PAPER

High-Fidelity Blind Separation of Acoustic Signals Using SIMO-Model-Based Independent Component Analysis

Tomoya TAKATANI[†a)], **Tsuyoki NISHIKAWA**[†], ***Student Members***, **Hiroshi SARUWATARI**[†], ***and*** **Kiyohiro SHIKANO**[†], ***Members***

SUMMARY We newly propose a novel blind separation framework for Single-Input Multiple-Output (SIMO)-model-based acoustic signals using an 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 the signal separation performance of the proposed SIMO-ICA is the same as that of the conventional ICA-based method, and that 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.

key words: *blind source separation, microphone array, independent component analysis, SIMO model*

1. Introduction

Source separation of acoustic signals is to estimate the original sound source signals from among the mixed signals observed in each input channel. This technique is applicable to the realization of noise-robust speech recognition and high-quality hands-free telecommunication systems. As a conventional source separation approach, the method based on array signal processing, e.g., a microphone array system, is one of the most effective techniques [1]. The delay-and-sum (DS) array and the adaptive beamformer (ABF) are popular microphone arrays currently used for source separation. However, these methods have the following drawbacks: The DS array requires a huge number of elements to achieve high performance, especially in the low frequency regions. In ABF, the directions of arrival (DOAs) of the separated source signals must be previously known. Also, the adaptation procedure should be performed during breaks in the target signal to avoid any distortion of separated signals, however, we cannot previously estimate the breaks in conventional use.

In recent years, alternative approaches have been proposed by researchers using information-geometry theory and neural networks [2]–[6]. Blind source separation (BSS) is the approach for estimating original source signals using only the information of the mixed signals observed in each input channel, where the independence among the source signals is mainly used for the separation. This technique is classified into unsupervised adaptive filtering approach [7], and provides us with extended flexibility in that the source-separation procedure requires no training sequences and no a priori information on the DOAs of the sound sources. In recent works on BSS based on independent component analysis (ICA) [5], various methods have been proposed to deal with a means of separation of acoustical sounds which corresponds to the convolutive mixture case [8]–[12]. 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 [13], [14] or high-fidelity sound reproduction systems [15].

In this paper, we propose a new blind separation technique using a Single-Input Multiple-Output (SIMO)-model-based ICA (SIMO-ICA). Here the term "SIMO" represents the specific transmission system in which the input is a single source signal and the outputs are its transmitted signals observed at multiple sensors. 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. 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 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, particularly for the spatial quality and the fidelity of the sound reproduction.

The rest of this paper is organized as follows. In Sect. 2, the sound mixing model and conventional ICA is explained.

Manuscript received March 13, 2003.
Manuscript revised November 2, 2003.
Final manuscript received April 22, 2004.

[†]The authors are with the Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630-0192 Japan.
a) E-mail: tomoya-t@is.naist.jp

In Sect. 3, the proposed SIMO-ICA is described in detail. In Sections 4 and 5, the signal-separation experiments are described and the results are compared with those of the conventional method. Following a discussion on the results of the experiments, we give conclusions in Sect. 6.

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 sources 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

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

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

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

$$\begin{aligned} \boldsymbol{A}(z) &= \begin{bmatrix} A_{11}(z) & \cdots & A_{1L}(z) \\ \vdots & \ddots & \vdots \\ A_{K1}(z) & \cdots & A_{KL}(z) \end{bmatrix} \\ &= \sum_{n=0}^{N-1} \boldsymbol{a}(n)z^{-n} = \left[\sum_{n=0}^{N-1} a_{ij}(n)z^{-n}\right]_{ij}, \end{aligned} \tag{5}$$

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 an FIR filter. In the TDICA, we optimize the separation matrix by using only the fullband observed signals without subband processing (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 Kullback-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 [16]

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

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

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

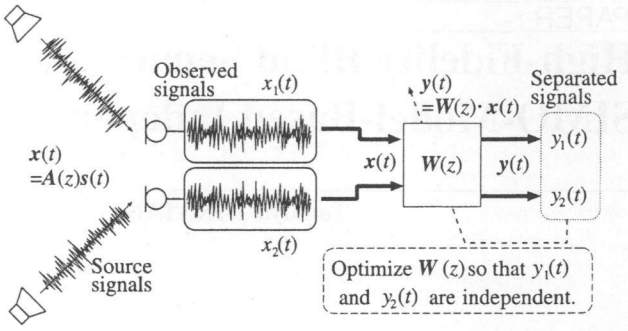


Fig. 1 Configuration of conventional TDICA.

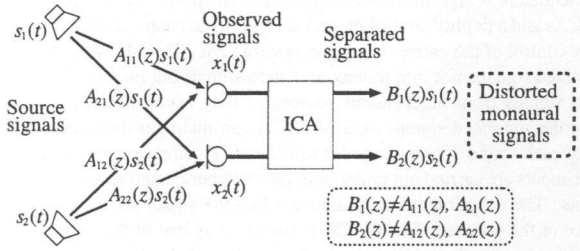


Fig. 2 Input and output relations in conventional ICA. Since $B_l(z)$ is possible to be an arbitrary filter ($B_l(z) \neq A_{kl}(z)$), the separated signals include the spectral distortions.

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. 2). 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. As shown in Fig. 2, $y_l(t) = B_l(z)s_l(t)$, where $B_l(z)$ ($\neq A_{kl}(z)$) is an arbitrary filter, is a possible solution obtained from the conventional ICA using Eq. (7). 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 [17]. However, 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. 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,

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

where $\| \boldsymbol{x} \|$ is the Euclidean norm of vector $\boldsymbol{x}$. The cost function Eq. (10) means a degree of similarity between the sum of all ICA's output $\sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t)$ and the sum of all SIMO components $[\sum_{l=1}^{L} \boldsymbol{A}_{kl}(t - D/2)]_{k1} (= \boldsymbol{x}(t - D/2)$. Here the delay of $D/2$ is used to deal with nonminimum phase systems. Using Eq. (9) and Eq. (10), we can obtain the appropriate separated signals and maintain their spatial qualities as follows.

Theorem: The output signals converge on unique solutions, Eq. (11), up to the permutation, if and only if the independent sound sources are separated by Eq. (9), and simultaneously Eq. (10) is minimized to be zero.

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

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

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

Proof of Theorem: First, the necessity is shown below. Obviously the solution Eq. (11) holds in Eq. (9) because the elements of Eq. (11) are mutually independent in each $\boldsymbol{y}_{\mathrm{ICA}l}(t)$ ($l = 1, \ldots, L$). Also, the following equation holds with Eq. (11).

$$\begin{aligned} \sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t) &= \sum_{l=1}^{L} \mathrm{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2) \\ &= \left[\sum_{l=1}^{L} \boldsymbol{A}_{kl}(z)s_l(t - D/2)\right]_{k1} \\ &= \boldsymbol{A}(z)\boldsymbol{s}(t - D/2). \end{aligned} \tag{13}$$

This results in $\boldsymbol{x}(t - D/2)$, and makes the cost function Eq. (10) be zero. This completes the proof of the necessity in Theorem.

Next, the sufficiency is shown below. Let $\boldsymbol{D}_l(z)$ ($l = 1, \ldots, L$) be arbitrary diagonal polynomial matrices and $\boldsymbol{Q}_l$ be arbitrary permutation matrices. If one of the preconditions ("independent sound sources have been separated by Eq. (9)") holds, the general expression of the l-th ICA's output is given by

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

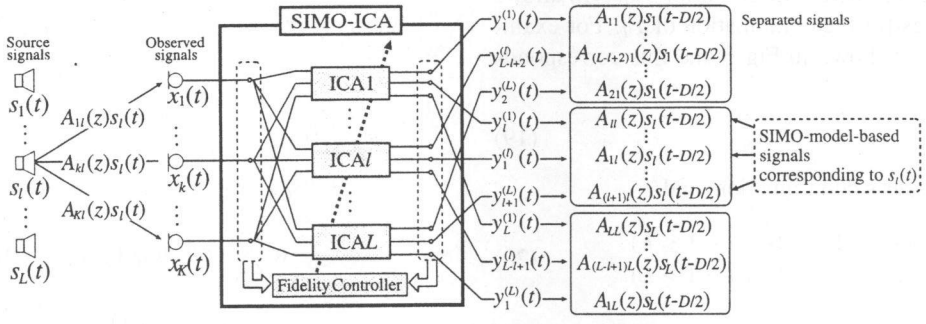


Fig. 3 Example of input and output relations in proposed SIMO-ICA, where exclusively-selected permutation matrices $\boldsymbol{P}_l$ are given by Eq. (19). The SIMO-ICA consists of multiple ICA parts and a fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system. In this system, the separated signals maintain their spatial qualities.

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 $s_l(t - D/2)$ ($l = 1, \ldots, L$). 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

$$\boldsymbol{y}_{\mathrm{ICA}l}(t) = \boldsymbol{D}_l(z)\boldsymbol{P}_l\boldsymbol{s}(t - D/2). \tag{15}$$

In Eq. (15) 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,

$$\boldsymbol{y}_{\mathrm{ICA}l}(t) = \mathrm{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2), \tag{16}$$

and consequently

$$\sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t) = \left[\sum_{l=1}^{L} B_{kl}(z)s_l(t - D/2)\right]_{k1}. \tag{17}$$

Substitution of Eq. (17) in Eq. (10) leads to the following equation.

$$\begin{aligned} &C(\boldsymbol{w}_{\mathrm{ICA}1}(n), \ldots, \boldsymbol{w}_{\mathrm{ICA}L}(n)) \\ &= \left\langle \left\| \left[\sum_{l=1}^{L} B_{kl}(z)s_l(t - D/2)\right]_{k1} \right. \right. \\ &\quad \left. \left. - \left[\sum_{l=1}^{L} A_{kl}(z)s_l(t - D/2)\right]_{k1} \right\|^2 \right\rangle_t \\ &= \sum_{l=1}^{L}\sum_{k=1}^{K} (B_{kl}(z) - A_{kl}(z))^2 \cdot \left\langle s_l(t - D/2)^2 \right\rangle_t, \end{aligned} \tag{18}$$

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. (18) becomes zero if and only if $B_{kl}(z) = A_{kl}(z)$ for all k and l. Thus, Eq. (16) results in Eq. (11). This completes the proof of the sufficiency in 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. However, the condition Eq. (12) allows multiple possibilities for the combination of $\boldsymbol{P}_l$. For example, one possibility is shown in Fig. 3 and this corresponds to

$$\boldsymbol{P}_l = [\delta_{im(k,l)}]_{ki}, \tag{19}$$

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

$$m(k, l) = \begin{cases} k + l - 1 & (k + l - 1 \leq L) \\ k + l - 1 - L & (k + l - 1 > L) \end{cases} \tag{20}$$

In this case, Eq. (11) yields

$$\boldsymbol{y}_{\mathrm{ICA}l}(t) = [A_{km(k,l)}s_{m(k,l)}(t - D/2)]_{k1}, \tag{21}$$

In order to obtain Eq. (11), the natural gradient [6], [18] of Eq. (10) with respect to $\boldsymbol{w}_{\mathrm{ICA}l}(n)$ should be added to the iterative learning rule of the separation filter. The natural gradient of Eq. (10) is given as (see Appendix)

$$\begin{aligned} &\left\{ \frac{\partial}{\partial \boldsymbol{w}_{\mathrm{ICA}l}(n)} \left\langle \left\| \sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}\left(t - \frac{D}{2}\right) \right\|^2 \right\rangle_t \right\} \\ &\quad \cdot \boldsymbol{W}_{\mathrm{ICA}l}(z^{-1})^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ICA}l}(z) \\ &= 2\sum_{d=0}^{D-1} \left\langle \left(\sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}(t) - \boldsymbol{x}\left(t - \frac{D}{2}\right) \right) \right. \\ &\quad \left. \cdot \boldsymbol{y}_{\mathrm{ICA}l}(t - n + d)^{\mathrm{T}} \right\rangle_t \cdot \boldsymbol{w}_{\mathrm{ICA}l}(d). \end{aligned} \tag{22}$$

By combining Eq. (7) with Eq. (22), we can obtain the new iterative algorithm of SIMO-ICA as

$$\begin{aligned} &\boldsymbol{w}_{\mathrm{ICA}1}^{[j+1]}(n) \\ &\quad = \boldsymbol{w}_{\mathrm{ICA}1}^{[j]}(n) \\ &\qquad - \alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}\left(\boldsymbol{y}_{\mathrm{ICA}1}^{[j]}(t)\right) \right. \right. \\ &\qquad \left. \cdot \boldsymbol{y}_{\mathrm{ICA}1}^{[j]}(t - n + d)^{\mathrm{T}} \right\rangle_t \\ &\qquad + \beta \left\langle \left(\sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t) - \boldsymbol{x}\left(t - \frac{D}{2}\right) \right) \right. \\ &\qquad \left. \left. \cdot \boldsymbol{y}_{\mathrm{ICA}1}^{[j]}(t - n + d)^{\mathrm{T}} \right\rangle_t \right\} \cdot \boldsymbol{w}_{\mathrm{ICA}1}^{[j]}(d), \end{aligned} \tag{23}$$

$$\vdots$$

$$\begin{aligned} &\boldsymbol{w}_{\mathrm{ICA}l}^{[j+1]}(n) \\ &\quad = \boldsymbol{w}_{\mathrm{ICA}l}^{[j]}(n) \\ &\qquad - \alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}\left(\boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t)\right) \right. \right. \\ &\qquad \left. \cdot \boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t - n + d)^{\mathrm{T}} \right\rangle_t \\ &\qquad + \beta \left\langle \left(\sum_{l=1}^{L} \boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t) - \boldsymbol{x}\left(t - \frac{D}{2}\right) \right) \right. \\ &\qquad \left. \left. \cdot \boldsymbol{y}_{\mathrm{ICA}l}^{[j]}(t - n + d)^{\mathrm{T}} \right\rangle_t \right\} \cdot \boldsymbol{w}_{\mathrm{ICA}l}^{[j]}(d), \end{aligned} \tag{24}$$

$$\vdots$$

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

$$\cdot \boldsymbol{y}_{\mathrm{ICA}L}^{[j]}(t-n+d)^{\mathrm{T}}\Big\rangle_t\Big\} \cdot \boldsymbol{w}_{\mathrm{ICA}L}^{[j]}(d), \tag{25}$$

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 Esq. (23)–(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 $\boldsymbol{y}_{\mathrm{ICA}l}^{[j]} = \boldsymbol{W}_{\mathrm{ICA}l}^{[j]}(z)\boldsymbol{x}(t)$. Also, the initial values of $\boldsymbol{w}_{\mathrm{ICA}l}(n)$ 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, $\max_n \langle y_k^{(l)}(t) y_{k'}^{(l')}(t-n)\rangle_t$, where $l \neq l'$ and $k \neq k'$. The large value of the correlation indicates that $y_k^{(l)}(t)$ and $y_{k'}^{(l')}(t)$ are SIMO components of the same sources.

4. Experiment and Results for Two-Source Case

4.1 Conditions for Experiment

In this section, we consider a case of $K = L = 2$. 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 [19], 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 (time lag between microphones only is considered) and 150 ms. The impulse response in the case of RT=150 ms is recorded in the experimental room as shown in Fig. 4. These sound data which are artificially convolved with the real impulse responses have the following advantages: (1) we can use the realistic mixture model of two sources neglecting the affection of background noise, (2) since the mixing condition is explicitly measured, we can easily calculate a reliable objective score to evaluate the separation performance as described in the next section.

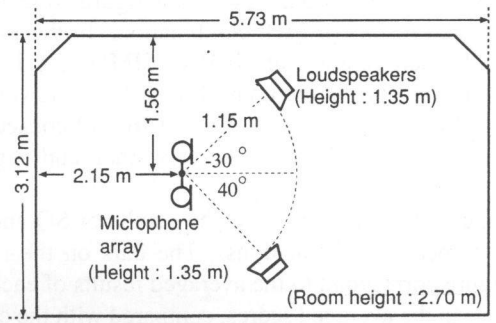


Fig. 4 Layout of reverberant room used in experiments ($K = L = 2$).

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 [11] whose directional null is steered to ±60°. The number of iterations in ICA is 5000. Regarding the conventional ICA given for comparison, we used Esq. (23)–(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

$$\mathrm{NRR} \equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{k=1}^{2}\left(\mathrm{OSNR}_l^{(\mathrm{ICA}k)} - \mathrm{ISNR}_l^{(\mathrm{ICA}k)}\right), \tag{26}$$

$$\mathrm{OSNR}_l^{(\mathrm{ICA}1)} = 10\log_{10}\frac{\sum_t |\, H_{ll}^{\mathrm{ICA}1}(z)s_l(t)\,|^2}{\sum_t |\, H_{ln}^{\mathrm{ICA}1}(z)s_n(t)\,|^2},$$

$$\mathrm{ISNR}_l^{(\mathrm{ICA}1)} = 10\log_{10}\frac{\sum_t |\, A_{ll}(z)s_l(t)\,|^2}{\sum_t |\, A_{ln}(z)s_n(t)\,|^2},$$

$$\mathrm{OSNR}_l^{(\mathrm{ICA}2)} = 10\log_{10}\frac{\sum_t |\, H_{ln}^{\mathrm{ICA}2}(z)s_n(t)\,|^2}{\sum_t |\, H_{ll}^{\mathrm{ICA}2}(z)s_l(t)\,|^2},$$

$$\mathrm{ISNR}_l^{(\mathrm{ICA}2)} = 10\log_{10}\frac{\sum_t |\, A_{ln}(z)s_n(t)\,|^2}{\sum_t |\, A_{ll}(z)s_l(t)\,|^2},$$

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,

$$\mathrm{SQ} \equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{n=1}^{2}\mathrm{SQ}_{y_l^{(n)}}, \tag{27}$$

$$\mathrm{SQ}_{y_1^{(1)}} = 10\log_{10}\frac{\sum_t |\, A_{11}(z)s_1(t)\,|^2}{\sum_t |\, A_{11}(z)s_1(t) - H_{11}^{\mathrm{ICA}1}(z)s_1(t)\,|^2},$$

$$\mathrm{SQ}_{y_2^{(1)}} = 10\log_{10}\frac{\sum_t |\, A_{12}(z)s_2(t)\,|^2}{\sum_t |\, A_{12}(z)s_2(t) - H_{12}^{\mathrm{ICA}2}(z)s_2(t)\,|^2},$$

$$\mathrm{SQ}_{y_1^{(2)}} = 10\log_{10}\frac{\sum_t |\, A_{21}(z)s_1(t)\,|^2}{\sum_t |\, A_{21}(z)s_1(t) - H_{21}^{\mathrm{ICA}2}(z)s_1(t)\,|^2},$$

$$\mathrm{SQ}_{y_2^{(2)}} = 10\log_{10}\frac{\sum_t |\, A_{22}(z)s_2(t)\,|^2}{\sum_t |\, A_{22}(z)s_2(t) - H_{22}^{\mathrm{ICA}1}(z)s_2(t)\,|^2},$$

where $\mathrm{SQ}_{y_l^{(n)}}$ is the sound quality of the separated signal $y_l^{(n)}$.

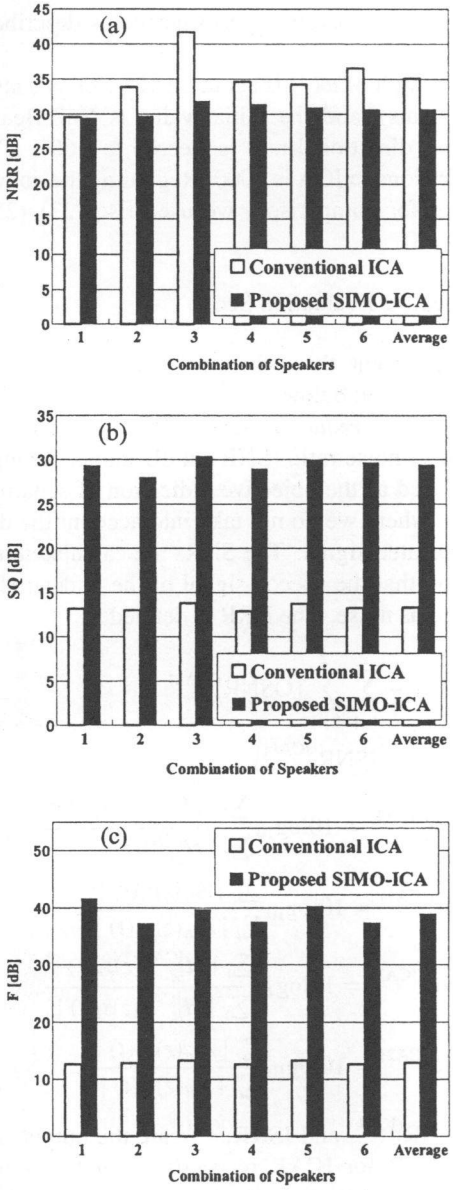


Fig. 5 Results of (a) NRR, (b) SQ, and (c) F in conventional ICA and proposed SIMO-ICA ($K = L = 2$). The reverberation time is 0 ms (time lag between microphones only is considered).

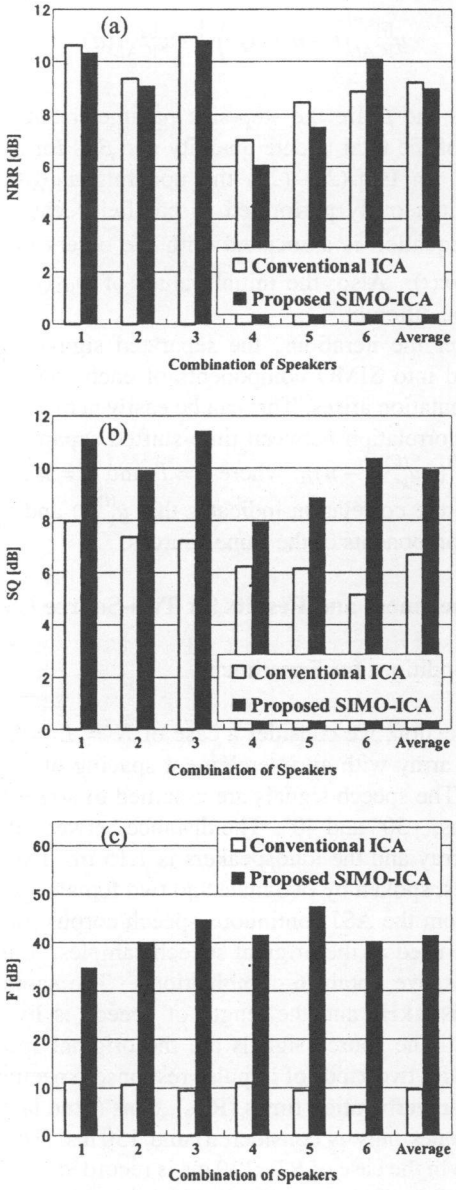


Fig. 6 Results of (a) NRR, (b) SQ, and (c) F in conventional ICA and proposed SIMO-ICA ($K = L = 2$). The reverberation time is 150 ms.

Lastly, *fidelity* (F) indicates the accuracy of the sound reproduction in the entire system. It is defined by

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

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 conventional 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 in 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; this 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. Experiment and Results for Three-Source Case

5.1 Conditions for Experiment

In this section, we consider a case of $K = L = 3$. A three-element array with an interelement spacing of 4 cm is assumed. The speech signals are assumed to arrive from three directions, −30°, 0°, and 40°. The distance between the microphone array and the loudspeakers is 1.15 m. The same speech samples (two males and two females) as described in the previous Sect. 4 are used, and we obtain 4 combinations. In order to generate the room impulse responses, we

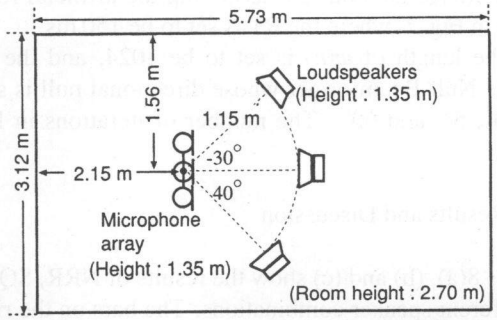


Fig. 7 Layout of artificial reverberant room used in experiments ($K = L = 3$).

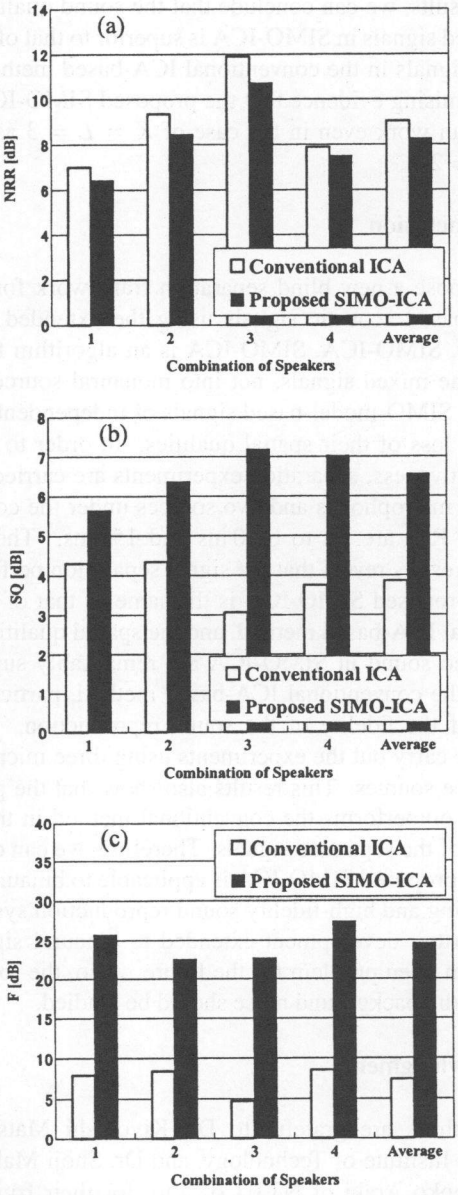


Fig. 8 Results of (a) NRR, (b) SQ, and (c) F in conventional ICA and proposed SIMO-ICA ($K = L = 3$). The reverberation time is 150 ms.

use the image method [20] assuming the artificial room as shown in Fig. 7, where the RT is set to be 150 ms.

The length of $w(n)$ is set to be 1024, and the initial value is Null-Beamformer whose directional null is steered to $-60°$, $5°$, and $60°$. The number of iterations in ICA is 20000.

5.2 Results and Discussion

Figures 8(a), (b) and (c) show the results of NRR, 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 deterioration of NRR is 0.8 dB, but the improvement of SQ is 2.7 dB, and that of F is 17.2 dB. From these results, we can conclude that the sound quality of the separated signals in SIMO-ICA is superior to that of the separated signals in the conventional ICA-based method. This is a promising evidence that the proposed SIMO-ICA algorithm can work even in the case of $K = L = 3$ as well as $K = L = 2$.

6. 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 two microphones and two sources under the conditions that the RTs are set to be 0 ms and 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 remarkably superior to that in the conventional ICA-based method, particularly in terms of the fidelity of the sound reproduction. In addition, we carry out the experiments using three microphones and three sources. This results also show that the proposed method outperforms the conventional method in the sound quality of the separated signals. 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. Also, the robustness against the background noise should be studied.

Acknowledgment

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

References

[1] G.W. Elko, "Microphone array systems for hands-free telecommunication," Speech Commun., vol.20, pp.229–240, 1996.
[2] J.F. Cardoso, "Eigenstructure of the 4th-order cumulant tensor with application to the blind source separation problem," Proc. ICASSP'89, pp.2109–2112, 1989.
[3] C. Jutten and J. Herault, "Blind separation of sources part I: An adaptive algorithm based on neuromimetic architecture," Signal Process., vol.24, pp.1–10, 1991.
[4] A.J. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," Neural Comput., vol.7, pp.1129–1159, 1995.
[5] P. Comon, "Independent component analysis, a new concept?," Signal Process., vol.36, pp.287–314, 1994.
[6] S. Amari, S. Douglas, A. Cichocki, and H.H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," Proc. IEEE international Workshop on Wireless Communication, pp.101–104, April 1997.
[7] S. Haykin, Unsupervised Adaptive Filtering, John Wiley & Sons, New York, NY, 2000.
[8] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," Neurocomputing, vol.22, pp.21–34, 1998.
[9] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," Proc. 1998 International Symposium on Nonlinear Theory and its Application (NOLTA'98), vol.3, pp.923–926, Sept. 1998.
[10] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," IEEE Trans. Speech Audio Process., vol.8, no.3, pp.320–327, May 2000.
[11] 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.
[12] T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of time-domain ICA, frequency-domain ICA and multistage ICA," 2002 European Signal Processing Conference (EUSIPCO2002), vol.II, pp.15–18, Sept. 2002.
[13] J. Blauert, Spatial Hearing (revised ed.), MIT Press, Cambridge, MA, 1997.
[14] J. Bauck and D.H. Cooper, "Generalized transaural stereo and applications," J. Audio Eng. Soc., vol.44, no.9, pp.683–705, 1996.
[15] 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.
[16] 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.
[17] 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.
[18] M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved non-stationary signals," Neurocomputing, vol.22, pp.157–171, Dec. 1998.
[19] T. Kobayashi, S. Itabashi, S. Hayashi, and T. Takezawa, "ASJ continuous speech corpus for research," J. Acoust. Soc. Jpn., vol.48, no.12, pp.888–893, 1992.
[20] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am., vol.65, no.4, pp.943–950, 1979.

Appendix: The Derivation of Eq. (22)

The (standard) gradient of Eq. (10) with respect to $\boldsymbol{w}_{\mathrm{ICA}l}(n)$ is given as

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

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

$$\boldsymbol{x}(t-n)^{\mathrm{T}} = \boldsymbol{y}_{\mathrm{ICA}l}(t-n)^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ICA}l}(z)^{-\mathrm{T}}, \tag{A·2}$$

where the superscript $-\mathrm{T}$ represents the transposed inverse matrix. By using Eq. (A·2), Eq. (A·1) is expanded as

$$\frac{\partial}{\partial \boldsymbol{w}_{\mathrm{ICA}l}(n)}\left\langle\left\|\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right\|^2\right\rangle_t = 2\left\langle\left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right)\cdot\boldsymbol{y}_{\mathrm{ICA}l}(t-n)^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ICA}l}(z)^{-\mathrm{T}}\right\rangle_t. \tag{A·3}$$

Here, we substitute $\boldsymbol{W}_{\mathrm{ICA}l}(z)^{-1}$ with $\boldsymbol{V}_{\mathrm{ICA}l}(z) = \sum_{d=0}^{D-1}\boldsymbol{v}(d)\,z^{-d}$, then Eq. (A·3) is rewritten as

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{w}_{\mathrm{ICA}l}(n)}\left\langle\left\|\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right\|^2\right\rangle_t \\
&= 2\left\langle\left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right)\cdot\boldsymbol{y}_{\mathrm{ICA}l}(t-n)^{\mathrm{T}}\boldsymbol{V}_{\mathrm{ICA}l}(z)^{\mathrm{T}}\right\rangle_t \\
&= 2\sum_{d=0}^{D-1}\left\langle\left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right)\cdot\boldsymbol{y}_{\mathrm{ICA}l}(t-n-d)^{\mathrm{T}}\right\rangle_t\boldsymbol{v}_{\mathrm{ICA}l}(d)^{\mathrm{T}} \\
&= 2\sum_{d=0}^{D-1}\boldsymbol{J}(n+d)\boldsymbol{v}_{\mathrm{ICA}l}(d)^{\mathrm{T}},
\end{aligned} \tag{A·4}$$

where $\boldsymbol{J}(n+d)$ represents the matrix in which each element is the time sequence of not the index t but the index n because the index t vanishes under the averaging $\langle\cdot\rangle_t$; this is defined as

$$\boldsymbol{J}(u) = \left\langle\left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right)\cdot\boldsymbol{y}_{\mathrm{ICA}l}(t-u)^{\mathrm{T}}\right\rangle_t. \tag{A·5}$$

Equation (A·4) is rewritten as

$$\begin{aligned}
&2\sum_{d=0}^{D-1}\boldsymbol{J}(n+d)\boldsymbol{v}_{\mathrm{ICA}l}(d)^{\mathrm{T}} \\
&= 2\sum_{d=0}^{D-1}\boldsymbol{J}(n)\left(\boldsymbol{v}_{\mathrm{ICA}l}(d)z^{d}\right)^{\mathrm{T}} \\
&= 2\boldsymbol{J}(n)\boldsymbol{V}_{\mathrm{ICA}l}(z^{-1})^{\mathrm{T}}
\end{aligned} \tag{A·6}$$

Therefore, the standard gradient of Eq. (10) with respect to $\boldsymbol{w}_{\mathrm{ICA}l}(n)$ is given as

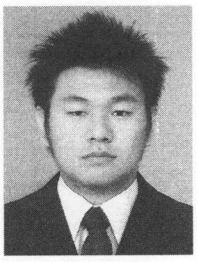
$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{w}_{\mathrm{ICA}l}(n)}\left\langle\left\|\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right\|^2\right\rangle_t \\
&= 2\left\langle\left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right)\cdot\boldsymbol{y}_{\mathrm{ICA}l}(t-n)^{\mathrm{T}}\right\rangle_t\boldsymbol{W}_{\mathrm{ICA}l}(z^{-1})^{-\mathrm{T}}.
\end{aligned} \tag{A·7}$$

From Eq. (A·7), the natural gradient [6], [18] of Eq. (10) is given as

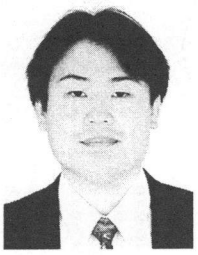
$$\begin{aligned}
&\left\{\frac{\partial}{\partial \boldsymbol{w}_{\mathrm{ICA}l}(n)}\left\langle\left\|\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right\|^2\right\rangle_t\right\}\cdot\boldsymbol{W}_{\mathrm{ICA}l}(z^{-1})^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ICA}l}(z) \\
&= 2\left\langle\left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right)\cdot\boldsymbol{y}_{\mathrm{ICA}l}(t-n)^{\mathrm{T}}\right\rangle_t\cdot\boldsymbol{W}_{\mathrm{ICA}l}(z) \\
&= 2\sum_{d=0}^{D-1}\left\langle\left(\sum_{l=1}^{L}\boldsymbol{y}_{\mathrm{ICA}l}(t)-\boldsymbol{x}\left(t-\frac{D}{2}\right)\right)\cdot\boldsymbol{y}_{\mathrm{ICA}l}(t-n+d)^{\mathrm{T}}\right\rangle_t\cdot\boldsymbol{w}_{\mathrm{ICA}l}(d).
\end{aligned} \tag{A·8}$$

Therefore, we have Eq. (22).

Tomoya Takatani was born in Hyogo, Japan on 1977. He received the B.E. degrees in electronics from Doshisha University in 2001 and received the M.E. degrees in information and science from Nara Institute of Science and Technology (NAIST) in 2003. He is now a Ph.D. student at Graduate School of Information Science, NAIST. His research interests include array signal processing and blind source separation. He is a member of the the Acoustical Society of Japan.



Tsuyoki Nishikawa was born in Mie, Japan on 1978. He received the B.E. degree in electronic system and information engineering from Kinki University in 2000 and received the M.E. degree in information and science from Nara Institute of Science and Technology (NAIST) in 2002. He is now a Ph.D. student at Graduate School of Information Science, NAIST. His research interests include acoustics signal processing, sensor array processing, and blind source separation. He is a member of the the Acoustical Society of Japan. He received TELECOM System Technology Award and TELECOM System Technology Student Award from the Telecommunications Advancement Foundation in 2004.



Hiroshi Saruwatari was born in Nagoya, Japan, on July 27, 1967. He received the B.E., M.E. and Ph.D. degrees in electrical engineering from Nagoya University, Nagoya, Japan, in 1991, 1993 and 2000, respectively. He joined Intelligent Systems Laboratory, SECOM CO., LTD., Mitaka, Tokyo, Japan, in 1993, where he engaged in the research and development on the ultrasonic array system for the acoustic imaging. He is currently an associate professor of Graduate School of Information Science, Nara Institute of Science and Technology. His research interests include array signal processing, blind source separation, and sound field reproduction. He received the Paper Award from IEICE in 2001 and TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2004. He is a member of the IEEE and the Acoustical Society of Japan.

Kiyohiro Shikano received the B.S., M.S., and Ph.D. degrees in electrical engineering from Nagoya University in 1970, 1972, and 1980, respectively. He is currently a professor of Nara Institute of Science and Technology (NAIST), where he is directing speech and acoustics laboratory. His major research areas are speech recognition, multi-modal dialog system, speech enhancement, adaptive microphone array, and acoustic field reproduction. From 1972, he had been working at NTT Laboratories, where he had been engaged in speech recognition research. During 1990–1993, he was the executive research scientist at NTT Human Interface Laboratories, where he supervised the research of speech recognition and speech coding. During 1986–1990, he was the Head of Speech Processing Department at ATR Interpreting Telephony Research Laboratories, where he was directing speech recognition and speech synthesis research. During 1984–1986, he was a visiting scientist in Carnegie Mellon University, where he was working on distance measures, speaker adaptation, and statistical language modeling. He received the Yonezawa Prize from IEICE in 1975, the Signal Processing Society 1990 Senior Award from IEEE in 1991, the Technical Development Award from ASJ in 1994, IPSJ Yamashita SIG Research Award in 2000, and Paper Award from the Virtual Reality Society of Japan in 2001. He is a member of the Information Processing Society of Japan, the Acoustical Society of Japan (ASJ), Japan VR Society, the Institute of Electrical and Electronics, Engineers (IEEE), and International Speech Communication Society.