^{-\mathrm{T}}, \tag{22}$$

where the superscript $-\mathrm{T}$ represents the transposed inverse matrix. By using Eq. (22), Eq. (21) is expanded as

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{w}_{\mathrm{ICA}l}(n)}\left\langle \| \sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t - \frac{D}{2}) \|^2 \right\rangle_t \\ &\quad = \quad 2\left\langle \left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t - \frac{D}{2})\right) \right. \\ &\qquad \left. \cdot \boldsymbol{y}_{\mathrm{ICA}l}(t - n)^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ICA}l}(z)^{-\mathrm{T}} \right\rangle_t \\ &\quad = \quad 2\left\langle \left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t - \frac{D}{2})\right) \right. \\ &\qquad \left. \cdot \boldsymbol{y}_{\mathrm{ICA}l}(t - n)^{\mathrm{T}} \right\rangle_t \boldsymbol{W}_{\mathrm{ICA}l}(z^{-1})^{-\mathrm{T}}. \end{aligned} \tag{23}$$

Note that $\boldsymbol{W}(z^{-1})^{-\mathrm{T}}$ is a convolution operation with respect not to the index t but to the index n because the index t vanishes under the averaging $\langle\cdot\rangle_t$. From Eq. (23), the natural gradient [9, 10] of Eq. (10) is given as

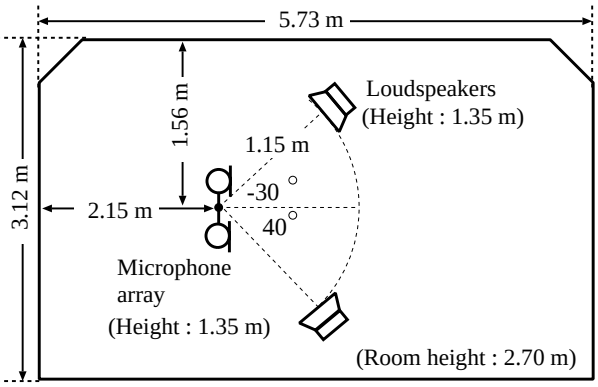

Fig. 4. Layout of reverberant room used in experiments.

$$
\begin{aligned}
&\left\{ \frac{\partial}{\partial \boldsymbol{w}_{\mathrm{ICA}l}(n)} \Big\langle \parallel \sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t-\frac{D}{2}) \parallel^2 \Big\rangle_t \right\} \\
&\quad \cdot \boldsymbol{W}_{\mathrm{ICA}l}(z^{-1})^{\mathrm{T}} \boldsymbol{W}_{\mathrm{ICA}l}(z) \\
&= 2\Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t-\frac{D}{2})) \\
&\quad \cdot \boldsymbol{y}_{\mathrm{ICA}l}(t-n)^{\mathrm{T}} \Big\rangle_t \cdot \boldsymbol{W}_{\mathrm{ICA}l}(z) \\
&= 2\sum_{d=0}^{D-1} \Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t-\frac{D}{2})) \\
&\quad \cdot \boldsymbol{y}_{\mathrm{ICA}l}(t-n+d)^{\mathrm{T}} \Big\rangle_t \cdot \boldsymbol{w}_{\mathrm{ICA}l}(d). \qquad (24)
\end{aligned}
$$

By combining Eq. (7) with Eq. (24), we can obtain the new iterative algorithm of SIMO-ICA as

$$
\begin{aligned}
&\boldsymbol{w}_{\mathrm{ICA}l}^{[j+1]}(n) \\
&= \boldsymbol{w}_{\mathrm{ICA}l}^{[j]}(n) \\
&\quad - \alpha \sum_{d=0}^{D-1} \Big\{ \text{off-diag} \Big\langle \boldsymbol{\varphi}(\boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t)) \boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t-n+d)^{\mathrm{T}} \Big\rangle_t \\
&\quad + \beta \Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t) - \boldsymbol{x}(t-\frac{D}{2})) \\
&\quad \cdot \boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t-n+d)^{\mathrm{T}} \Big\rangle_t \Big\} \cdot \boldsymbol{w}_{\mathrm{ICA}l}^{[j]}(d), \qquad (25)
\end{aligned}
$$

where α and β are the step-size parameters; α is for the control of the total update quantity and β is for the fidelity control. In Eq. (25), the updating $\boldsymbol{w}_{\mathrm{ICA}l}(n)$ should be simultaneously performed in parallel because each iterative equation is associated with the others via $\boldsymbol{y}_{\mathrm{ICA}l}^{[j]} = \boldsymbol{W}_{\mathrm{ICA}l}^{[j]}(z)\boldsymbol{x}(t)$.

4. EXPERIMENT AND RESULTS

4.1. Conditions for experiment

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 array and the loudspeakers is 1.15 m. Two kinds of sentences, spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research, 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 two kinds of impulse responses specified by the different reverberation times (RTs), 0 ms and 150 ms. The impulse response in the case of RT=150 ms is recorded in the experimental room (see Fig. 4). The length of $\boldsymbol{w}(n)$ is set to be 128 (RT=0 ms) or 512 (RT=150 ms), and the initial value is Null-Beamformer [3] whose directional null is steered to $\pm 60^\circ$. The number of iterations in ICA is 5000. Regarding the conventional ICA given for comparison, we used Eq. (25) in the case of $\beta = 0$.

4.2. Objective evaluation score

In this experiment, three objective evaluation scores 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 signal. The SNRs are calculated under the assumption that the speech signal of the undesired speaker is regarded as noise. The NRR is defined as

$$
\begin{aligned}
\mathrm{NRR} &\equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{k=1}^{2} \big(\mathrm{OSNR}_l^{(\mathrm{ICA}k)} - \mathrm{ISNR}_l^{(\mathrm{ICA}k)}\big), \qquad (26) \\
\mathrm{OSNR}_l^{(\mathrm{ICA}1)} &= 10\log_{10} \frac{\sum_t \mid H_{ll}^{\mathrm{ICA}1}(z) s_l(t) \mid^2}{\sum_t \mid H_{ln}^{\mathrm{ICA}1}(z) s_n(t) \mid^2}, \\
\mathrm{ISNR}_l^{(\mathrm{ICA}1)} &= 10\log_{10} \frac{\sum_t \mid A_{ll}(z) s_l(t) \mid^2}{\sum_t \mid A_{ln}(z) s_n(t) \mid^2}, \\
\mathrm{OSNR}_l^{(\mathrm{ICA}2)} &= 10\log_{10} \frac{\sum_t \mid H_{ln}^{\mathrm{ICA}2}(z) s_n(t) \mid^2}{\sum_t \mid H_{ll}^{\mathrm{ICA}2}(z) s_l(t) \mid^2}, \\
\mathrm{ISNR}_l^{(\mathrm{ICA}2)} &= 10\log_{10} \frac{\sum_t \mid A_{ln}(z) s_n(t) \mid^2}{\sum_t \mid A_{ll}(z) s_l(t) \mid^2},
\end{aligned}
$$

where $\mathrm{OSNR}_l^{(\mathrm{ICA}k)}$ and $\mathrm{ISNR}_l^{(\mathrm{ICA}k)}$ are the output SNR and the input SNR for ICAk, respectively, and $l \neq n$. Also, $H_{ij}^{\mathrm{ICA}k}(z)$ is the element in the i-th row and the j-th column of the matrix $\boldsymbol{H}^{\mathrm{ICA}k}(z) = \boldsymbol{W}_{\mathrm{ICA}k}(z)\boldsymbol{A}(z)$. Secondly, *sound quality* (SQ), defined as described below, indicates the sound quality of the separated signal,

$$
\begin{aligned}
\mathrm{SQ} &\equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{n=1}^{2} \mathrm{SQ}_{y_l^{(n)}}, \qquad (27) \\
\mathrm{SQ}_{y_1^{(1)}} &= 10\log_{10} \frac{\sum_t \mid A_{11}(z) s_1(t) \mid^2}{\sum_t \mid A_{11}(z) s_1(t) - H_{11}^{\mathrm{ICA}1}(z) s_1(t) \mid^2}, \\
\mathrm{SQ}_{y_2^{(1)}} &= 10\log_{10} \frac{\sum_t \mid A_{12}(z) s_2(t) \mid^2}{\sum_t \mid A_{12}(z) s_2(t) - H_{12}^{\mathrm{ICA}2}(z) s_2(t) \mid^2}, \\
\mathrm{SQ}_{y_1^{(2)}} &= 10\log_{10} \frac{\sum_t \mid A_{21}(z) s_1(t) \mid^2}{\sum_t \mid A_{21}(z) s_1(t) - H_{21}^{\mathrm{ICA}2}(z) s_1(t) \mid^2}, \\
\mathrm{SQ}_{y_2^{(2)}} &= 10\log_{10} \frac{\sum_t \mid A_{22}(z) s_2(t) \mid^2}{\sum_t \mid A_{22}(z) s_2(t) - H_{22}^{\mathrm{ICA}1}(z) s_2(t) \mid^2},
\end{aligned}
$$

where $\mathrm{SQ}_{y_l^{(n)}}$ is the sound quality of the separated signal $y_l^{(n)}$. Lastly, *fidelity* (F) indicates the accuracy of the sound reproduction in the entire system. It is defined by

$$
\mathrm{F} \equiv 10\log_{10} \frac{\Big\langle \parallel \boldsymbol{x}(t) \parallel^2 \Big\rangle_t}{\Big\langle \parallel \sum_{l=1}^{2} \boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}(t-\frac{D}{2}) \parallel^2 \Big\rangle_t}. \qquad (28)
$$

4.3. Results and discussion

4.3.1. Nonreverberant case (RT=0 ms)

The step-size parameter α is changed from 1.0×10^{-6} to 2.0×10^{-6} and β is changed from 2.0×10^{-3} to 4.0×10^{-4} in order to find the optima which minimize Eq. (10). Figure 5 (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 deterioration of NRR in SIMO-ICA is 4.4 dB compared with that in the conventional ICA. However, the absolute NRR score is more than 30 dB and consequently the deterioration of NRR is relatively small and negligible from the practical viewpoint.

Figures 5 (b) and (c) show the results of SQ and F 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 16.1 dB, and that of F is 26.1 dB. From these results, it is evident that the sound quality of the separated signals in SIMO-ICA is remarkably superior to that of the separated signals in the ICA-based method.

4.3.2. Reverberant case (RT=150 ms)

The step-size parameter α is changed from 5.0×10^{-8} to 1.0×10^{-6} and β is changed from 1.0×10^{-2} to 7.0×10^{-2} in order to find the optima which minimize Eq. (10). Figure 6 (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 deterioration of NRR in SIMO-ICA is 0.2 dB compared with that in the conventional ICA. From these results, it is evident that the signal separation performance of the proposed SIMO-ICA is almost the same as that of the conventional ICA-based method.

Figures 6 (b) and (c) show the results of SQ and F 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 3.3 dB, and that of F is 31.8 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, particularly in terms of the fidelity of the sound reproduction. Regarding the SQ score, the improvement in SIMO-ICA is not large compared with that of SIMO-ICA in the nonreverberant case described in the previous section. The main reason for this is the insufficiency of the source-separation performance. In order to improve this, the separation filter should be lengthened beyond the length of the reverberation time; it remains an open problem for future study.

Overall, the results indicate the following points. (1) In SIMO-ICA, the addition of a fidelity controller is effective in compensating for the spatial qualities of the separated SIMO-model-based signals. (2) 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 newly propose a novel 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 conditions that the RTs are set to be 0 ms and 150 ms. The experimental results reveal that (1) the signal separation performance of the proposed SIMO-ICA is the same as that of the conventional ICA-based method, and (2) the spatial qualities of the separated sound in SIMO-ICA are remarkably superior to that in the conventional ICA-based method, particularly in terms of the fidelity of the sound reproduction. Therefore, we can conclude that the proposed SIMO-ICA is applicable to binaural signal processing and high-fidelity sound reproduction systems. Further development extended to binaural signals remains an open problem for the future.

6. ACKNOWLEDGEMENT

The authors are grateful to Dr. Kiyotoshi Matsuoka of Kyusyu Institute of Technology, and Dr. Shoji Makino and Ms. Shoko Araki of NTT Co., Ltd. for their discussions. This work was partly supported by CREST (Core Research for Evolutional Science and Technology) in Japan.

7. REFERENCES

[1] P. Common, "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] T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of time-domain ICA, frequency-domain ICA and multistage ICA," *The 2002 European Signal Processing Conference (EUSIPCO2002)*, vol.II, pp.15–18, Sept. 2002.

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

[6] 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.

[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 Independent Component Analysis and Blind Signal Separation (ICA'99)*, pp.371–376, 1999.

[8] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. International Conference on Independent Component Analysis and Blind Signal Separation*, pp.722–727, Dec. 2001.

[9] S. Amari, S. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient, *Proc. of IEEE international Workshop on Wireless Communication*, pp.101–104, April 1997.

[10] M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved non-stationary signals," *Neurocomputing*, 22, pp.157–171, Dec. 1998.

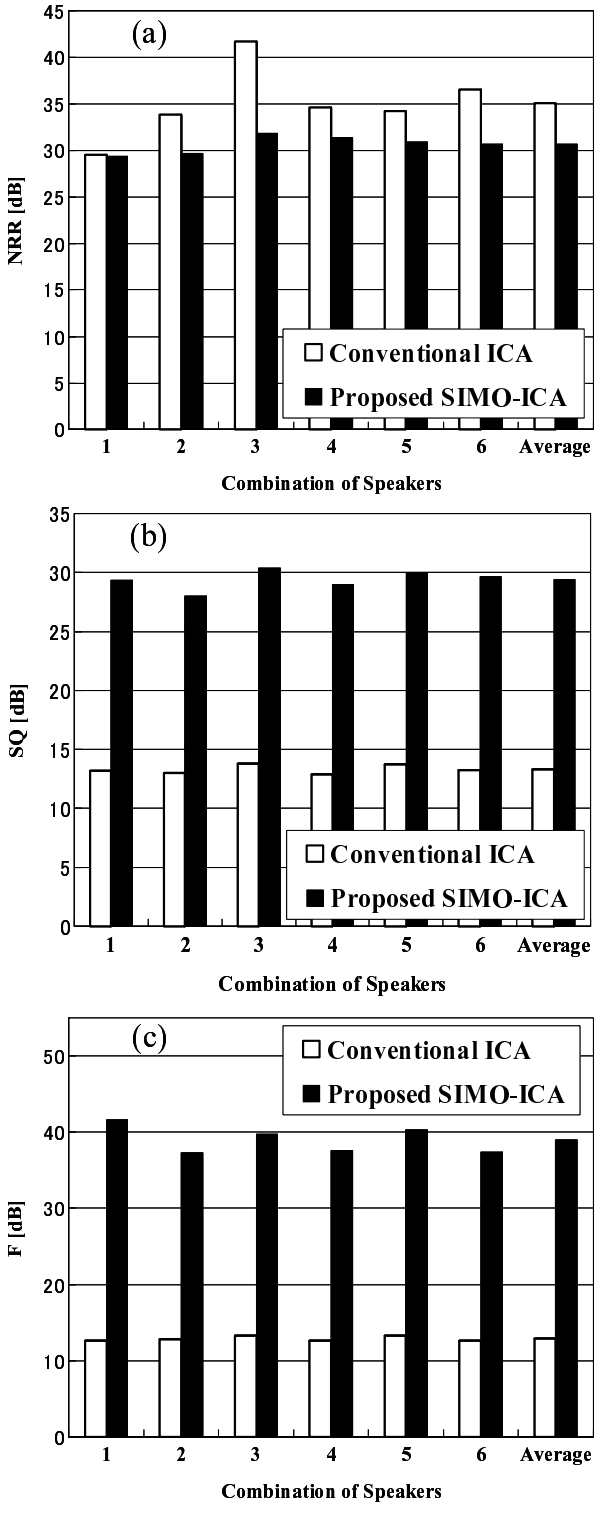

Fig. 5. Results of (a) NRR, (b) SQ, and (c) F in the case that the reverberation time is 0 ms (time lag between microphones only is considered).

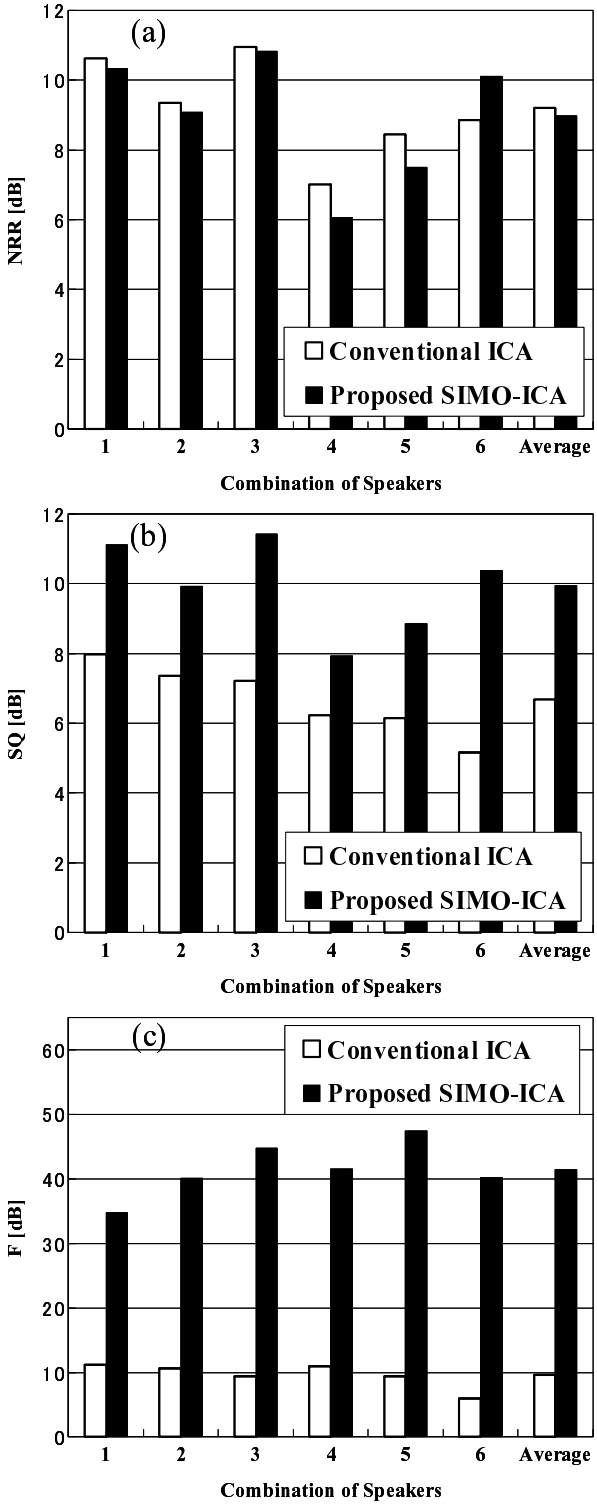

Fig. 6. Results of (a) NRR, (b) SQ, and (c) F in the case that the reverberation time is 150 ms.