NAIST-IS-DD0361208

Doctoral Dissertation

High-Fidelity Blind Source Separation Using Single-Input-Multiple-Output-Model-Based Independent Component Analysis

Tomoya Takatani

March 24, 2006

Department of Information Processing
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Tomoya Takatani

Thesis Committee:

Professor Kiyohiro Shikano	(Supervisor)
Professor Kenji Sugimoto	(Co-supervisor)
Associate Professor Hiroshi Saruwatari	(Co-supervisor)

High-Fidelity Blind Source Separation Using Single-Input-Multiple-Output-Model-Based Independent Component Analysis*

Tomoya Takatani

Abstract

Blind source separation (BSS) technique using independent component analysis (ICA) for acoustic signals has been developed over the last decade. This technique assumes that the source signals are mutually independent, and can estimate the source signals from the mixed signals without a priori information. Thus, this technique is highly applicable in high-quality hands-free telecommunication system. The conventional ICA-based BSS method is a means of extracting the independent sound source signals as the monaural signals from the mixed signals observed in each input channel, and the separated signals include arbitrary spectral distortions. Consequently, they have a serious drawback in that the separated sounds cannot maintain information about the directivity, localization, reverberation, or spatial qualities of each sound source. These problems prevent any BSS methods from being applied to binaural signal processing or high-fidelity sound reproduction system.

In this thesis, firstly, in order to solve the above-mentioned fundamental problems, we propose a new ICA algorithm, which is called Single-Input Multiple-Output (SIMO)-model-based ICA (SIMO-ICA) with least squares criterion (SIMO-ICA-LS). Here the term "SIMO" denotes specific transmission system in which the input is a single source signal and the outputs are its transmitted signals observed at multiple sensors. SIMO-ICA-LS can separate the mixed signals, not into

* Doctoral Dissertation, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0361208, March 24, 2006.

monaural source signals but into the SIMO-model-based signals from independent sources as they are at the microphones. The experimental results revealed that the performance of the proposed SIMO-ICA-LS is superior to that in the conventional ICA-based method.

The SIMO-ICA consists of multiple ICA parts and a fidelity controller. In SIMO-ICA-LS, each of the ICA parts are driven by information-geometry theory, whereas the fidelity controller is based on the least squares criterion which is a different class of criterion from the information-geometric metric. Due to the mismatch between two kinds of criteria, there is an inherent disadvantage in that an additional balancing parameter is needed, but the parameter is very sensitive to the convergence in the iterative learning. To solve the above-mentioned fundamental problem, we newly propose an SIMO-ICA algorithm with an information-geometric learning algorithm (SIMO-ICA-IG). In this method, all of the procedures for optimization of the separation filters are conducted by the information-geometric learning algorithm. The experimental results reveal that the performance of the proposed SIMO-ICA-IG is almost the same as that of the previously proposed SIMO-ICA-LS. In addition, it is confirmed that the internal parameter setting in the proposed SIMO-ICA-IG does not depend on the source signals' properties, unlike that of SIMO-ICA-LS.

Finally, in order to apply the SIMO-ICA algorithm to the blind decomposition problem of mixed binaural signals, we propose a self-generator for initial filters (SG) of SIMO-ICA. Although the above-mentioned attractive feature of SIMO-ICA is beneficial to the binaural sound separation, the original SIMO-ICA has a serious drawback in its high sensitivity to the initial settings of the separation filter. In the proposed SIMO-ICA with SG, SG functions as the preprocessor of SIMO-ICA, and thus it can provide a valid initial filter for SIMO-ICA. The experimental results reveal that the separation performance of the proposed method is superior to those of the conventional methods.

Keywords:

blind source separation, independent component analysis, SIMO-model, high-fidelity, binaural signal

Single-Input-Multiple-Output モデルに基づく独立成分分析を用いた高品質ブラインド音源分離*

高谷 智哉

内容梗概

独立成分分析 (ICA) を用いた音響信号のブラインド音源分離 (BSS) 技術が近年進展している．この技術は音源信号が互いに独立であると仮定し, 事前情報を用いることなく，観測信号から音源信号を推定する技術である．従って，この技術は高品質ハンズフリー音声通信システムに応用が可能である．従来の ICA に基づく BSS 手法の多くは，音源信号を各チャネルで観測された混合信号からモノラル信号として抽出する手法であり，任意のスペクトル歪みを持っている．従って，従来法には各音源の方位・位置情報，残響感などの空間的な情報を維持することができないという重要な欠点をもっている．この問題点により，バイノーラル信号処理や高忠実音場再現システムへの応用が非常に困難となっている．

本論文では，はじめに，上記問題点を改善するために，最小二乗型Single-Input Multiple-Output (SIMO) モデルに基づく ICA (SIMO-ICA) (SIMO-ICA-LS) と呼ばれる新しい ICA アルゴリズムを提案する．ここで，”SIMO” とは入力信号が単一で且つ出力信号が各マイクロホンで観測される，入力信号が伝送された信号であるような伝達システムのことを示している．SIMO-ICA-LS は観測された混合信号をモノラル信号に分離するのではなく，各マイクロホンで観測された各音源からの SIMO モデルに基づく信号に分解する．従来の ICA を用いた BSS 手法と提案手法を比較をするために，残響環境下において分離実験を行った．実験結果より，SIMO-ICA-LS の性能は従来法より優れていることが示された．

SIMO-ICA は複数の ICA 部と fidelity controller から構成される．SIMO-ICA-LS では，各 ICA 部が情報幾何規範に基づいて動作するが，fidelity controller は情報幾何規範とは異なる最小二乗規範に基づいて動作する．両規範のミスマッチに

* 奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士論文, NAIST-IS-DD0361208, 2006 年 3 月 24 日.

より，それらのバランスを決定するパラメータが必要となるが，そのパラメータの設定は学習更新式における収束に影響する．この問題点を改善するために，新しく情報幾何学習型SIMO-ICA アルゴリズム (SIMO-ICA-IG) を提案する．SIMO-ICA-IG では，フィルタの最適化のための処理がすべて情報幾何学習アルゴリズムによって行われる．有効性の検証のため，残響環境下において分離実験を行った．実験結果より，SIMO-ICA-IG の性能はSIMO-ICA-LS より優れている事が示された．更に，SIMO-ICA-IG におけるパラメータはSIMO-ICA-LS とは事なり，音源の性質に依存しないことが示された．

最後に，SIMO-ICA アルゴリズムを混合バイノーラル信号のブラインド分解問題に適用するために，SIMO-ICA の初期フィルタ自己生成器 (SG) を提案する．SIMO-ICA の魅力的な特長はバイノーラル音源分離問題に有益であるが，SIMO-ICA アルゴリズム単体の収束は分離フィルタの初期設定に大きく依存するという欠点も持っている．SG を用いたSIMO-ICA-IG (SIMO-ICA-SG) では，SG はSIMO-ICA の前処理として機能し，SIMO-ICA-IG に最適な初期フィルタを提供することができる．有効性の検証のため，残響環境下で混合バイノーラル信号の分離実験を行った．実験結果より，提案法の性能は従来法より優れていることが示された．

キーワード

ブラインド音源分離, 独立成分分析, SIMO モデル, 高品質, バイノーラル信号

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

Source separation of acoustic signals is to estimate the original sound source signals from among the mixed signals observed at 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) [2, 3] array and the adaptive beamformer (ABF) [4, 5, 6] 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 [7, 8, 9, 10, 11]. 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 [12], 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) [9], various methods have been proposed to deal with a means of separation of acoustical sounds which corresponds to the convolutive mixture case [13, 14, 15, 16, 17, 18]. 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 [24, 25] or high-fidelity sound reproduction systems [26, 27].

1.2 Research purpose

This dissertation deals with the following three subjects.

SIMO-Model-Based ICA(SIMO-ICA) with Least Squares Criterion (SIMO-ICA-LS) (Chapter 4)

The aim of this topic is to obtain the signals which can maintain the spatial qualities. To achieve this aim, in this topic, we propose a novel ICA algorithm, which is called Single-Input Multiple-Output (SIMO)-model-based ICA (SIMO-ICA) with least squares criterion (SIMO-ICA-LS). Here the term "SIMO" denotes 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-LS minimize the sum of the cost functions about separation and fidelity control. The output signals of SIMO-ICA converge on unique solutions, if and only if the total cost function is minimized to be zero. This solution of SIMO-ICA-LS is mathematically proved.

SIMO-ICA with Information Geometric Criterion (SIMO-ICA-IG) (Chapter 5)

In this topic, our purpose is to improve the mismatch between two kinds of criteria of cost functions in SIMO-ICA-LS. The SIMO-ICA consists of multiple ICA parts and a fidelity controller. In SIMO-ICA-LS, each of the ICA parts are driven by information-geometry theory, whereas the fidelity controller is based on the least squares criterion which is a different class of criterion from the information-geometric metric. To solve the above-mentioned fundamental problem, we newly propose an SIMO-ICA algorithm with an information-geometric learning algorithm (SIMO-ICA-IG). In SIMO-ICA-IG, the fidelity controller as well as each of the ICA parts is designed on the basis of information-geometry theory. Namely, all of the procedures for optimization of the separation filters are conducted by the information-geometric learning algorithm.

Blind Decomposition of Mixed Binaural Sound Using SIMO-ICA with Self-Generator for Initial Filter (Chapter 6)

The purpose of this topics is to apply the SIMO-ICA algorithm to the blind decomposition problem of mixed binaural signals. Although the attractive feature of SIMO-ICA, whose output signals can maintain the spatial qualities, is beneficial to the binaural sound separation, the original SIMO-ICA has a serious drawback in its high sensitivity to the initial settings of the separation filter. In the proposed SIMO-ICA with SG, SG functions as the preprocessor of SIMO-ICA, and thus it can provide a valid initial filter for SIMO-ICA. Also, since the SG is still a blind process total proposed system is also blind processing.

1.3 Thesis overview

The thesis is organized as follows.

First, the transfer function of the room and sound mixing model is defined in Section 2

In Section 3, we introduce the conventional ICA algorithm to design the separation filter matrix for the sound mixtures, and represent the disadvantage of the conventional methods.

In Section 4, we propose a novel ICA algorithm, single-input multiple-output (SIMO)-model-base ICA with least square criterion (SIMO-ICA-LS). Also, we revealed the effectiveness of SIMO-ICA-LS through the separation experiment under the reverberant condition.

In Section 5, we propose a novel SIMO-ICA algorithm, SIMO-ICA with information geometric criterion (SIMO-ICA-IG), in which the balancing parameter can be almost negligible. Moreover, we show the experimental results to compare SIMO-ICA-IG with the conventional ICA and SIMO-ICA-LS.

In Section 6, we propose self-generator for initial filter for SIMO-ICA, which consists of frequency-domain ICA, single talk detection, direction of arrival estimation, and head related transfer function matrix bank. Moreover, we clarify the effectiveness of SIMO-ICA-SG through the experimental results.

In Section 7, we discuss the application of SIMO-ICA, and we evaluate the novel acoustic augmented reality system using SIMO-ICA. Moreover, we clarify

the effectiveness of SIMO-ICA-SG through the subjective and objective evaluation.

2. Mixing and demixing process

2.1 Introduction

In this chapter, we explain the sound mixing model to define the sound separation problem. In Section 2.2, we describe the acoustic transfer system, and define the sound mixing model. In Section 2.3, the novel demixing setting is defined to realize the main approach.

2.2 Sound mixing process

Single-Input Single-Output (SISO) system

Firstly, we consider the simplest situation which have one sound source and one microphone without interference (see Figure 1). This situation is called as Single-Input Single-Output (SISO) system. In this system, the sound detected at the microphone includes not only the direct sound from the source but also the multiple early reflections and reverberation. The composition among these sounds depends on the various acoustic settings, for example the recording room size, reflection coefficient of the wall, floor, and ceiling, and the distance between the source and the microphone, and is expressed as the impulse response between the source and the microphone. The observed signal is defined as

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

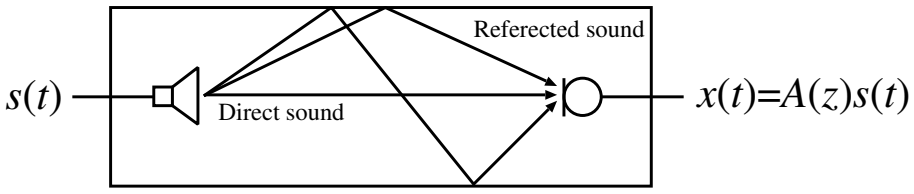

Figure 1. Single-Input Single-Output (SISO) system. This system has one sound source and one microphone. The observed signal $x(t)$ in this system include the direct sound, the early reflections, and reverberation.

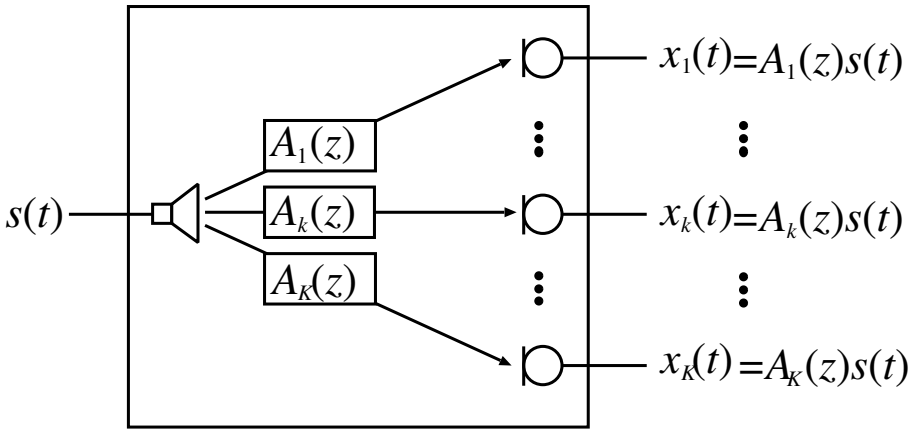

Figure 2. Single-Input Multiple-Output (SIMO) system. In this figure, $A_k(z)$ means the z-transform of impulse response $a_k(n)$ between the source and k-th microphone.

$$= \quad A(z)s(t), \tag{2}$$

where $x(t)$ is the observed signal, $a(n)$ is the impulse response between the sound source and the microphone with the length of N, $s(t)$ is the source signal, $A(z)$ is the z-transform of $a(n)$, and z^{-1} is used as the unit-delay operator, i.e., $z^{-n} \cdot x(t) = x(t-n)$.

Single-Input Multiple-Output (SIMO) system

We can easily extend the SISO system to multidimensional system which involve the multiple microphones. This system is called Single-Input Multiple-Output (SIMO) system. Figure 2 shows the configuration of SIMO system. In SIMO system, the observed signals are expressed as

$$\begin{aligned} \boldsymbol{x}(t) &= [x_1(t), \cdots, x_k(t), \cdots, x_K(t)]^{\mathrm{T}} && (3) \\ &= [x_k(t)]_k && (4) \\ &= [\sum_{n=1}^{N-1} a_1(n)s(t-n), \cdots, \sum_{n=1}^{N-1} a_k(n)s(t-n), \cdots, \sum_{n=1}^{N-1} a_K(n)s(t-n)]^{\mathrm{T}} && (5) \end{aligned}$$

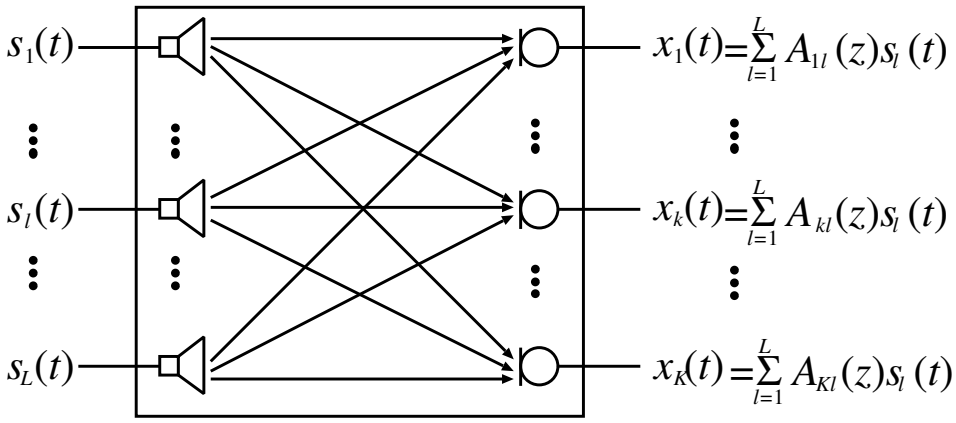

Figure 3. Multiple-Input Multiple-Output (MIMO) system.

$$= [A_1(z)s(t), \cdots, A_k(z)s(t), \cdots, A_K(z)s(t)]^{\mathrm{T}} \quad (6)$$

where $\boldsymbol{x}(t)$ is the observed signal vector, $[X]_k$ denotes the vector which includes the element X in the k-th row, $a_k(n)$ is the impulse response between the k-th microphone and the sound source with the length of N, and $A_k(z)$ is the z-transform of $a_k(n)$. Since all microphones are set on the different locations in the room, the impulse response $a_k(n)$ recorded at k-th microphone is different from the other impulse response $a_{k'}(n)$ $(k' \neq k)$.

Hereafter, the signals observed at the microphones in SIMO system are called the SIMO-model-based signals.

Multiple-Input Multiple-Output (MIMO) system

In more general case, we consider that the number of array elements (microphones) is K and the number of multiple sound sources is L. This system is called Multiple-Input Multiple-Output (MIMO) system. Figure 3 describes the relation between input and output signals in MIMO system. In this system, the observed signals in which sources are linearly mixed are expressed as

$$\boldsymbol{x}(t) = [x_1(t), \cdots, x_k(t), \cdots, x_K(t)]^{\mathrm{T}} \quad (7)$$

$$= [x_k(t)]_k \quad (8)$$

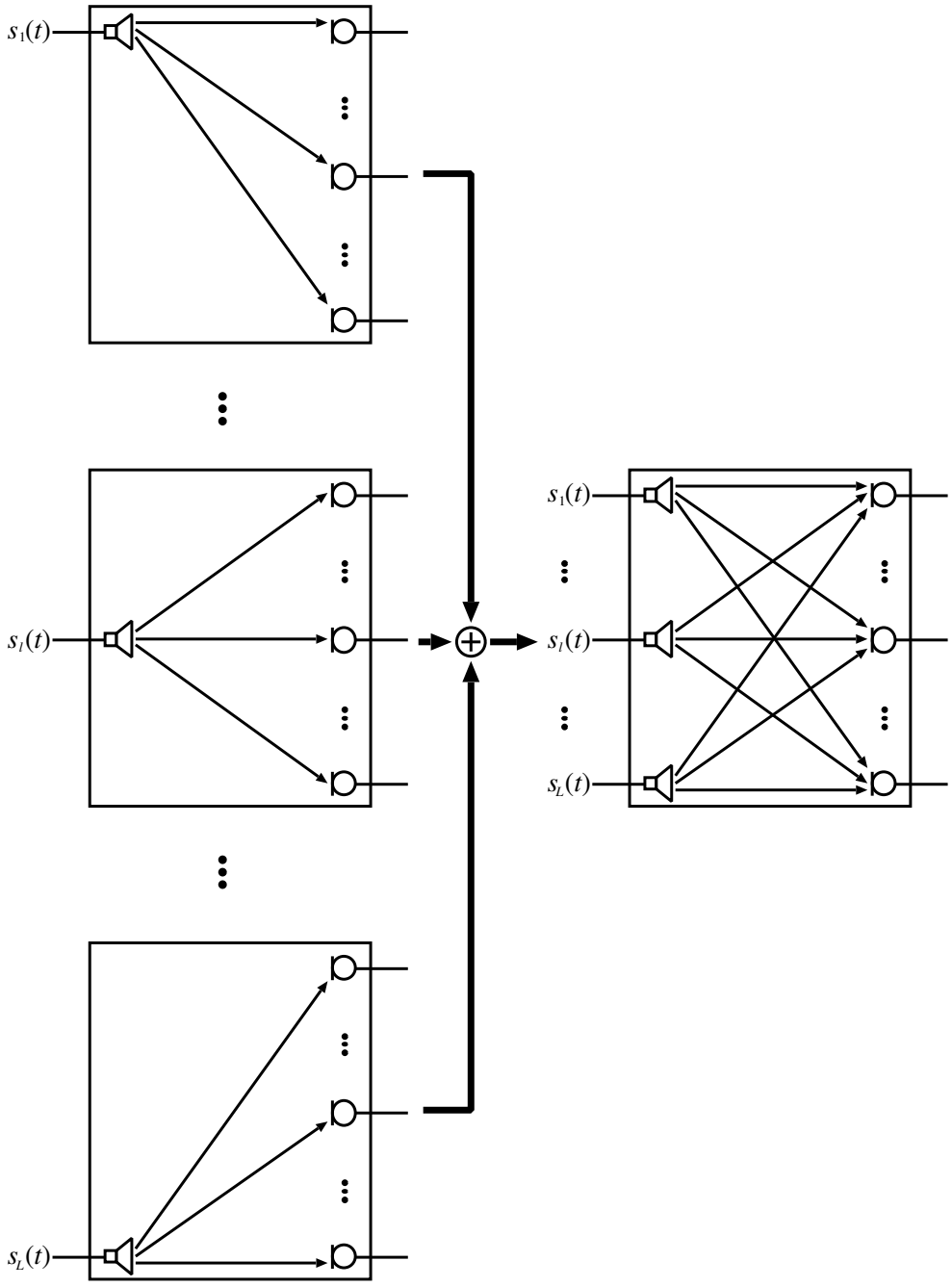

Figure 4. MIMO system is superposition of the SIMO systems. Thus, the mixed signals are superposition of SIMO-model-based signals $A_{kl}(z)s_l(t)$ with respect to each source $s_l(t)$

$$= \sum_{n=0}^{N-1} \boldsymbol{a}(n)\boldsymbol{s}(t-n) \tag{9}$$

$$= \boldsymbol{A}(z)\boldsymbol{s}(t) \tag{10}$$

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

$$\boldsymbol{s}(t) = [s_1(t), \cdots, s_L(t)]^{\mathrm{T}} \tag{11}$$

$$\boldsymbol{a}(n) = \begin{bmatrix} a_{11}(n) & \cdots & a_{1L}(n) \\ \vdots & \ddots & \vdots \\ a_{K1}(n) & \cdots & a_{KL}(n) \end{bmatrix} \tag{12}$$

$$\boldsymbol{A}(z) = \begin{bmatrix} A_{11}(z) & \cdots & A_{1L}(z) \\ \vdots & \ddots & \vdots \\ A_{K1}(z) & \cdots & A_{KL}(z) \end{bmatrix} \tag{13}$$

$$= \begin{bmatrix} a_{11}(n)z^{-n} & \cdots & a_{1L}(n)z^{-n} \\ \vdots & \ddots & \vdots \\ a_{K1}(n)z^{-n} & \cdots & a_{KL}(n)z^{-n} \end{bmatrix} \tag{14}$$

$$= \left[\sum_{n=0}^{N-1} a_{kl}(n)z^{-n}\right]_{kl} \tag{15}$$

$$= \sum_{n=0}^{N-1} \boldsymbol{a}(n)z^{-n} \tag{16}$$

where, a_{kl} is the impulse response between the k-th microphone and the l-th sound source, and $[X]_{kl}$ denotes the matrix which includes the element X in the k-th row and the l-th column.

Also, Eq. (10) is rewritten as

$$\boldsymbol{x}(t) = \left[\sum_{l=1}^{L} A_{1l}(z)s_l(t), \cdots, \sum_{l=1}^{L} A_{kl}(z)s_l(t), \cdots, \sum_{l=1}^{L} a_{Kl}(z)s_l(t)\right]^{\mathrm{T}} \tag{17}$$

This means that the observed signals in MIMO system are superposition of the SIMO-model-based signals for all l sound sources (see Figure 4).

In acoustic MIMO system, the element $a_{kl}(n)$ of the mixing matrix $\boldsymbol{a}(n)$ corresponds to the impulse responses between the l-th source and the $k-th$ microphone. If all impulse responses are known, the source signals can be easily

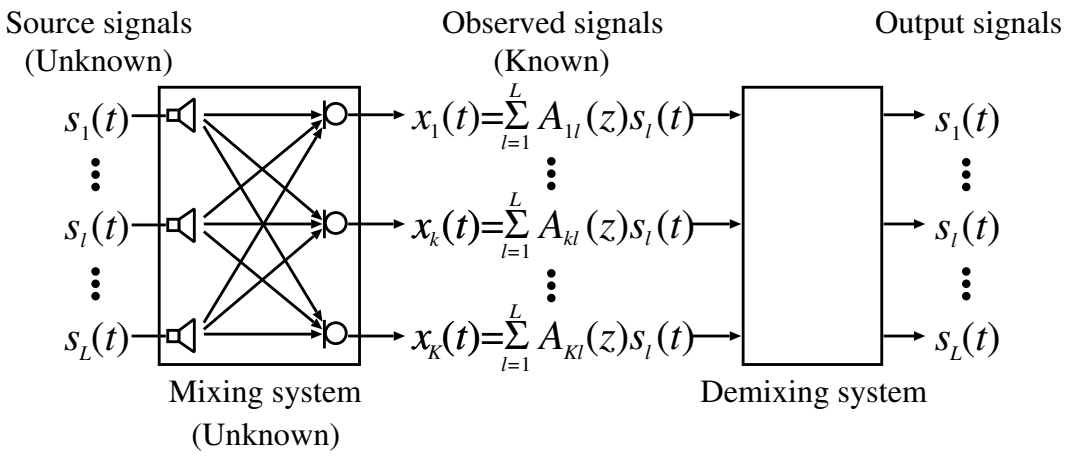

Figure 5. Configuration of the conventional demixing system. This demixing system estimates the source signals. Thus, these output signals are monaural signals with respect to each sound source, and do not maintain information about directivity, localization, and spatial qualities of each sound source.

estimated using the inverse processing [26]. However, the impulse responses depend on the room conditions and is difficult to measure them under the real environments. Thus, in this study, we assume the sound mixing matrix $\boldsymbol{a}(n)$ and sound source $\boldsymbol{s}(t)$ are unknown.

2.3 Main approach and sound demixing process

2.3.1 Conventional demixing system

To extract the source signals from the mixed signals, the demixing system is designed as the inversion system of the mixing system (see Figure. 5). This system reduce not only the interference but also the influence of the transfer function. Thus, this system is applicable to the realization of noise-robust speech recognition and high-quality hands-free telecommunication systems. However, the latter reduction cause the lack of information about localization, directivity, and spatial qualities of each sound source, and this demixing system is not applicapable to the binaural signal processing and high-fidelity sound reproduction systems.

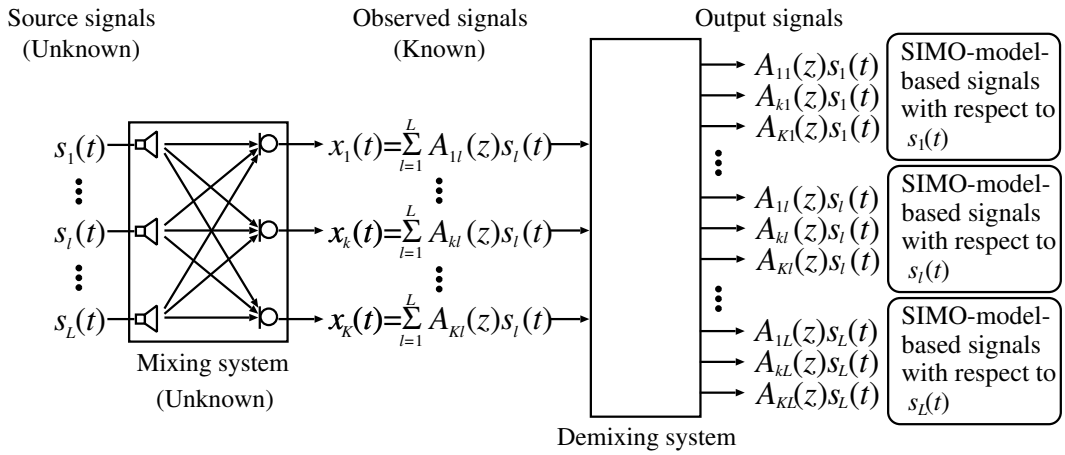

Figure 6. Configuration of the proposed demixing (decomposing) system. This demixing system estimates the SIMO-model-based components of the observed signals. Thus the output signals maintain information about directivity, localization, or spatial qualities of each sound source.

2.3.2 Main approach

In this study, our main research purpose is to estimate the multidimensional source signals which maintain information about directivity, localization, and spatial qualities. To realize this purpose, we consider the novel demixing system which can execute only the former reduction (reduction of interference sound). From Equation (17), the mixed signals are superposition of the multiple SIMO-model-based signals with respect to each sound source. Thus, we should construct the decomposition system which can decompose the observed signals into SIMO-model-based signals from each source at microphone points (see Figure 6), not into *monaural* signals (see Figure 5).

2.4 Conclusion

In this chapter, we described the sound transfer function, the sound mixing model, and the conventional demixing system. However, the conventional demixing method can not apply the binaural signal processing and high-fidelity sound

reproduction systems. To solve this problem, we introduced the novel demixing system which can maintain information which is required by the binaural signals processing.

3. Independent component analysis (ICA)

3.1 Introduction

In this chapter, we explain the blind source separation (BSS) using the conventional independent component analysis (ICA) algorithm. In ICA algorithm, we assume that the source signals are mutually independent, and optimize the demixing system so that the output signals are mutually independent using the alternative learning rule based on the information-geometry theory and neural networks. In Section 3.2, we explain the conventional ICA methods and present development of a new SIMO-output-type method in which all SIMO components can be simultaneously estimated in the ICA updating with stable FIR filters, is a problem demanding prompt attention.

3.2 Conventional ICA algorithm

3.2.1 Holonomic ICA algorithm by Amari [11]

In the BSS method, we consider the time-domain ICA (TDICA), in which each element of the separation filter matrix is represented as an FIR filter. In the TDICA, we optimize the separation filter matrix by using only the fullband observed signals without subband processing (see Figure 7). The separated signal vector $\boldsymbol{y}(t) = [y_1(t), \cdots, y_K(t)]^{\mathrm{T}}$ is expressed as

$$\begin{aligned} \boldsymbol{y}(t) &= [y_k(t)]_{k1} && (18)\\ &= \sum_{n=0}^{D-1} \boldsymbol{w}(n)\boldsymbol{x}(t-n) && (19)\\ &= \boldsymbol{W}(z)\boldsymbol{x}(t) && (20)\\ &= \boldsymbol{W}(z)\boldsymbol{A}(z)\boldsymbol{s}(t), && (21) \end{aligned}$$

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 the ICA-based BSS framework assuming independently identically distributed (i.i.d.) sources, Amari [11] proposed the holonomic TDICA algorithm which optimizes the separation filter by minimizing the Kullback-Leibler divergence (KLD) between the joint probability

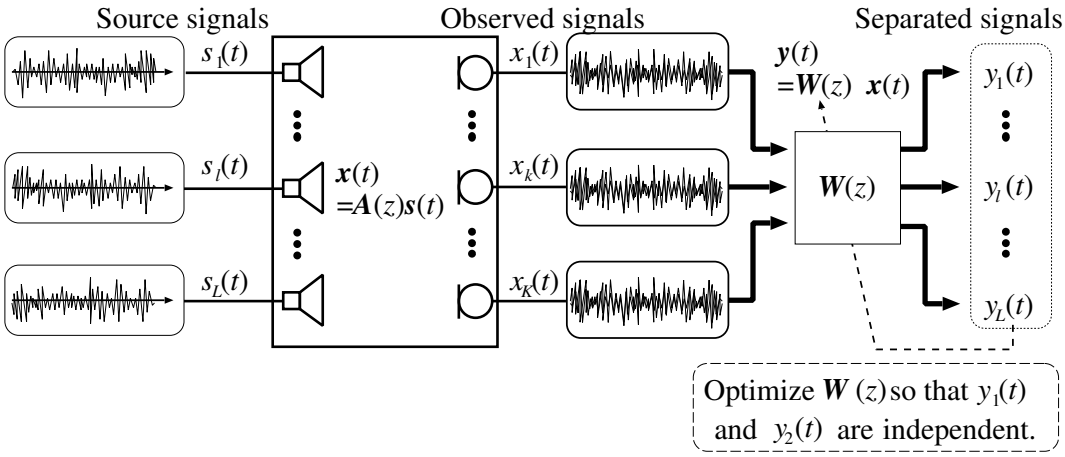

Figure 7. Configuration of the conventional TDICA.

density function (PDF) of $\boldsymbol{y}(t)$ and the product of marginal PDFs of $y_l(t)$. The KLD of Eq. (21) is defined as

$$\mathrm{KLD}(\boldsymbol{y}(t)) = \int p(\hat{\boldsymbol{y}}(t))\log\frac{p(\hat{\boldsymbol{y}}(t))}{\prod_{l=1}^{L}\prod_{t=0}^{T-1}p(y_l(t))}d\hat{\boldsymbol{y}}(t), \tag{22}$$

where $p(y_l(t))$ is the marginal PDF of $y_l(t)$, $p(\hat{\boldsymbol{y}}(t))$ is the joint probability density function of $\boldsymbol{y}(t)$, and $\hat{\boldsymbol{y}}(t)$ is defined as

$$\begin{aligned}\hat{\boldsymbol{y}}(t) = [&y_1(0),\cdots,y_1(T-1),\cdots,y_l(0),\cdots,y_l(T-1),\cdots,\\ &y_K(0),\cdots,y_K(T-1)]^{\mathrm{T}}.\end{aligned} \tag{23}$$

The natural gradient [11, 28] of KLD with respect to $\boldsymbol{w}(n)$ is given by

$$\begin{aligned}&\frac{\partial\mathrm{KLD}(\boldsymbol{y}(t))}{\partial\boldsymbol{w}(n)}\cdot\boldsymbol{W}(z^{-1})^{\mathrm{T}}\boldsymbol{W}(z)\\ &= -\sum_{d=0}^{D-1}\Big\{\boldsymbol{I}\delta(n-d)-\Big\langle\varphi(\boldsymbol{y}(t))\boldsymbol{y}(t-n+d)^{\mathrm{T}}\Big\rangle_t\Big\}\boldsymbol{w}(d),\end{aligned} \tag{24}$$

where $\langle\cdot\rangle_t$ denotes the time-averaging operator and $\boldsymbol{I}$ is the identity matrix. Also, $\delta(n)$ is a delta function, i.e., $\delta(0)=1$ and $\delta(n)=0$ $(n\neq 0)$, and $\boldsymbol{\varphi}(\cdot)$ is an appropriate nonlinear vector function, e.g., given as

$$\boldsymbol{\varphi}(\boldsymbol{y}(t)) = [\tanh(y_1(t)),\cdots,\tanh(y_K(t))]^{\mathrm{T}}. \tag{25}$$

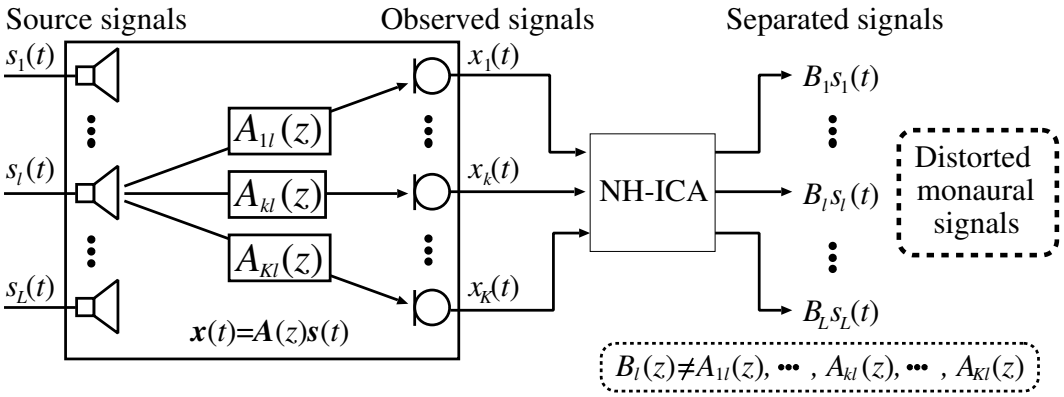

Figure 8. Input and output relations in non-holonomic ICA (NH-ICA). Since it is possible for $B_l(z)$ to be an arbitrary filter, the separated signals include the spectral distortions.

The iterative learning rule is given as

$$\begin{aligned}&\boldsymbol{w}^{[j+1]}(n)\\ &= \boldsymbol{w}^{[j]}(n)+\alpha\sum_{d=0}^{D-1}\left\{\boldsymbol{I}\delta(n-d)-\left\langle\boldsymbol{\varphi}(\boldsymbol{y}^{[j]}(t))\boldsymbol{y}^{[j]}(t-n+d)^{\mathrm{T}}\right\rangle_t\right\}\boldsymbol{w}^{[j]}(d),\end{aligned} \tag{26}$$

where α is the step-size parameter and the superscript $[j]$ is used to express the value of the j-th step in the iterations.

3.2.2 Non-holonomic ICA (NH-ICA) by Choi [29]

It is known that the separated signals of the algorithm Eq. (26) are forced to be i.i.d. and thus become white. To overcome this, Choi et al. proposed the following iterative learning rule with the non-holonomic constraint [29]:

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

$$= \boldsymbol{w}^{[j]}(n) - \alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}^{[j]}(t)) \boldsymbol{y}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}^{[j]}(d), \quad (27)$$

where off-diag$\boldsymbol{W}(z)$ is the operation for setting every diagonal element of the matrix $\boldsymbol{W}(z)$ as zero.

This ICA method is basically a means of extracting each of the independent sound sources as a monaural signal (see Figure 8). In addition, the quality of the separated sound cannot be guaranteed, i.e., the separated signals often include spectral distortions because the modified separated signals which are convolved with arbitrary linear filters are also mutually independent. As shown in Figure 8, $y_l(t) = B_l(z)s_l(t)$, where $B_l(z)$ is an arbitrary filter, is a possible solution obtained from the conventional ICA using Eq. (27). Here it should be noted that the specific solution in the case of $B_l(z) = A_{kl}(z)$ is still admissible in Eq. (27) as one possible *nondistorted* solution, but the problem is that we cannot limit the space of solutions to such a nondistortion case by only using Eq. (27). Overall, 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.

3.2.3 ICA based on the minimal distortion principle (MDP-ICA) by Matsuoka [30]

In order to solve the problem, particularly with respect to the sound quality, Matsuoka et al. have proposed a modified ICA based on the Minimal Distortion Principle (MDP). The iterative learning rule using this principle is given as

$$\begin{aligned} \boldsymbol{w}^{[j+1]}(n) &= \boldsymbol{w}^{[j]}(n) \\ &\quad -\alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}^{[j]}(t)) \boldsymbol{y}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}^{[j]}(d) \\ &\quad + \alpha\beta \sum_{d=0}^{D-1} \left\{ \text{diag} \left\langle (\boldsymbol{y}^{[j]}(t) - \boldsymbol{x}(t-\frac{D}{2})) \boldsymbol{y}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}^{[j]}(d), \end{aligned} \quad (28)$$

Since Matsuoka's method should simultaneously minimize the two different kinds of cost functions, namely, the KLD and the Euclidean distance between

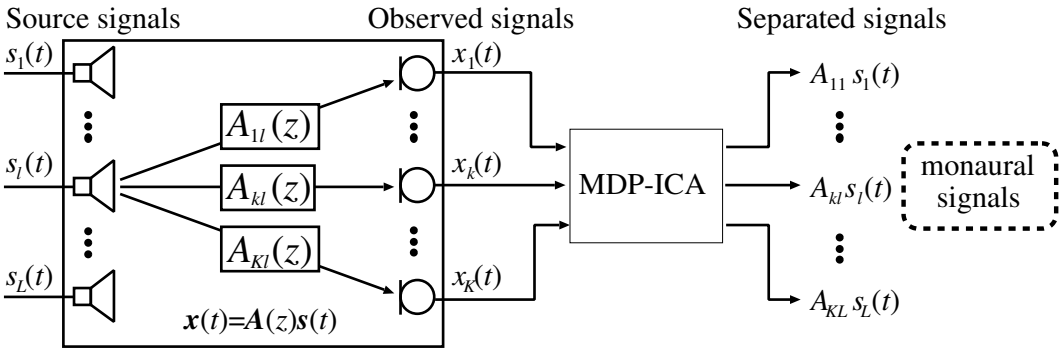

Figure 9. Input and output relations in ICA using minimal distortion principle (MDP-ICA). This algorithm can extract the acoustic source signals from the observed signals without the spectral distortion caused by separation processing. However, since the output signals are monaural with respect to each source, information about directivity, localization, spatial qualities of each sound source is lost.

the separated signal vector and the observed signal vector, the method has the problem that an additional parameter for balancing the cost functions should be required, but is excessively sensitive to the convergence in the iterative learning of the separation filter matrix. The theoretical indication for 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.

3.2.4 2nd-order ICA by Parra [15]

Parra et al., have proposed 2nd-order ICA which can separate the mixing signals of non-stationary signals in the frequency domain. The cost function of this system contains the constraint about the sound qualities of the output signals. The iterative learning rule using this method is given as

$$\boldsymbol{W}^{[j+1]}(f) = \boldsymbol{W}^{[j]}(f) - \alpha\text{off-diag}\{\boldsymbol{R}_{YY}(f,t)\}\boldsymbol{W}^{[j]}(f)\boldsymbol{R}_{XX}(f,t), \quad (29)$$

$$\boldsymbol{R}_{XX}(f,t) = \left\langle \boldsymbol{X}(f,t,b)\boldsymbol{X}(f,t,b)^{\mathrm{T}} \right\rangle_b, \quad (30)$$

$$\boldsymbol{R}_{YY}(f,t) = \left\langle \boldsymbol{Y}^{[j]}(f,t,b)\boldsymbol{Y}^{[j]}(f,t,b)^{\mathrm{T}} \right\rangle_b, \quad (31)$$

$$\boldsymbol{Y}^{[j]}(f,t,b) = \boldsymbol{W}^{[j]}(f)\boldsymbol{X}(f,t,b), \quad (32)$$

where $\boldsymbol{R}_{XX}(f,t)$ and $\boldsymbol{R}_{YY}(f,t)$ mean the correlation matrix of the observed and separated signals, $\boldsymbol{X}(f,t,b)$ is the observed signal vector which is calculated by means of a frame-by-frame discrete Fourier transform (DFT), and $\boldsymbol{Y}(f,t,b)$ is the separated signal vector. This method optimize the output signals so that 2nd-order criterial cost function is minimized. Thus, to obtain the high-performance, this method needs the long time sequence. Also, this method is valid only for monaural outputs, and the fidelity of the output signals as SIMO-model-based signals cannot be guaranteed.

3.2.5 Multistage ICA by Nishikawa [17]

In order to improve the separation performance, Nishikawa have proposed multi-stage ICA (MSICA) which combine the frequency domain ICA (FDICA) and time domain ICA (TDICA). Since this method's concept can be reflected in all kinds of ICA methods, using this method, we can always execute the high separation performance.

3.2.6 Other conventional ICA methods

In order to resolve the problems, several methods which project the separated signal $y_l(t)$ onto observed signals' space have been proposed. The first example is a method which utilizes the inverse of $\boldsymbol{w}(n)$ (see, e.g., [14]). In this method, the nonsingularity of the separation filter matrix must be required, but it cannot be always guaranteed [18, 21]. The second example is a deflation-type method (see, e.g., [22, 23]), in which we extract a specific monaural source signal $y_l(t)$, and then $y_l(t)$ is projected back onto the k-th microphone. However, the method often fails if the separation filters produce a null space because the invertibility of the filters cannot hold. Also, IIR-filter-based ICA [20] has been proposed, but the problem is how to guarantee the separation filters' stability. Therefore, development of a new SIMO-output-type method in which all SIMO components can be simultaneously estimated in the ICA updating with stable FIR filters, is a problem demanding prompt attention. It is also worth mentioning that Cardoso

has proposed a Multidimensional ICA (MICA) [31] as an extended ICA for decomposing the mixed signals into non-monaural contributions at the observation points. However, MICA theory and the algorithm were developed only in the case of instantaneous mixtures, and hence we cannot apply the algorithm to BSS for the acoustic sound mixtures addressed in this paper.

3.3 Conclusion

In this section, we describe the conventional ICA methods and we explain the disadvantage of these methods. Furthermore, we describe that development of a new SIMO-output-type method, in which all SIMO components can be simultaneously estimated in the ICA updating with stable FIR filters, is a problem demanding prompt attention.

4. SIMO-model-based ICA (SIMO-ICA) with least squares criterion (SIMO-ICA-LS)

4.1 Introduction

In this and next chapter, 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 this section, we propose SIMO-ICA with least squares criterion (SIMO-ICA-LS), which have the fidelity controller designed on the basis of least squares criterion. (In Section 5, we propose the another fidelity controller which is designed using the information geometric criterion.)

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

The rest of this chapter is organized as follows. In the Section 4.2, the proposed SIMO-ICA-LS is described in detail. In Sections 4.3 and 4.4, the unique solution and iterative learning rule in the proposed SIMO-ICA-LS are described. In Section 4.5, the experimental results are presented. Finally, Section 4.7 conclude this chapter.

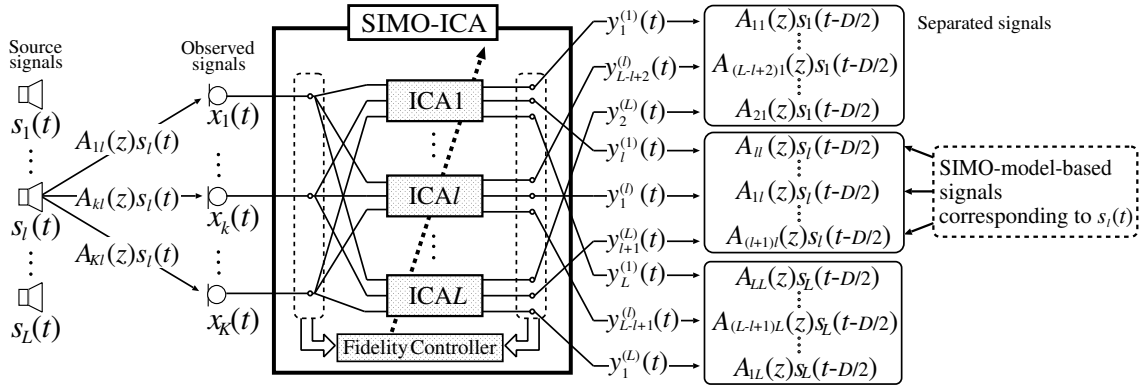

Figure 10. Example of input and output relations in proposed SIMO-ICA, where exclusively-selected permutation matrices $\boldsymbol{P}_l$ are given by Eq. (48)

.

4.2 Fidelity controller in SIMO-ICA-LS

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 Figure 10). The separated signals of the l-th ICA in SIMO-ICA are defined by

$$\begin{aligned} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) &= [y_k^{(l)}(t)]_{k1} && (33)\\ &= \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(n)\boldsymbol{x}(t-n) && (34)\\ &= \boldsymbol{W}_{(\mathrm{ICA}l)}(z)\boldsymbol{x}(t), && (35) \end{aligned}$$

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 $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$, and D is the filter length of the separation filter $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ and should be set to enough large (see Appendix A). Regarding the fidelity controller, we newly introduce the following cost function to be minimized,

$$C(\boldsymbol{w}_{(\mathrm{ICA}1)}(n), ..., \boldsymbol{w}_{(\mathrm{ICA}L)}(n)) \equiv \left\langle \left\| \sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t - D/2) \right\|^2 \right\rangle_t, \quad (36)$$

where $\| \boldsymbol{x} \|$ is the Euclidean norm of vector $\boldsymbol{x}$. The cost function Eq. (36) 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 $[\sum_{l=1}^{L} A_{kl}(t - D/2)]_k (= \boldsymbol{x}(t - D/2)$. Here the delay of $D/2$ is used to deal with nonminimum phase systems. Using Eqs. (35) and (36), we can obtain the appropriate separated signals and maintain their spatial qualities as follows.

4.3 Unique solution in SIMO-ICA-LS

Theorem: The output signals converge on unique solutions, Eq. (37), up to the permutation, if and only if the independent sound sources are separated by Eq. (35), and simultaneously Eq. (36) is minimized to be zero.

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \mathrm{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2), \tag{37}$$

where $l = 1, \cdots, L$ and $\boldsymbol{P}_l$ $(l = 1,\ ...,\ L)$ are exclusively selected permutation matrices [35] which satisfy

$$\sum_{l=1}^{L} \boldsymbol{P}_l = [1]_{ij}. \tag{38}$$

Proof of Theorem: First, the necessity is shown below. Obviously the solution Eq. (37) holds in Eq. (35) because the elements of Eq. (37) are mutually independent in each $\boldsymbol{y}_{(\mathrm{ICA}l)}(t)$ $(l = 1,\ ...,\ L)$. Also, the following equation holds with Eq. (37).

$$\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \sum_{l=1}^{L} \mathrm{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2) \tag{39}$$

$$= \left[\sum_{l=1}^{L} A_{kl}(z)s_l(t - D/2)\right]_k \tag{40}$$

$$= \boldsymbol{A}(z)\boldsymbol{s}(t - D/2). \tag{41}$$

This results in $\boldsymbol{x}(t - D/2)$, and makes the cost function Eq. (36) be zero. This completes the proof of the necessity in Theorem.

Next, 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. If one of the preconditions ("independent sound sources have been separated by Eq. (35)") holds, the general expression of the l-th ICA's output is given by

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) \;=\; \boldsymbol{D}_l(z)\boldsymbol{Q}_l\boldsymbol{s}(t - D/2). \tag{42}$$

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 $s_l(t - D/2)$ $(l = 1,\ ...,\ L)$. This obviously makes the cost function Eq. (36) 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. (38), and we obtain

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) \;=\; \boldsymbol{D}_l(z)\boldsymbol{P}_l\boldsymbol{s}(t - D/2). \tag{43}$$

In Eq. (43) under Eq. (38), 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,

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) \;=\; \mathrm{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t - D/2), \tag{44}$$

and consequently

$$\sum_{l=1}^{L}\boldsymbol{y}_{(\mathrm{ICA}l)}(t) \;=\; \left[\sum_{l=1}^{L}B_{kl}(z)s_l(t - D/2)\right]_k. \tag{45}$$

Substitution of Eq. (45) in Eq. (36) leads to the following equation.

$$\begin{aligned} &C(\boldsymbol{w}_{(\mathrm{ICA}1)}(n), ..., \boldsymbol{w}_{(\mathrm{ICA}L)}(n)) \\ &= \left\langle \left\| \left[\sum_{l=1}^{L}B_{kl}(z)s_l(t - D/2)\right]_k - \left[\sum_{l=1}^{L}A_{kl}(z)s_l(t - D/2)\right]_k \right\|^2 \right\rangle_t \qquad (46) \\ &= \sum_{l=1}^{L}\sum_{k=1}^{K}(B_{kl}(z) - A_{kl}(z))^2\langle s_l(t - D/2)^2\rangle_t, \qquad (47) \end{aligned}$$

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. (47) becomes zero if and

only if $B_{kl}(z) = A_{kl}(z)$ for all k and l. Thus, Eq. (44) results in Eq. (37). This completes the proof of the sufficiency in Theorem.

Obviously the solutions given by Eq. (37) provide necessary and sufficient SIMO components, $A_{kl}(z)S_l(t-D/2)$, for each l-th source. However, the condition Eq. (38) allows multiple possibilities for the combination of $\boldsymbol{P}_l$. For example, one possibility is shown in Figure 10 and this corresponds to

$$\boldsymbol{P}_l \;=\; [\delta_{im(k,l)}]_{ki}, \tag{48}$$

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

$$m(k,l) \;=\; \begin{cases} k+l-1 & (k+l-1 \le L) \\ k+l-1-L & (k+l-1 > L) \end{cases} \tag{49}$$

In this case, Eq. (37) yields

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = [A_{km(k,l)} s_{m(k,l)}(t - D/2)]_k, \tag{50}$$

4.4 Iterative learning rule in SIMO-ICA-LS

In order to obtain Eq. (37), the natural gradient [11, 28] of Eq. (36) 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. (36) is given as (see Appendix B)

$$\begin{aligned} &\left\{ \frac{\partial}{\partial \boldsymbol{w}_{(\mathrm{ICA}l)}(n)} \Big\langle \Big\| \sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t - \frac{D}{2}) \Big\|^2 \Big\rangle_t \right\} \boldsymbol{W}_{(\mathrm{ICA}l)}(z^{-1})^{\mathrm{T}} \boldsymbol{W}_{(\mathrm{ICA}l)}(z) \\ &\quad = \; 2 \sum_{d=0}^{D-1} \Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t - \frac{D}{2})) \boldsymbol{y}_{(\mathrm{ICA}l)}(t-n+d)^{\mathrm{T}} \Big\rangle_t \boldsymbol{w}_{(\mathrm{ICA}l)}(d). \end{aligned} \tag{51}$$

By combining Eq. (27) with Eq. (51), we can obtain the new iterative algorithm of SIMO-ICA as

$$\begin{aligned} \boldsymbol{w}^{[j+1]}_{(\mathrm{ICA1})}(n) \;=\; & \boldsymbol{w}^{[j]}_{(\mathrm{ICA})1}(n) \\ & -\alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}^{[j]}_{(\mathrm{ICA1})}(t)) \boldsymbol{y}^{[j]}_{(\mathrm{ICA})1}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}^{[j]}_{(\mathrm{ICA})1}(d) \\ & +\alpha\beta \sum_{d=0}^{D-1} \left\{ \Big\langle \Big(\sum_{l=1}^{L} \boldsymbol{y}^{[j]}_{(\mathrm{ICA})l}(t) - \boldsymbol{x}(t - \frac{D}{2}) \Big) \boldsymbol{y}^{[j]}_{(\mathrm{ICA1})}(t-n+d)^{\mathrm{T}} \Big\rangle_t \right\} \end{aligned}$$

$$\boldsymbol{w}^{[j]}_{(\mathrm{ICA}1)}(d), \tag{52}$$

$$\vdots$$

$$\begin{aligned}\boldsymbol{w}^{[j+1]}_{(\mathrm{ICA}l)}(n) &= \boldsymbol{w}^{[j]}_{(\mathrm{ICA}l)}(n) \\ &\quad -\alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}^{[j]}_{(\mathrm{ICA}l)}(t)) \boldsymbol{y}^{[j]}_{(\mathrm{ICA}l)}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}^{[j]}_{(\mathrm{ICA}l)}(d) \\ &\quad +\alpha\beta \sum_{d=0}^{D-1} \left\{ \left\langle (\sum_{l=1}^{L} \boldsymbol{y}^{[j]}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2})) \boldsymbol{y}^{[j]}_{(\mathrm{ICA}l)}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \\ &\quad \boldsymbol{w}^{[j]}_{(\mathrm{ICA}l)}(d),\end{aligned} \tag{53}$$

$$\vdots$$

$$\begin{aligned}\boldsymbol{w}^{[j+1]}_{(\mathrm{ICA}L)}(n) &= \boldsymbol{w}^{[j]}_{(\mathrm{ICA}L)}(n) \\ &\quad -\alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}^{[j]}_{(\mathrm{ICA}L)}(t)) \boldsymbol{y}^{[j]}_{(\mathrm{ICA}L)}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}^{[j]}_{(\mathrm{ICA}L)}(d) \\ &\quad +\alpha\beta \sum_{d=0}^{D-1} \left\{ \left\langle (\sum_{l=1}^{L} \boldsymbol{y}^{[j]}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2})) \boldsymbol{y}^{[j]}_{(\mathrm{ICA}L)}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \\ &\quad \boldsymbol{w}^{[j]}_{(\mathrm{ICA}L)}(d),\end{aligned} \tag{54}$$

where α is the step-size parameter which is for the control of the total update quantity, and β is balancing parameter which is for the control of the balancing between the term for separation and fidelity. In Eqs. (52)–(54), the updating $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ should be simultaneously performed in parallel because each iterative equation is associated with the others via $\boldsymbol{y}^{[j]}_{(\mathrm{ICA}l)} = \boldsymbol{W}^{[j]}_{(\mathrm{ICA}l)}(z)\boldsymbol{x}(t)$. 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 possibly arises. This can be easily achieved by using a cross correlation between time-shifted separated signals,

$$CC(l, l', k, k') = \max_n \langle y_k^{(\mathrm{ICA}l)}(t) y_{k'}^{(\mathrm{ICA}l')}(t-n) \rangle_t, \tag{55}$$

where $l \neq l'$ and $k \neq k'$. The large value of $CC(l, l', k, k')$ indicates that $y_k^{(\mathrm{ICA}l)}(t)$ and $y_{k'}^{(\mathrm{ICA}l')}(t)$ are SIMO components from the same source.

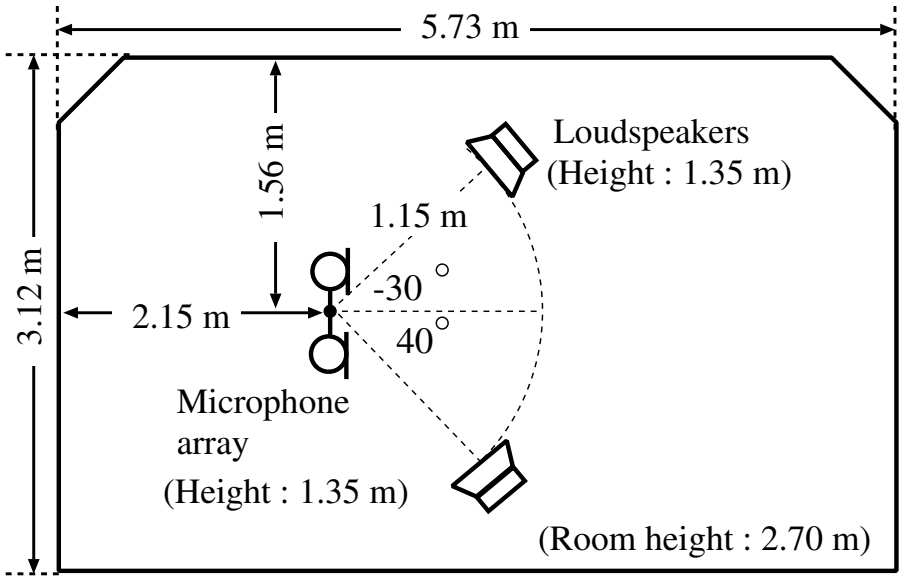

Figure 11. Layout of reverberant room used in experiment. The reverberation time of this room is 150 ms.

4.5 Experiments and results for two-source case

4.5.1 Experimental conditions

Two kinds of sentences, spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research [36], are used as the original speech samples $\boldsymbol{s}(t)$. Using these sentences, we obtain six combinations. The sampling frequency is 8 kHz and the length of speech is limited to 3 seconds. To evaluate the feasibility of the proposed SIMO-ICA-LS in a real environment, we used the impulse response $\boldsymbol{a}(n)$ which is recorded in the experimental room as shown in Figure 11. The reverberation time of this impulse response is 150 ms. 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. The length of the separation filter, D, is set to be 512 taps, and the initial value is the null-beamformer [16] with the interelement spacing of 4 cm, whose directional null is steered to $\pm 60°$. As the conventional ICA-based BSS method for comparison,

we use NH-ICA algorithm given by Eq. (27). (See Section 3.2.2) The step-size parameter α was set in the range of $5 \times 10^{-8} \sim 5 \times 10^{-6}$. Also, the balancing parameter β was set in the range of $1 \times 10^{-4} \sim 1 \times 10^{-1}$ in the proposed SIMO-ICA-LS. We select the optima which give the best performance in each method.

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.

4.5.2 Objective evaluation score

In this experiment, three objective evaluation scores are defined as described below. Note that the unit of all scores is the *decibel (dB)*, but hereafter we omit the unit in equations.

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

$$\mathrm{NRR} \equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{k=1}^{2}(\mathrm{OSNR}_k^{(\mathrm{ICA}l)} - \mathrm{ISNR}_k^{(\mathrm{ICA}l)}), \tag{56}$$

$$\mathrm{OSNR}_k^{(\mathrm{ICA}l)} = 10\log_{10}\frac{\langle\mid H_{kk'}^{(\mathrm{ICA}l)}(z)s_{k'}(t)\mid^2\rangle_t}{\langle\mid y_k^{(\mathrm{ICA}l)}(t) - H_{kk'}^{(\mathrm{ICA}l)}(z)s_{k'}(t)\mid^2\rangle_t}, \tag{57}$$

$$\mathrm{ISNR}_k^{(\mathrm{ICA}l)} = 10\log_{10}\frac{\langle\mid A_{kk'}(z)s_{k'}(t)\mid^2\rangle_t}{\langle\mid x_k(t-\frac{D}{2}) - A_{kk'}(z)s_{k'}(t)\mid^2\rangle_t}, \tag{58}$$

$$k' = \underset{\kappa}{\mathrm{argmax}}(\langle\mid H_{k\kappa}^{(\mathrm{ICA}l)}(z)s_\kappa(t)\mid^2\rangle_t), \tag{59}$$

where $\mathrm{OSNR}_k^{(\mathrm{ICA}l)}$ and $\mathrm{ISNR}_k^{(\mathrm{ICA}l)}$ are the output SNR and the input SNR for ICAl, respectively. Also, $H_{ij}^{(\mathrm{ICA}l)}(z)$ is the element in the i-th row and the j-th column of the matrix $\boldsymbol{H}_{(\mathrm{ICA}l)}(z) = \boldsymbol{W}_{(\mathrm{ICA}l)}(z)\boldsymbol{A}(z)$. For example, $\boldsymbol{y}(t) = [B_1(z)s_1(t), B_2(z)s_2(t)]^{\mathrm{T}}$ gains an NRR of infinity because the separation performance is prefect, even if $B_l(z)$ $(l = 1, 2)$ represent any arbitrary distortions.

Secondly, *sound quality* (SQ), defined as described below, indicates the sound quality of the separated signal,

$$\mathrm{SQ} \equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{k=1}^{2}\mathrm{SQ}_{y_k^{(\mathrm{ICA}l)}}, \tag{60}$$

$$\mathrm{SQ}_{y_k^{(\mathrm{ICA}l)}} = 10\log_{10}\frac{\langle|\ A_{kk'}(z)s_{k'}(t-\frac{D}{2})\ |^2\rangle_t}{\langle|\ A_{kk'}(z)s_{k'}(t-\frac{D}{2}) - H_{kk'}^{(\mathrm{ICA}l)}(z)s_{k'}(t)\ |^2\rangle_t}, \tag{61}$$

where $\mathrm{SQ}_{y_k^{(\mathrm{ICA}l)}}$ is the sound quality of the separated signal $y_k^{(\mathrm{ICA}l)}(t)$. In SQ, we do not take into account the degree of the separation performance in the output signals. For example, $\boldsymbol{y}(t) = \boldsymbol{A}(z)\boldsymbol{s}(t - D/2)$ gains an SQ of infinity due to the absence of target signals' distortions, whereas there is no source separation (i.e., NRR is quite low).

Lastly, *SIMO-model accuracy* (SA) indicates the degree of similarity between the SIMO-ICA's outputs and the original SIMO-model-based signals. It is defined by

$$\mathrm{SA} \equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{k=1}^{2}\mathrm{SA}_{y_k^{(\mathrm{ICA}l)}}, \tag{62}$$

$$\mathrm{SA}_{y_k^{(\mathrm{ICA}l)}} = 10\log_{10}\frac{\langle|\ A_{kk'}(z)s_{k'}(t-\frac{D}{2})\ |^2\rangle_t}{\langle|\ A_{kk'}(z)s_{k'}(t-\frac{D}{2}) - y_k^{(\mathrm{ICA}l)}(t)\ |^2\rangle_t}, \tag{63}$$

where $\mathrm{SA}_{y_k^{(\mathrm{ICA}l)}}$ is the similarity between the separated signal $y_k^{(\mathrm{ICA}l)}(t)$ and the original SIMO-model-based signal $A_{kk'}(z)s_{k'}(t)$. The SA corresponds to the total indication which combines NRR and SQ, that is, both the separation performance and the sound quality are taken into account.

4.5.3 Experimental results and discussion for two-source case

Figure 12 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-LS 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-LS is almost the same as that of the conventional ICA-based method.

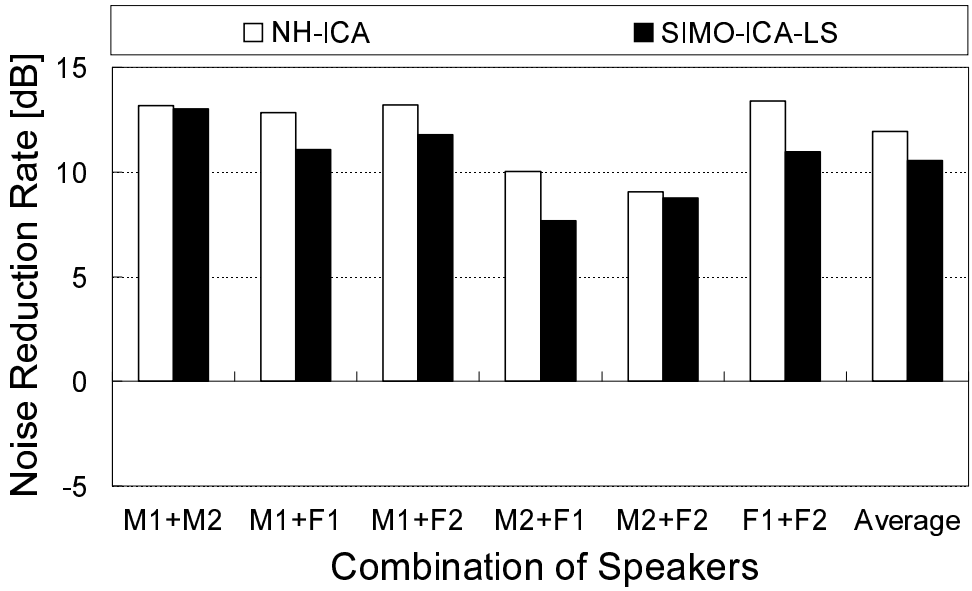

Figure 12. Results of noise reduction rate in the conventional ICA and the proposed SIMO-ICA-LS, where the reverberation time is 150 ms. The length of the filter, D, in all methods is set to be 512 taps. In the horizontal axis, the symbols "M1+M2"–"F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers. Thus, for example, "M1+F2" corresponds to the male-female combination.

Figure 13 and Figure 14 show the results of SQ and SA 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 SA is 5.0 dB. From these results, it is evident that the sound quality of the separated signals in SIMO-ICA-LS is obviously superior to that of the separated signals in the conventional ICA-based method, particularly in terms of the accuracy of the sound reproduction. Regarding the SQ score, the improvement in SIMO-ICA-LS is not large compared with that in SIMO-ICA-LS in the nonreverberant case described in the previous section. The main reason for this is the insufficiency of the source-

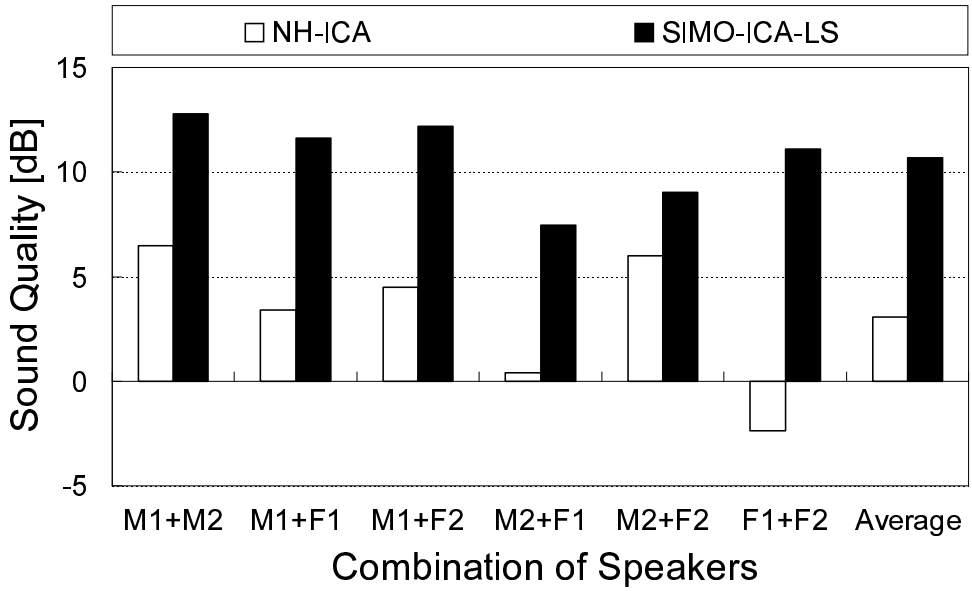

Figure 13. Results of noise reduction rate in the conventional ICA and the proposed SIMO-ICA-LS, where the reverberation time is 150 ms. The length of the filter, D, in all methods is set to be 512 taps. In the horizontal axis, the symbols "M1+M2"–"F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers. Thus, for example, "M1+F2" corresponds to the male-female combination.

separation performance.

Overall, the results indicate the following points. (1) In SIMO-ICA-LS, 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-LS. Therefore, we can conclude that the proposed SIMO-ICA-LS is applicable to binaural signal processing and high-fidelity sound reproduction systems.

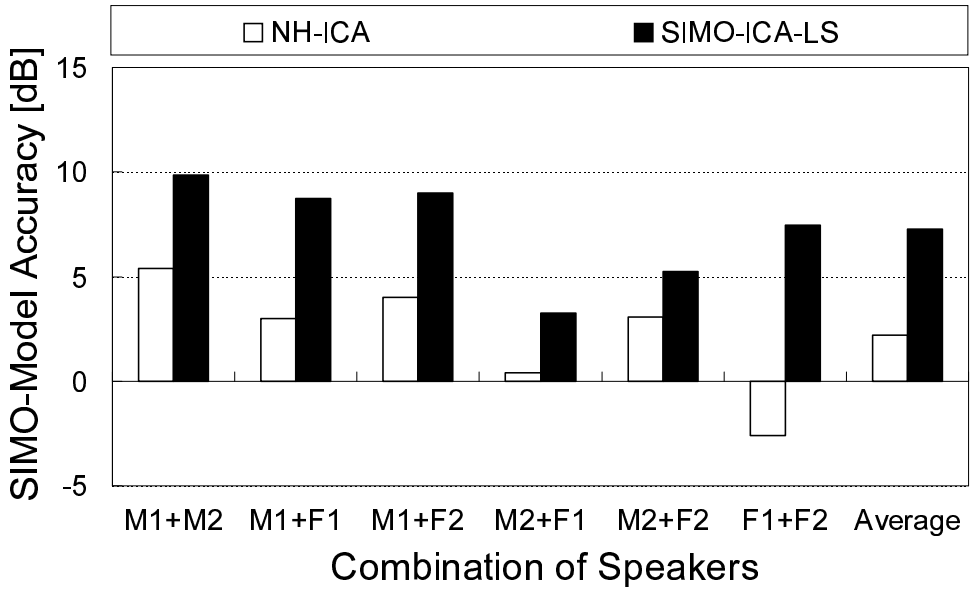

Figure 14. Results of noise reduction rate in the conventional ICA and the proposed SIMO-ICA-LS, where the reverberation time is 150 ms. The length of the filter, D, in all methods is set to be 512 taps. In the horizontal axis, the symbols "M1+M2"–"F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers. Thus, for example, "M1+F2" corresponds to the male-female combination.

4.6 Experiments and results for three-source case

4.6.1 Experimental conditions

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 Section 4 are used, and we obtain 4 combinations. In order to generate the room impulse responses, we use the image method [37] assuming the artificial room as shown in Fig. **??**, where the RT is

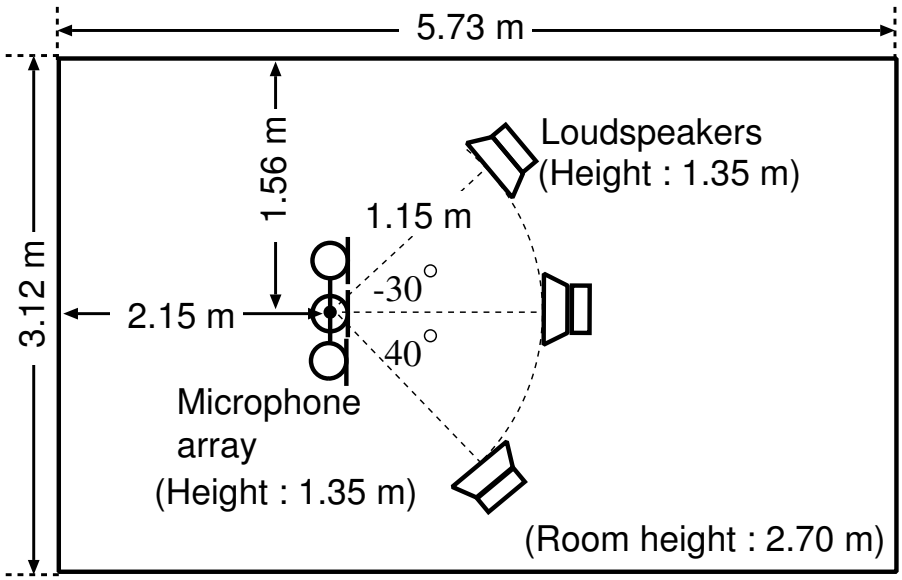

Figure 15. Layout of reverberant room used in experiment. The reverberation time of this room is 150 ms.

set to be 150 ms.

The length of $\boldsymbol{w}(n)$ 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.

4.6.2 Experimental results and discussion for two-source case

Figures 16–18, show the results of NRR, SQ and SA 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 SA is 0.5 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 $K = L = 2$.

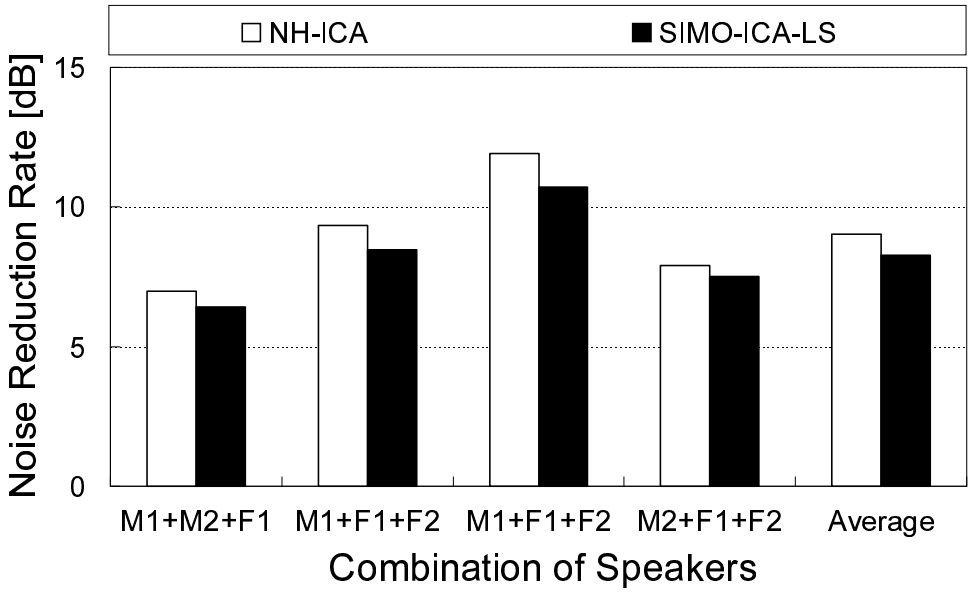

Figure 16. Results of noise reduction rate in the conventional ICA and the proposed SIMO-ICA-LS, where the reverberation time is 150 ms. The length of the filter, D, in all methods is set to be 1024 taps. In the horizontal axis, the symbols "M1+M2"–"F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers.

4.7 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 this chapter, we design the fidelity controller of SIMO-ICA using the least squares criterion. In order to evaluate its effectiveness, separation experiments are carried out using two microphones and two sources under the condition that the RTs 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

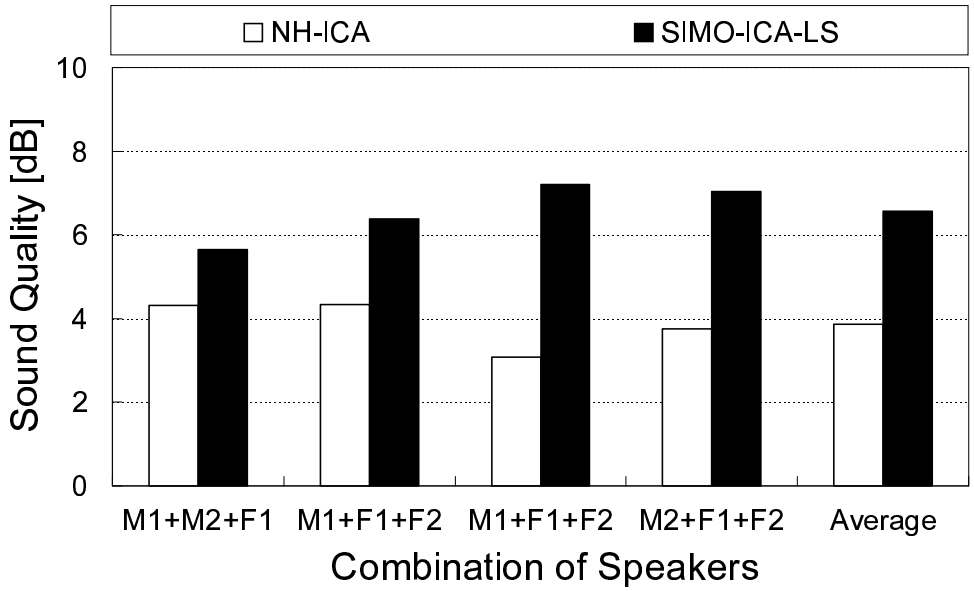

Figure 17. Results of noise reduction rate in the conventional ICA and the proposed SIMO-ICA-LS, where the reverberation time is 150 ms. The length of the filter, D, in all methods is set to be 1024 taps. In the horizontal axis, the symbols "M1+M2"–"F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers.

in SIMO-ICA are remarkably 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.

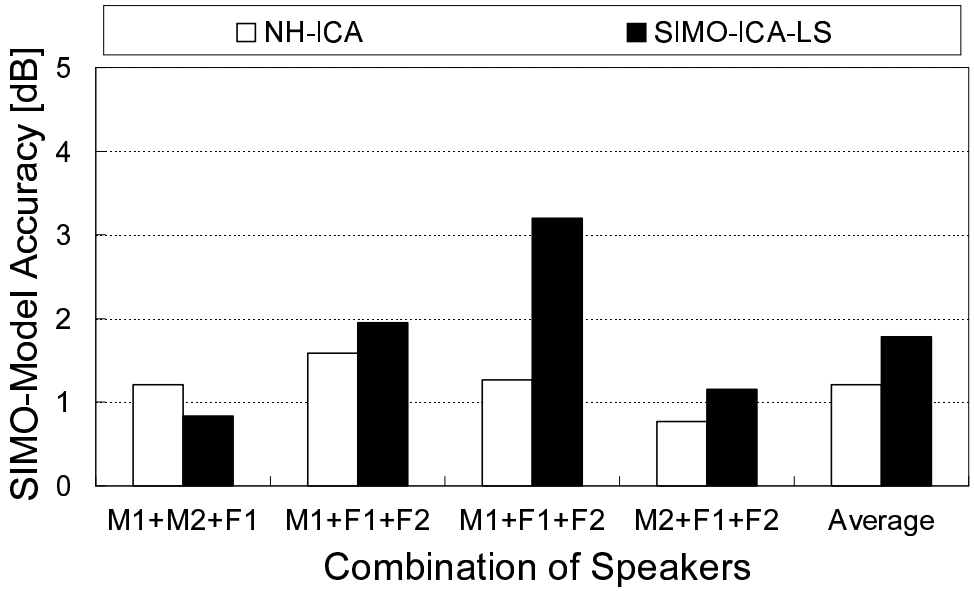

Figure 18. Results of noise reduction rate in the conventional ICA and the proposed SIMO-ICA-LS, where the reverberation time is 150 ms. The length of the filter, D, in all methods is set to be 1024 taps. In the horizontal axis, the symbols "M1+M2+F1"–"M2+F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers.

5. SIMO-ICA with information geometric criterion (SIMO-ICA-IG)

5.1 Introduction

In the previous chapter, we have discussed one type of SIMO-ICA algorithm, the so-called SIMO-ICA with the least squares criterion (SIMO-ICA-LS). The SIMO-ICA-LS consists of multiple ICA parts and a fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system. Each of the ICA parts are driven by information-geometry theory, whereas the fidelity controller is based on the least squares criterion which is a different class of criterion from the information-geometric metric. Due to the mismatch between two kinds of criteria, there is an inherent disadvantage in that an additional balancing parameter is needed, but the parameter is very sensitive to the convergence in the iterative learning. To solve the above-mentioned fundamental problem, in this chapter, we newly propose an SIMO-ICA with an information-geometric learning algorithm (SIMO-ICA-IG). In this method, the fidelity controller as well as each of the ICA parts is designed on the basis of information-geometry theory. Namely, all of the procedures for optimization of the separation filters are conducted by the information-geometric learning algorithm. 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-IG is almost the same as those of the conventional ICA and the previously proposed SIMO-ICA-LS, and the sound quality of the separated signals in the SIMO-ICA-IG is markedly superior to that in the conventional ICA, particularly with regard to the spatial quality. In addition, it is confirmed that the internal parameter setting in the proposed SIMO-ICA-IG does not depend on the source signals' properties, unlike that of SIMO-ICA-LS.

The rest of this chapter is organized as follows. In Section 5.2, the proposed SIMO-ICA-IG is described in detail. In Sections 5.3 and 5.4, the unique solution and iterative learning rule in the proposed SIMO-ICA-IG are described. In Sections 5.5 and 5.6, the experimental results are presented. Finally, Section 5.7 conclude this chapter.

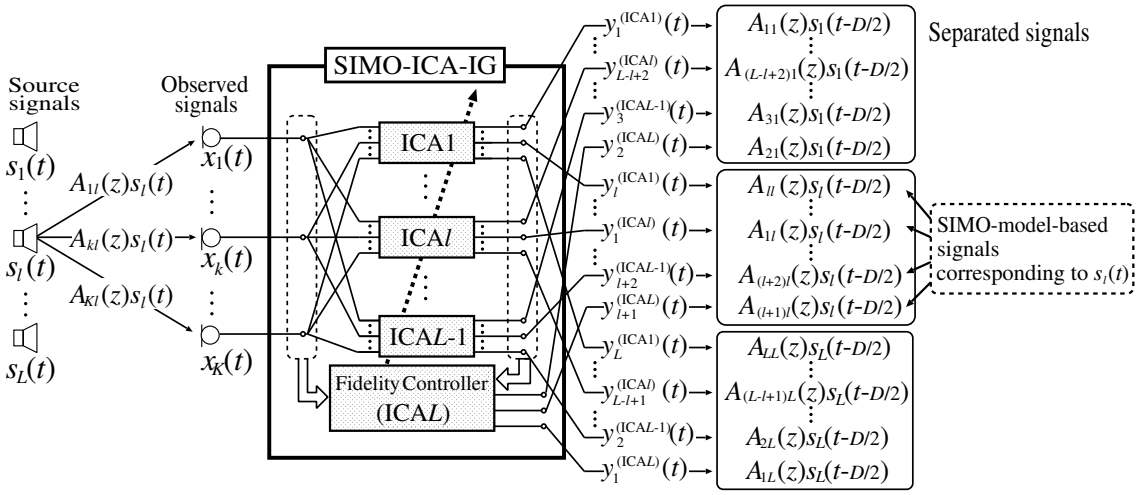

Figure 19. Example of input and output relations in the proposed SIMO-ICA-IG, where exclusively selected permutation matrices $\boldsymbol{P}_l$ are given by Eq. (48). The SIMO-ICA-IG consists of $(L-1)$ ICA parts and an information-geometry-based fidelity controller (virtual ICAL), and each ICA runs in parallel under the fidelity control of the entire separation system. In this system, the separated signals maintain their spatial qualities.

5.2 Fidelity controller in SIMO-ICA-IG

In order to improve the SIMO-ICA-LS, we propose a new blind separation method for SIMO-model-based acoustic signals using the SIMO-ICA-IG. The SIMO-ICA-IG consists of "$(L-1)$" ICA parts and an *information-geometry-based* fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system (see Figure 19). In SIMO-ICA-IG, all of the procedures for optimization of the separation filters are conducted by the information-geometric learning algorithm, and this can solve the problem of the balancing-parameter selection (see Section 5.5.3).

The separated signals of the l-th ICA ($l = 1, \cdots, L-1$) in the SIMO-ICA-IG are defined by

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) \;=\; [y_k^{(\mathrm{ICA}l)}(t)]_k \tag{64}$$

$$= \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(n)\boldsymbol{x}(t-n) \quad (65)$$

$$= \boldsymbol{W}_{(\mathrm{ICA})l}(z)\boldsymbol{x}(t), \quad (66)$$

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

$$\boldsymbol{y}_{(\mathrm{ICA}L)}(t) = [y_k^{(\mathrm{ICA}L)}(t)]_k \quad (67)$$

$$= \boldsymbol{x}(t-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}(t). \quad (68)$$

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

$$\boldsymbol{w}_{(\mathrm{ICA}L)}(n) = \boldsymbol{I}\delta(n-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(n), \quad (69)$$

$$\boldsymbol{W}_{(\mathrm{ICA}L)}(z) = \boldsymbol{I}z^{-\frac{D}{2}} - \sum_{l=1}^{L-1} \boldsymbol{W}_{(\mathrm{ICA}l)}(z). \quad (70)$$

On the basis of Eq. (69) and Eq. (70) we can rewrite Eq. (68) as

$$\boldsymbol{y}_{(\mathrm{ICA}L)}(t) = \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA}L)}(n)\boldsymbol{x}(t-n) \quad (71)$$

$$= \boldsymbol{W}_{(\mathrm{ICA}L)}(z)\boldsymbol{x}(t). \quad (72)$$

The reason we use the word "*virtual* " here is that the L-th ICA does not have its own separation filters, unlike 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. (68) as

$$\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}) = [0]_k. \quad (73)$$

Equation (73) expresses a constraint to force the sum of all of the ICAs' output vectors $\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t)$ to be the sum of all of the SIMO components $[\sum_{l=1}^{L} A_{kl}(z)s_l(t-D/2)]_k (= \boldsymbol{x}(t-D/2))$. Using Eq. (66) and Eq. (68), we can obtain the appropriate separated signals and maintain their spatial qualities as follows.

5.3 Unique solution in SIMO-ICA-IG

Theorem: The output signals converge on unique SIMO solutions up to the permutation $\boldsymbol{P}_l$ ($l = 1,\ ...,\ L$) given by following equation, if and only if the independent sound sources are separated by Eq. (66) and simultaneously the signals obtained by Eq. (68) are mutually independent.

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) \ = \ \mathrm{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t-D/2), \tag{74}$$

where $l = 1, \cdots, L$ and $\boldsymbol{P}_l$ ($l = 1,\ ...,\ L$) are exclusively selected permutation matrices [35] which satisfy

$$\sum_{l=1}^{L}\boldsymbol{P}_l \ = \ [1]_{ij}. \tag{75}$$

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

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) \ = \ \boldsymbol{D}_l(z)\boldsymbol{Q}_l\boldsymbol{s}(t-\frac{D}{2}). \tag{76}$$

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 of the components of $s_l(t-D/2)$ ($l = 1,\ ...,\ L$). This obviously makes the left-hand side of Eq. (73) nonzero because the observed signal vector $\boldsymbol{x}(t-D/2)$ includes all of the components of $s_l(t-D/2)$ in each element. Accordingly, $\boldsymbol{Q}_l$ should be $\boldsymbol{P}_l$ specified by Eq. (75), and we obtain

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \boldsymbol{D}_l(z)\boldsymbol{P}_l\boldsymbol{s}(t-\frac{D}{2}). \tag{77}$$

In Eq. (77) under Eq. (75), 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,

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \mathrm{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t-\frac{D}{2}). \tag{78}$$

The substitution of Eq. (78) into Eq. (68) leads to the following equation:

$$\begin{aligned}
&\mathrm{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_L^{\mathrm{T}}\right]\boldsymbol{P}_L\boldsymbol{s}(t-\frac{D}{2}) \\
&\quad = \quad \boldsymbol{x}(t-\frac{D}{2}) - \sum_{l=1}^{L-1}\mathrm{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t-\frac{D}{2}), \qquad (79)
\end{aligned}$$

and consequently

$$\begin{aligned}
&\sum_{l=1}^{L}\mathrm{diag}\left[\boldsymbol{B}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t-\frac{D}{2}) - \boldsymbol{x}(t-\frac{D}{2}) \\
&\quad = \quad \left[\sum_{l=1}^{L}B_{kl}(z)s_l(t-\frac{D}{2})\right]_k - \left[\sum_{l=1}^{L}A_{kl}(z)s_l(t-\frac{D}{2})\right]_k \qquad (80) \\
&\quad = \quad \left[\sum_{l=1}^{L}\Big\{B_{kl}(z)-A_{kl}(z)\Big\}s_l(t-\frac{D}{2})\right]_k \qquad (81) \\
&\quad = \quad [0]_k. \qquad (82)
\end{aligned}$$

Equation (82) is satisfied if and only if $B_{kl}(z) = A_{kl}(z)$ for all k and l. Thus, Eq. (78) results in Eq. (74). This completes the Proof of Theorem.

Obviously, the solutions given by Eq. (74) provide the necessary and sufficient SIMO components, $A_{kl}(z)S_l(t-D/2)$, for each l-th source. However, condition (75) allows multiple possibilities for the combination of $\boldsymbol{P}_l$. For example, one possibility is shown in Figure 19 and this corresponds to

$$\boldsymbol{P}_l \quad = \quad [\delta_{im(k,l)}]_{ki}, \qquad (83)$$

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

$$m(k,l) \quad = \quad \begin{cases} k+l-1 & (k+l-1 \le L) \\ k+l-1-L & (k+l-1 > L) \end{cases}. \qquad (84)$$

In this case, Eq. (74) yields

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) \quad = \quad \left[A_{km(k,l)}(z)s_{m(k,l)}(t-\frac{D}{2})\right]_k. \qquad (85)$$

5.4 Iterative learning rule in SIMO-ICA-IG

In order to obtain Eq. (74), the gradient of KLD of Eq. (68) with respect to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ should be added to the iterative learning rule of the separation filter

in l-th ICA ($l = 1, \cdots, L-1$). Using Eq. (68), we obtain the partial differentiation (standard gradient) of $\mathrm{KLD}(\boldsymbol{y}_{(\mathrm{ICA}L)}(t))$ with respect to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ ($l = 1, \ldots, L-1$) as

$$\frac{\partial \mathrm{KLD}(\boldsymbol{y}_{(\mathrm{ICA}L)}(t))}{\partial \boldsymbol{w}_{(\mathrm{ICA}l)}(n)} = \left[\frac{\partial \mathrm{KLD}(\boldsymbol{y}_{(\mathrm{ICA}L)}(t))}{\partial w_{ij}^{(\mathrm{ICA}L)}(n)} \cdot \frac{\partial w_{ij}^{(\mathrm{ICA}L)}(n)}{\partial w_{ij}^{(\mathrm{ICA}l)}(n)}\right]_{ij} \tag{86}$$

$$= \left[\frac{\partial \mathrm{KLD}(\boldsymbol{y}_{(\mathrm{ICA}L)}(t))}{\partial w_{ij}^{(\mathrm{ICA}L)}(n)} \cdot (-1)\right]_{ij}, \tag{87}$$

where $w_{ij}^{(\mathrm{ICA}L)}(n)$ is the element of $\boldsymbol{w}_{(\mathrm{ICA}L)}(n)$. By replacing $\partial \mathrm{KLD}(\boldsymbol{y}_{(\mathrm{ICA}L)}(t))$ $/\partial \boldsymbol{w}_{(\mathrm{ICA}L)}(n)$ with its natural gradient [28], we modify Eq. (87) as

$$-\frac{\partial \mathrm{KLD}(\boldsymbol{y}_{(\mathrm{ICA}L)}(t))}{\partial \boldsymbol{w}_{(\mathrm{ICA}L)}(n)} \boldsymbol{W}_{(\mathrm{ICA}L)}(z^{-1})^{\mathrm{T}} \boldsymbol{W}_{(\mathrm{ICA}L)}(z)$$
$$= \sum_{d=0}^{D-1} \left\{ \left(\boldsymbol{I}\delta(n-d) - \left\langle \boldsymbol{\varphi}(\boldsymbol{y}_{(\mathrm{ICA}L)}(t)) \boldsymbol{y}_{(\mathrm{ICA}L)}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}_{(\mathrm{ICA}L)}(d). \tag{88}$$

By inserting Eq. (68) and Eq. (72) into Eq. (88), we obtain

$$-\frac{\partial \mathrm{KLD}(\boldsymbol{y}_{(\mathrm{ICA}L)}(t))}{\partial \boldsymbol{w}_{(\mathrm{ICA}L)}(n)} \boldsymbol{W}_{(\mathrm{ICA}L)}(z^{-1})^{\mathrm{T}} \boldsymbol{W}_{(\mathrm{ICA}L)}(z)$$
$$= \sum_{d=0}^{D-1} \left\{ \left(\boldsymbol{I}\delta(n-d) - \left\langle \boldsymbol{\varphi}(\boldsymbol{x}(t-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}(t)) \right.\right.$$
$$\left.\left. \left(\boldsymbol{x}(t-n+d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}(t-n+d) \right)^{\mathrm{T}} \right\rangle_t \right\}$$
$$\left(\boldsymbol{I}\delta(d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(d) \right). \tag{89}$$

In order to deal with non-i.i.d. signals, we apply the non-holonomic constraint to Eq. (89). The natural gradient with the non-holonomic constraint is given as

$$-\sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{x}(t-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}(t)) \right.\right.$$
$$\left.\left. \left(\boldsymbol{x}(t-n+d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}(t-n+d) \right)^{\mathrm{T}} \right\rangle_t \right\}$$
$$\left(\boldsymbol{I}\delta(d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(d) \right). \tag{90}$$

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

$$
\begin{aligned}
\boldsymbol{w}_{(\mathrm{ICA}1)}^{[j+1]}(n) &= \boldsymbol{w}_{(\mathrm{ICA}1)}^{[j]}(n) \\
&\quad -\alpha \sum_{d=0}^{D-1} \left[\left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}_{(\mathrm{ICA}1)}^{[j]}(t)) \boldsymbol{y}_{(\mathrm{ICA}1)}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \right. \\
&\quad \left. \boldsymbol{w}_{(\mathrm{ICA}1)}^{[j]}(d) \right. \\
&\quad -\alpha\beta \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{x}(t-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t)) \right. \right. \\
&\quad \left. \left. \left(\boldsymbol{x}(t-n+d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t-n+d) \right)^{\mathrm{T}} \right\rangle_t \right\} \\
&\quad \left. \left(\boldsymbol{I}\delta(d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{w}_{(\mathrm{ICA}l)}^{[j]}(d) \right) \right], \qquad (91) \\
&\quad \vdots \\
\boldsymbol{w}_{(\mathrm{ICA}l)}^{[j+1]}(n) &= \boldsymbol{w}_{(\mathrm{ICA}l)}^{[j]}(n) \\
&\quad -\alpha \sum_{d=0}^{D-1} \left[\left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t)) \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \right. \\
&\quad \left. \boldsymbol{w}_{(\mathrm{ICA}l)}^{[j]}(d) \right. \\
&\quad -\alpha\beta \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{x}(t-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t)) \right. \right. \\
&\quad \left. \left. \left(\boldsymbol{x}(t-n+d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t-n+d) \right)^{\mathrm{T}} \right\rangle_t \right\} \\
&\quad \left. \left(\boldsymbol{I}\delta(d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{w}_{(\mathrm{ICA}l)}^{[j]}(d) \right) \right], \qquad (92) \\
&\quad \vdots \\
\boldsymbol{w}_{(\mathrm{ICA}L-1)}^{[j+1]}(n) &= \boldsymbol{w}_{(\mathrm{ICA}L-1)}^{[j]}(n) \\
&\quad -\alpha \sum_{d=0}^{D-1} \left[\left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}_{(\mathrm{ICA}L-1)}^{[j]}(t)) \boldsymbol{y}_{(\mathrm{ICA}L-1)}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \right. \\
&\quad \left. \boldsymbol{w}_{(\mathrm{ICA}L-1)}^{[j]}(d) \right. \\
&\quad -\alpha\beta \left\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{x}(t-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t)) \right. \right.
\end{aligned}
$$

$$\Big(\boldsymbol{x}(t-n+d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{y}^{[j]}_{(\mathrm{ICA}l)}(t-n+d)\Big)^{\mathrm{T}}\Big\rangle_t\Big\}$$
$$\Big(\boldsymbol{I}\delta(d-\frac{D}{2}) - \sum_{l=1}^{L-1} \boldsymbol{w}^{[j]}_{(\mathrm{ICA}l)}(d)\Big)\Big], \tag{93}$$

where α is the step-size parameter, and β is the balancing parameter. Unlike Eqs. (52)–(54) in the SIMO-ICA-LS, we can expect that β on the right-hand side of Eqs. (91)–(93) will result in nearly 1 because the gradient of KLD (the second term) and the gradient of the fidelity controller (the third term) are comparable as regards the generation of independent components. This will be shown and discussed later in Section 5.5 and 5.6

In Eqs. (91)–(93), the updating of $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ for all l should be simultaneously performed in parallel in terms of l because each iterative equation is associated with the others via $\sum_{l=1}^{L-1} \boldsymbol{y}^{[j]}_{(\mathrm{ICA}l)}(t)$. Also, the initial values of $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ for all l should be different. If not, each ICA has the same set of inputs and will produce the same outputs, and this results in an undesired solution. Note that the distinct initial values of $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ are indispensable, but do not always guarantee a complete convergence of the proposed algorithm because of the existence of the local minimum.

After the iterations, the separated signals should be classified into SIMO components of each source because the permutation possibly arises. This can be easily achieved by using a cross correlation $CC(l, l', k, k')$ between time-shifted separated signals, which is given as Eq. (55).

5.5 Experiments and results using microphone array signals

5.5.1 Experimental conditions

We carried out the experiments using the same conditions as Section 4.5.1. As the conventional ICA-based BSS method for comparison, we use 2nd-order ICA by Parra, NH-ICA by Choi, MDP-ICA by Matsuoka, and SIMO-ICA-LS. The step-size parameter α was set in the range of $5 \times 10^{-8} \sim 5 \times 10^{-6}$. Also, the balancing parameter β was set in the range of $1 \times 10^{-4} \sim 1 \times 10^{-1}$ in MDP-ICA and SIMO-ICA-LS, and $1 \times 10^{-1} \sim 1 \times 10^{2}$ in SIMO-ICA-IG. We select the

Table 1. The range of the step-size parameter α and the balancing parameter β in the 2nd-order ICA by Parra, NH-ICA by Choi, ICA using Minimal Distortion Principle (MDP-ICA) by Matsuoka, SIMO-ICA-LS, and SIMO-ICA-IG. The step-size parameter control the convergence speed, and the balancing parameter control the balance between the cost function about separation and the cost function about sound quality

Method	step-size parameter α	balancing parameter β
2nd-order ICA	$1 \times 10^{-8} \sim 1 \times 10^{-2}$	–
NH-ICA	$5 \times 10^{-8} \sim 5 \times 10^{-6}$	–
MDP-ICA	$5 \times 10^{-8} \sim 5 \times 10^{-6}$	$1 \times 10^{-4} \sim 1 \times 10^{-1}$
SIMO-ICA-LS	$5 \times 10^{-8} \sim 5 \times 10^{-6}$	$1 \times 10^{-4} \sim 1 \times 10^{-1}$
SIMO-ICA-IG	$5 \times 10^{-8} \sim 5 \times 10^{-6}$	$1 \times 10^{-1} \sim 1 \times 10^{2}$

optima which give the best performance.

5.5.2 Experimental results and discussion on SA

Figures 20 (a)–(g) show the results of SA for different speaker combinations and average of them in the conventional 2nd-order ICA by Parra, NH-ICA by Choi, MDP-ICA by Matsuoka, MDP-ICA ($\beta = 2.0 \times 10^{-4}$), SIMO-ICA-LS, SIMO-ICA-LS ($\beta = 2.0 \times 10^{-4}$), the proposed SIMO-ICA-IG, SIMO-ICA-IG ($\beta = 1$). In these figures, the symbols "M1+M2"–"F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers. Thus, for example, "M1+F2" corresponds to the male-female combination.

In the averaged scores, compared with the conventional 2nd-order ICA, NH-ICA, and MDP-ICA, the improvement of SA in the SIMO-ICA-IG is 5.4 dB, 5.0 dB, and 2.4 dB respectively. Thus, it is revealed that the performance of SIMO-ICA-IG is superior to those of conventional methods.

Regarding the comparison between the proposed SIMO-ICA-IG and SIMO-ICA-LS, we can only confirm a minimal superiority of the proposed SIMO-ICA-IG

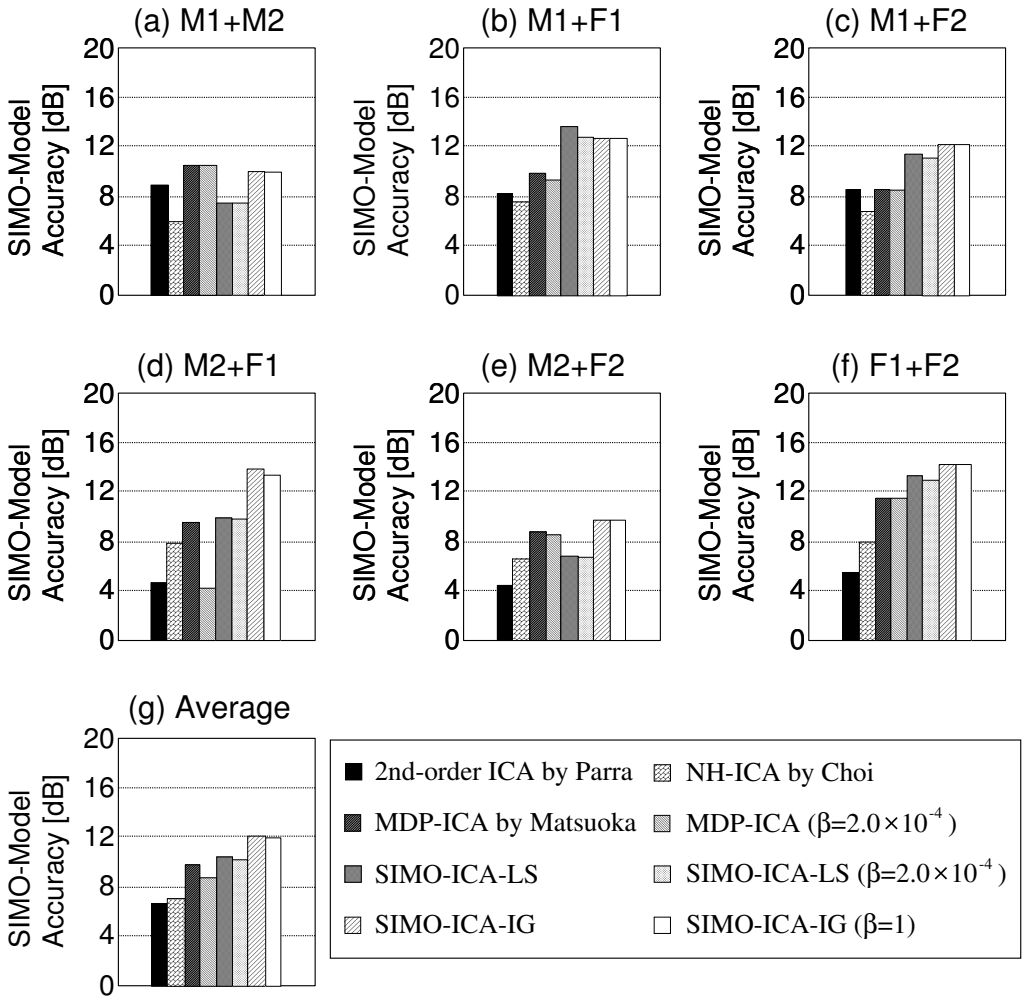

Figure 20. Results of SIMO-model accuracy in the conventional 2nd-order ICA by Parra, NH-ICA by Choi, MDP-ICA by Matsuoka, SIMO-ICA-LS, and the proposed SIMO-ICA-IG, where the reverberation time is 150 ms. The length of the filter, D, in all methods is set to be 512 taps. In this figure (a)–(f), the symbols "M1+M2"–"F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers. Thus, for example, "M1+F2" corresponds to the male-female combination.

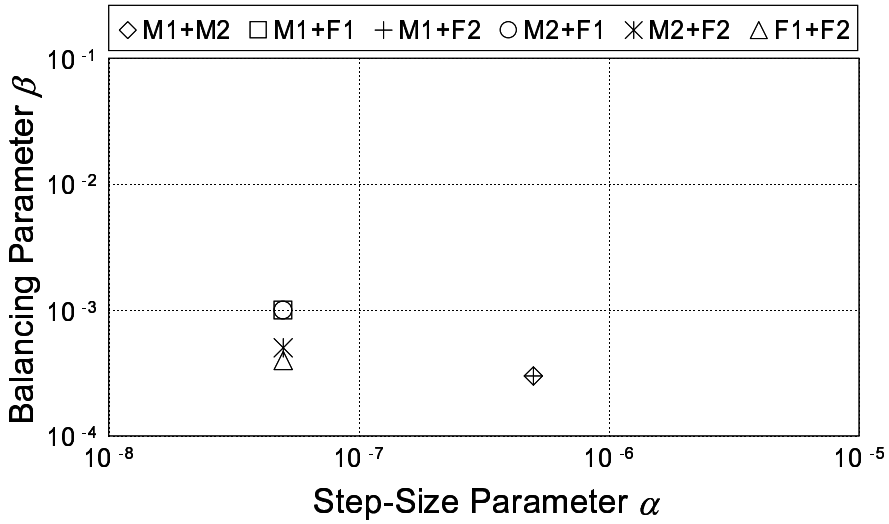

Figure 21. Optimum combination of step-size parameters α and balancing parameter β in MDP-ICA for different speaker combinations. Plotted symbols correspond to six speaker combinations as depicted in Figures. 20 (a)–(f).

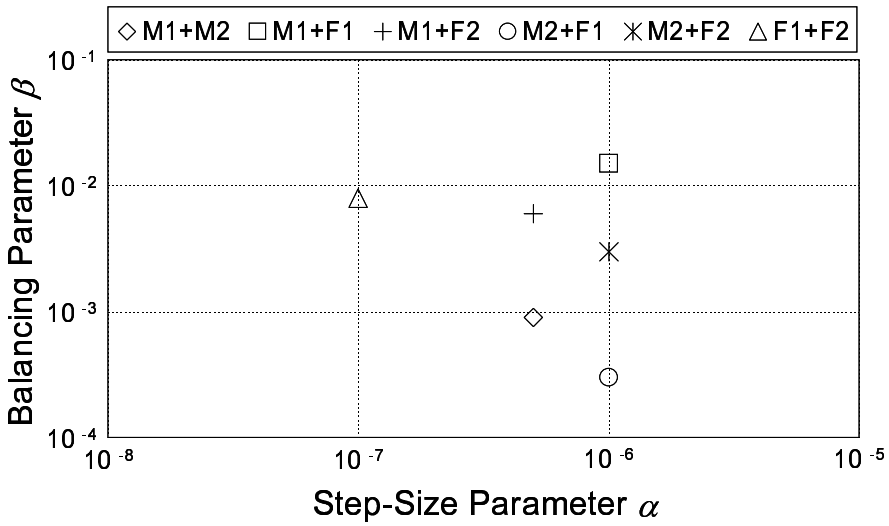

Figure 22. Optimum combination of step-size parameters α and balancing parameter β in SIMO-ICA-LS for different speaker combinations. Plotted symbols correspond to six speaker combinations as depicted in Figures. 20 (a)–(f).

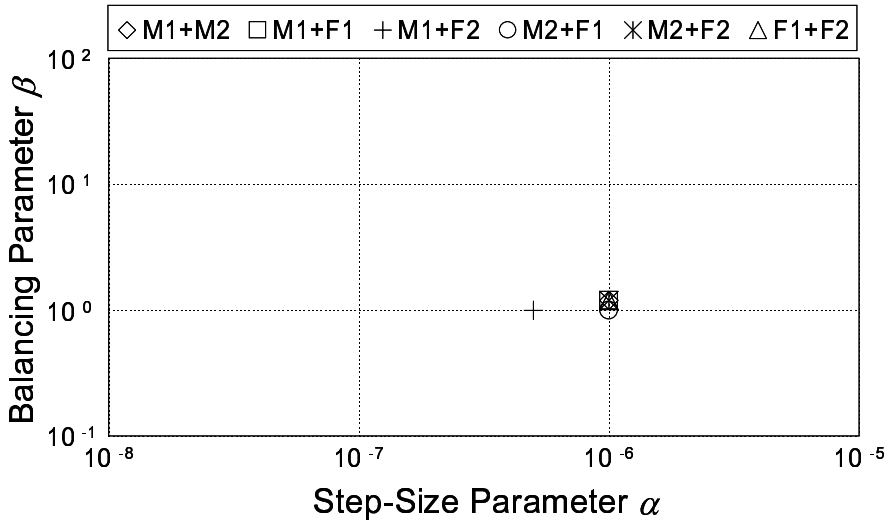

Figure 23. Optimum combination of step-size parameters α and balancing parameter β in SIMO-ICA-IG for different speaker combinations. Plotted symbols correspond to six speaker combinations as depicted in Figures. 20 (a)–(f).

as far as SA is concerned. In the next section, however, we show another considerable advantage of the SIMO-ICA-IG related to the sensitivity of a balancing parameter setting.

5.5.3 Results and discussion on balancing parameter settings

Figures 21–23 show the combinations of optimum step-size parameter α and the balancing parameter β for different speaker combinations in MDP-ICA, SIMO-ICA-LS, and SIMO-ICA-IG. From these results, the optimum step-size parameter α in all methods is within the range between 5.0×10^{-8} and 1.0×10^{-6}.

In Figure 21, the optimum balancing parameter β in the MDP-ICA is dispersed in a wide range between 3.0×10^{-4} and 1.0×10^{-3}. Also, in Figure 22, the optimum balancing parameter β in the SIMO-ICA-LS is dispersed in a wide range between 3.0×10^{-4} and 1.5×10^{-2}. In Figure 23, however, the optimum balancing parameter β in the SIMO-ICA-IG is within the range between 1.0 and 1.2, i.e., it is almost 1. From this result, the range of balancing parameter β in

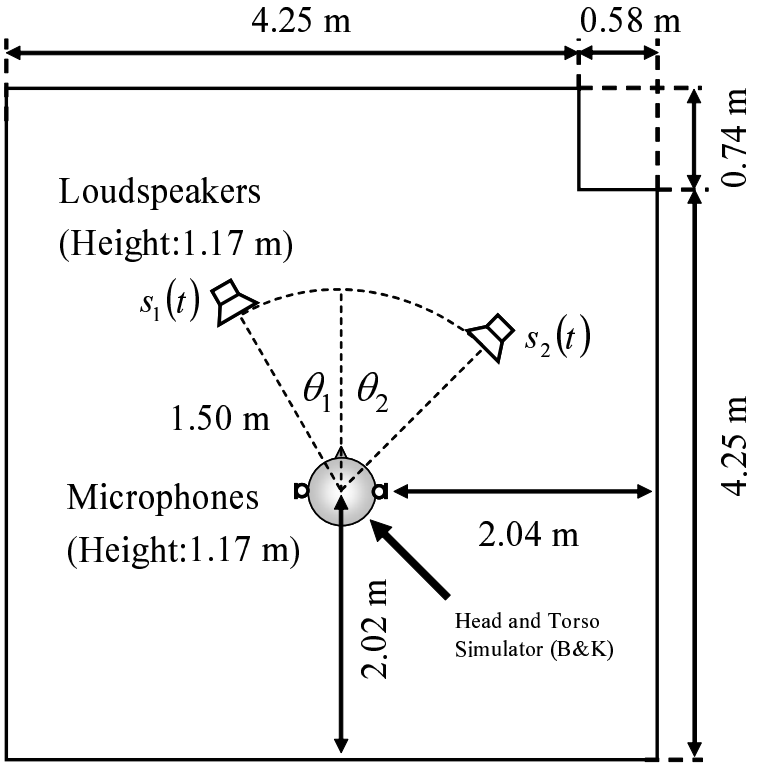

Figure 24. Layout of reverberant room used in experiments.Reverberation time of this room is 200 ms.

the SIMO-ICA-IG is narrower than that in the MDP-ICA and SIMO-ICA-LS, and consequently the parameter setting in the SIMO-ICA-IG does not depend on the source signals' properties. In addition, we can confirm an attractive feature that β in the SIMO-ICA-IG is negligible. In fact, SA in the SIMO-ICA-IG with $\beta = 1$ were almost the same as those in Figures 20 (a)–(g) unlike MDP-ICA and SIMO-ICA-LS. That is to say, the proposed SIMO-ICA-IG is a more stable and easy-to-use algorithm than the MDP-ICA and SIMO-ICA-LS.

Table 2. The range of the step-size parameter α and the balancing parameter β in the 2nd-order ICA by Parra, NH-ICA by Choi, ICA using Minimal Distortion Principle (MDP-ICA) by Matsuoka, SIMO-ICA-LS, and SIMO-ICA-IG. The step-size parameter control the convergence speed, and the balancing parameter control the balance between the cost function about separation and the cost function about sound quality

Method	step-size parameter α	balancing parameter β
2nd-order ICA	$1 \times 10^{-8} \sim 1 \times 10^{-2}$	–
NH-ICA	$5 \times 10^{-8} \sim 5 \times 10^{-6}$	–
MDP-ICA	$5 \times 10^{-8} \sim 5 \times 10^{-6}$	$1 \times 10^{-4} \sim 1 \times 10^{-1}$
SIMO-ICA-LS	$5 \times 10^{-8} \sim 5 \times 10^{-6}$	$1 \times 10^{-4} \sim 1 \times 10^{-1}$
SIMO-ICA-IG	$5 \times 10^{-8} \sim 5 \times 10^{-6}$	$1 \times 10^{-1} \sim 1 \times 10^{2}$

5.6 Experiments and results using mixed binaural signals

5.6.1 Experimental conditions

We carried out binaural-sound-separation experiments using source signals which are convolved with impulse responses recorded with a head and torso simulator (HATS) (Brüel & Kjær) in the experimental room illustrated in Figure 24. The reverberation time in this room is 200 ms. Two speech signals are assumed to arrive from different directions, θ_1 and θ_2; $\theta_1 = -30^\circ$ and $\theta_2 = 45^\circ$. The distance between the HATS and the sound source is 1.5 m. The sampling frequency is 8 kHz and the length of each speech sample is limited to 3 seconds. The length of $\boldsymbol{w}(n)$ in each method is 1024, and the initial values are given by the inverse of the HRTF matrix (see Section. 6.3) whose directions of sources, $\hat{\theta}_{\text{init1}}$ and $\hat{\theta}_{\text{init2}}$, are -60° and 60°, respectively. The step-size parameters α is 5.0×10^{-2} and 1.0×10^{-6}, respectively. SIMO-model accuracy (SA) is used as an evaluation score. In each method, the step-size parameter α and balancing parameter β are changed in the range as following table 2.

5.6.2 Experimental results and discussion on SA

Figures 25 (a)–(g) provides the results of SIMO-model accuracy for each speaker combination and average of them in 2nd-order ICA by Parra, NH-ICA by Choi, MDP-ICA by Matsuoka, SIMO-ICA-LS, SIMO-ICA-IG,and SIMO-ICA-IG ($\beta = 1$). As shown in this figure, the proposed SIMO-ICA-IG consistently outperforms the conventional ICA. In the averaged scores, compared with the conventional 2nd-order ICA, NH-ICA, MDP-ICA, and SIMO-ICA-LS, the improvement of SA in the SIMO-ICA-IG is 8.4 dB, 7.6 dB, 3.4 dB, and 4.4 dB, respectively.

5.6.3 Results and discussion on balancing parameter settings

In Figures. 26–27, the optimum balancing parameter β in MDP-ICA and SIMO-ICA-LS is dispersed in the huge range between 1.0×10^{-4} and 1.0×10^{-3}, and 1.0×10^{-5} and 5.0×10^{-4}. In Figure 28, however, the optimum balancing parameter β in SIMO-ICA-IG is within the range between 0.8 to 2.0, i.e., almost around 1. From these results, the range of balancing parameter β in SIMO-ICA-IG is narrower than that in MDP-ICA and SIMO-ICA-LS, and consequently the parameter setting in SIMO-ICA-IG does not depend on the source signals' properties. In addition, we can mention the attractive feature that β in SIMO-ICA-IG is negligible because the separation performance of SIMO-ICA-IG with $\beta = 1$ is almost the same as that of optimal SIMO-ICA-IG (see white bars in Figure 25).

Overall, the results indicate that the proposed SIMO-ICA-IG outperforms other methods, and there is no deterioration in performance of SIMO-ICA-IG even if the balancing parameter β is set to 1.

5.7 Conclusion

We propose a new single-input multiple-output (SIMO)-model-based ICA with an information-geometric learning algorithm for high-fidelity blind source separation. The SIMO-ICA-IG is an algorithm for separating mixed signals, not into monaural source signals but into SIMO-model-based signals of independent sources without loss of their spatial qualities. The SIMO-ICA-IG consists of multiple ICA parts and a fidelity controller, and all of the procedures are conducted on the basis of the information-geometry theory. Hence the proposed SIMO-

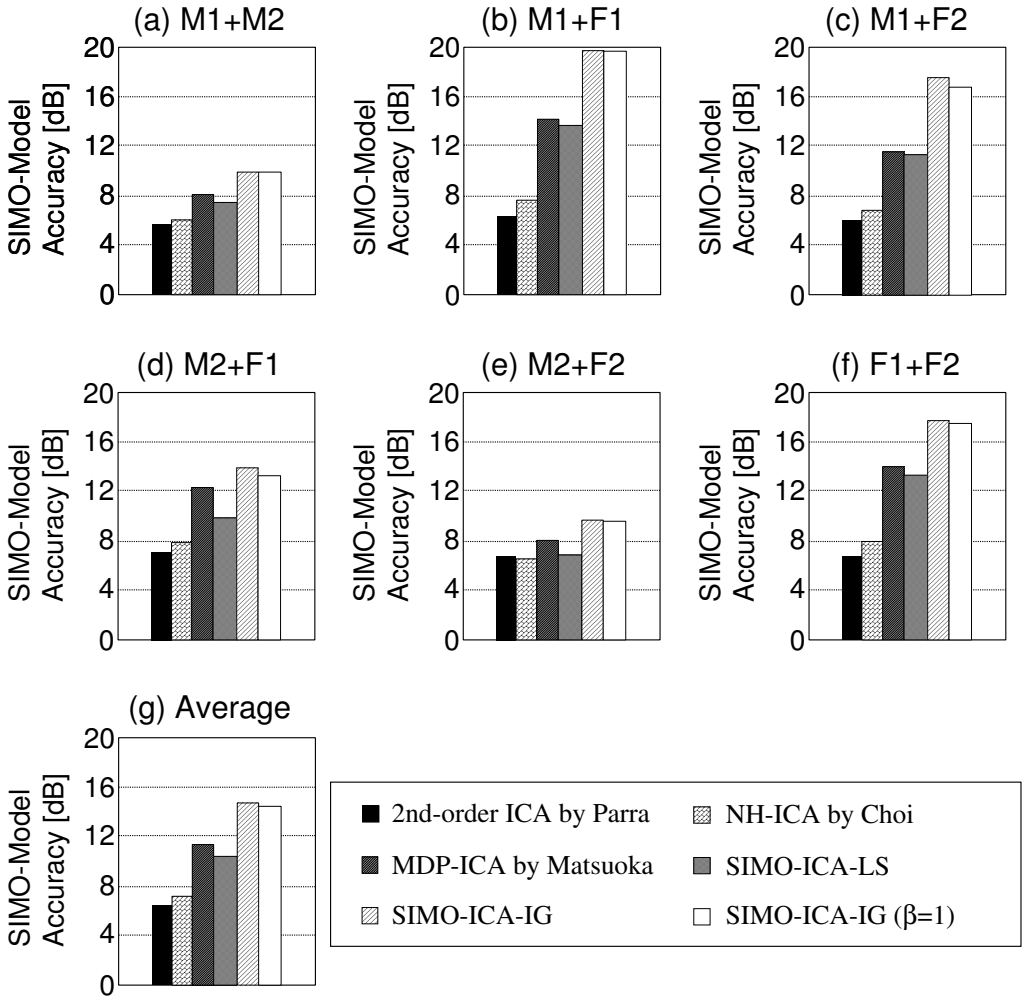

Figure 25. Results of SIMO-model accuracy in the 2nd-order ICA by Parra, NH-ICA by Choi, MDP-ICA by Matsuoka, SIMO-ICA-LS, and the proposed SIMO-ICA-IG, where the reverberation time is 200 ms. The length of the filter, D, in all methods is set to be 1024 taps. In the horizontal axis, the symbols "M1+M2"–"F1+F2" denote the combinations of speakers, e.g., "M1" and "M2" correspond to two different male speakers, and "F1" and "F2" correspond to two different female speakers. Thus, for example, "M1+F2" corresponds to the male-female combination.

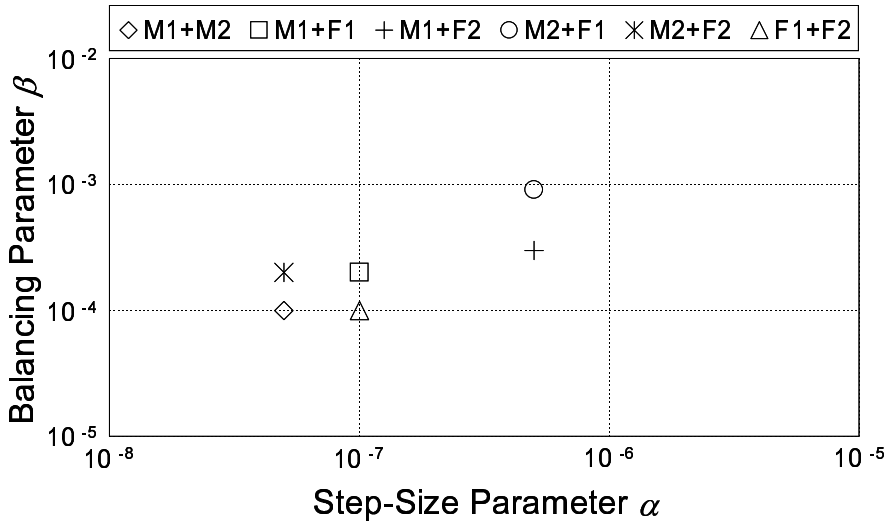

Figure 26. Optimum combination of step-size parameters α and balancing parameter β in MDP-ICA for different speaker combinations. Plotted symbols correspond to six speaker combinations as depicted in Figs. 25.

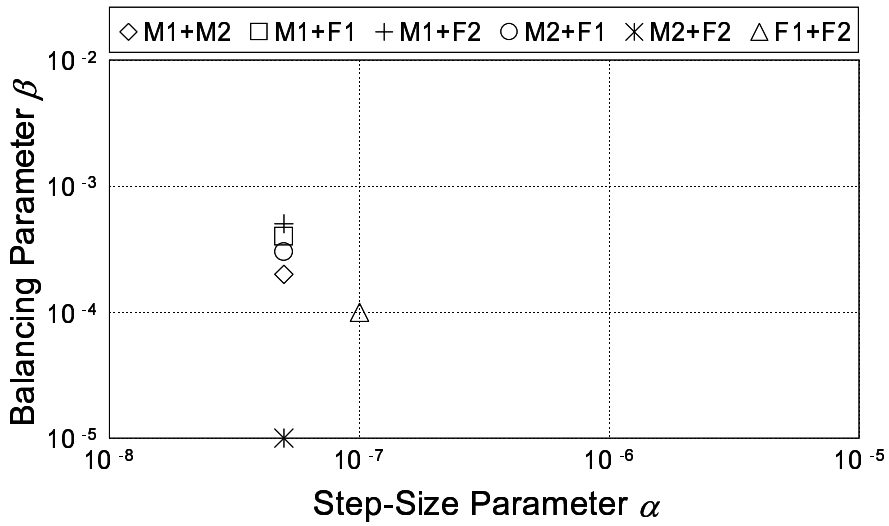

Figure 27. Optimum combination of step-size parameters α and balancing parameter β in SIMO-ICA-LS for different speaker combinations. Plotted symbols correspond to six speaker combinations as depicted in Figs. 25.

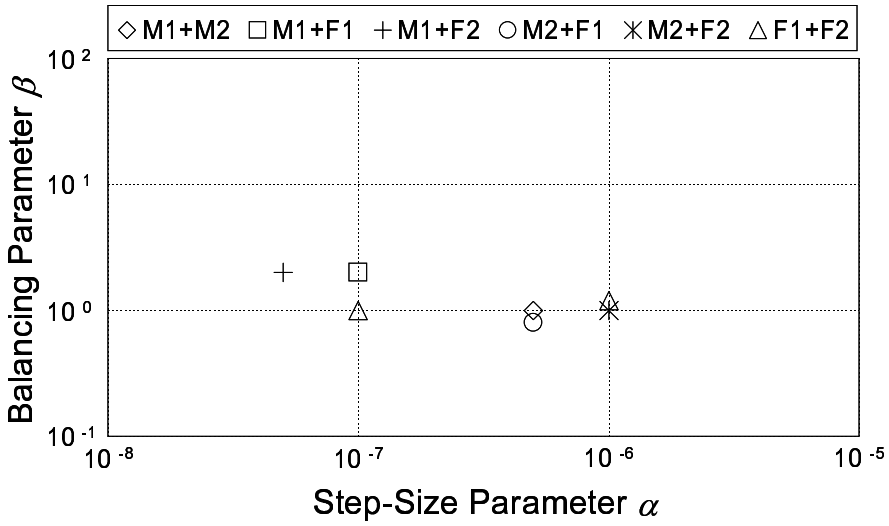

Figure 28. Optimum combination of step-size parameters α and balancing parameter β in SIMO-ICA-IG for different speaker combinations. Plotted symbols correspond to six speaker combinations as depicted in Figs. 25.

ICA-IG is free from deterioration due to mismatching among multiple criteria in ICAs and fidelity controller. In order to evaluate its effectiveness, separation experiments are carried out using two kinds of the mixed signals; one is mixed microphone array signals under a reverberant condition and the other is mixed binaural signals under another reverberant condition. The experimental results reveal the following.

- The SIMO-model accuracy of the proposed SIMO-ICA-IG is superior to those of the conventional 2nd-order ICA, NH-ICA, MDP-ICA.
- The SIMO-model accuracy of the proposed SIMO-ICA-IG is almost the same as that of SIMO-ICA-LS.
- It is confirmed that the balancing parameter setting in the proposed SIMO-ICA-IG does not depend on the source signals' properties. Although SIMO-ICA-LS is very sensitive to the balancing parameter setting, the balancing parameter in the proposed SIMO-ICA-IG can be almost negligible.

From these findings, we can conclude that the proposed SIMO-ICA-IG has great potential for application to high-fidelity signal processing systems, and has a good robustness against parameter setting.

6. Blind decomposition of mixed binaural sound using SIMO-ICA with self-generator for initial filter

6.1 Introduction

In this section, we address the blind separation problem of mixed binaural signals, and we propose a novel blind separation method, in which a self-generator for initial filters of SIMO-ICA is implemented. The original SIMO-ICA which is discussed in Section 4–5 can separate mixed signals, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Although this attractive feature of SIMO-ICA is beneficial to the binaural sound separation, the original SIMO-ICA has a serious drawback in its high sensitivity to the initial settings of the separation filter. In the proposed method, the self-generator for the initial filter (SG) functions as the preprocessor of SIMO-ICA, and thus it can provide a valid initial filter for SIMO-ICA. The SG is still a blind process because it mainly consists of a frequency-domain ICA (FDICA) part and the direction of arrival (DOA) estimation part which is driven by the separated outputs of the FDICA. To evaluate its effectiveness, binaural sound separation experiments are carried out under a reverberant condition. The experimental results reveal that the separation performance of the proposed method is superior to those of the conventional methods.

6.2 Motivation and strategy

The SIMO-ICA algorithm has the drawback of arbitrariness with respect to the initial value of the separation filter, and also that the separation performance of this method is deteriorated by an irrelevant initial value. Thus, the development of the self-generation of effective initial filters is a problem demanding attention.

Meanwhile, the transfer function in binaural mixing is roughly divided into the room transfer function and HRTF. Since the former depends on the components of the direct sound, the reflection sound, and reverberation, it is generally unknown in a blind setup. However, the latter depends on the only DOAs of

sources and can be previously known because HRTF is an inherent feature in the recording apparatus itself and can be approximately measured by using, for example, a dummy head. Thus, HRTF is an important factor which solves the blind separation problem of mixed binaural signals, and we can use HRTF as the valid initial value in the separation filter matrix if the DOAs of sources can be known in advance.

These findings motivated us to combine FDICA with projection back (FDICA-PB) and DOA estimation as the self-generation system for the initial value of SIMO-ICA. First, we set the initial value of the separation filter in FDICA-PB using arbitrary DOAs of sources, and perform the FDICA-PB to separate the observed signals to certain extent. After FDICA-PB, we estimate the DOAs of sources on the basis of the output of FDICA-PB *blindly*. Then, this generator resets the valid initial value of the separation filter on the basis of the estimated DOAs of sources, and re-optimizes the filter via FDICA-PB. In this procedure, a matrix bank of previously measured HRTFs for multiple DOAs is supplied to generate the valid initial filter. The important methodology required here is an accurate DOA estimation of sources. The SIMO-output algorithm offers the advantage that the output signals maintain the spatial qualities of each source. Thus, the output signals of FDICA-PB, which is one of the SIMO-output BSS, are synchronized with the observed signals, i.e., the time alignment has been taken. We can detect the single talk segments in the observed signals by looking at the SIMO-output signals of FDICA-PB, and estimate the DOAs of sources by using the specific observed signals' durations correspond to these segments.

6.3 Algorithm

The proposed algorithm is conducted with the following steps.

[**Step 0: Early Initialization**] Set DOAs of sources $\hat{\theta}_i$ to early initial (arbitrary) values, $\hat{\theta}_{\mathrm{init}i}$.

[**Step 1: HRTF Matrix Bank**] The HRTF matrix bank consists of multiple HRTF matrices. The single HRTF matrix for θ_1 and θ_2 is given as

$$\boldsymbol{H}(\theta_1, \theta_2, f) = \begin{bmatrix} H_L(\theta_1, f) & H_L(\theta_2, f) \\ H_R(\theta_1, f) & H_R(\theta_2, f) \end{bmatrix}, \tag{94}$$

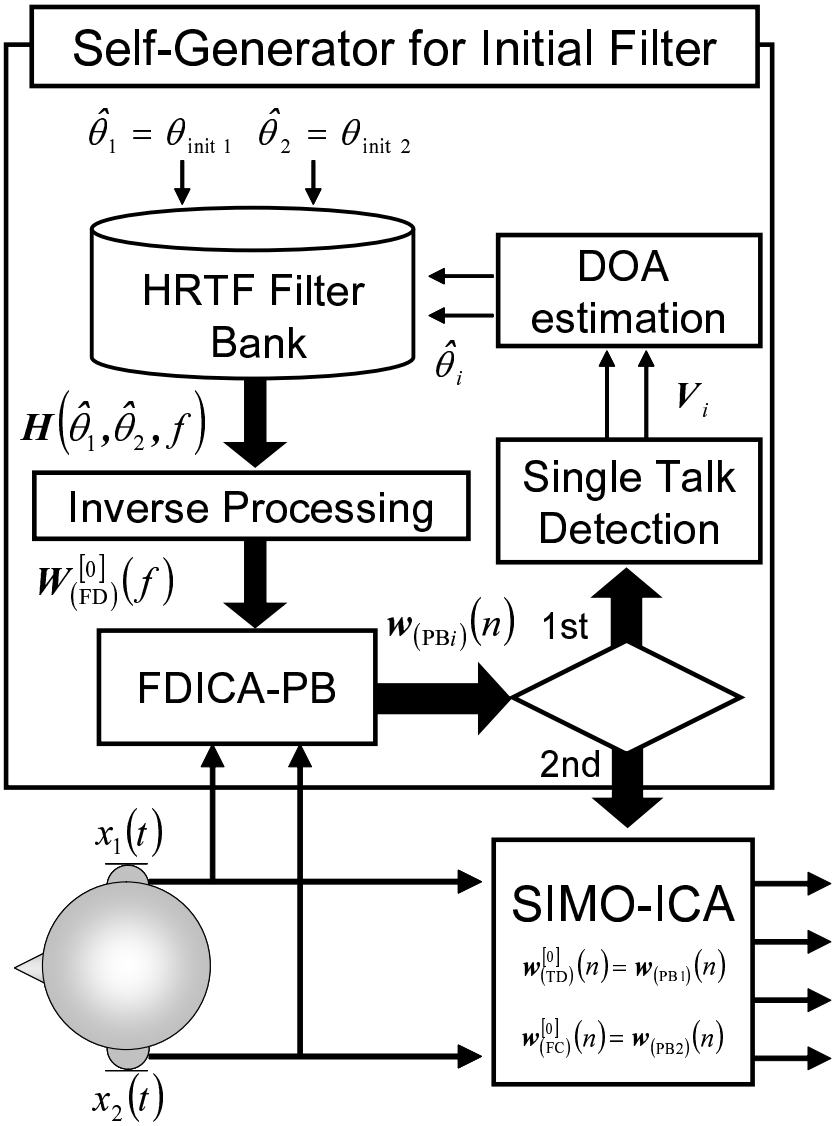

Figure 29. Configuration of SIMO-ICA with self-generator for initial filer.

where $H_L(\theta, f)$ is the HRTF between the left ear and the source whose direction is θ, and $H_R(\theta, f)$ is that between the right ear and the same source. To construct the HRTF matrix bank, we prepare the multiple HRTF matrices in advance by changing θ_1 and θ_2. Using the HRTF matrix bank and the DOAs of sources, we can automatically generate the valid initial value for FDICA as follows:

$$\boldsymbol{W}_{(\mathrm{FD})}^{[0]}(f) \;=\; \boldsymbol{H}^{-1}(\hat{\theta}_1, \hat{\theta}_2, f). \tag{95}$$

Note that the initial value is not an optimal separation filter matrix under a reverberant condition, but the separation filter matrix can be finally optimized through ICA iterations.

[**Step 2: FDICA-PB [14]**] Murata et al. have proposed an FDICA-PB method which can estimate the SIMO components of the observed signals on the basis of the monaural outputs of FDICA. In this method, first, the separation filter matrix $\boldsymbol{W}_{(\mathrm{FD})}(f)$ in the frequency domain is optimized to the separate source signals to obtain the monaural signals. The separated signals $\boldsymbol{Y}_{(\mathrm{FD})}(f,t)$ in the time-frequency domain are expressed as

$$\boldsymbol{Y}_{(\mathrm{FD})}(f,t) \;=\; \boldsymbol{W}_{(\mathrm{FD})}(f)\boldsymbol{X}(f,t), \tag{96}$$

where $\boldsymbol{X}(f,t)$ is the observed signal vector which is calculated by means of a frame-by-frame discrete Fourier transform (DFT). The iterative learning algorithm is expressed as

$$\boldsymbol{W}_{(\mathrm{FD})}^{[i+1]}(f) \;=\; \boldsymbol{W}_{(\mathrm{FD})}^{[i]}(f) + \eta\left\{\boldsymbol{I} - \left\langle \Phi(\boldsymbol{Y}_{(\mathrm{FD})}^{[i]}(f,t))\boldsymbol{Y}_{(\mathrm{FD})}^{[i]}(f,t)^{\mathrm{H}}\right\rangle_t\right\}\boldsymbol{W}_{(\mathrm{FD})}^{[i]}(f), \tag{97}$$

where the initial value of $\boldsymbol{W}_{(\mathrm{FD})}(f)$ is given by Eq. (95). However, the output signals of FDICA given by Eq. (96) are monaural signals with respect to the sound sources, not SIMO-model-based signals. Thus, using the following equations, we project the separation filter onto the SIMO separation filters.

$$\boldsymbol{w}_{(\mathrm{PB1})}(n) \;=\; \mathrm{IDFT}\Big[\mathrm{diag}\left\{\boldsymbol{W}_{(\mathrm{FD})}^{-1}(f)\right\}\boldsymbol{W}_{(\mathrm{FD})}(f)\Big], \tag{98}$$

$$\boldsymbol{w}_{(\mathrm{PB2})}(n) \;=\; \mathrm{IDFT}\Big[\text{off-diag}\left\{\boldsymbol{W}_{(\mathrm{FD})}^{-1}(f)\right\}\boldsymbol{W}_{(\mathrm{FD})}(f)\Big], \tag{99}$$

where IDFT$[\cdot]$ represents an inverse DFT with the circular time shift of the $D/2$ samples. The separated signals of FDICA-PB in the time domain are expressed

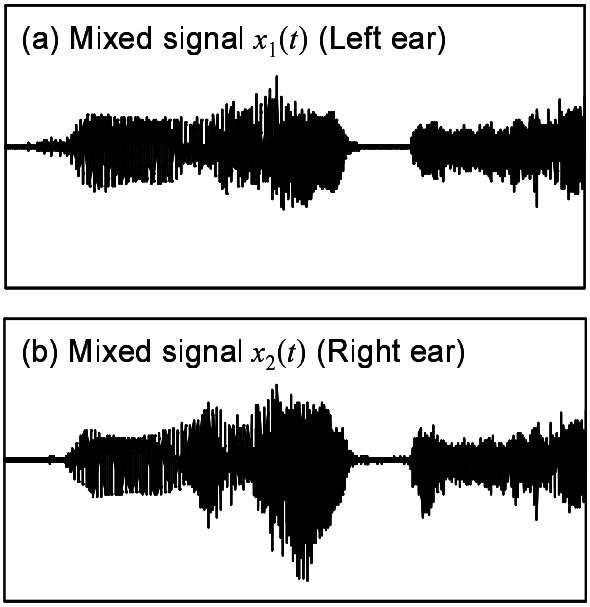

Figure 30. Example of the observed signals recorded at the microphones. From these signals, it is difficult to estimate the single talk segments.

as

$$\begin{aligned} \boldsymbol{y}_{(\mathrm{PB}i)}(t) &= [y_k^{(\mathrm{PB}i)}(t)]_k & (100) \\ &= \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{PB}i)}(n)\boldsymbol{x}(t-n) & (101) \\ &= \boldsymbol{W}_{(\mathrm{PB}i)}(z)\boldsymbol{x}(t). & (102) \end{aligned}$$

$\boldsymbol{y}_{(\mathrm{PB}i)}(t)$ is a good approximation of the SIMO solution in Eq. (37) without permutation

[**Step 3: Single Talk Detection**] In order to detect the single talk segments of the observed signals, we divide the observed signals and output signals of FDICA-PB into multiple frames. Each frame of these signals is expressed as

$$\begin{aligned} \boldsymbol{x}(u,v) &= \boldsymbol{x}(u+(v-1)\times U), & (103) \\ \boldsymbol{y}_{(\mathrm{PB}i)}(u,v) &= \boldsymbol{y}_{(\mathrm{PB}i)}(u+(v-1)\times U), & (104) \end{aligned}$$

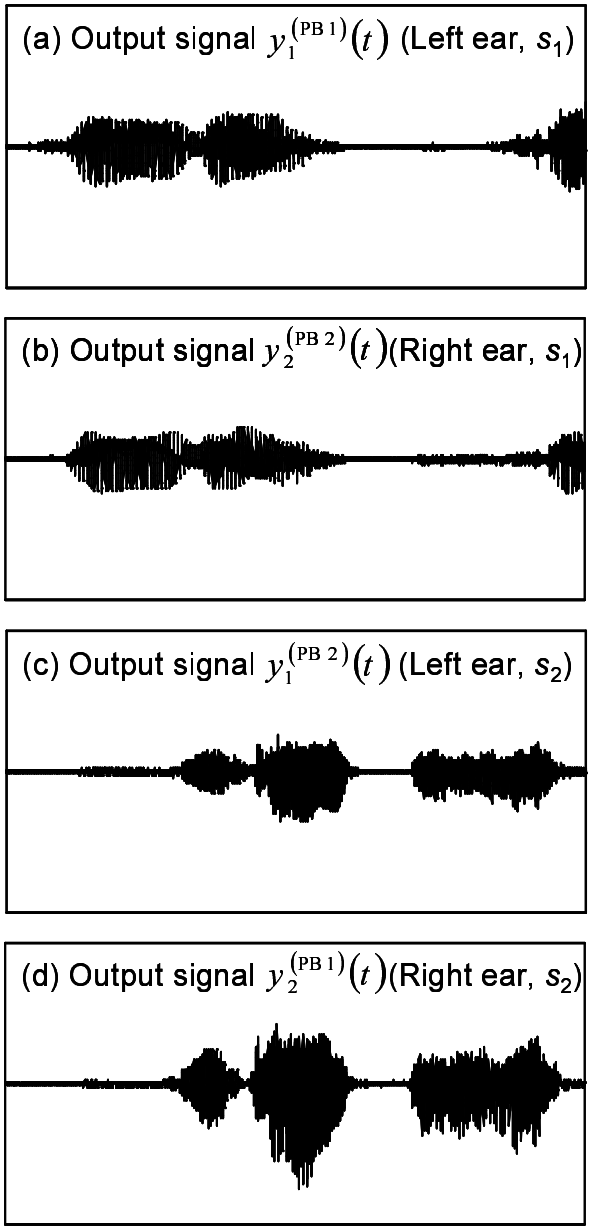

Figure 31. Example of output signals in the FDICA-PB. The output signals (a) and (b) is the estimated binaural signals from the source 1 at left and right ear, and the output signals (c) and (d) is the estimated binaural signals from the source 2 at left and right ear. From these signals, it is easy to estimate the single talk segments of the observed signals.

where u is the time index in a frame, U is the number of samples in a frame, v is the frame index. Each single talk segment $\boldsymbol{V}_i$ of the observed signals is detected on the basis of the following criteria:

$$\boldsymbol{V}_1 = \left\{ v | Q_1^{(\mathrm{PB1})}(v) > T; Q_2^{(\mathrm{PB2})}(v) > T; Q_2^{(\mathrm{PB1})}(v) < T; Q_1^{(\mathrm{PB2})}(v) < T \right\}, \tag{105}$$

$$\boldsymbol{V}_2 = \left\{ v | Q_1^{(\mathrm{PB1})}(v) < T; Q_2^{(\mathrm{PB2})}(v) < T; Q_2^{(\mathrm{PB1})}(v) > T; Q_1^{(\mathrm{PB2})}(v) > T \right\}, \tag{106}$$

$$Q_k^{(\mathrm{PB}i)}(v) = 10 \log_{10} \frac{\sum_{u=1}^{U} |y_k^{(\mathrm{PB}i)}(u,v)|^2}{\max\limits_{v} \left\{ \sum_{u=1}^{U} |y_k^{(\mathrm{PB}i)}(u,v)|^2 \right\}}, \tag{107}$$

where T is a threshold which is experimentally determined.

[**Step 4: DOA Estimation Using Single Talk Segments**] We can obtain the DOAs $\hat{\theta}_i$ of sources by using the single talk segments. The estimated angle $\hat{\theta}_i$ is given as

$$\hat{\theta}_i = \arg\max_{\theta} \left\{ \left\langle \sum_f X_1(f,v) X_2(f,v)^{\mathrm{H}} e^{-\frac{j2\pi f d \sin\theta}{c}} \right\rangle_{v \in V_i} \right\}, \tag{108}$$

where $\langle \cdot \rangle_{v \in V}$ is the frame-averaging operator which is composed of elements v in single talk segments V. Thus, we can obtain the valid initial value of the separation filter matrix using these estimated values, $\hat{\theta}_1$ and $\hat{\theta}_2$.

[**Step 5: Re-Optimization**] Using the DOAs of the sources estimated with Eq. (108), we reset the initial value of the separation filter, and re-optimize that in FDICA-PB (execute [**Steps 1 and 2**] again).

[**Step 6: SIMO-ICA**] Optimize the separation filter matrices $\boldsymbol{w}_{(\mathrm{ICA})}(n)$ and $\boldsymbol{w}_{(\mathrm{FC})}(n)$ in the time domain, by using Eqs. (91)–Eqs. (93) to enhance the target components further. The separation filter matrices, Eqs. (98) and (99), are used as the initial values of the separation filter matrices $\boldsymbol{w}_{(\mathrm{ICA})}(n)$ and $\boldsymbol{w}_{(\mathrm{FC})}(n)$ in SIMO-ICA, i.e.,

$$\boldsymbol{w}_{(\mathrm{ICA})}^{[0]}(n) = \boldsymbol{w}_{(\mathrm{PB1})}(n) \tag{109}$$

$$\boldsymbol{w}_{(\mathrm{FC})}^{[0]}(n) = \boldsymbol{w}_{(\mathrm{PB2})}(n) \tag{110}$$

If the early initialization, HRTF matrix bank, FDICA-PB, and SIMO-ICA ([**Step 0–2, 6**]) are executed without single talk detection, DOA estimation,

Table 3. The number of iterations in SIMO-ICA-IG, FDICA-PB, MS-SIMO-ICA, proposed SIMO-ICA with self-generator for initial filter

	FDICA part	TDICA part
SIMO-ICA-IG	–	5000
FDICA-PB	1000	–
MS-SIMO-ICA	1000	100
Proposed SIMO-ICA with self-generator for initial filter	500 (in **Step 2**) + 500 (in **Step 5**)	100

nor re-optimization ([**Step 3**–**5**]), this algorithm corresponds to the multistage SIMO-ICA (MS-SIMO-ICA) algorithm [32] which has previously been proposed by one of the present authors. The difference between the proposed method and MS-SIMO-ICA will be discussed in the next section.

6.4 Experiments and results

6.4.1 Experimental conditions

We carried out binaural-sound-separation experiments using source signals which are convolved with impulse responses recorded with a head and torso simulator (HATS) (Brüel & Kjær) in the experimental room illustrated in Figure 24. Two speech signals are assumed to arrive from different directions, θ_1 and θ_2; $\theta_1 = \{-90^\circ, -75^\circ, -60^\circ, -45^\circ, -30^\circ, -15^\circ, 0^\circ\}$ and $\theta_2 = \{0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ, 90^\circ\}$. The early initial values are given by the inverse of the HRTF matrix whose directions of sources, $\hat{\theta}_{\text{init1}}$ and $\hat{\theta}_{\text{init2}}$, are -60° and 60°, respectively. The step-size parameters α and η are 5.0×10^{-2} and 1.0×10^{-6}, respectively. SIMO-model accuracy (SA) is used as an evaluation score.

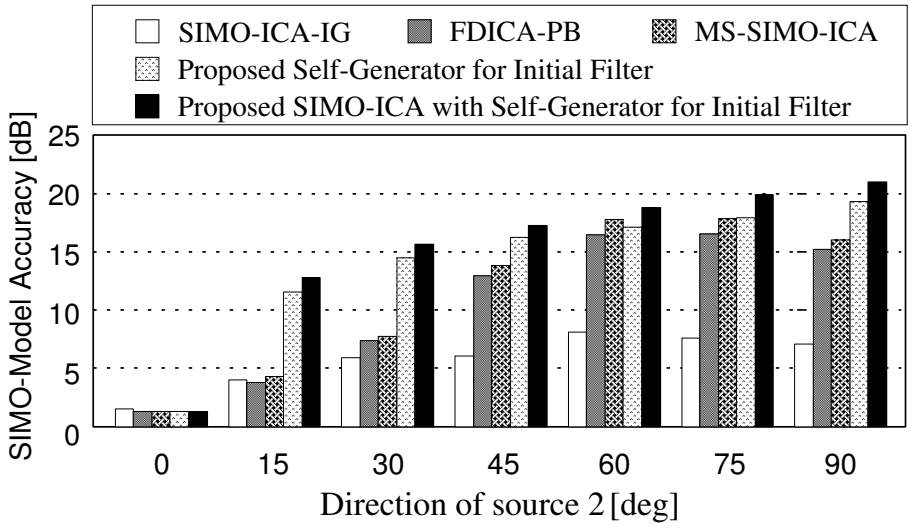

Figure 32. Experimental results of SIMO-model accuracy in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self-generator for two-speech mixture separation, where the direction of sound source 1 is 0°.

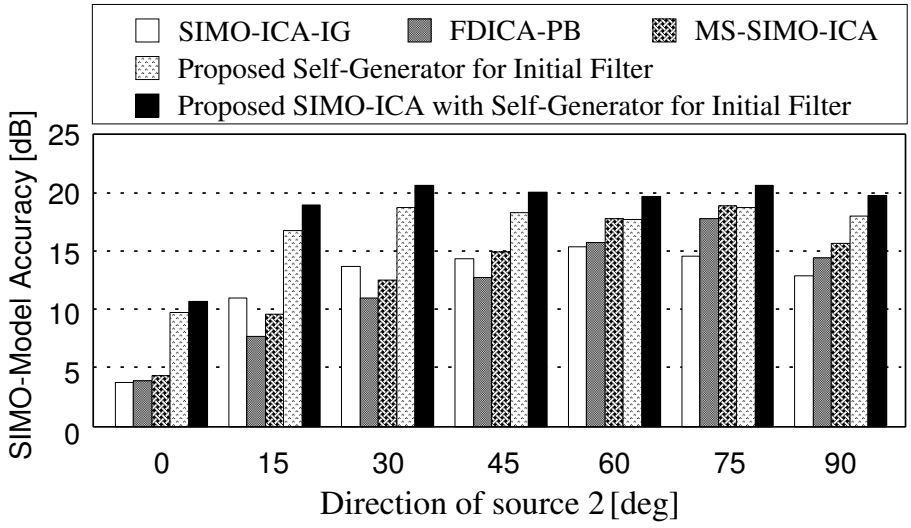

Figure 33. Experimental results of SIMO-model accuracy in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self-generator for two-speech mixture separation, where the direction of sound source 1 is $-15°$.

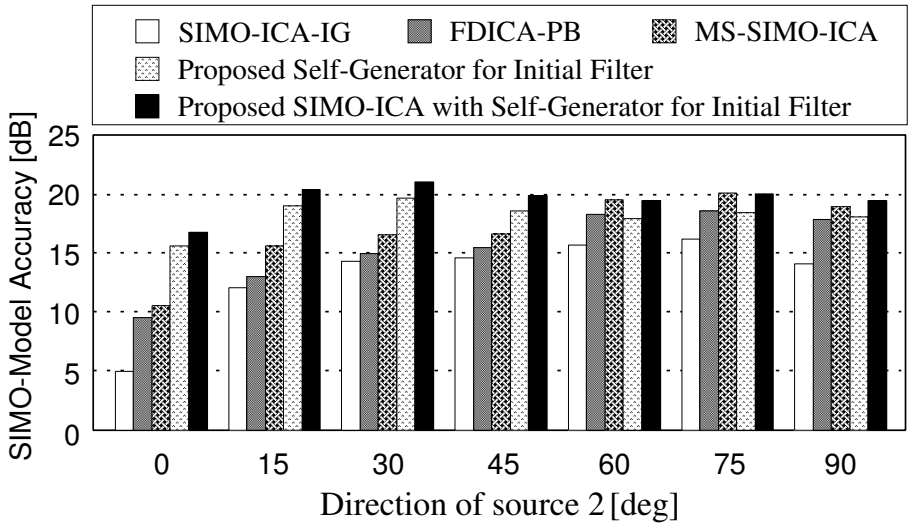

Figure 34. Experimental results of SIMO-model accuracy in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self-generator for two-speech mixture separation, where the direction of sound source 1 is $-30°$.

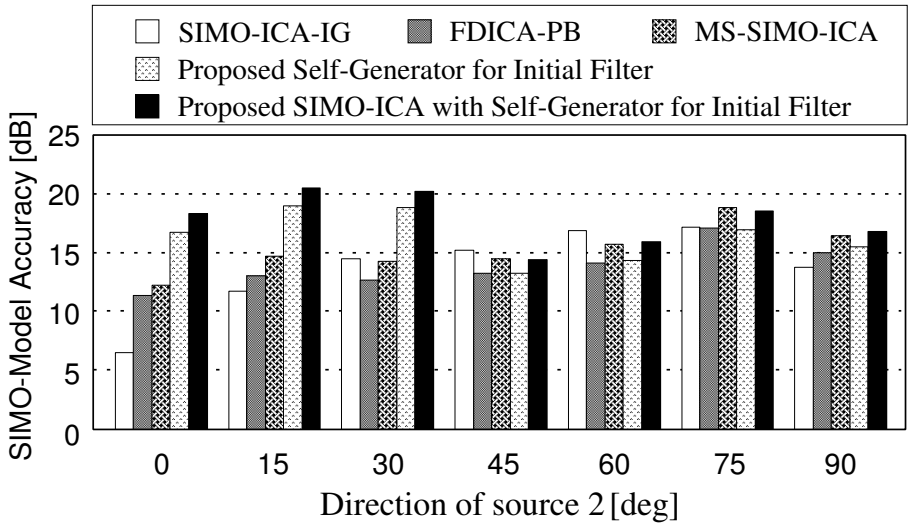

Figure 35. Experimental results of SIMO-model accuracy in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self-generator for two-speech mixture separation, where the direction of sound source 1 is $-45°$.

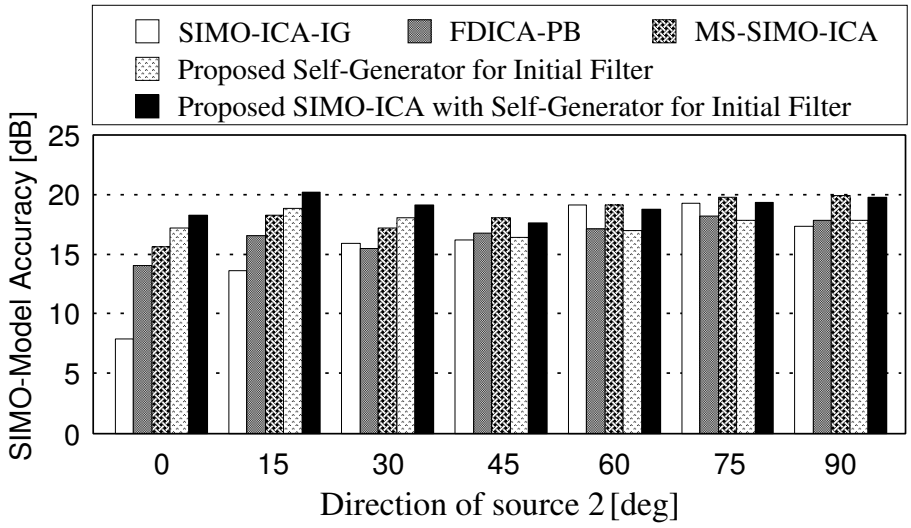

Figure 36. Experimental results of SIMO-model accuracy in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self-generator for two-speech mixture separation, where the direction of sound source 1 is $-60°$.

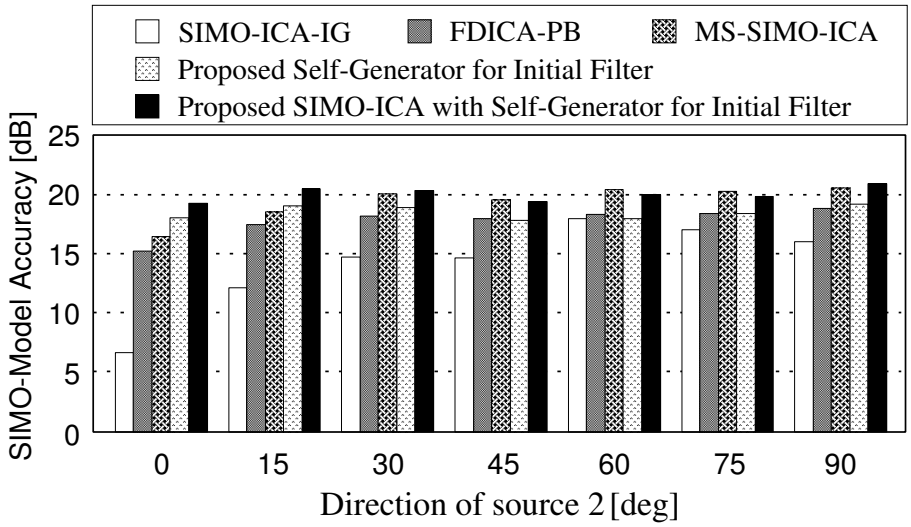

Figure 37. Experimental results of SIMO-model accuracy in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self-generator for two-speech mixture separation, where the direction of sound source 1 is $-75°$n.

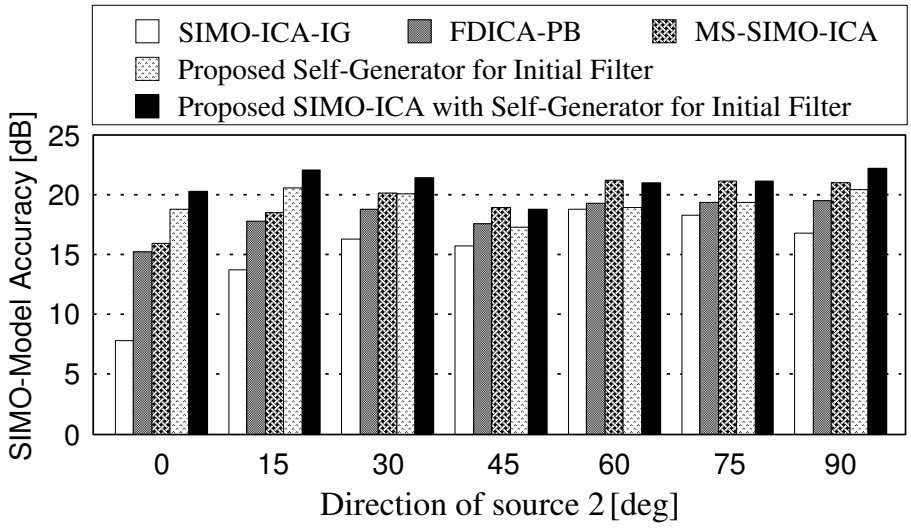

Figure 38. Experimental results of SIMO-model accuracy in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self-generator for two-speech mixture separation, where the direction of sound source 1 is $-90°$.

6.4.2 Experiments for separation of two speech mixtures

We used the speech signals spoken by two male and two female speakers as nonstationary source samples. Using these sentences, we obtain 12 combinations. With respect to the conventional ICA, for comparison, we use SIMO-ICA, FDICA-PB, and MS-SIMO-ICA. The number of iterations in each method is listed in Table 3; these are sufficient for filter updating in ICA.

Figures 32–38 show the results of SA for different θ_1 and θ_2. These are the averaged scores for all speaker combinations. The following points are revealed.

- When θ_2 is smaller than $45°$, separation performances in the proposed SIMO-ICA with self-generator for the initial filter are superior to those in the conventional methods regardless of θ_1, except for the trivial case of $\theta_1 = \theta_2 = 0$.

- The performances of the proposed SIMO-ICA with the self-generator can be remarkably improved, particularly when the angle between the speakers

is narrow, for example, $\theta_1 = 0^\circ \sim -45^\circ$ and $\theta_2 = 0^\circ \sim 30^\circ$. Also, the performances in the self-generator itself are superior to those of FDICA-PB. This means that the self-generator can contribute to the improvement of performances in the whole separation system.

- When θ_2 is larger than 45°, the separation performances of the proposed SIMO-ICA with the self-generator are almost the same as those of the conventional method.

DOA estimation for the separated sources is carried out in each method to evaluate the spatial qualities directly. The DOAs of the sources are detected by using a cross-correlation-based DOA estimation technique which is applied to the extracted SIMO-model-based signals. Figures 39–41 show the errors of DOA estimation for different θ_1 and θ_2; $\theta_1 = \{-30^\circ, -15^\circ, 0^\circ\}$ and $\theta_2 = \{0^\circ, 15^\circ, 30^\circ\}$. The scores are the averages in terms of all speaker combinations. These results indicate that the errors of the DOA estimation in outputs of the proposed self-generator is less than those of FDICA-PB. Indeed, the averaged error of DOA estimation using single talk segments (in **Step. 4**) is 0.6 degree. This finding is well consistent with the result of SA.

Overall it can be asserted that the proposed algorithm for the self-generation of initial values works effectively and increases the SIMO-ICA's separation performance.

6.4.3 Experiments for separation of speech and noise mixtures

In order to evaluate the effectiveness of the proposed method with respect to various source conditions, we carried out separation experiments using the binaural mixing of speech and human-speech-like noise (HSLN) [40]. HSLN is a kind of *bubble noise* and is generated by superposing independent speech signals. By changing the number of superpositions, we can simulate various noise conditions. When the number of superpositions is set to be around 10, HSLN becomes a nonstationary signal which sounds like bubble noise. When the number of superpositions is set to be more than 100, HSLN results in colored stationary noise while preserving the long-time spectrum of human speech. In this experiment, we set the numbers of superpositions to 8 (bubble noise) and 128 (stationary noise),

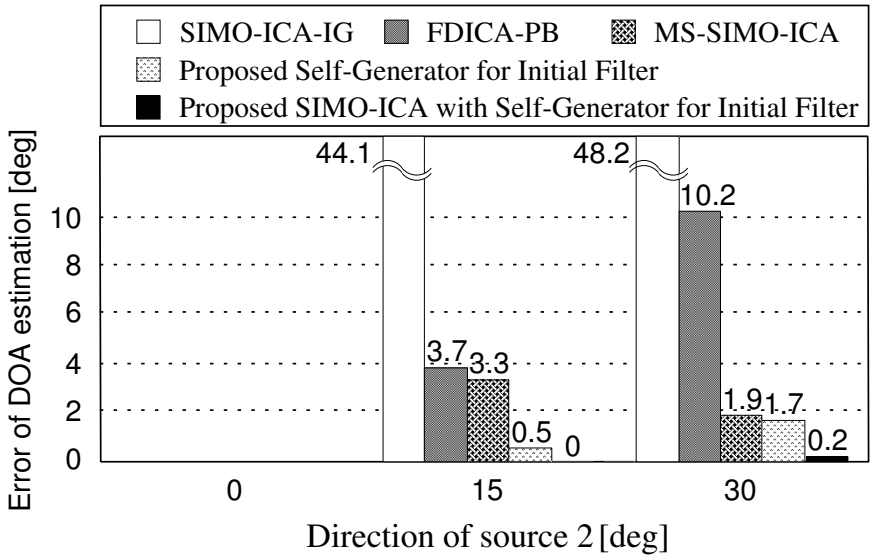

Figure 39. Error of DOA estimation in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self generator for two-speech mixture separation, where the direction of sound source 1 is 0°.

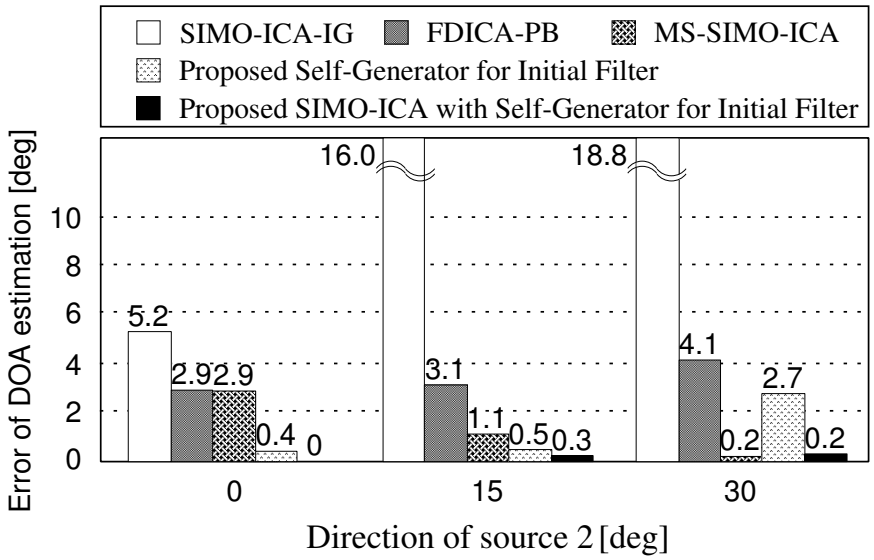

Figure 40. Error of DOA estimation in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self generator for two-speech mixture separation, where the direction of sound source 1 is $-15°$.

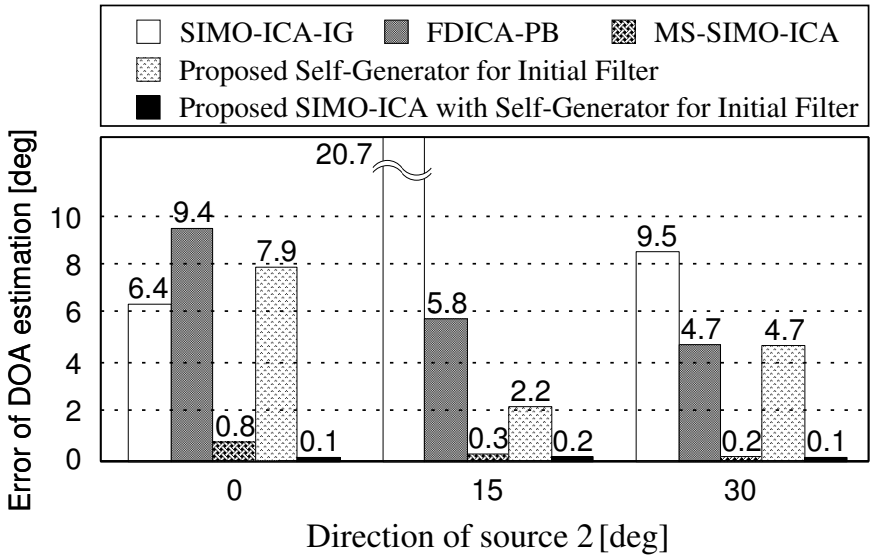

Figure 41. Error of DOA estimation in SIMO-ICA, FDICA-PB, MS-SIMO-ICA, the proposed self-generator for the initial filter, and the proposed SIMO-ICA with the self generator for two-speech mixture separation, where the direction of sound source 1 is $-30°$.

and we obtain the eight combinations for each type of noise.

Figures 42–48 and 49–55 show the results of SA for the conventional MS-SIMO-ICA and the proposed method, where the numbers of superpositions in HSLN are 8 and 128. These figures show the averaged scores of eight combinations. From Figure 42–48 (bubble noise case), if θ_1 is larger than $-30°$ and simultaneously θ_2 is smaller than $30°$, the separation performances in the proposed method are superior to those in the conventional method, except for the trivial case of $\theta_1 = \theta_2 = 0$. Otherwise, the separation performances of the proposed method are almost the same as those of the conventional method. Meanwhile, from Figure 49–55 (stationary noise case), the separation performances of the proposed method are almost the same as those of the conventional method under all conditions of θ_1 and θ_2. These results indicate that the proposed self-initial-value generator is still applicable to and effective in the nonstationary noise reduction, where single-talk detection for the noise is likely to be successful. As for the stationary noise reduction, there are no obvious improvements in the proposed

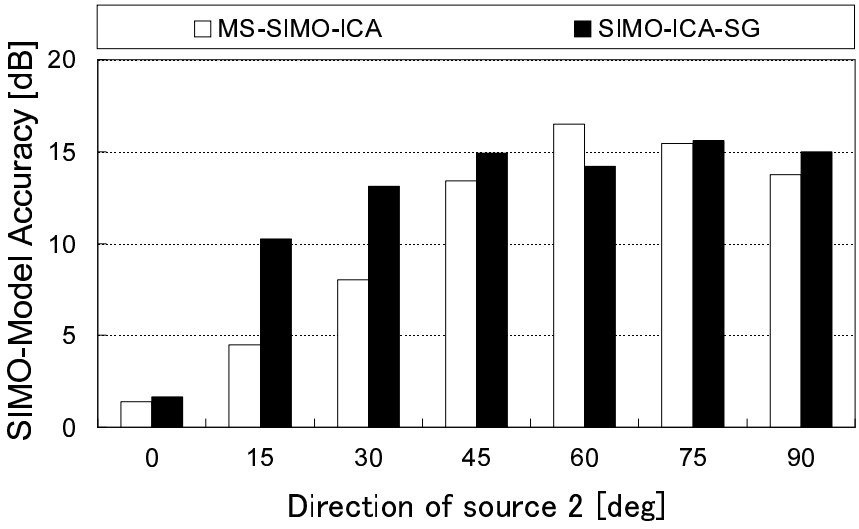

Figure 42. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is 0°. The source signals are speech and HSLN. The number of superpositions in HSLN is 8 (bubble noise case).

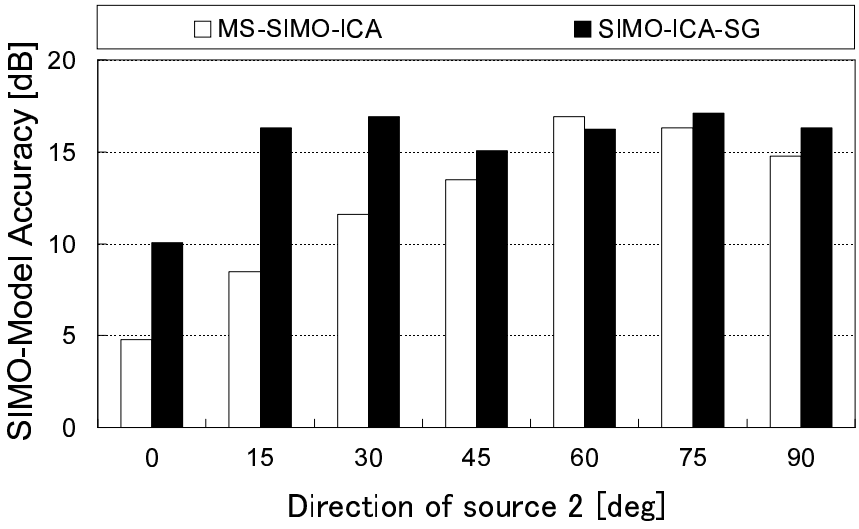

Figure 43. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is −15°. The source signals are speech and HSLN. The number of superpositions in HSLN is 8 (bubble noise case).

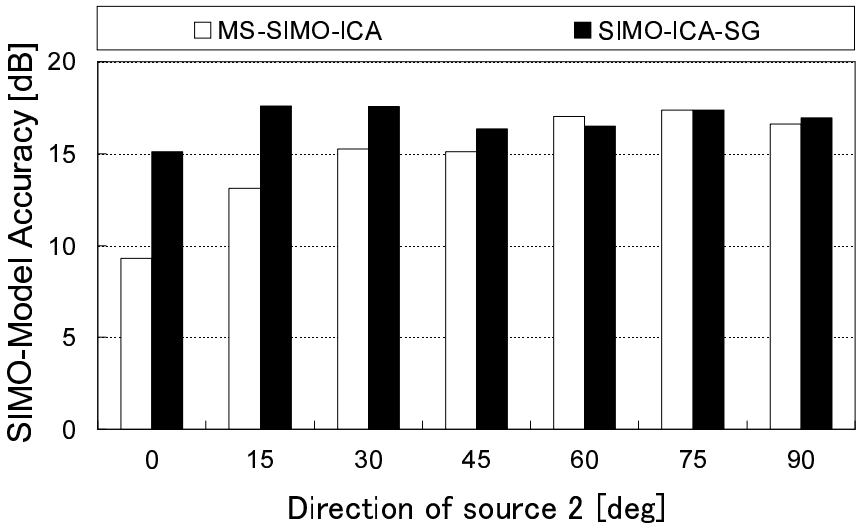

Figure 44. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is $-30°$. The source signals are speech and HSLN. The number of superpositions in HSLN is 8 (bubble noise case).

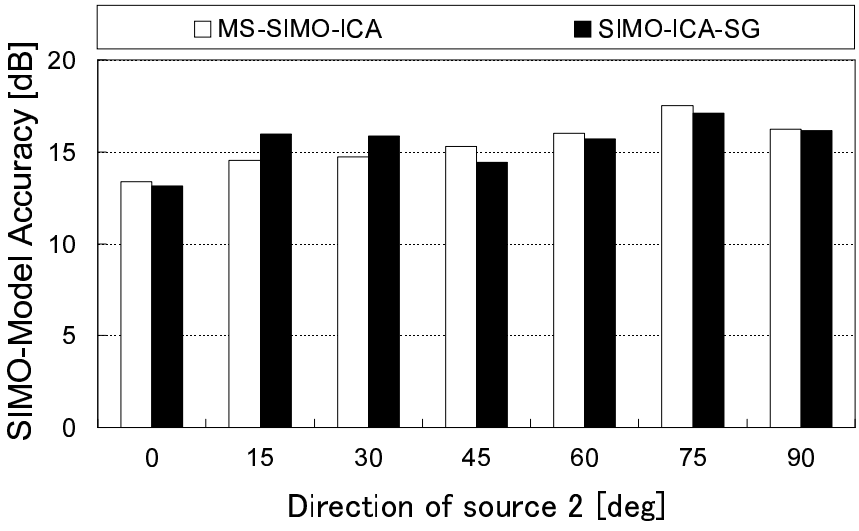

Figure 45. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is $-45°$. The source signals are speech and HSLN. The number of superpositions in HSLN is 8 (bubble noise case).

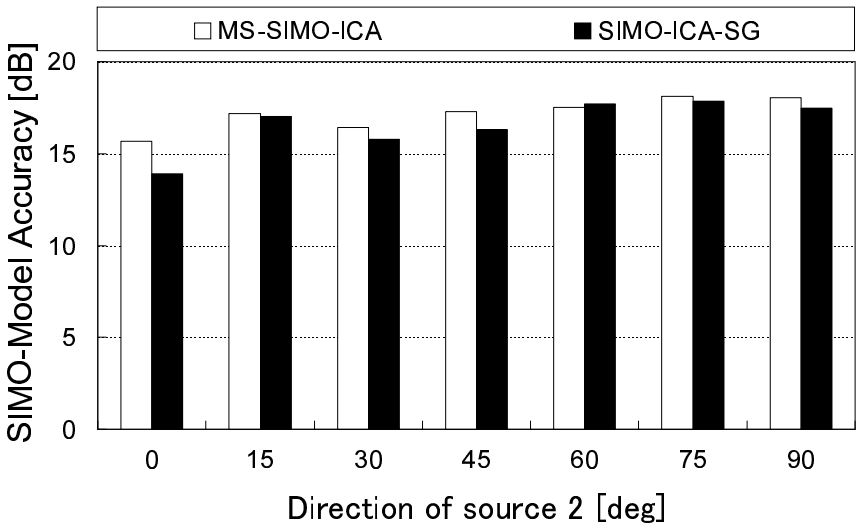

Figure 46. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is $-60°$. The source signals are speech and HSLN. The number of superpositions in HSLN is 8 (bubble noise case).

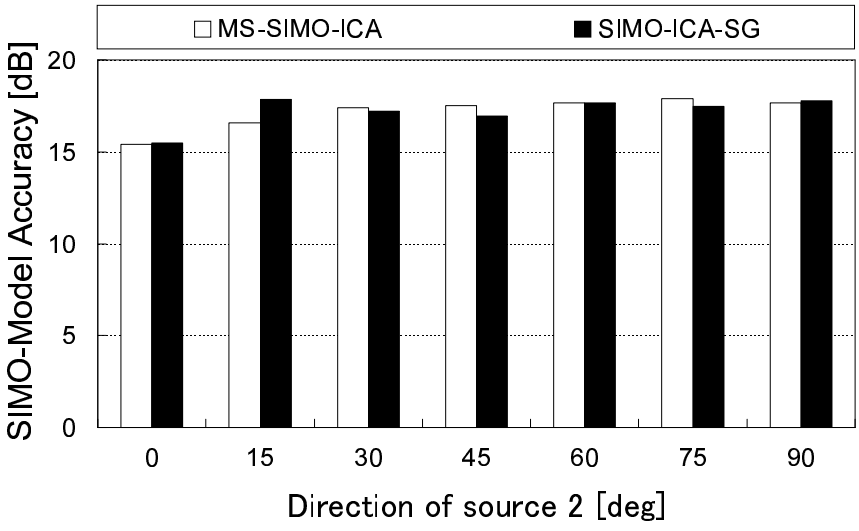

Figure 47. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is $-75°$. The source signals are speech and HSLN. The number of superpositions in HSLN is 8 (bubble noise case).

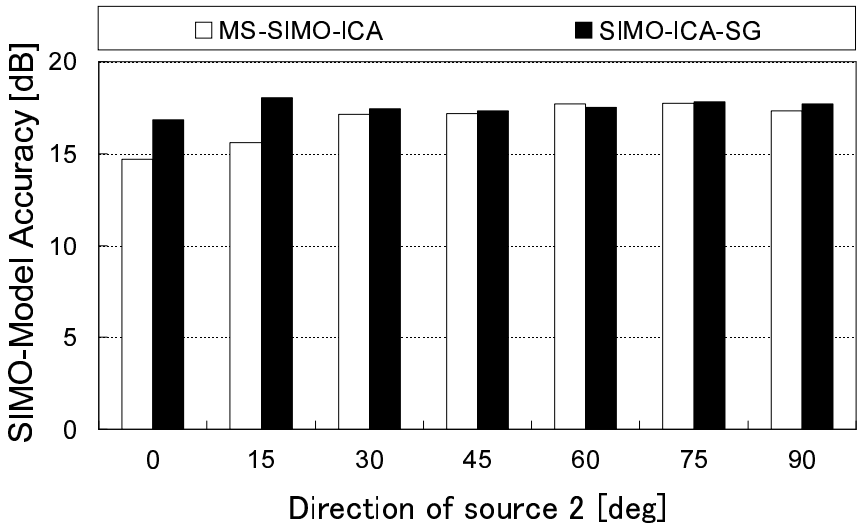

Figure 48. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is $-90°$. The source signals are speech and HSLN. The number of superpositions in HSLN is 8 (bubble noise case).

method, but also no deterioration; this means that the proposed method has no serious side-effects. In summary, the proposed idea of the self-generation for initial values is widely beneficial for application to various sounds.

6.4.4 Robustness against HRTF mismatch

Here we evaluate the robustness of the proposed method against *mismatch* between the mixing process and the HRTF matrix bank in the proposed system. Our measured transfer function which contains both HRTF of the HATS and the room reverberation (see Figure 3) is used as the mixing system in this experiment, like the previous experiments. With respect to the HRTFs used in Step 2 of the proposed method, we replace the previous HRTF database measured by ourselves (hereafter we call this the *NAIST HRTF database*) with the alternative *MIT HRTF database* [41] which is recorded with KEMAR dummy head. The NAIST HRTF database is matched with the mixing system, but the MIT HRTF database is a mismatched one.

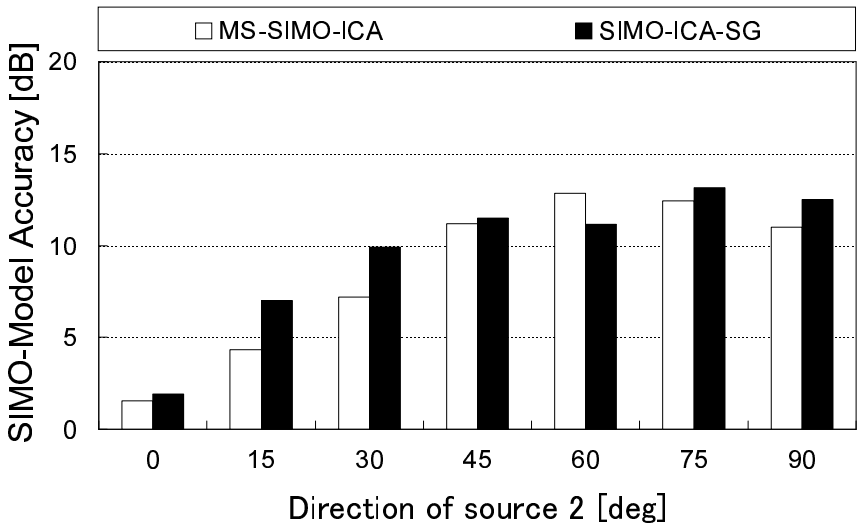

Figure 49. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is 0°. The source signals are speech and HSLN. The number of superpositions in HSLN is 128 (stationary noise case).

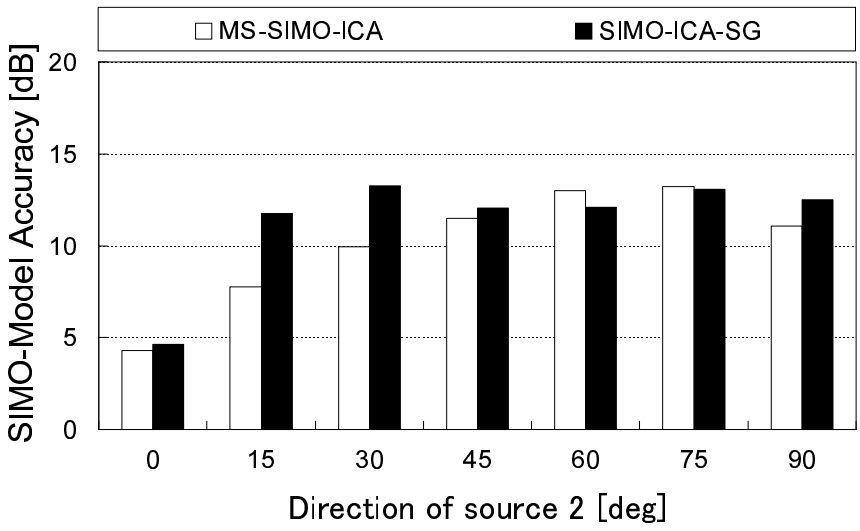

Figure 50. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is 0°. The source signals are speech and HSLN. The number of superpositions in HSLN is 128 (stationary noise case).

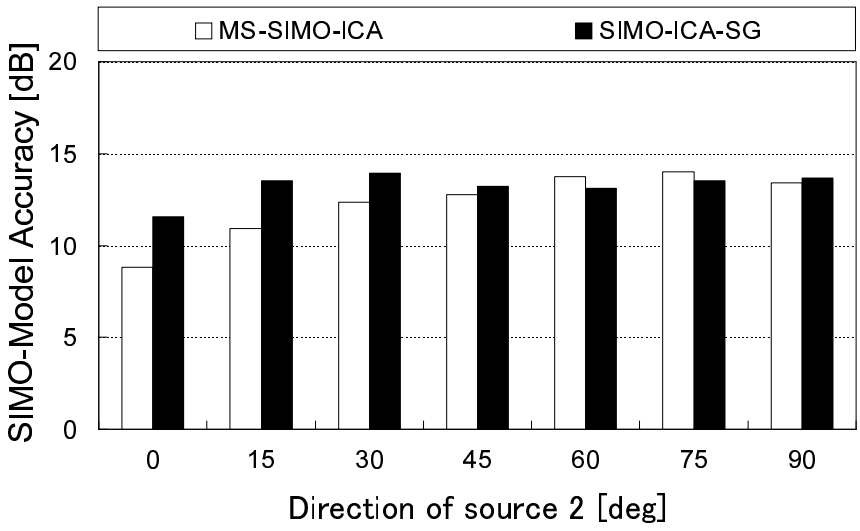

Figure 51. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is 0°. The source signals are speech and HSLN. The number of superpositions in HSLN is 128 (stationary noise case).

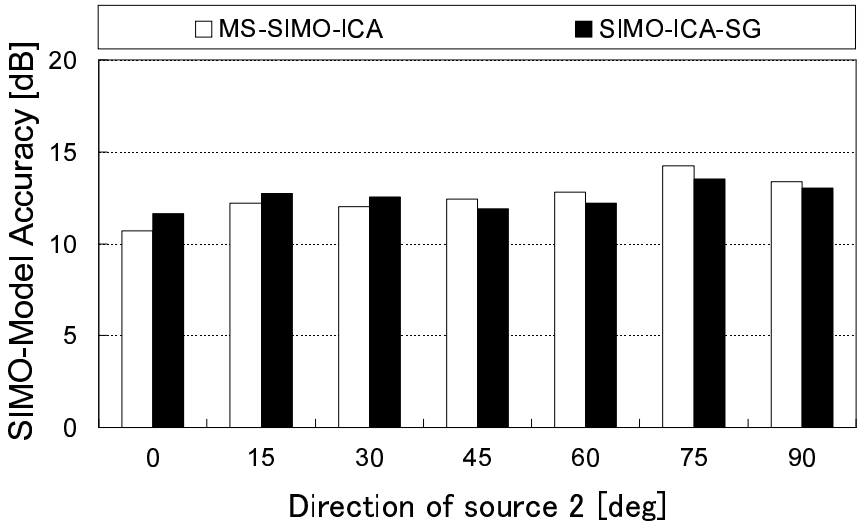

Figure 52. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is 0°. The source signals are speech and HSLN. The number of superpositions in HSLN is 128 (stationary noise case).

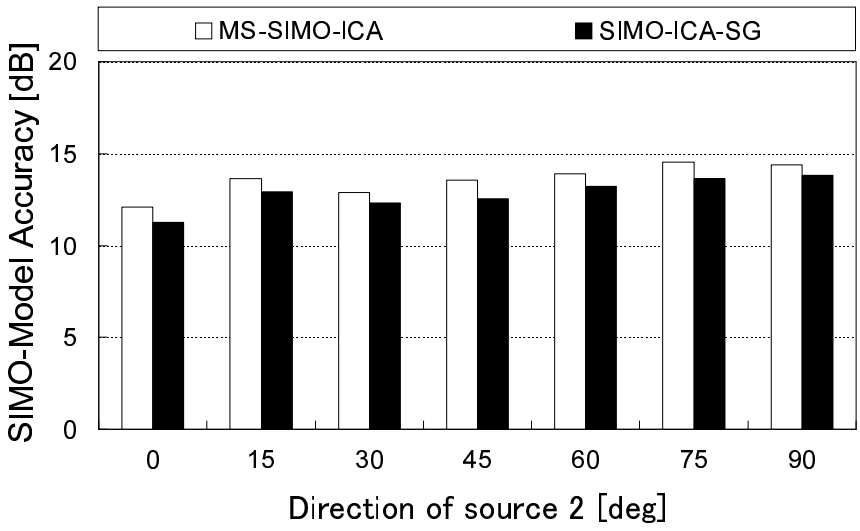

Figure 53. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is 0°. The source signals are speech and HSLN. The number of superpositions in HSLN is 128 (stationary noise case).

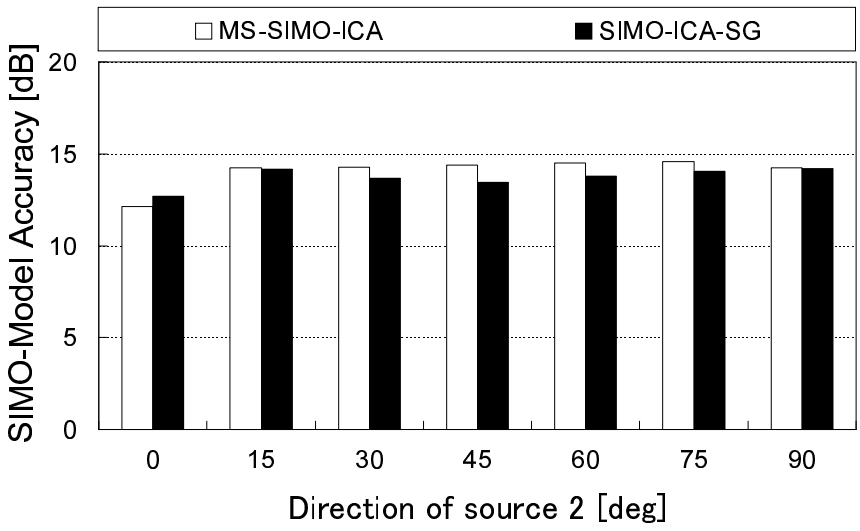

Figure 54. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is 0°. The source signals are speech and HSLN. The number of superpositions in HSLN is 128 (stationary noise case).

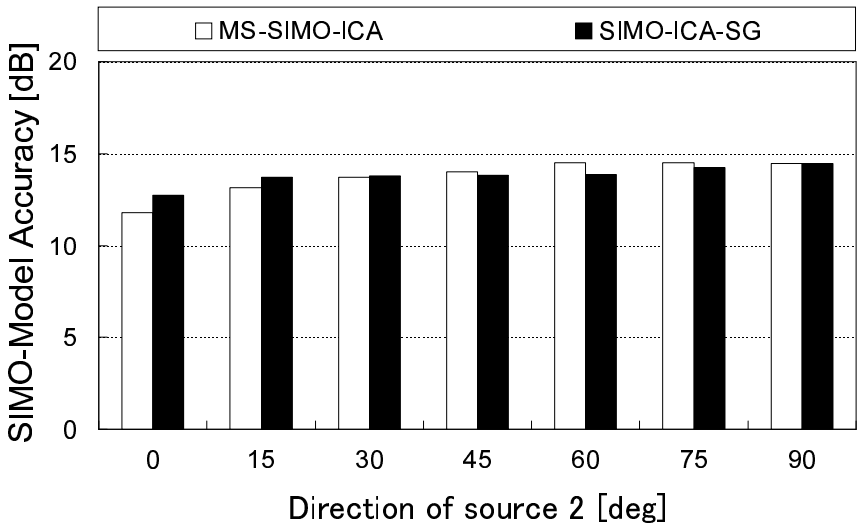

Figure 55. Experimental results of MS-SIMO-ICA and the proposed method, where the direction of sound source 1 is 0°. The source signals are speech and HSLN. The number of superpositions in HSLN is 128 (stationary noise case).

Figure 56–62 shows the results of SA for different θ_1 and θ_2, where the following methods are compared.

(a) The conventional FDICA-PB, for reference,

(b) The conventional MS-SIMO-ICA, for reference,

(c) the proposed method with the NAIST HRTF database (matched case),

(d) the proposed method with the MIT HRTF database (mismatched case).

According to results of (a) the conventional FDICA-PB, (b) the conventional MS-SIMO-ICA, and (d) the proposed SIMO-ICA-SG with the mismatched HRTF database, the performances of the proposed method with the mismatched HRTF database are still superior to those of the conventional methods. According to results of (c) the proposed SIMO-ICA-SG with the matched HRTF database, (d) the proposed SIMO-ICA-SG with the mismatched HRTF database, the performances of the proposed method with the mismatched HRTF database are almost

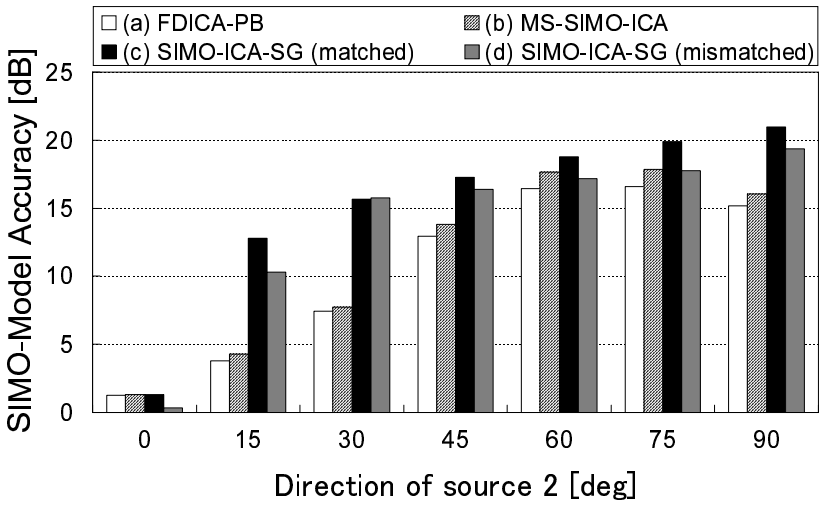

Figure 56. Experimental results of SA using the different HRTF database, where the direction of sound source 1 is 0°. The mixing system is our measured transfer function which contains both HRTF of HATS and room reverberation (see Figure 24). In this experiment, we use the MIT HRTF database as the mismatched one.

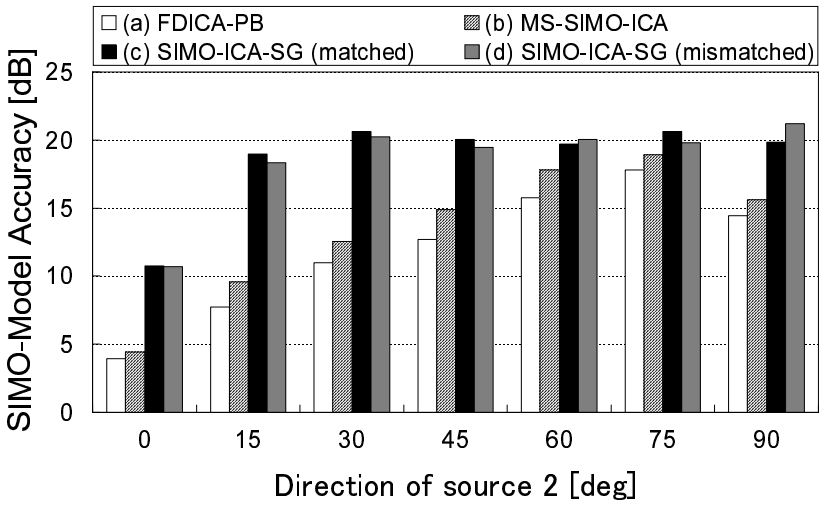

Figure 57. Experimental results of SA using the different HRTF database, where the direction of sound source 1 is −15°. The mixing system is our measured transfer function which contains both HRTF of HATS and room reverberation (see Figure 24). In this experiment, we use the MIT HRTF database as the mismatched one.

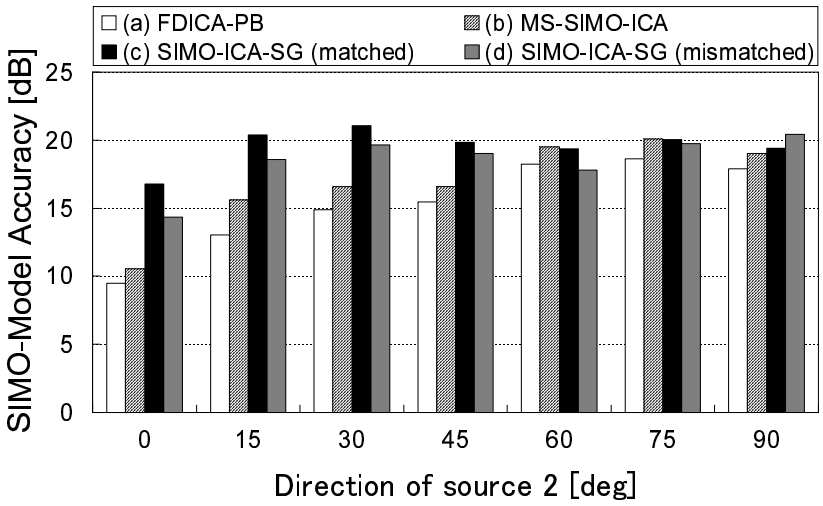

Figure 58. Experimental results of SA using the different HRTF database, where the direction of sound source 1 is $-30°$. The mixing system is our measured transfer function which contains both HRTF of HATS and room reverberation (see Figure 24). In this experiment, we use the MIT HRTF database as the mismatched one.

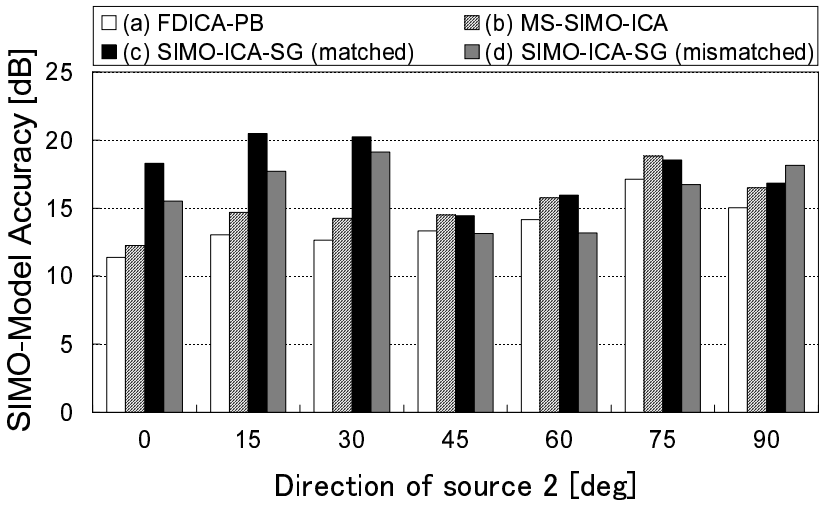

Figure 59. Experimental results of SA using the different HRTF database, where the direction of sound source 1 is $-45°$. The mixing system is our measured transfer function which contains both HRTF of HATS and room reverberation (see Figure 24). In this experiment, we use the MIT HRTF database as the mismatched one.

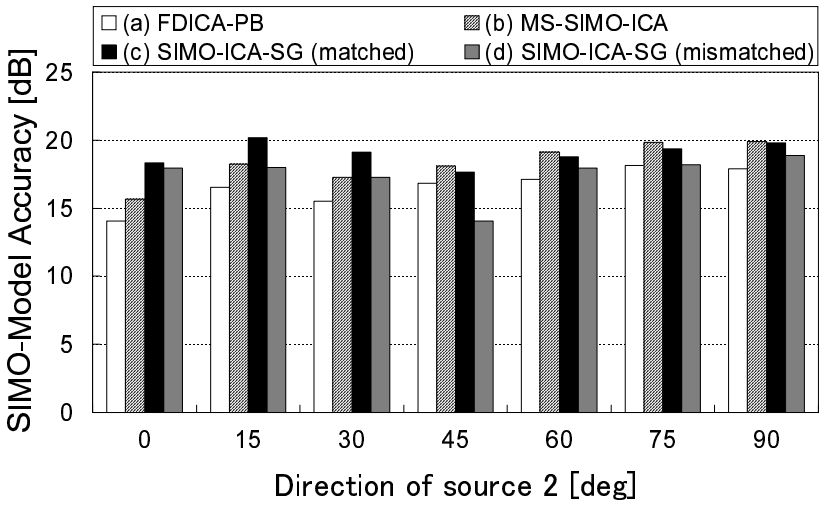

Figure 60. Experimental results of SA using the different HRTF database, where the direction of sound source 1 is $-60°$. The mixing system is our measured transfer function which contains both HRTF of HATS and room reverberation (see Figure 24). In this experiment, we use the MIT HRTF database as the mismatched one.

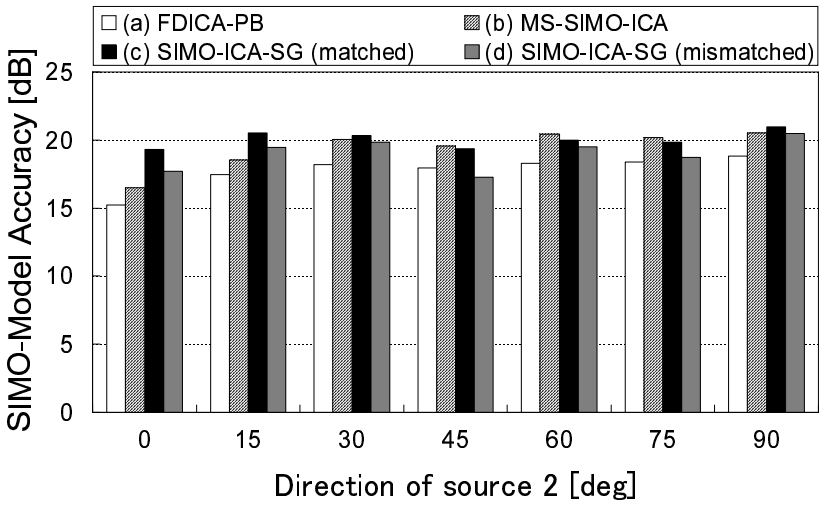

Figure 61. Experimental results of SA using the different HRTF database, where the direction of sound source 1 is $-75°$. The mixing system is our measured transfer function which contains both HRTF of HATS and room reverberation (see Figure 24). In this experiment, we use the MIT HRTF database as the mismatched one.

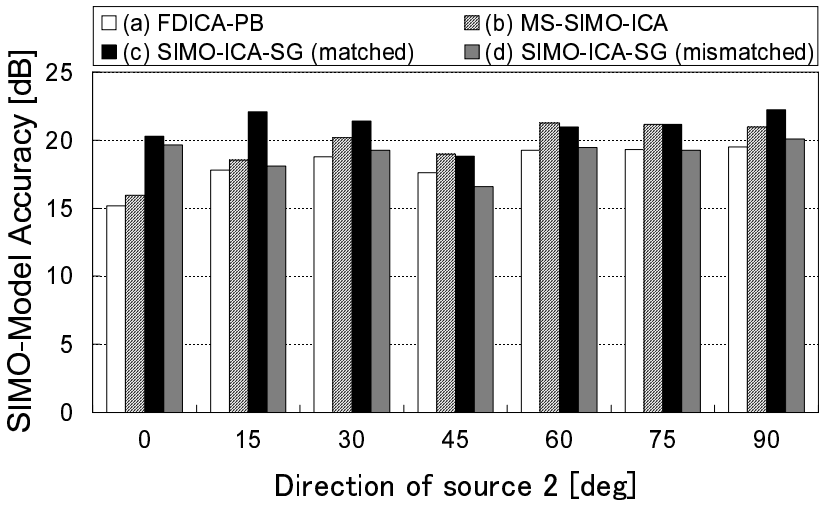

Figure 62. Experimental results of SA using the different HRTF database, where the direction of sound source 1 is $-90°$. The mixing system is our measured transfer function which contains both HRTF of HATS and room reverberation (see Figure 24). In this experiment, we use the MIT HRTF database as the mismatched one.

the same as those of the proposed method with the matched HRTF database. These results imply that the proposed method has sufficient robustness against the HRTF mismatch.

6.5 Conclusion

We newly propose an improved blind separation method based on a self-generator for initial filter of SIMO-ICA. The self-generator for initial filter consists of FDICA with projection back processing and DOA estimation, and blindly provides valid initial values for SIMO-ICA. The generated initial value can improve the separation performance of SIMO-ICA. To evaluate its effectiveness, separation experiments are carried out under a reverberant condition. The experimental results reveal the following.

- In the separation of two speech mixture, the performance of the proposed method is superior to those of the conventional methods, particularly when

the angle between the sound sources is narrow.

- In the speech-noise separation, the proposed method can effectively decompose the observed signals into SIMO-model-based signals under the bubble noise condition. With respect to the stationary noise condition, although there are no remarkable improvements, there are no serious side-effects (i.e., no deterioration).

- The proposed method can run robustly even if a mismatch arises between the mixing process and the HRTF matrix bank in the proposed method.

On the basis of these findings, we can conclude that the proposed SIMO-ICA with a self-generator for the initial filter is a versatile blind separation algorithm.

7. Application of SIMO-ICA

7.1 Introduction

In this section, we describe the application of SIMO-ICA. As a paradigm case of application, in Section 7.2, we present the acoustic augmented reality system using SIMO-ICA. Also, we verify its effectiveness through objective and subjective evaluation experiments for multiple users.

7.2 Acoustic augmented reality

Generally speaking, human beings listen to the sounds by their two ears. These sounds detected at both ears called "binaural sounds." This binaural sound involves information about the localization, directivity, and spatial qualities of each sound source. Also, in real environments, several *undesired* sources exist around the target sound, for example noise, music, interference speech. Thus, we always listen to the mixed binaural sound from the multiple sources, not the binaural sound from the single source. In this section, we apply SIMO-ICA algorithm to acoustic augmented reality system which can extract the target binaural sound component of the mixed binaural sound without the loss of information about the spatial qualities. To realize this system, we use the special apparatus, earphone-microphone system, shown in Figure 63, for picking up the sounds at the entrance of ear canal (cf. Figure 64). We verify its effectiveness through objective and subjective evaluation experiments for multiple users. From these experimental results, we can confirm that the decomposition performance of the proposed method is superior to those of the conventional method.

7.2.1 Experimental conditions

We carried out the experiments using the same room (see Figure 24) as Section 4.5.1. We carried out binaural-sound-separation experiments using source signals which are convolved with impulse responses recorded at user's ears in the experimental room. Two speech signals are assumed to arrive from different directions, θ_1 and θ_2; $\theta_1 = \{-90°, -60°, -30°, 0°\}$ and $\theta_2 = \{0°, 30°, 60°, 90°\}$. The distance between a user and the sound source is 1.5 m. The number of users is

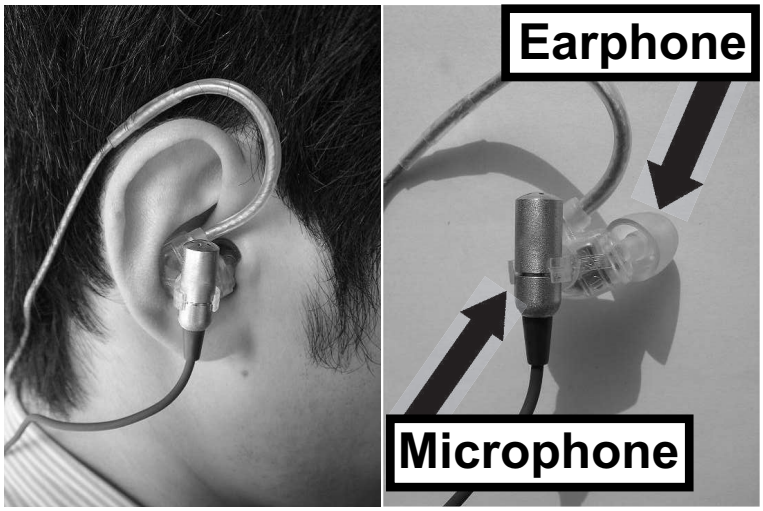

Figure 63. Detail of the earphone-microphone. This apparatus picks up the binaural sounds detected at both ears.

3. The sampling frequency is 16 kHz and the length of each speech sample is limited to 3 seconds. The length of $\boldsymbol{w}(n)$ in each method is 2048. The step-size parameters α and η are 1.0×10^{-6} and 1.0×10^{-2}. With respect to the conventional method, for comparison, we use the MSICA [17] with HRTF matrix bank. MS-ICA is the method which cascade the time domain ICA to the frequency domain ICA, the separation performance of this method is superior to that of each domain method. Since the output signals of this method are monaural signals with respect to each source, we give information about HRTF of head and torso simulator (HATS) (see Section 6.3) with respect to the accurate target direction after ICA's optimization. Also, Noise reduction rate (NRR) and SIMO-model accuracy (SA) are used as an objective evaluation scores.

7.2.2 Objective evaluation

Figures 65–67 show the results of NRR by each **USER1**–**3**, and Figure 68 show the average of them From these results, the separation performance of proposed SIMO-ICA-SG is superior to that of conventional MS-ICA with HRTF matrix bank, and we can confirm that the self-generator of SIMO-ICA can generate the

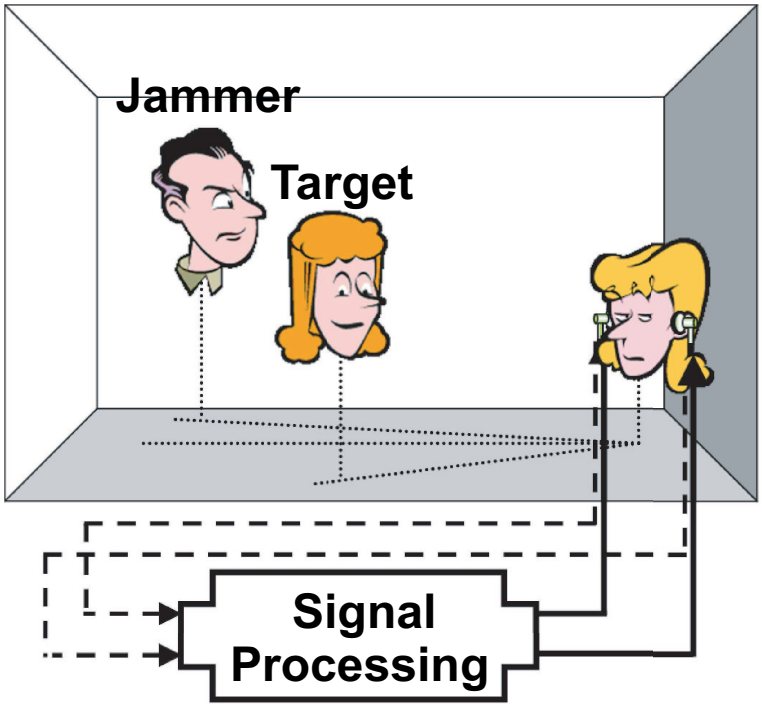

Figure 64. The concept of acoustic augmented reality which can reproduce only the target sound. This system aims to extract the target component of the mixed sounds detected at both ears without the loss of information about directivity, localization, and the spatial qualities of target source.

valid initial filter (see Section 6).

Figures 69–71 show the results of SA by each **USER1**–**3**, and Figure 72 show the average of them. From these results, it is revealed that the SIMO-model accuracy of proposed SIMO-ICA-SG is superior to that of the conventional MS-ICA with HRTF matrix bank. In the conventional method, since the accurate HRTF matrices are unknown, we approximately used the HRTF of HATS as that of USER. Due to this approximation, the approach which cascade the binaural processing to monaural-output ICA can not reconstruct the accurate spatial information.

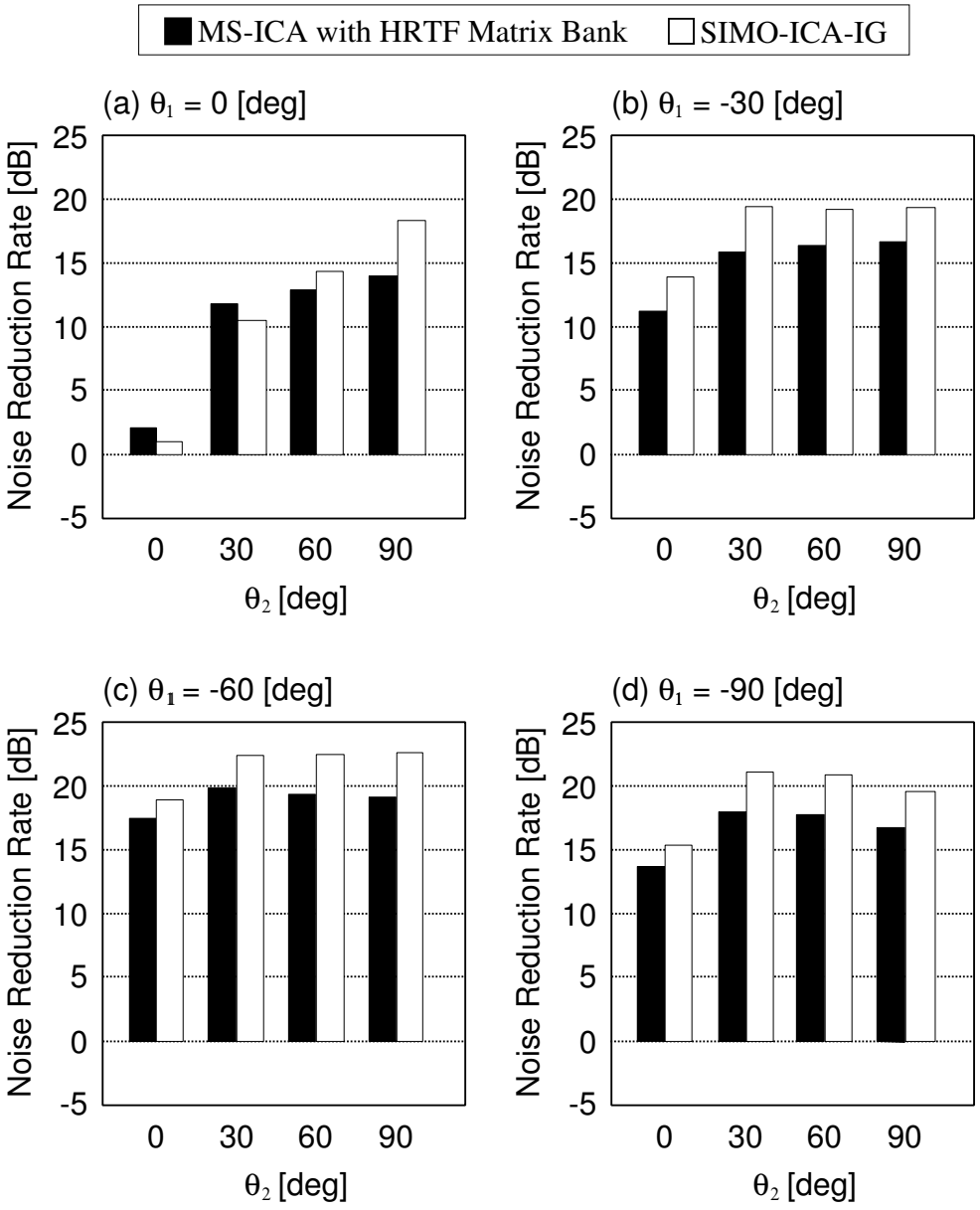

Figure 65. Experimental results of NRR for two-speech mixture separation. In this experiment, we use the impulse responses recorded at **USER1**'s ear point under the experimental room (see Figure 24. The direction of sound source 1 is (a) 0°, (b) −30°, (c) −60°, (d) −90°. The horizontal axis means the direction of sound source 2.

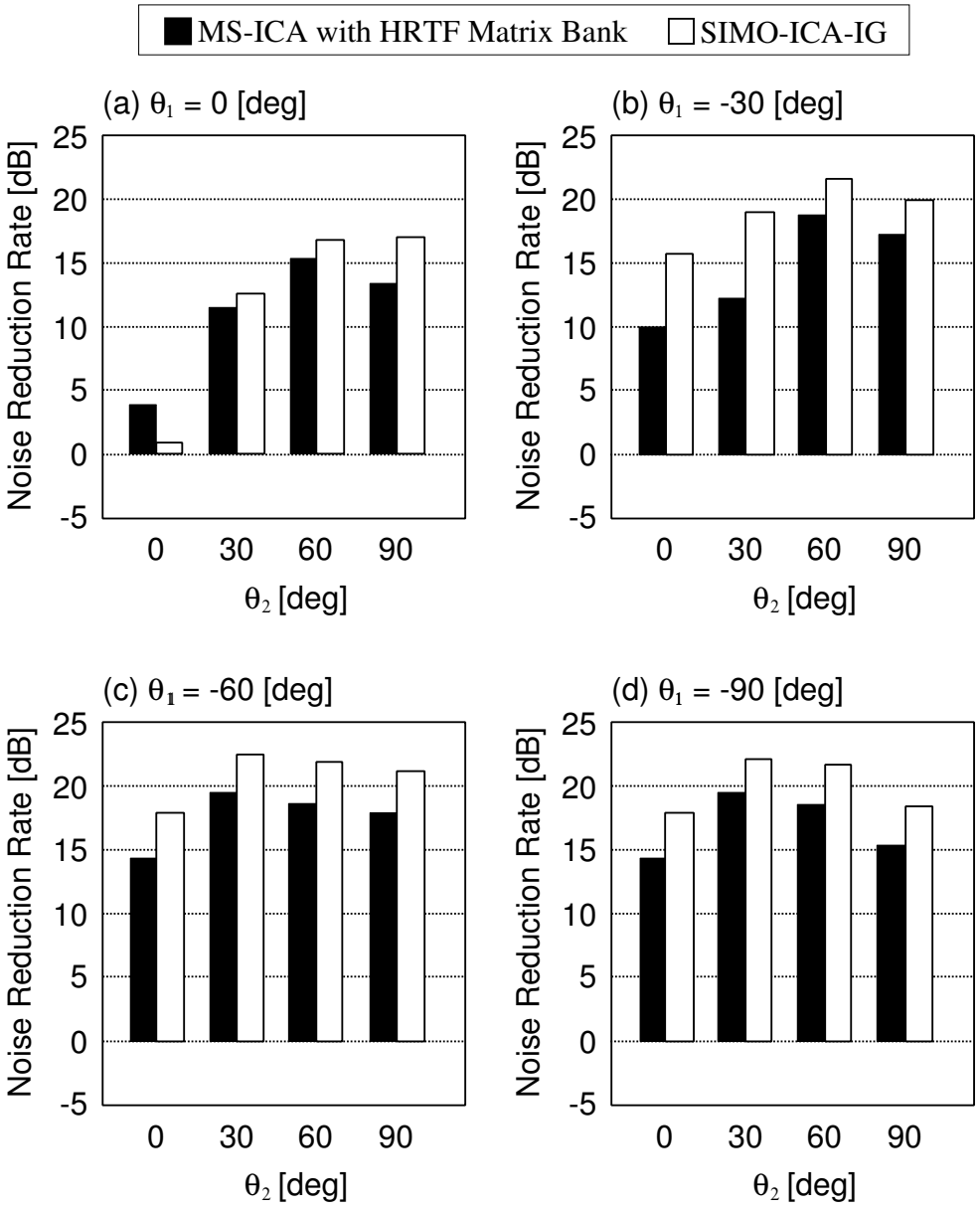

Figure 66. Experimental results of NRR for two-speech mixture separation. In this experiment, we use the impulse responses recorded at **USER2**'s ear point under the experimental room (see Figure 24. The direction of sound source 1 is (a) 0°, (b) −30°, (c) −60°, (d) −90°. The horizontal axis means the direction of sound source 2.

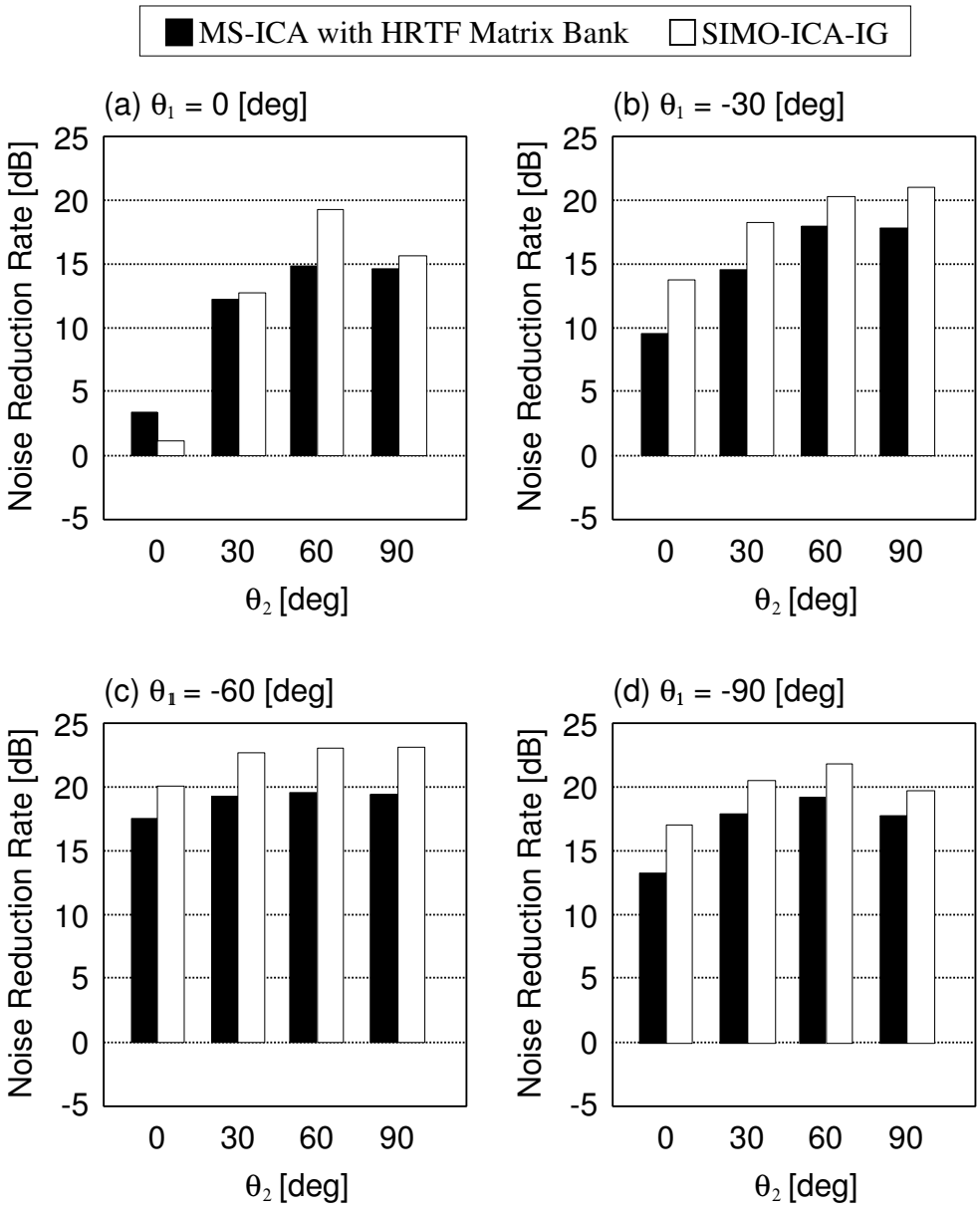

Figure 67. Experimental results of NRR for two-speech mixture separation. In this experiment, we use the impulse responses recorded at **USER3**'s ear point under the experimental room (see Figure 24. The direction of sound source 1 is (a) 0°, (b) −30°, (c) −60°, (d) −90°. The horizontal axis means the direction of sound source 2.

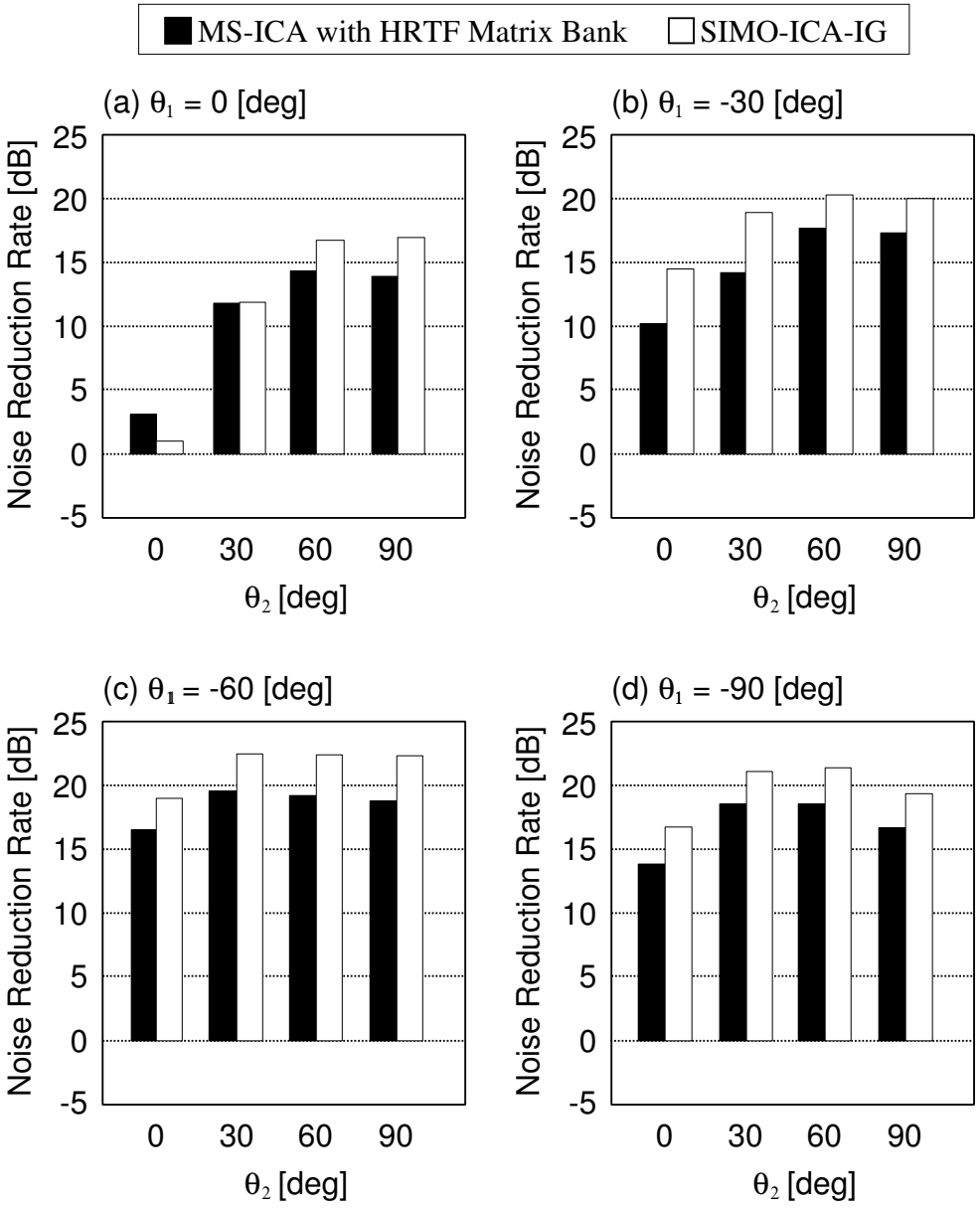

Figure 68. Average of NRR for two-speech mixture separation. The direction of sound source 1 is (a) 0°, (b) −30°, (c) −60°, (d) −90°. The horizontal axis means the direction of sound source 2.

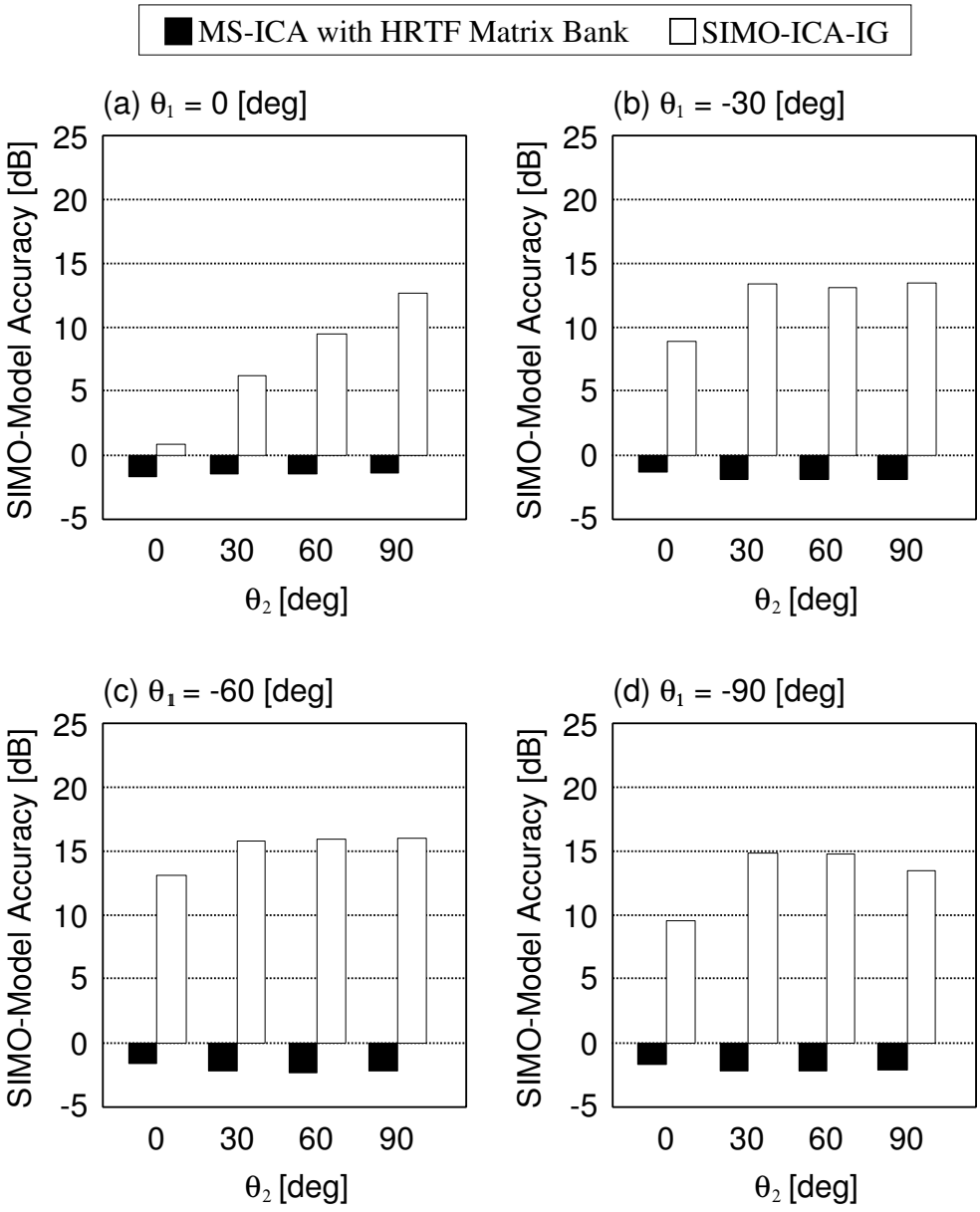

Figure 69. Experimental results of SA for two-speech mixture separation. In this experiment, we use the impulse responses recorded at **USER1**'s ear point under the experimental room (see Figure 24. The direction of sound source 1 is (a) 0°, (b) −30°, (c) −60°, (d) −90°. The horizontal axis means the direction of sound source 2.

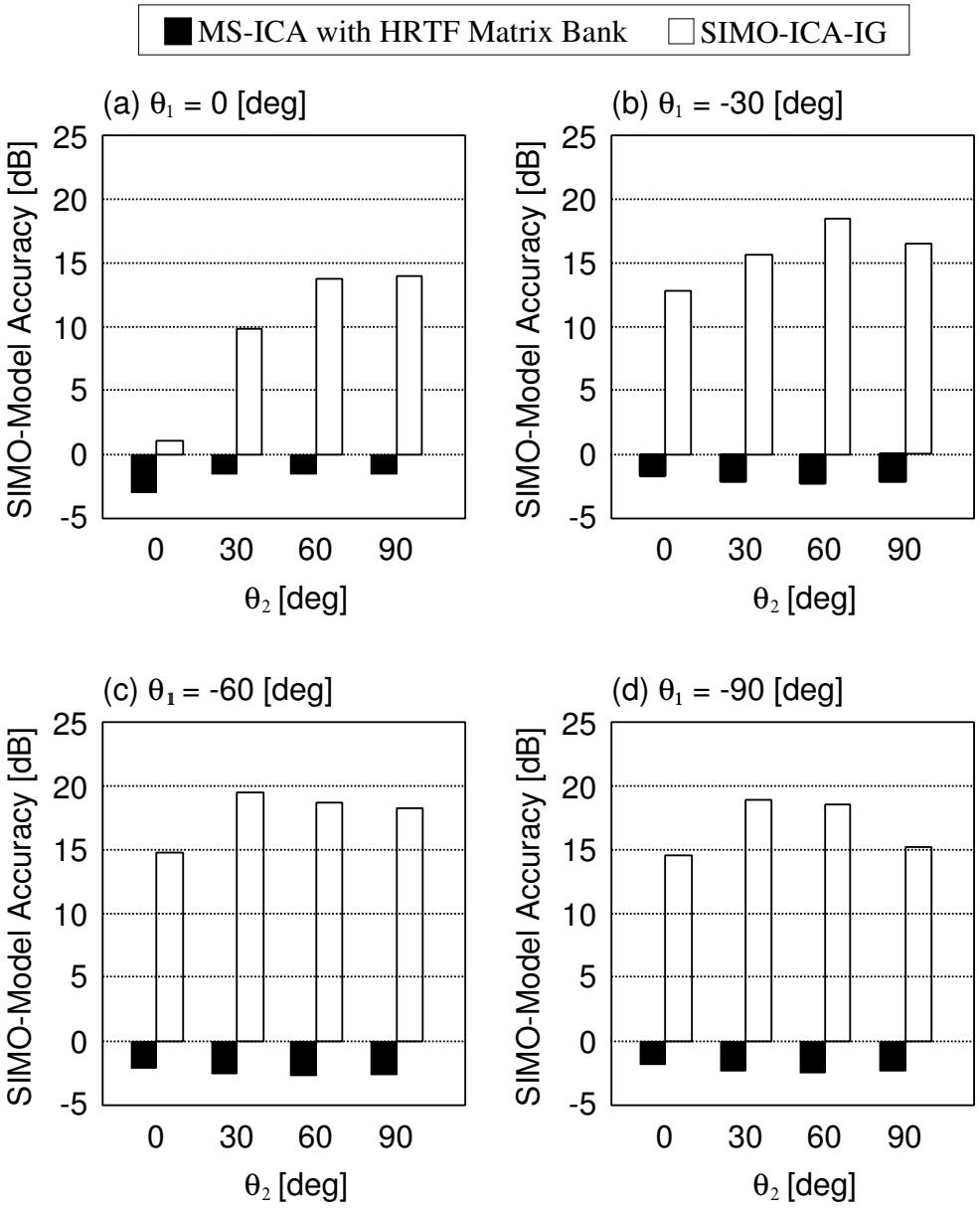

Figure 70. Experimental results of SA for two-speech mixture separation. In this experiment, we use the impulse responses recorded at **USER2**'s ear point under the experimental room (see Figure 24. The direction of sound source 1 is (a) 0°, (b) −30°, (c) −60°, (d) −90°. The horizontal axis means the direction of sound source 2.

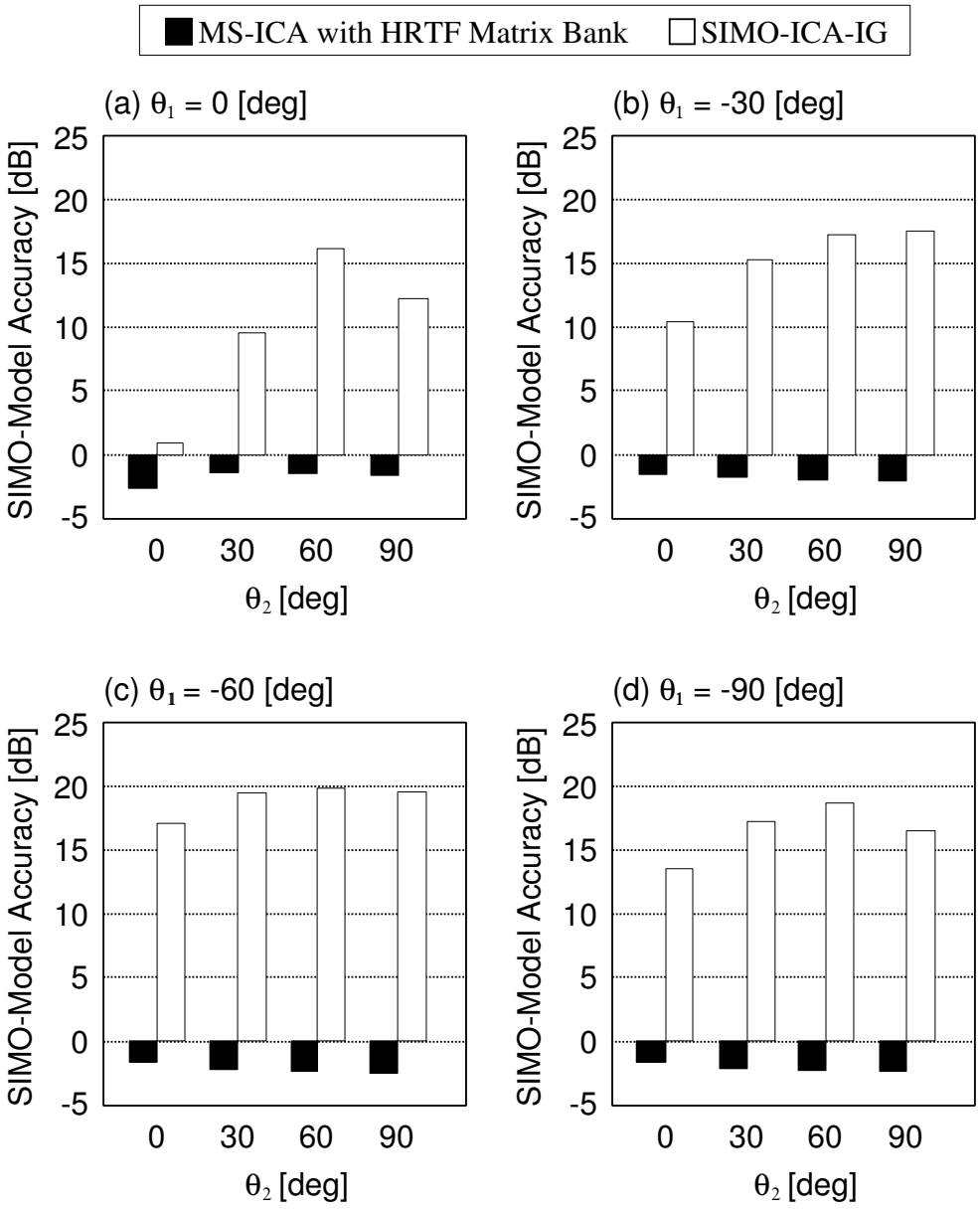

Figure 71. Experimental results of SA for two-speech mixture separation. In this experiment, we use the impulse responses recorded at **USER3**'s ear point under the experimental room (see Figure 24. The direction of sound source 1 is (a) 0°, (b) −30°, (c) −60°, (d) −90°. The horizontal axis means the direction of sound source 2.

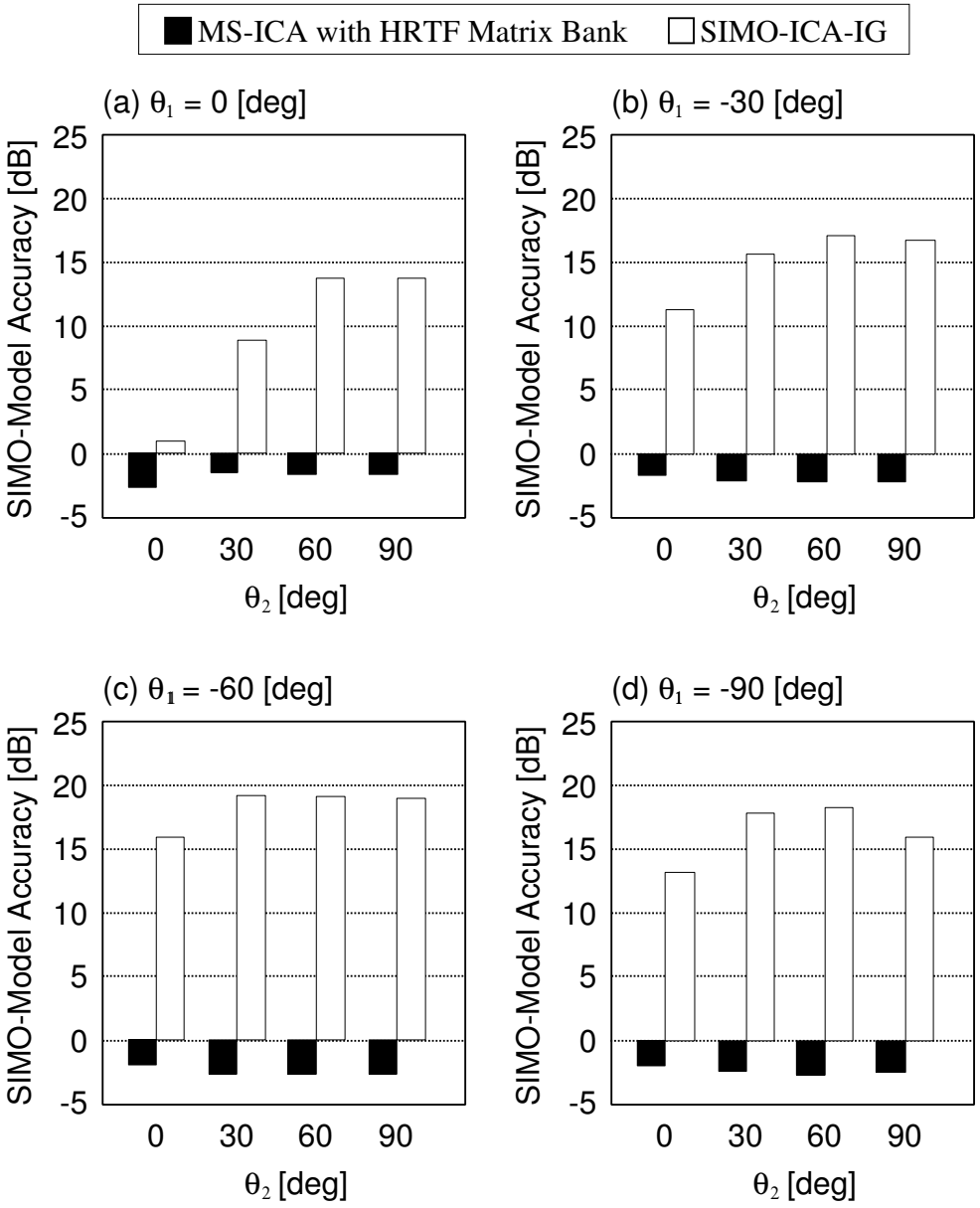

Figure 72. Average of SA for two-speech mixture separation. The direction of sound source 1 is (a) 0°, (b) −30°, (c) −60°, (d) −90°. The horizontal axis means the direction of sound source 2.

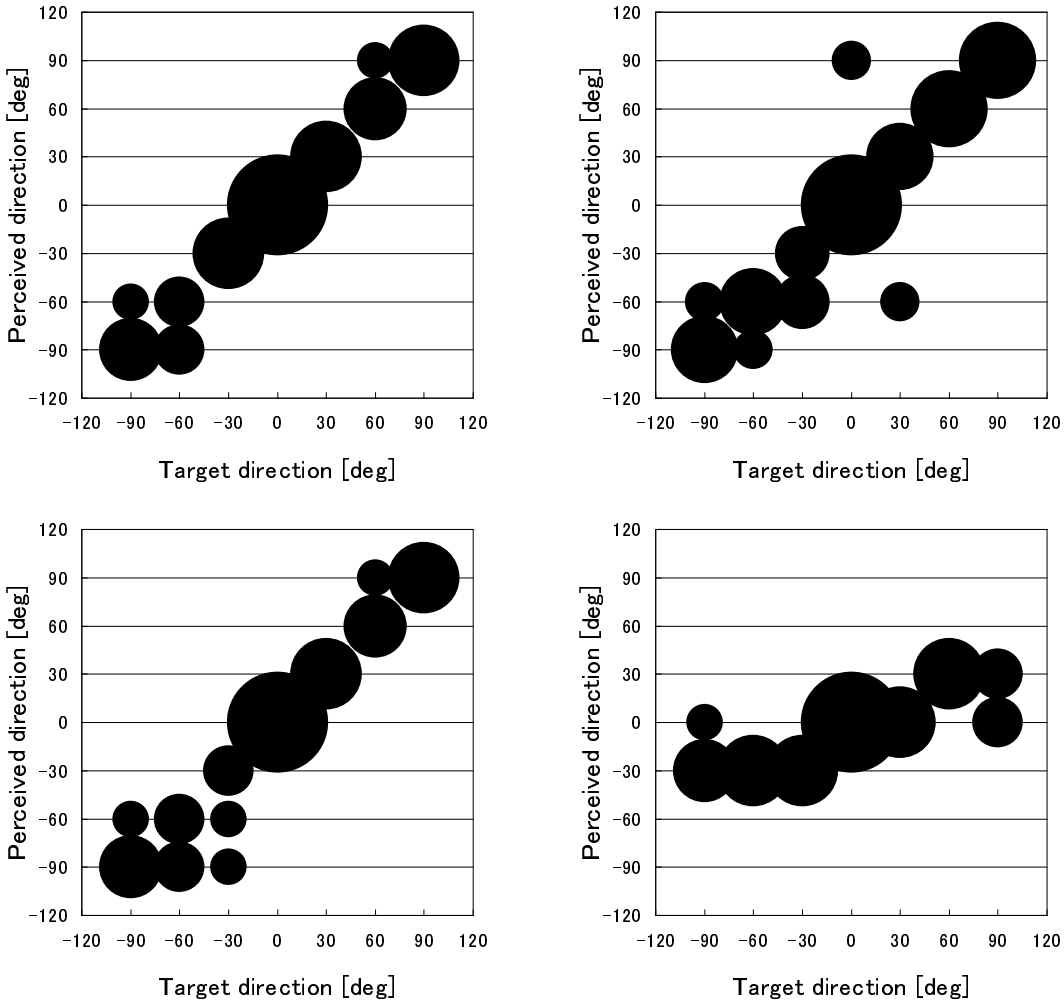

Figure 73. The results of subjective evaluation by **USER1**. This figure shows the relations between perceptual directions and target directions in (a) real SIMO-model-based signals (no interference signals), (b) observed signals (c) output signals of SIMO-ICA, (d) output signals of MS-ICA with HRTF matrix bank. The objective evaluation results is shown in Figures 65 and 69

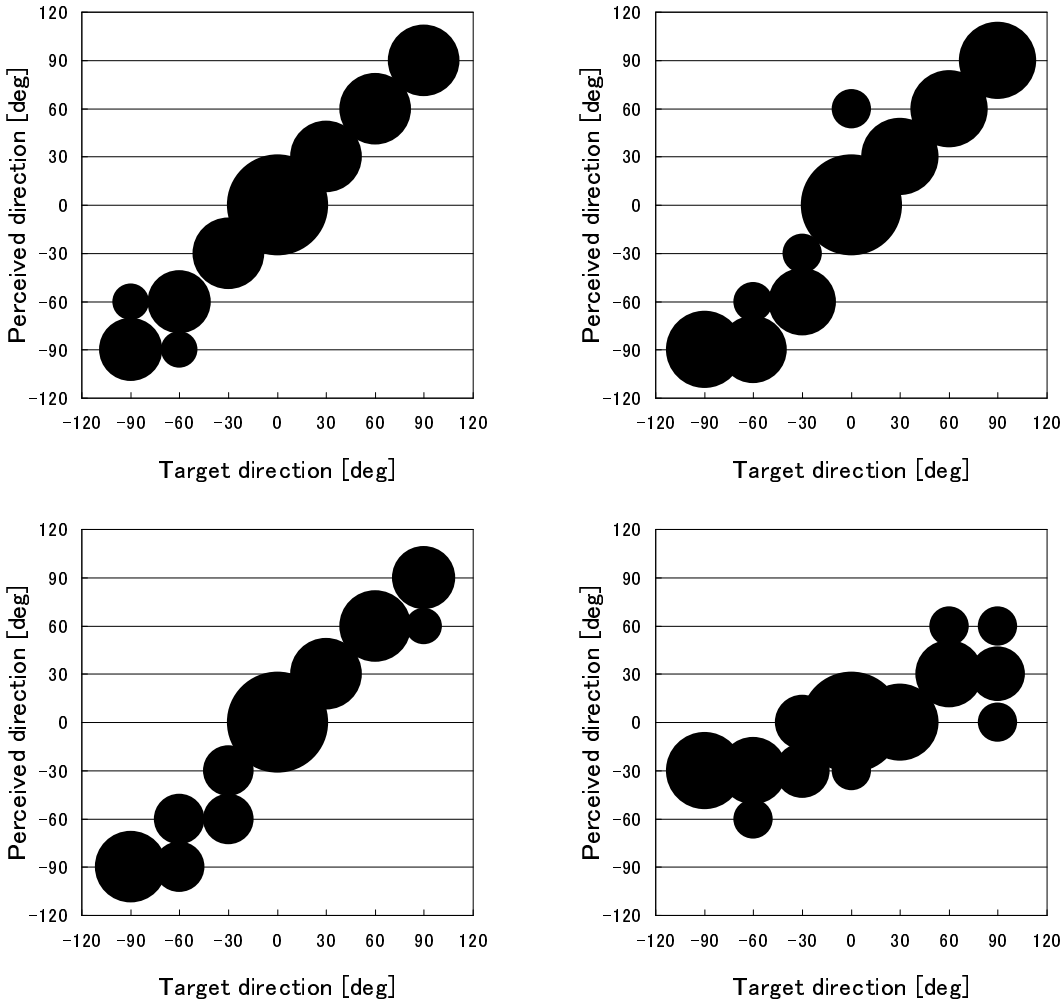

Figure 74. The results of subjective evaluation by **USER2**. This figure shows the relations between perceptual directions and target directions in (a) real SIMO-model-based signals (no interference signals), (b) observed signals (c) output signals of SIMO-ICA, (d) output signals of MS-ICA with HRTF matrix bank. The objective evaluation results is shown in Figures 66 and 70

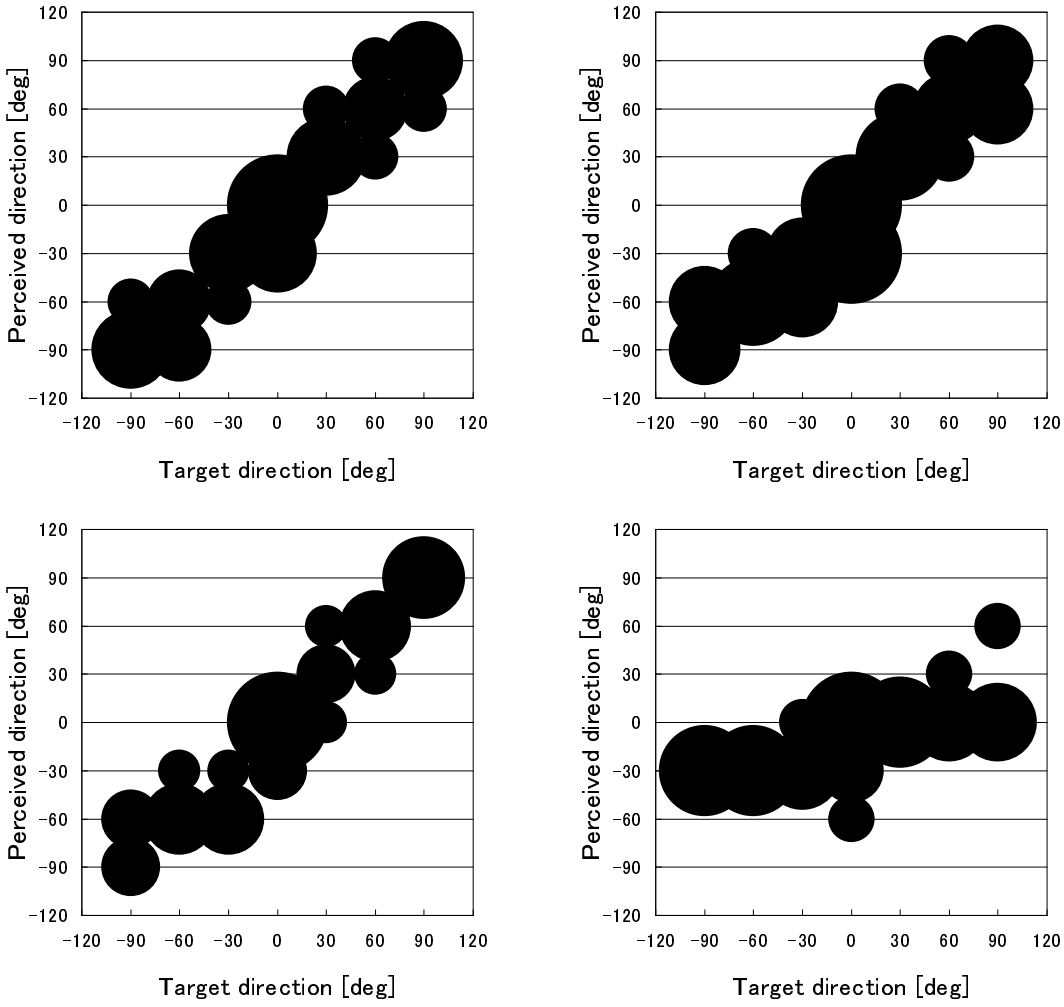

Figure 75. The results of subjective evaluation by **USER3**. This figure shows the relations between perceptual directions and target directions in (a) real SIMO-model-based signals (no interference signals), (b) observed signals (c) output signals of SIMO-ICA, (d) output signals of MS-ICA with HRTF matrix bank. The objective evaluation results is shown in Figures 67 and 71

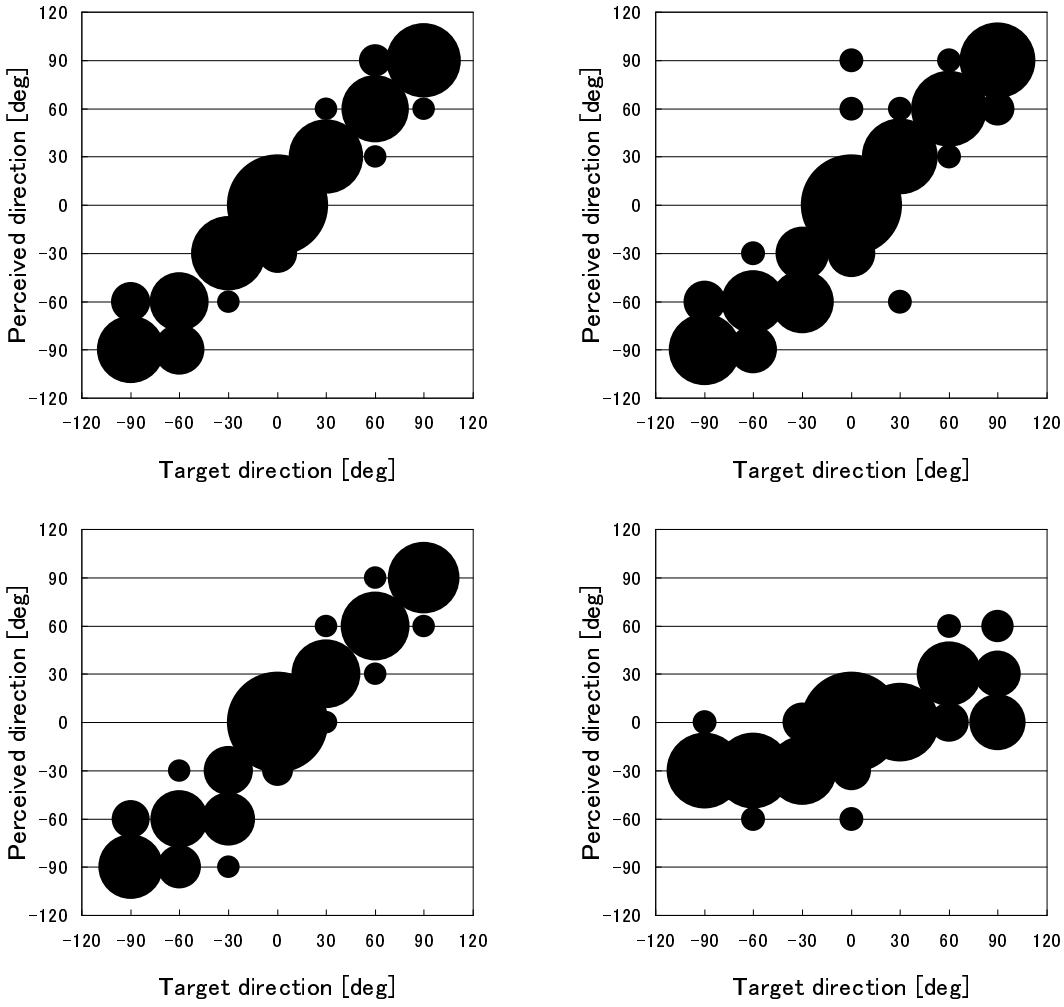

Figure 76. The average of subjective evaluation (relations between perceptual directions and target directions) in (a) real SIMO-model-based signals (no interference signals), (b) observed signals (c) output signals of SIMO-ICA, (d) output signals of MS-ICA with HRTF matrix bank. The objective evaluation results is shown in Figures 67 and 71

7.2.3 Subjective evaluation

In order to confirm that the output signals in SIMO-ICA maintain the spatial qualities of each sound source, we carried out the subjective evaluation experiments by 3 users. Figures 73 –75 show the the relations between perceptual directions and target directions in (a) the real SIMO-model-based signals (no interference signals), (b) the observed signals, (c) output signals of SIMO-ICA, (d) output signals of conventional ICA which are convolved HRTF with respect to the accurate target direction by **USER1**–**3**.

From these results, even if we convolve the head related transfer function corresponds to the accurate source direction with the monaural separated signals of ICA, we can not generate accurate sound image. Also, we can confirm that sound image in SIMO-ICA is almost the same as that in real SIMO-model-based signals.

Overall it is asserted that SIMO-ICA can reduce the interference component of the mixed binaural signals and reproduce the target component while keeping information about the directivity and spatial qualities of target source.

7.3 Conclusion

In this section, we describe the application of SIMO-ICA, and we discuss the acoustic augmented reality system using SIMO-ICA algorithm. To evaluate its effectiveness, separation experiments are carried out under a reverberant condition for different 3 users. The objective and subjective experimental results reveal the performance of the proposed method is superior to that of the conventional method. Thus, we can conclude that the proposed system using SIMO-ICA is applicable to the hearing aids [24] and acoustic tele-presence system using the humanoid robots [38].

Also, we should mention combination methods which cascade other array processings to SIMO-ICA algorithm because the output signals of SIMO-ICA are still multiple array signals. we enumerate the combination methods which we addressed.

(a) blind separation and deconvolution

Mr. Yamajo, who was a member of our laboratory in 2002–2004, propose the blind separation and deconvolution method combining SIMO-ICA and Multi-channel Inverse Filtering [21]. This research can achieve the MIMO deconvolution processing for convolutive mixture of speech.

(b) high-performance blind speech signal separation

Mr. Ukai, who was a member of our laboratory in 2003–2005, propose the blind separation method combining SIMO-ICA and adaptive beamforming (ABF) [33]. In this algorithm, using the output signals of SIMO-ICA, the information which is required in ABF processing is blindly generated, and ABF can reduce the residual component of the interference effectively.

(c) real-time blind speech signal separation

Mr. Mori, who is a member of our laboratory now, propose the blind separation method combining SIMO-ICA and binary masking processing [34]. Using this method, we can realize the real-time blind source separation system.

8. Conclusion

8.1 Thesis summary

We addressed to the blind source separation (BSS) to realize a high-fidelity hands-free telecommunication system. BSS technique using independent component analysis (ICA) for acoustic signals has been developed over the last decade. This technique assumes that the source signals are mutually independent, and can estimate the source signals from the mixed signals without information about the source direction, source's properties, room transfer function, and mixing process. Thus, this technique is highly applicable in high-quality hands-free telecommunication system. The conventional ICA-based BSS method is a means of extracting the independent sound source signals as the monaural signals from the mixed signals observed in each input channel, and the separated signals include arbitrary spectral distortions. Consequently, they have a serious drawback in that the separated sounds cannot maintain information about the directivity, localization, reverberation, or spatial qualities of each sound source. These problems prevent any BSS methods from being applied to binaural signal processing or high-fidelity sound reproduction system.

In order to solve the above-mentioned problems, we proposed BSS using the novel extended ICA algorithm, i.e., (a) BSS using single-input multiple-output (SIMO)-model-based ICA (SIMO-ICA) with least squares criterion, and (b) BSS using SIMO-ICA with information geometric criterion. Also, in order to apply SIMO-ICA algorithm to blind decomposition problems of the mixed binaural sounds, we proposed the self-generator method for initial filter of SIMO-ICA.

In Section 2, first, the sound mixing model is explained. Next, we introduced the novel demixing process of the acoustic source signals to apply to the binaural processing.

In Section 3, we explained the conventional ICA methods and describe the disadvantage of these methods. Furthermore we presented that development of a new SIMO-output-type method in which all SIMO components can be simultaneously estimated in the ICA updating with stable FIR filters, is a problem demanding prompt attention.

In Section 4, We proposed 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 this section, we designed the fidelity controller of SIMO-ICA using the least squares criterion. In order to evaluate its effectiveness, separation experiments were carried out using two microphones and two sources under the condition that the RTs is set to be 150 ms. The experimental results revealed that the signal separation performance of the proposed SIMO-ICA was 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. Therefore, we can conclude that the proposed SIMO-ICA is applicable to binaural signal processing and high-fidelity sound reproduction systems.

In Section 5, We proposed a new single-input multiple-output (SIMO)-model-based ICA with an information-geometric learning algorithm for high-fidelity blind source separation. The SIMO-ICA-IG is an algorithm for separating mixed signals, not into monaural source signals but into SIMO-model-based signals of independent sources without loss of their spatial qualities. The SIMO-ICA-IG consists of multiple ICA parts and a fidelity controller, and all of the procedures are conducted on the basis of the information-geometry theory. Hence the proposed SIMO-ICA-IG is free from deterioration due to mismatching among multiple criteria in ICAs and fidelity controller. In order to evaluate its effectiveness, separation experiments were carried out using two microphones and two sources under a reverberant condition. The experimental results revealed the following.

- The signal separation performance of the proposed SIMO-ICA-IG was almost the same as those of the conventional ICA and the previously proposed SIMO-ICA-LS.

- The spatial qualities of the separated sound in the SIMO-ICA-IG were superior to those in the conventional ICA, and almost the same as those in the SIMO-ICA-LS.

- It was confirmed that the balancing parameter setting in the proposed SIMO-ICA-IG did not depend on the source signals' properties. Although

SIMO-ICA-LS was very sensitive to the balancing parameter setting, the balancing parameter in the proposed SIMO-ICA-IG could be almost negligible.

From these findings, we can conclude that the proposed SIMO-ICA-IG has great potential for application to high-fidelity signal processing systems, and has a good robustness against parameter setting.

In Section 6, We newly proposed an improved blind separation method based on a self-generator for initial filter of SIMO-ICA. The self-generator for initial filter consists of FDICA with projection back processing and DOA estimation, and blindly provides valid initial values for SIMO-ICA. The generated initial value can improve the separation performance of SIMO-ICA. To evaluate its effectiveness, separation experiments were carried out under a reverberant condition. The experimental results revealed the following.

- In the separation of two speech mixture, the performance of the proposed method was superior to those of the conventional methods, particularly when the angle between the sound sources was narrow.

- In the speech-noise separation, the proposed method can effectively decompose the observed signals into SIMO-model-based signals under the bubble noise condition. With respect to the stationary noise condition, although there were no remarkable improvements, there were no serious side-effects (i.e., no deterioration).

- The proposed method could run robustly even if a mismatch arises between the mixing process and the HRTF matrix bank in the proposed method.

On the basis of these findings, we can conclude that the proposed SIMO-ICA with a self-generator for the initial filter is a versatile blind separation algorithm.

In Section 7, we applied SIMO-ICA algorithm to acoustic augmented reality system which could extract the target binaural sound component of the mixed binaural sound without the loss of information about the spatial qualities. To realize this system, we used the special apparatus, earphone-microphone system, for picking up the sounds at the entrance of ear canal. We verified its effectiveness through objective and subjective evaluation experiments for multiple users. From

these experimental results, we can conclude that the decomposition performance of the proposed SIMO-ICA-based method is superior to those of the conventional ICA-based method.

In summary, we confirmed that the proposed SIMO-ICA-based BSS is effective for improving the sound qualities. In addition, SIMO-ICA-SG using the self-generator method is also effective for achieving the high-convergence.

8.2 Future research

Although we could estimate the SIMO-model-based signals which maintain information about the directivity, localization, or spatial qualities of each sound source, following points still remain to be solved.

Identifiability in the SIMO decomposition processing

BSS using SIMO-ICA cannot identify SIMO-model-based signals, but can reduce the residuals by setting the separation filter length to be sufficiently long (see Section A). Thus, in the high-reverberation conditions, the length of separation filter matrix which is required is so long that it is difficult to optimize the separation filter matrix. In order to improve the high separation performance and obtain robustness against the reverberation, the strategy which can solve this problem is required.

Estimation of the number of source signals

BSS using SIMO-ICA could decompose the observed signals without spectral distortion. However, in SIMO-ICA algorithm, we need the number of sound sources because the iterative learning rule is designed on the basis of the superposition principle. Thus, we should combine the estimation method of the sound sources [19] and SIMO-ICA algorithm.

Extension of overdetermined SIMO-ICA

The separation performance of BSS using SIMO-ICA is almost the same as

the conventional BSS based on ICA. To improve the separation performance, we should extend SIMO-ICA to the overdetermined algorithm to obtain more source information.

Reduction of the calculation time

In stead of improving the performance in SIMO-ICA-LS and SIMO-ICA-IG algorithms, the complexity is large. In the application of acoustic augmented reality system, we should achieve the real-time separation. To realize this system, we should combine it with other array signal methods and subband signal processings.

Acknowledgements

This dissertation is a summary of studies for four years and a half carried out at Graduate School of Information Science, Nara Institute of Science and Technology, Japan.

I would like to express my sincere thanks to Professor Kiyohiro Shikano, my thesis adviser, for his valuable guidance and constant encouragement.

I would also like to express my gratitude to Professor Kenji Sugimoto for his valuable comments to the master and doctoral thesis.

I would especially like to express genuine gratitude to Associate Professor Hiroshi Saruwatari for his continuous support and valuable advice. Without his guidance, this work could not have been completed.

I would also like to thank to Assistant Professor Tomoki Toda and Assistant Professor Hiromichi Kawanami for their beneficial comments.

I want to thank all members of the Speech and Acoustics Laboratory in Nara Institute of Science and Technology for providing fruitful discussions. I would especially like to express my appreciation to Dr. Akinobu Lee, who is currently Associate Professor at Nagoya Institute of Technology, Dr. Yosuke Tatekura, who is currently Assistant Professor at Shizuoka University, Dr. Ryuichi Nishimura, who is currently Assistant Professor at Wakayama University, Dr. Tsuyoki Nishikawa, who is researcher in Matsushita Electric Industrial Co., Ltd.,

I would sincerely like to thank Dr. Shoji Makino, who is currently Executive Manager at Media Information Laboratories of NTT Communication and Science Laboratories, Ms. Shoko Araki, Dr. Ryo Mukai, Dr. Hiroshi Sawada, who are members of Signal Processing Research Group, for providing fruitful discussions about blind source separation. I would also like to thank Dr. Masato Miyoshi, Group Leader of Signal Processing Research Group at Media Information Laboratories of NTT Communication and Science Laboratories, and Dr. Hiroshi Sawada for their support when I was a student intern at NTT Communication and Science Laboratories. I would sincerely like to thank Mr. Takashi Hiekata, Mr. Takashi Morita, who are currently researcher at Kobe Steel, Ltd for the discussions about real-time blind source separation.

I would like to thank Mr. Hiroaki Yamajo, who is currently a researcher at Sony Corporation Home Electronics Network Company, Mr. Satoshi Ukai,

who is currently researcher at Yamaha Corporation, Mr. Akira Baba, who is a researcher at Matsushita Electric Works, Ltd., Mr. Shigeki Miyabe, who is Doctoral candidate of Nara Institute of Science and Technology, Mr. Yoshimitsu Mori, Ms. Chie Kiuchi, Mr. Masayuki Shimada, Mr. Kazuyuki Hayashi, Mr. Yuu Takahasi, and Mr. Tadashi Mihashi, who are currently master's course of Nara Institute of Science and Technology, for providing valuable discussion and suggestions about blind source separation and microphone array processing.

I would also like to express my gratitude to all of my friends for their support. I would especially like to express my gratitude to Mr. Yohei Iwami, Mr. Shigefumi Urata, Mr. Yasunori Ohno, Mr. Hidekazu Kato, Mr. Tatsuya Shiraishi, Mr. Ryosuke Tsurumi, Mr. Yoichi Hinamoto, Mr. Yohei Matsumoto, Ms. Hiroko Yamada, and Mr. Shingo Yamade,

Finally, I would like to thank my family, Madoka, Yoshiki, Sachiko, Tetsuo, Masako, Naoko, and Teppei, for their supports in daily life to carry on this study.

References

[1] G. W. Elko, "Microphone array systems for hands-free telecommunication," *Speech Communication*, vol.20, pp.229–240, 1996.

[2] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Am.*, vol.78, pp.1508–1518, 1985.

[3] M. Omologo, M. Matassoni, P. Svaizer, and D. Giuliani, "Microphone array based speech recognition with different talker-array positions," *Proc. ICASSP '97*, pp.227–230, April 1997.

[4] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol.60, pp.926–935, 1972.

[5] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas & Propag.*, vol.30, pp.27–34, 1982.

[6] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoust., Speech & Signal Process.*, vol.ASSP-34, pp.1391–1400, 1986.

[7] 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.

[8] C. Jutten and J. Herault, "Blind separation of sources part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol.24, pp.1–10, 1991.

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

[10] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol.7, pp.1129–1159, 1995.

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

[12] S. Haykin, *Unsupervised Adaptive Filtering*, New York, NY: John Wiley & Sons, Inc., 2000.

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

[14] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proceedings of 1998 International Symposium on Nonlinear Theory and its Application (NOLTA '98)*, vol.3, pp.923–926, Sep. 1998.

[15] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech and Audio Processing*, vol.8, no.3, pp.320–327, May 2000.

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

[17] T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA," *IEICE Trans. Fundamentals*, vol.E86–A, no.4, pp846–858, April 2003.

[18] T. Nishikawa, H. Saruwatari, K. Shikano, "Stable leaning algorithm for blind separation of temporally correlated acoustic signals combining multistage ICA and linear prediction," *IEICE Trans. Fundamentals*, vol.E86-A, no.8, pp.2028–2036, August 2003.

[19] H. Sawada, R. Mukai, S. Araki, S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. on Speech and Audio Processing*, vol.12, no.5, pp.530–538, September 2005.

[20] Charkani, N. and Deville, Y., "A convolutive separation method with self-optimizing non-linearities," *Proc. ICASSP '99,* pp.2909–2912, Mar. 1999.

[21] H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa, K. Shikano, "Blind Separation and Deconvolution for Convolutive Mixture of Speech Combining SIMO-Model-Based ICA and Multichannel Inverse Filtering," *IEICE Transactions on Fundamentals*. Vol.E88-A, No.9, pp.2387–2400, September 2005.

[22] Simon, C., Loubaton, P., Vignat, C., Jutten, C., d'Urso, G., "Separation of a class of convolutive mixturesa contrat: a contrast function approach," *Proc. ICASSP'99* 1429–1432, March 1999.

[23] Tugnait, J.K., "Identification and deconvolution of multichannel linear non-gaussian processes using higher order statistics and inverse filter criteria," *IEEE trans. on signal processing.* Vol. 45. pp 658–672, 1997.

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

[25] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc.*, vol.44, no.9, pp.683–705, 1996.

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

[27] Y. Nagata, Y. Tatekura, H. Saruwatari, and K. Shikano, "Iterative inverse filter relaxation algorithm for adaptation to acoustic fluctuation in sound reproduction system," Electronics and Communications in Japan, Part 3, J. Wiley & Sons, Inc., vol.87, no.7, pp.15–26, 2004.

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

[29] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a non-holonomic constraint for blind deconvolution of multiple channels,"

Proceedings of International Workshop on Independent Component Analysis and Blind Signal Separation (ICA&BSS '99), pp.371–376, 1999.

[30] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA&BSS '01)*, pp.722–727, Dec. 2001.

[31] J.-F. Cardoso, "Multidimensional independent component analysis," *Proc. ICASSP '98*, vol.4, pp.1941–1944, May 1998.

[32] S. Ukai, H. Saruwatari, T. Takatani, R. Mukai, H. Sawada, "Multistage SIMO-model-based blind source separation combining frequency-domain ICA and time-domain ICA," *Proceedings of ICASSP2004,* pp.IV-109–IV-112, 2004.

[33] S. Ukai, T. Takatani, H. Saruwatari, T. Nishikawa, K. Shikano, "Blind Source Separation Combining SIMO-model-based ICA and Adaptive Beamforming," *Proceedings of ICASSP2005*, pp.III-85–III-88, 2005.

[34] Y. Mori, H. Saruwatari, T. Takatani, S. Ukai, K. Shikano, T. Hiekata, T. Morita, "Real-Time Implementation of Two-Stage Blind Source Separation Combining SIMO-ICA and Binary Masking," *Proceedings of IWAENC2005*, pp.229 − 232, September 2005

[35] A. Poularikas, *The Handbook of Formulas and Tables for Signal Processing*, CRC Press, Boca Raton, 1999.

[36] 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 (in Japanese).

[37] 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.

[38] I. Toshima, H. Uematsu, T. Hirahara, "A steerable dummy head that tracks three-dimensional head movement: TeleHead," *Acoustical Science and Technology*, vol.24, no.5, pp.327–329, 2003.

[39] S. Tachi, K. Komoriya, K. Sawada, T. Nishiyama, T. Itoko, M. Kobayashi and K. Inoue, "Telexistence cockpit for humanoid robot control," *Advanced Robotics*, vol.17, no.3, pp.199–217, 2003.

[40] D. Kobayashi, S. Kajita, K. Takeda, and F. Itakura, "Extracting speech features from human speech like noise," *Proceedings of ICSLP'96,* vol.1, pp.418–421, 1996.

[41] http://sound.media.mit.edu/KEMAR.html

[42] K. Furuya and Y. Kaneda, "Two-channel blind deconvolution of nonminimum phase FIR system," *IEICE Trans. Fundamentals,* vol.E80-A, no.5, pp.804–808, 1997.

[43] Y. Huang, J. Benesty, J. Chen, "Optimal step size of the adaptive multichannel LMS algorithm for blind SIMO identification," *IEEE Trans. on Signal Processing Letter,* vol.12, pp.173–176, Mar. 2005.

[44] Y. Inouye and K. Hirano, "Cumulant-based blind identification of linear multi-input-multi-output systems driven by colored inputs," *IEEE Trans. Signal Processing,* vol.45, no.6, pp.1543–1552, June 2000.

Appendix

A. Identifiability in the SIMO decomposition processing

A.1 Introduction

In this chapter, we describe the identifiability in the SIMO decomposition processing by SIMO-ICAs and discuss the filter length D used in the proposed SIMO-ICAs.

A.2 Identifiability in the SIMO decomposition processing

In Chapter 4 and 5, it is revealed that SIMO-ICA can decompose the mixed signals $x_k(t) = \sum_{l=1}^{L} A_{kl}(z)s_l(t)$ into the SIMO-model-based signals $A_{kl}(z)s_l(t)$. Recently, SIMO deconvolution processing which can estimate the original source signals $s_l(t)$ from the SIMO-model-based signals $A_{kl}(z)s_l(t)$ has been proposed [42, 43]. Especially, blind multichannel inverse filtering method [42] which have been proposed by Furuya can blindly identify the source signals using FIR filter without errors. Thus, the combination method cascading SIMO-ICA to blind multichannel inverse filtering processing can serve as MIMO deconvolution processing which estimate the original source signals from the mixed signals. However, Inouye revealed that MIMO deconvolution processing cannot identify the original source signals using only an FIR-type separation filter matrix in the case of $L = K$ if we do not assume special constraints [44]. This instance means that MIMO deconvolution cascading SIMO-ICA to blind multichannel inverse filtering processing using FIR filter cannot hold the identifiability, and SIMO-ICA cannot identify SIMO-model-based signals because the blind multichannel inverse filtering processing has the source identifiability.

In practice, however, we can reduce the residuals by setting the separation filter length D to be sufficiently long in the SIMO-ICA; this can be shown in the next simulation. Thus, the SIMO-model-based signals are approximately reproduced in this case. Overall, the identifiability almost holds under the assumption that we are allowed to use the long FIR filters in SIMO-ICA.

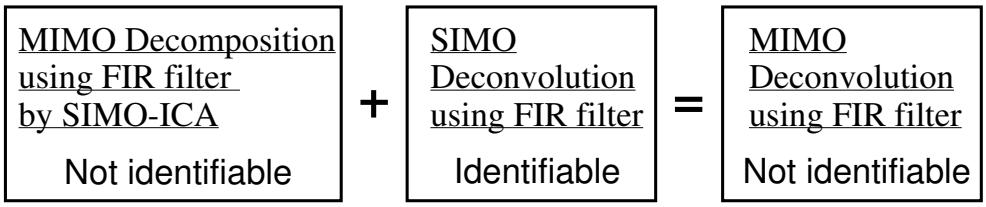

Figure 77. Relation between SIMO deconvolution, MIMO-deconvolution, and SIMO-ICA for the square case (the number of sources = the number of microphones).

A.3 Simulation in Artificial Mixing Condition

A.3.1 Conditions for Experiment

The mixing filter matrix $\boldsymbol{A}(z)$ is taken to be

$$A_{11}(z) = 1 - 0.7z^{-1} - 0.3z^{-2}, \tag{111}$$
$$A_{21}(z) = z^{-1} + 0.7z^{-2} + 0.4z^{-3}, \tag{112}$$
$$A_{12}(z) = z^{-1} + 0.7z^{-2} + 0.4z^{-3}, \tag{113}$$
$$A_{22}(z) = 1 - 0.7z^{-1} - 0.3z^{-2}. \tag{114}$$

The impulse responses $a_{ij}(n)$ corresponding to $A_{ij}(z)$ are shown in Fig. 78. Two sentences spoken by two male speakers are used as the original speech samples $\boldsymbol{s}(t)$. The sampling frequency is 8 kHz and the length of speech is limited to 7 seconds. The number of iterations in ICA is 15000. We evaluate the accuracy of the reproduced SIMO-model-based signals in SIMO-ICA-IG by changing the length of the separation filter, D, from 4 to 64 taps. The step-size parameter α was set in the range of 1×10^{-6}–2×10^{-6}, and we select the optimum step size which gives the best performance.

A.3.2 Results and Discussion in Experiment

Figures 79 and 80 show the elements of a *composite* filter matrix of $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ and $\boldsymbol{a}(n)$. The composite filter matrix represents the whole system characteristics of

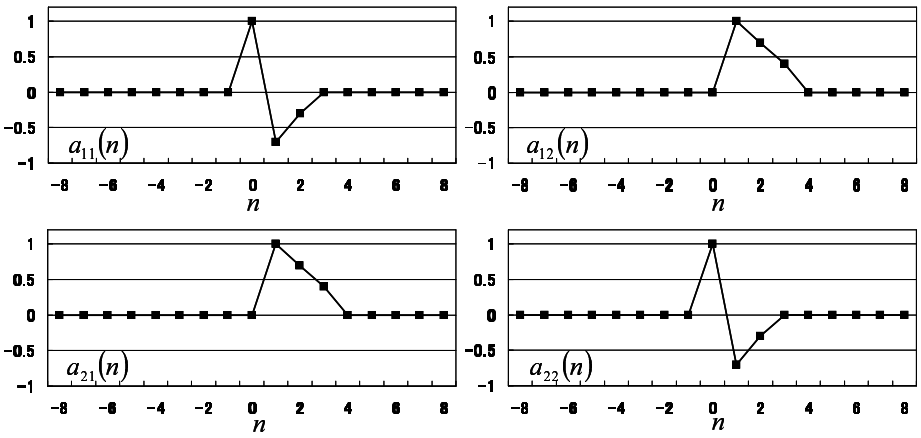

Figure 78. Elements of mixing matrix $\boldsymbol{a}(n)$ used in simulation, which are given by inverse z-transform of (111)–(114).

the mixing and separation processes in SIMO-ICA; this is defined as

$$\left[\tilde{h}_{ij}^{(\mathrm{ICA1})}(n)\right]_{ij} = z^{D/2}\sum_{d=0}^{D-1}\boldsymbol{w}_{(\mathrm{ICA1})}(d)\boldsymbol{a}(n-d), \tag{115}$$

$$\left[\tilde{h}_{ij}^{(\mathrm{ICA2})}(n)\right]_{ij} = z^{D/2}\sum_{d=0}^{D-1}\boldsymbol{w}_{(\mathrm{ICA2})}(d)\boldsymbol{a}(n-d), \tag{116}$$

where $z^{D/2}$ is used to cancel the time delay of $D/2$ in $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ for centering the impulse responses in each of the figures. From these figures, it is confirmed that the composite filter matrix for the first ICA, $[\tilde{h}_{ij}^{(\mathrm{ICA1})}(n)]_{ij}$, becomes diag$[\boldsymbol{a}(n)]$ as the length of the separation filter, D, is increased to more than the length of the mixing system. Also, the composite filter matrix for the second ICA, $[\tilde{h}_{ij}^{(\mathrm{ICA2})}(n)]_{ij}$, becomes off-diag$[\boldsymbol{a}(n)]$ as the length of the separation filter is increased. This obviously indicates that each ICA part in the SIMO-ICA-IG can work so as to separately extract each SIMO component.

Figure 81 shows the results of SA, where the SA increases monotonically as the length of the separation filter, D, becomes large. In particular, the SA of more than 35 dB is achieved when the filter length is set to 64 taps. Thus, the SIMO-ICA can reproduce the SIMO-model-based signals using the sufficiently long filter.

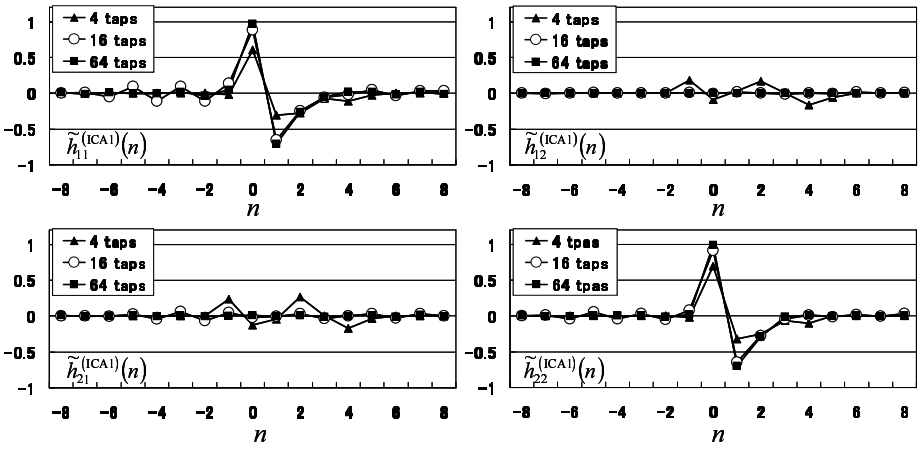

Figure 79. Elements of composite filter matrix $[\tilde{h}_{ij}^{(\mathrm{ICA1})}(n)]_{ij} = z^{D/2}\sum_{d=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA1})}(d)\boldsymbol{a}(n-d)$ for different filter lengths D. This represents the whole system characteristics of the mixing and separation processes in the first ICA part of the SIMO-ICA.

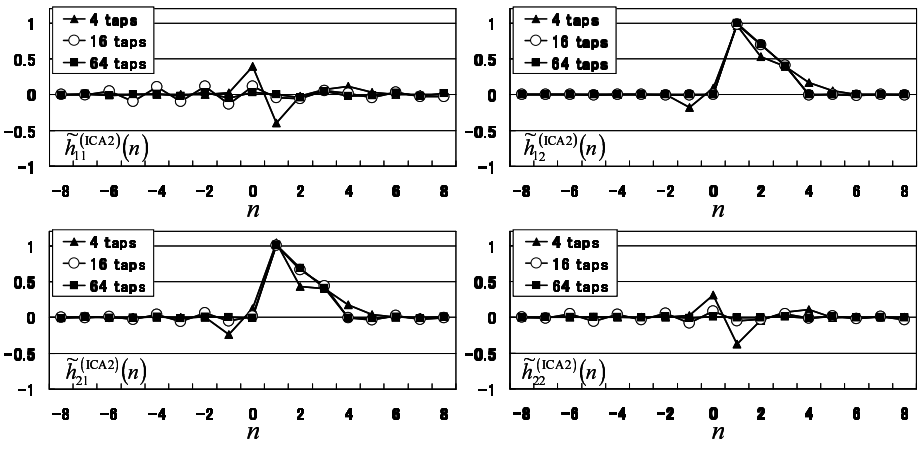

Figure 80. Elements of composite filter matrix $[\tilde{h}_{ij}^{(\mathrm{ICA2})}(n)]_{ij} = z^{D/2}\sum_{d=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA2})}(d)\boldsymbol{a}(n-d)$ for different filter lengths D. This represents the whole system characteristics of the mixing and separation processes in the second ICA part of the SIMO-ICA.

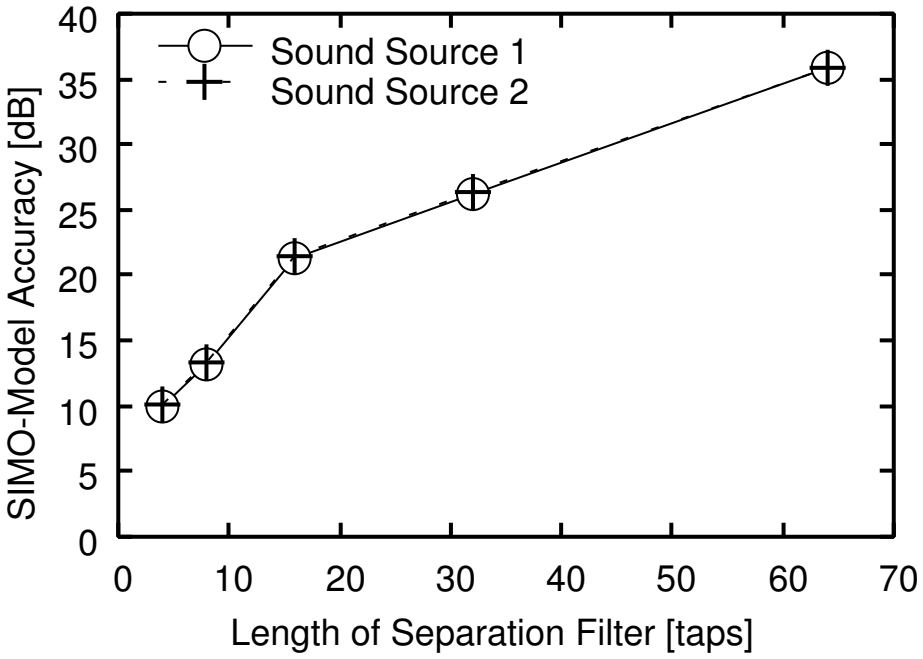

Figure 81. SIMO-model accuracy (SA) of the SIMO-ICA with different filter lengths D. The SA is used as the indication of a degree of similarity between the SIMO-ICA's outputs and the original SIMO-model-based signals.

This result supports the discussion on the identifiability in the proposed method as described in Section A.

B. The derivation of Iterative learning rule in SIMO-ICA-LS

The (standard) gradient of Eq. (36) with respect to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ is given as

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{w}_{(\mathrm{ICA}l)}(n)} \Big\langle \Big\| \sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t - \frac{D}{2}) \Big\|^2 \Big\rangle_t \\
&= \; 2 \Big\langle \Big(\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t - \frac{D}{2}) \Big) \boldsymbol{x}(t-n)^{\mathrm{T}} \Big\rangle_t . \qquad (117)
\end{aligned}$$

Here $\boldsymbol{x}(t-n)^{\mathrm{T}}$ is expressed as the following equation from Eq. (35):

$$\boldsymbol{x}(t-n)^{\mathrm{T}} = \boldsymbol{y}_{\mathrm{ICA}l}(t-n)^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ICA}l}(z)^{-\mathrm{T}}, \tag{118}$$

where the superscript $-\mathrm{T}$ represents the transposed inverse matrix. By using Eq. (118), Eq. (117) is expanded as

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{w}_{(\mathrm{ICA}l)}(n)}\Big\langle \Big\| \sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2})\Big\|^2 \Big\rangle_t \\
&= 2\Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}))\boldsymbol{y}_{(\mathrm{ICA}l)}(t-n)^{\mathrm{T}}\boldsymbol{W}_{(\mathrm{ICA}l)}(z)^{-\mathrm{T}}\Big\rangle_t. \qquad (119)
\end{aligned}$$

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. (119) is rewritten as

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{w}_{(\mathrm{ICA}l)}(n)}\Big\langle \Big\| \sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2})\Big\|^2 \Big\rangle_t \\
&= 2\Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}))\boldsymbol{y}_{(\mathrm{ICA}l)}(t-n)^{\mathrm{T}}\boldsymbol{V}_{(\mathrm{ICA}l)}(z)^{\mathrm{T}}\Big\rangle_t \\
&= 2\sum_{d=0}^{D-1}\Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}))\boldsymbol{y}_{(\mathrm{ICA}l)}(t-n-d)^{\mathrm{T}}\Big\rangle_t \boldsymbol{v}_{(\mathrm{ICA}l)}(d)^{\mathrm{T}} \qquad (120) \\
&= 2\sum_{d=0}^{D-1}\boldsymbol{J}(n+d)\boldsymbol{v}_{(\mathrm{ICA}l)}(d)^{\mathrm{T}}, \qquad (121)
\end{aligned}$$

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

$$\boldsymbol{J}(u) = \Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}))\Big)\boldsymbol{y}_{(\mathrm{ICA}l)}(t-u)^{\mathrm{T}}\Big\rangle_t. \tag{122}$$

Equation (121) is rewritten as

$$\begin{aligned}
2\sum_{d=0}^{D-1}\boldsymbol{J}(n+d)\boldsymbol{v}_{(\mathrm{ICA}l)}(d)^{\mathrm{T}} &= 2\sum_{d=0}^{D-1}\boldsymbol{J}(n)(\boldsymbol{v}_{(\mathrm{ICA}l)}(d)z^{d})^{\mathrm{T}} \qquad (123) \\
&= 2\boldsymbol{J}(n)\boldsymbol{V}_{(\mathrm{ICA}l)}(z^{-1})^{\mathrm{T}} \qquad (124)
\end{aligned}$$

Therefore, the standard gradient of Eq. (36) with respect to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ is given as

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{w}_{(\mathrm{ICA}l)}(n)}\Big\langle \Big\| \sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}) \Big\|^2 \Big\rangle_t \\
&= 2\Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}))\boldsymbol{y}_{(\mathrm{ICA}l)}(t-n)^{\mathrm{T}} \Big\rangle_t \boldsymbol{W}_{(\mathrm{ICA}l)}(z^{-1})^{-\mathrm{T}}. \quad (125)
\end{aligned}$$

From Eq. (125), the natural gradient [11, 28] of Eq. (36) is given as

$$\begin{aligned}
&\left\{ \frac{\partial}{\partial \boldsymbol{w}_{(\mathrm{ICA}l)}(n)}\Big\langle \Big\| \sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}) \Big\|^2 \Big\rangle_t \right\} \boldsymbol{W}_{(\mathrm{ICA}l)}(z^{-1})^{\mathrm{T}} \boldsymbol{W}_{(\mathrm{ICA}l)}(z) \\
&= 2\Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}))\boldsymbol{y}_{(\mathrm{ICA}l)}(t-n)^{\mathrm{T}} \Big\rangle_t \boldsymbol{W}_{(\mathrm{ICA}l)}(z) \quad (126) \\
&= 2\sum_{d=0}^{D-1} \Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-\frac{D}{2}))\boldsymbol{y}_{(\mathrm{ICA}l)}(t-n+d)^{\mathrm{T}} \Big\rangle_t \boldsymbol{w}_{(\mathrm{ICA}l)}(d). \\
& \quad (127)
\end{aligned}$$

Therefore, we have Eq. (4.4) in SIMO-ICA-LS.

List of Publication

Journal Papers

1. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "High-fidelity blind separation using SIMO-model-based ICA," *IEICE Transactions on Fundamentals*, Vol.E87-A, No.8, pp.2063–2072, August 2004.

2. Satoshi Ukai, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, Ryo Mukai, Hiroshi Sawada, "Multistage SIMO-Model-Based Blind Source Separation Combining Frequency-Domain ICA and Time-Domain," *IEICE Transactions on Fundamentals*, Vol.E88-A, No.3, pp.642–650, March 2005.

3. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "A Self-Generator Method for Initial Filters of SIMO-ICA Applied to Blind Separation of Binaural Sound Mixtures," *IEICE Transactions on Fundamentals*, Vol.E88-A, No.7, pp.1673–1682, July 2005.

4. Hiroshi Saruwatari, Hiroaki Yamajo, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Blind Separation and Deconvolution for Convolutive Mixture of Speech Combining SIMO-Model-Based ICA and Multichannel Inverse Filtering," *IEICE Transactions on Fundamentals.* Vol.E88-A, No.9, pp.2387–2400, September 2005.

International Conference

1. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, "Blind Source Separation based on Binaural ICA", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2003)*, pp.V-321–V-324, April 2003.

2. Hiroshi Saruwatari, Tomoya Takatani, Hiroaki Yamajo, Tsuyoki Nishikawa, Kiyohiro Shikano, "Blind Separation and Deconvolution for Real Convolutive Mixture of Temporally Correlated Acoustic Signals Using SIMO-Model-Based ICA," *Proceedings of Fourth International Symposium on Indepen-*

dent Component Analysis and Blind Signal Separation (ICA & BSS 2003), pp.549–554, Nara, April 2003.

3. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "SIMO-Model-Based Independent Component Analysis for High-Fidelity Blind Separation of Acoustic Signals," *Proceedings of Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA & BSS 2003)*, pp.993–998, Nara, April 2003.

4. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "High-Fidelity Blind Separation for Convolutive Mixture of Acoustic Signals Using SIMO-Model-Based Independent Component Analysis," *Proceedings of Seventh International Symposium on Signal Processing and its Applications (ISSPA2003)*, ThuPmPO2, Paris, July 2003.

5. Hiroshi Saruwatari, Hiroaki Yamajo, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Parallel Structured Independent Component Analysis for SIMO-Model-Based Blind Separation and Deconvolution of Convolutive Speech Mixture," *Proceedings of International Joint Conference on Neural Networks (IJCNN2003)*, pp.714–719, July 2003.

6. Hiroaki Yamajo, Hiroshi Saruwatari, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Blind Separation and Deconvolution for Convolutive Mixture of Speech Using SIMO-Model-Based ICA and Multichannel Inverse Filtering," *Proceedings of 8th European Conference on Speech Communication and Technology (Eurospeech2003)*, pp.I-537-540, September 2003.

7. Hiroshi Saruwatari, Hiroaki Yamajo, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Blind Separation and Deconvolution of MIMO System Driven by Colored Inputs Using SIMO-Model-Based ICA with Information-Geometric Learning," *Proceedings of IEEE Neural Network for Signal Processing Workshop 2003 (NNSP2003)*, pp.379-388, September 2003.

8. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "High-Fidelity Blind Separation of Acoustic Signals using SIMO-Model-Based ICA with Information-Geometric Learning," *Proceedings of Interna-*

tional Workshop on Acoustic Echo and Noise Control 2003 (IWAENC2003), pp.251–254, Kyoto, September 2003.

9. Hiroaki Yamajo, Hiroshi Saruwatari, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Evaluation of Blind Separation and Deconvolution for Convolutive Speech Mixture Using SIMO-Model-Based ICA," *Proceedings of International Workshop on Acoustic Echo and Noise Control 2003 (IWAENC2003)*, pp.299-302, September 2003.

10. Hiroshi Saruwatari, Hiroaki Yamajo, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Blind Separation and Deconvolution of MIMO-FIR System with Colored Sound Inputs Using SIMO-Model-Based ICA," *Proceedings of 2003 IEEE Workshop on Statistical Signal Processing (SSP2003)*, pp.421-424, September 2003.

11. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "SIMO-Model-Based ICA with Information-Geometric Learning for High-Fidelity Blind Source Separation," *Proceedings of 2003 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2003)*, pp.108–113, Hyogo, December 2003.

12. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Comparison between SIMO-ICA with least mean squares criterion and SIMO-ICA with information-geometric learning," *Proceedings of 18th International Congress on Acoustics (ICA2004)*, pp.I-329–I-332, Kyoto, April 2004.

13. Satoshi Ukai, Hiroshi Saruwatari, Tomoya Takatani, Ryo Mukai, Hiroshi Sawada, "Multistage SIMO-Model-Based Blind Source Separation Combining Frequency-Domain ICA and Time-Domain ICA," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2004)*, pp.IV-109–IV-112, Montreal, May 2004.

14. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind Separation of Binaural Sound Mixtures Using SIMO-Model-Based

Independent Component Analysis," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2004)*, pp.IV-113–IV-116, Montreal, May 2004.

15. Hiroaki Yamajo, Hiroshi Saruwatari, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Evaluation of Blind Separation and Deconvolution for Binaural-Sound Mixtures Using SIMO-Model-Based ICA," *Proceedings of 12th European Signal Processing Conference (EUSIPCO2004)*, pp.1709–1712, September 2004.

16. Satoshi Ukai, Hiroshi Saruwatari, Tomoya Takatani, Kiyohiro Shikano, Ryo Mukai, Hiroshi Sawada, "Evaluation of Multistage SIMO-Model-Based Blind Source Separation Combining Frequency-Domain ICA and Time-Domain ICA," *Proceedings of 5th International Conference on Independent Component Analysis and Blind Signal Separation* pp.626–633, September 2004.

17. Yoshitsugu Manabe, Hiroshi Saruwatari, Muneyuki Sakata, Tomoya Takatani, Yoshihiro Yasumuro, Masataka Imura, Kiyohiro Shikano, Kunihiro Chihara, "Wearable Computing for Virtualizing Real Space," *3rd CREST/ISWC Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing*, pp.75–83, Arlington, October 2004.

18. Tomoya Takatani, Satoshi Ukai, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind Sound Scene Decomposition for Audio Virtual Reality," *3rd CREST/ISWC Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing*, pp.94–98, Arlington, October 2004.

19. Hiroshi Saruwatari, Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, "Two-Stage Blind Source Separation Using SIMO-ICA and Binary Masking," *2005 Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA2005).*

20. Satoshi Ukai, Tomoya Takatani, Hiroshi Saruwatari, Tsuyoki Nishikawa, Kiyohiro Shikano, "Blind Source Separation Combining SIMO-model-based ICA and Adaptive Beamforming," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2005 (ICASSP2005).*

21. Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, "Blind Separation of Convolutive Speech Mixtures Using SIMO-Model-Based ICA and Binary Mask Processing," *International Workshop on Nonlinear Signal and Image Processing (NSIP 2005)*, pp.187 − 192, Sapporo, May 2005.

22. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind Decomposition of Binaural Mixed Signals Using High-Convergence Algorithm Combining SIMO-ICA and DOA estimation," *Proceedings of International Workshop on Nonlinear Signal and Image Processing (NSIP2005)*, pp.197–202, Sapporo, May 2005.

23. Hiroshi Saruwatari, Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Kiyohiro Shikano,Takashi Hiekata, Takashi Morita, "Two-Stage Blind Source Separation Based on ICA and Binary Masking for Real-Time Robot Audition System," *IEEE/RSJ International Conference on Inteligence Robots and Systems (IROS2005)*, pp.209 − 214, Edmonton, August 2005.

24. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind Sound Scene Decomposition for Robot Audition Using SIMO-Model-Based ICA," *Proceedings of IEEE/RSJ International Conference on Inteligence Robots and Systems (IROS2005)*, pp.215–220, Edmonton, August 2005.

25. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind Separation of Binaural Sound Mixtures Using SIMO-ICA with Self-Generator for Initial Filter," *Proceedings of European Signal Processing Conference (EUSIPCO2005)*, TueAmPO2-7, Antalya, September 2005.

26. Hiroshi Saruwatari, Satoshi Ukai, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Two-Stage Blind Source Separation Combining SIMO-Model-Based ICA and Adaptive Beamforming," *13th European Signal Processing Conference (EUSIPCO2005)*, TueAmPO2-12, Antalya, September 2005

27. Yoshimitsu Mori, Hiroshi Saruwatari, Tomoya Takatani, Satoshi Ukai, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, "Real-Time Implementation of Two-Stage Blind Source Separation Combining SIMO-ICA and Binary Masking," *International Workshop on Acoustic Echo and Noise Control (IWAENC2005)*, pp.229 − 232, September 2005

28. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of SIMO Separation Methods for Blind Decomposition of Binaural Mixed Signals," *Proceedings of 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC2005)*, pp.233 − 236, Eindhoven, September 2005.

29. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of A Self-Generator Method for Initial Filters of SIMO-ICA Applied to Blind Separation of Binaural Sound Mixtures," *Proceedings of 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA2005)*, pp.13 − 16, New Paltz, October 2005.

Invited Talk

1. Hiroshi Saruwatari, Tomoya Takatani, Kiyohiro Shikano, "SIMO-Model-Based Blind Acoustic Signal Separation: Concept and Its Application," *Proceedings of Special Workshop in MAUI (SWIM)*, Hawai, Janurary 2004.

2. Hiroshi Saruwatari, Tsuyoki Nishikawa, Tomoya Takatani, Kiyohiro Shikano, "Recent advance in acoustic blind separation," *2005 IEICE General Conference of the Institute of Electrics, Information and Communication Engineers*, March 2005 (In Japanese).

3. Hiroshi Saruwatari, Hiroaki Yamajo, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Blind Separation and Deconvolution of MIMO-FIR System with Colored Inputs Based on SIMO-Model-Based ICA," *2005 IEEE/URSI AP-S International Symposium*, July 2005.

Technical Reports

1. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "High-Fidelity Blind Source Separation Using SIMO-Model-Based ICA," *IEICE Technical Report*, no.EA2002-108, pp.19–24, Osaka, January 2003.

2. Hiroaki Yamajo, Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind Separation and Deconvolution of Convolutive Speech Mixture Using SIMO-Model-Based ICA and Multichannel Inverse Filtering," *IEICE Technical Report*, no.EA2003-11, pp.19-24, Tokyo, April 2003.

3. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "SIMO-Model-Based ICA with Infomax constraint for High-Fidelity Blind Source Separation," *IEICE Technical Report*, no.EA2003-12, pp.25–30, Tokyo, April 2003.

4. Hiroshi Saruwatari, Tomoya Takatani, Kiyohiro Shikano, "Blind Sound Scene Decomposition and Reconstruction for Audio Virtual Reality," *Proceedings of 2nd CREST Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing*, pp.90–96, May 2003.

5. Hiroshi Saruwatari, Hiroaki Yamajo, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Blind Separation and Deconvolution of Convolutive Speech Mixture Using SIMO-Model-Based ICA with Information-Geometric Learning," *IEICE Technical Report*, no.EA2003-46, pp.25–32, Kyoto, June 2003.

6. Tomoya Takatani, "Blind Source Separation of Acoustic Signals Using SIMO-ICA for Audio Augmented Reality," *2003 NAIST COE Symposium*, pp.69–70, Nara, October 2003.

7. Satoshi Ukai, Hiroshi Saruwatari, Tomoya Takatani, Kiyohiro Shikano, Ryo Mukai, Hiroshi Sawada "Evaluation of Blind SIMO-Model-Signal Extraction combining Frequency-Domain ICA and Time-Domain ICA," *IEICE Technical Report*, no.EA2004-23, pp.37–42, Tokyo, June 2004.

8. Satoshi Ukai, Hiroshi Saruwatari, Tomoya Takatani, Kiyohiro Shikano "Blind Source Separation Using SIMO-Model-Signal Extraction and Adaptive Beamforming," *IEICE Technical Report*, no.EA2004-24, pp.43–48, Tokyo, June 2004.

9. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "High-Convergence Algorithm Combining SIMO-ICA and DOA estimation for Blind Binaural Signal Separation," *IEICE Technical Report*, no.EA2004-43, pp.31–36, Sendai, August 2004.

10. Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, "Blind Source Separation Combining SIMO-Model-Based ICA and Binary Mask Processing," *IEICE Technical Report*, no.EA2004-115, pp.71–76, Tokyo, December 2004.

11. Satoshi Ukai, Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of Blind Source Separation Combining SIMO-Model-Based ICA and Adaptive Beamforming", *Smart Infomedia System (SIS) Symposium*, no.SIS2004-48, pp.33–38, Osaka, December 2004.

Meetings

1. Tsuyoki Nishikawa, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, Shoko Araki, Shoji Makino, "Comparison of time-domain ICA methods based on minimization of KL divergence and simultaneous decorrelation of nonstationary signal," The Meeting of ASJ, 2-5-14, pp.545–546, September 2002 (in Japanese).

2. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Speech enhancement using blind source separation based on minimal distortion principle," The Meeting of ASJ, 2-5-15, pp.547–548, September 2002 (in Japanese).

3. Hiroaki Yamajo, Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind source separation and deconvolution using SIMO-model-based ICA and blind multichannel inverse filtering," The Meeting of ASJ, 2-8-11, pp.673–674, March 2003 (in Japanese).

4. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind source separation using SIMO-model-based ICA," The Meeting of ASJ, 2-8-13, pp.677–678, March 2003 (in Japanese).

5. Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind source separation using SIMO-model-based ICA with information-geometric learning," The Meeting of ASJ, 2-5-14, pp.537–538, September 2003 (in Japanese).

6. Hiroaki Yamajo, Hiroshi Saruwatari, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Evaluation of blind source separation and deconvolution using SIMO-model-based ICA with information-geometric learning," The Meeting of ASJ, 2-5-15, pp.539–540, September 2003 (in Japanese).

7. Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, "Wearable Audio Interface Based on Audio Augmented Reality," SICE System Integration Division Annural Conference (SI2003), 3G4-3, pp.251, December 2003 (in Japanese).

8. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind source separation for binaural mixed signals using SIMO-model-based independent component analysis," The Meeting of ASJ, 1-6-18, pp.497–498, March 2004 (in Japanese).

9. Hiroaki Yamajo, Hiroshi Saruwatari, Tomoya Takatani, Tsuyoki Nishikawa, Kiyohiro Shikano, "Separation and Equalization for colored signal mixtures with HRTF using two-stage blind separation and deconvolution," The Meeting of ASJ, 2-10-1, pp.545–546, March 2004 (in Japanese).

10. Satoshi Ukai, Hiroshi Saruwatari, Tomoya Takatani, Kiyohiro Shikano, Ryo Mukai, Hiroshi Sawada, "Extraction of SIMO-model-based signals using frequency-domain ICA and time-domain ICA," The Meeting of ASJ, 2-10-4, pp.551–552, March 2004 (in Japanese).

11. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "High-convergence algorithm using SIMO-ICA and DOA

estimation for blind binaural signal separation," The Meeting of ASJ, 3-Q-13, pp.739–740, September 2004 (in Japanese).

12. Satoshi Ukai, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind Speech signal separation combining SIMO-model-based signal extraction and supervised adaptive beamforming," The Meeting of ASJ, 3-Q-18, pp.749–750, September 2004 (in Japanese).

13. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of SIMO-ICA with self-generator for initial filter," The Meeting of ASJ, 1-6-10, pp.427–428, March 2005 (in Japanese).

14. Satoshi Ukai, Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of blind source separation using SIMO-model-based extraction and adaptive beamforming," The Meeting of ASJ, 1-6-11, pp.429–430, March 2005 (in Japanese).

15. Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, "Two-stage blind source separation using SIMO-model-based ICA and binary masking processing," The Meeting of ASJ, 1-6-12, pp.431–432, March 2005 (in Japanese).

16. Shigeki Miyabe, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, "Speech recognition using barge-in free spoken dialogue interface based on response sound cancellation," The Meeting of ASJ, 3-Q-23, pp.551–552, March 2005 (in Japanese).

17. Tomoya Takatani, Satoshi Ukai, Tsuyoki Nishikawa, Hiroshi Saruwatari, Kiyohiro Shikano, "Audio Augmented Reality System Using SIMO-ICA," The Meeting of ASJ, 2-2-5, pp.595–596, September 2005 (in Japanese).

18. Masayuki Shimada, Shigeki Miyabe, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, "Robust Sound Field Reproduction against User s Move Based on Multi-Channel Inverse Filtering with Selected and Enhanced Secondary Sources," The Meeting of ASJ, 3-Q-12, pp.671–672, September 2005 (in Japanese).

19. Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita "The Evaluation of Real-Time Blind Source Separation Combining SIMO-ICA and Binary Mask Processing," The Meeting of ASJ, 3-Q-27, pp.701–702, September 2005 (in Japanese).

20. Kazuyuki Hayashi, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, "Overdetermined blind separation of acoustic signals using SIMO-model-based ICA," The Meeting of ASJ, 3-Q-28, pp.703–704, September 2005 (in Japanese).

21. Shigeki Miyabe, Tomoya Takatani, Yoshimitsu Mori, Hiroshi Saruwatari, Kiyohiro Shikano, Yosuke Tatekura "Barge-in free spoken dialogue interface using sound field control and blind source separation," The Meeting of ASJ, 3-Q-30, pp.707–708, September 2005 (in Japanese).