NAIST-IS-DD0261017

Doctoral Dissertation

Blind Source Separation Based on Multistage Independent Component Analysis

Tsuyoki Nishikawa

March 24, 2005

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

Tsuyoki Nishikawa

Thesis Committee:

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

Blind Source Separation Based on Multistage Independent Component Analysis*

Tsuyoki Nishikawa

Abstract

A hands-free speech recognition system and a hands-free telecommunication system are essential for realizing an intuitive, unconstrained, and stress-free human-machine interface. In real acoustic environments, however, the speech recognition performance and a speech recording performance significantly degraded because we cannot detect the user's speech with a high *signal-to-noise ratio (SNR)* owing to the interference signals such as noise. In this thesis, we introduce blind source separation (BSS), which is an approach for estimating original source signals only from the information of the mixed signals observed in each input channel. Many BSS methods based on independent component analysis (ICA) have been proposed for the acoustic signal separation. However, the performances of these methods degrade seriously particularly under extreme reverberant conditions.

The ICA-based BSS can be classified into two groups in terms of the processing domain, i.e., frequency-domain ICA (FDICA) and time-domain ICA (TDICA) From the experimental study using the conventional FDICA, the source-separation performance is saturated because the independence assumption collapses in each narrow-band. In TDICA, the convergence degrades because the iterative learning rule becomes more complicated as the reverberation increases. In order to resolve the problems, I newly propose multistage ICA (MSICA), in which FDICA and TDICA are cascaded. In the proposed method, the separated signals of FDICA

* Doctoral Dissertation, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0261017, March 24, 2005.

are regarded as the input signals for TDICA, and we can remove the residual crosstalk components of FDICA by using TDICA. The experimental results in the convolutive speech mixtures reveal that the separation performance and the speech recognition performance of the proposed method are superior to those of TDICA- and FDICA-based BSS methods.

In the original MSICA, we assume the specific mixing model, where the number of microphones is equal to that of sources. However, additional microphones are required to achieve an improved separation performance. This leads to alternative problems, e.g., a complication of the permutation problem. In order to solve them, we propose a new extended MSICA using subarray processing, where the number of microphones and that of sources are set to be the same in every subarray. The experimental results reveal that the separation performance of the proposed MSICA using subarray processing is improved as the number of microphones is increased.

In the speech recognition system and telecommunication system, not only a high SNR but also a high speech quality is required. For speech signals, we must use TDICA with a nonholonomic constraint to avoid the decorrelation effect caused by the holonomic constraint. However, the stability cannot be guaranteed in the nonholonomic case. To solve the problem, the linear predictors estimated from the roughly separated signals by FDICA are inserted before the holonomic TDICA as a prewhitening processing, and the dewhitening is performed after TDICA. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the pre/dewhitening processing prevents the decorrelation. Moreover, to achieve a stable learning and low-distortion in the model where the number of microphones is larger than that of sources, an extended learning algorithm is newly proposed. The experimental results revealed that the proposed algorithm provides the higher stability and the higher separation performance.

Keywords:

hands-free, microphone array, blind source separation, frequency-domain independent component analysis, time-domain independent component analysis, convolutive mixture

多段型独立成分分析に基づくブラインド音源分離*

西川 剛樹

内容梗概

ユーザーに優しいヒューマンマシンインタフェースとして，ハンズフリー音声認識システムやハンズフリー通話システムの実現が期待されている．実環境下では，種々の背景雑音や反射・残響が存在するため，接話型マイクロホンを用いた場合に比べユーザーの音声を高い信号対雑音比 *(Sigal-to-Noise Ratio: SNR)* で集音できなくなる．そのため高精度な音声認識を実現することが困難になる．本論文では，目的音の到来方位や適応処理のための非発話区間情報といった事前情報が不要であるという利点を有するブラインド音源分離 (BSS) を導入する．BSS とはマイクロホンアレーで受音された観測信号のみから目的音を分離する技術であり，音源信号同士の独立性を用いた独立成分分析 (ICA) に基づく手法が広く用いられている．しかし，ICA の研究では通常の室内環境下での評価が十分にされていなかった．そこで本論文では，従来 ICA の実環境下における性能評価を行い適用限界を調査する．そしてさらに高精度かつ高品質に目的音を分離するための新たな分離手法の提案を行う．

BSS は分離フィルタを推定する領域の違いで，周波数領域において分離する周波数領域 ICA（FDICA）と時間領域において分離する時間領域 ICA（TDICA）に分類される．実環境下における実験結果より，FDICA に基づく BSS では，帯域分割数を過度に増やすと狭帯域信号間の独立性の仮定が成立しなくなるという問題により，長い残響を含む音の分離は困難であることが新たに確認された．一方，TDICA に基づく BSS においては，分離フィルタの反復学習における低収束性により，長い残響時間を有する混合系へ適用することは非常に困難であることが確認された．そこで，本提案手法においては，FDICA によって分離された信号を TDICA の入力とみなし，FDICA における残留クロストーク成分を TDICA によって分離することによりこれらの問題を解決する．実環境下での音源分離実

* 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-DD0261017, 2005 年 3 月 24 日.

験より，提案手法はTDICA及びFDICAに比べ高い音源分離性能及び音声認識率を実現できた．よって，残響環境下において，MSICAに基づくBSSはTDICA及びFDICAに基づくBSSよりも有効であることが確認された．

提案した元のMSICAでは，マイクロホンと音源は同数であるというモデルを仮定していた．残響環境下において，さらに高精度な音源分離を実現するためには，マイクロホン数を増やす必要がある．しかし，素子数を増やすことで，FDICA部において音源の入れ替わり問題が複雑になる，などの問題が生じる．この問題を解決するために，音源と同数のマイクロホンにより構成されたサブアレーを用いたMSICAに基づく優決定BSSを提案する．提案法では，サブアレーは音源とマイクロホンは同数であるのでこの問題が生じない．実環境下における音源分離実験より，提案法はマイクロホン数を増やすことで分離性能が向上するということが確認された．

ハンズフリー音声認識やハンズフリー通話では，妨害音の抑圧性能だけでなく分離信号の音質もまた重要となる．音声信号のように時間的に相関のある信号に対しては，ホロノミック拘束による無相関化を避けるために，非ホロノミック拘束のTDICAを適用しなければならない．しかしながら，非ホロノミックの場合，安定性は保証されない．この問題を解決するために，ホロノミックTDICAの前にFDICAによってある程度分離された信号から推定された線形予測器を挿入しprewhiteningを行う．そしてTDICAの出力信号に対し，dewhiteningを行う．提案手法はホロノミック拘束により安定性が保証されており，pre/dewhitening処理により無相関化を防ぐことができる．本提案手法では，マイクロホンと音源は同数であるというモデルを仮定していたため優決定BSSに適用することが困難であった．この問題を解決するために，ホロノミック拘束のTDICAによる分離信号の白色化に寄与する成分を推定し，推定された処理歪み成分を用いて分離信号の音質を改善する．残響環境下における実験結果より，提案手法は従来のホロノミック拘束やノンホロノミック拘束に基づくTDICAに比べ高い安定性を有し，高い分離性能を実現できることが確認された．

キーワード

ハンズフリー，マイクロホンアレー，ブラインド音源分離，周波数領域独立成分分析，時間領域独立成分分析，畳み込み混合

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

A hands-free speech recognition system [1] (see Fig. 1) and a hands-free telecommunication system (see Fig. 2) are essential for realizing an intuitive, unconstrained, and stress-free human-machine interface. In the system, not only the user's speech but also interference such as interfering speech, background noise and music, is detected by the microphone of the system. Thus, we cannot detect the user's speech with a high *signal-to-noise ratio (SNR)* compared with the case in that we use a close-talking microphone such as a headset microphone (see Fig. 3 (a)) and a hand microphone (see Fig. 3 (b)).

The methods for establishing a noise-robust speech recognition system can be classified into two groups: methods based on a single-channel input, and those based on multichannel inputs. As single-channel types of source separation and speech enhancement techniques [2], a method of tracking a formant structure [3], the organization technique for hierarchical perceptual sounds [4], and a method based on auditory scene analysis [5] have been proposed. As multichannel-type source separation, the method based on array signal processing, e.g., a microphone array system (see Fig. 4), is one of the most effective techniques [6]. In this system, the directions of arrival (DOAs) of the sound sources are estimated and then each of the source signals is separately obtained using the directivity of the array. The delay-and-sum (DS) array [7] and the adaptive beamformer (ABF) [8, 9, 10] are conventional and popular microphone arrays currently used for source separation and noise reduction.

While the DS array has a simple structure, it nevertheless requires a large number of microphones to achieve high performance, particularly in the low-frequency regions. Thus, the degradation of separated signals at low frequencies cannot be avoided in these array systems. The ABF has the following drawbacks. (1) The look direction for each signal separated is necessary in the adaptation process. Thus, the DOAs of the separated sound source signals must be previously known. (2) The adaptation procedure should be performed during breaks in the target signal to avoid any distortion of separated signals. However, we cannot previously estimate signal breaks in conventional use. The above-mentioned

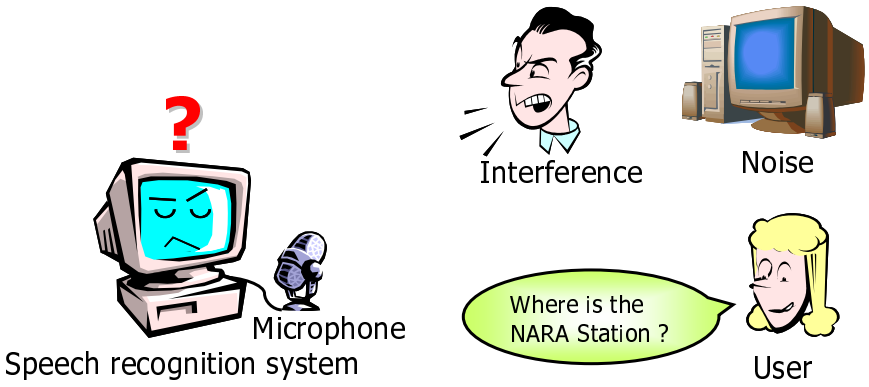


Figure 1. Configuration of hands-free speech recognition system.

requirements arise from the fact that the conventional ABF is based on *supervised* adaptive filtering, and this significantly limits the applicability of the ABF to source separation in practical applications.

In recent years, alternative source-separation approaches have been proposed by researchers using not array signal processing but a specialized branch of information theory, i.e., information-geometry theory [11, 12]. 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 based on *unsupervised* adaptive filtering [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. The early contributory works on BSS were performed by Cardoso and Jutten [13, 14], where high-order statistics of the signals are used for measuring the independence. Common has clearly defined the term *independent component analysis* (ICA), and presented an algorithm that measures independence among the source signals [15]. This report on ICA was later followed by Bell and Sejnowski, where ICA extended to the informax (or the maximum-entropy) algorithm for BSS which is based on a minimization of mutual information of the signals [16].

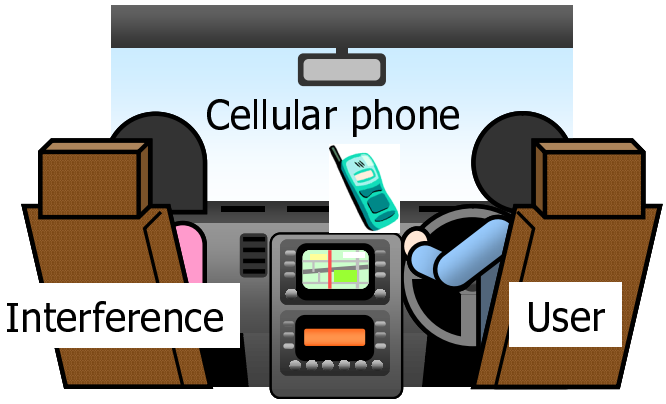
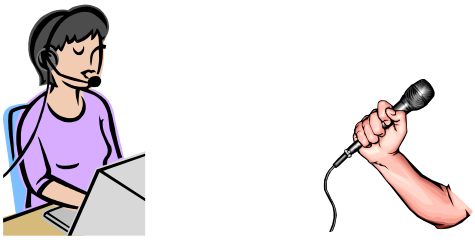


Figure 2. Configuration of hands-free speech telecommunication system.

1.2 Thesis Scope

1.2.1 Problems of Conventional ICAs

The BSS methods based on ICA [15, 16] can be classified into two groups in terms of the processing domain, i.e., frequency-domain ICA (FDICA) in which the complex-valued inverse of the mixing matrix is calculated in the frequency domain [17, 18, 19, 30, 31, 20], and time-domain ICA (TDICA) in which the separation system of the FIR-filter matrix is calculated in the time domain [11, 21, 22, 23, 24]. The recently developed BSS techniques can achieve a good source-separation performance under artificial or short reverberant conditions. However, the performances of these methods under extreme reverberant conditions significantly degrade because of the following problems. (1) In conventional FDICA, the source-separation performance is saturated before reaching a sufficient performance because we transform the fullband signals into the narrow-band signals and the independence assumption collapses in each narrow-band [25]. (2) In TDICA, the convergence degrades because the iterative learning rule becomes more complicated as the reverberation increases.



(a) Headset microphone (b) Hand microphone

Figure 3. Close-talking microphones. (a) a headset microphone, (b) a hand microphone.

1.2.2 New combination framework using FDICA and TDICA

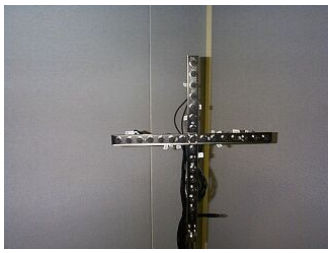
In order to achieve a superior separation performance, we propose a new BSS algorithm called multistage ICA (MSICA), in which FDICA and TDICA are combined. In the proposed method, the separated signals of FDICA are regarded as the input signals for TDICA, and we can remove the residual crosstalk components of FDICA by using TDICA. By using the proposed method, we can achieve a superior source-separation performance and an improvement of the speech recognition performance even under extreme reverberant conditions.

In the original MSICA, we assumed the specific mixing model, where the number of microphones is equal to that of sources. However, additional microphones are required to achieve an improved separation performance under reverberant environments. This leads to alternative problems, e.g., a complication of the permutation problem. In order to solve them, we propose a new extended MSICA using subarray processing, where the number of microphones and that of sources are set to be the same in every subarray. By using the proposed method, the separation performance of the proposed MSICA is improved as the number of microphones is increased.

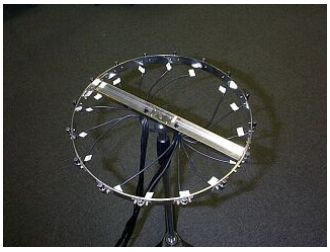
1.2.3 Improvement of Stability of Learning and Sound Quality

For temporally correlated signals such as speech signals, we must use TDICA with a nonholonomic constraint to avoid the decorrelation effect from the holonomic

(a) Linear type microphone array



(b) Cross type microphone array



(c) Circle type microphone array

(b) Plain type microphone array

Figure 4. Various types of microphone arrays. (a) a linear type microphone array, (b) a cross type microphone array, (c) a circle type microphone array, and (d) a plain type microphone array.

constraint. However, the stability cannot be guaranteed in the nonholonomic case. To solve this problem, linear predictors estimated from the roughly separated signals by FDICA are inserted before the holonomic TDICA as a prewhitening processing, and the dewhitening is performed after TDICA. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the pre/dewhitening processing prevents the decorrelation. By using the proposed method, we can achieve higher stability and separation performance.

We cannot apply the original BSS combining MSICA and linear prediction to overdetermined BSS based on MSICA because the specific mixing model, where the number of microphones is equal to that of sources, was assumed. To solve the problem, we estimate the distortion components by the holonomic constraint and we compensate the sound qualities by using the estimated components. By using the proposed method, we can achieve higher stability and higher separation

performance compared with the conventional TDICA algorithm even when we use many microphones.

1.3 Thesis Overview

The thesis is organized as follows.

First, the sound mixing model of the microphone array is introduced in Sect. 2. Next, we introduce two types of ICA for the convolutive mixture i.e., FDICA and TDICA. In addition to the advantages and disadvantages in FDICA and TDICA are explained.

In Sect. 3, the experimental results of FDICA and TDICA in the real acoustic condition are presented and the experimental analyses of both ICAs are also described.

In Sect. 4, we propose a novel algorithm for BSS, in which FDICA and TDICA are combined to achieve a superior source-separation performance under reverberant conditions. Moreover, we provide a comparison results for the separation performance of FDICA, TDICA, and the proposed method under the real acoustic condition.

In Sect. 5, we proposed a novel extended MSICA using subarray processing, where the number of microphones and that of sources are set to be the same in every subarray to achieve an improved separation performance. Also, we provide comparison results for the separation performance of the conventional FDICA, MSICA, and the proposed method under the real acoustic condition.

In Sect. 6, we newly proposed a stable and low-distortion algorithm combining MSICA and linear prediction for BSS in the case where the number of microphones is equal to that of sources. Moreover, we proposed a novel algorithm with a stability and low-distortion for overdetermined BSS based on MSICA using subarray processing in the case where the number of microphones to be larger than that of sources. Also, we provide comparison results for the separation performance and the sound quality performance of the conventional MSICA including H-TDICA or NH-TDICA and the proposed method under the real acoustic condition.

Finally, we summarize the contributions of this thesis and provide suggestions for future work in Sect. 7.

2. Principle of Blind Source Separation

2.1 Introduction

In this section, first, we explain the sound mixing model of microphone array to define the problem of BSS. Next, we introduce two types of BSS methods based on ICA, i.e., FDICA and TDICA.

In the Sect. 2.2, we explain the sound mixing model of microphone array in a real acoustic environment with a long reverberation. In the Sect. 2.3, first, we describe ICA based on a minimization of Kullback-Leibler divergence (KLD) for instantaneous mixture models. Next, we extend this algorithm into FDICA for convolutive mixture models. Furthermore, we represent the advantages and disadvantages of FDICA. In the Sect. 2.4, we describe two types of TDICA, i.e., TDICA methods based on minimization of KL divergence and simultaneous decorrelation of nonstationary signal. Furthermore, we represent the advantages and disadvantages of TDICA.

2.2 Sound Mixing Model of Microphone Array

In this study, a straight-line array (see Fig. 4 (a)) is assumed. In general, the observed signals in which multiple source signals are convoluted with room impulse responses are obtained by the following equation:

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

where $\boldsymbol{x}(t)$ is the observed signal vector, $\boldsymbol{s}(t)$ is the source signal vector, and $\boldsymbol{a}(\tau)$ is the mixing filter (the impulse response) matrix; these are given as

$$\boldsymbol{x}(t) = [x_1(t), \cdots, x_K(t)]^{\mathrm{T}}, \tag{2}$$

$$\boldsymbol{s}(t) = [s_1(t), \cdots, s_L(t)]^{\mathrm{T}}, \tag{3}$$

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

where P is the length of the impulse response which is assumed to be an FIR-filter of thousands of taps because we introduce a model to deal with the arrival lags

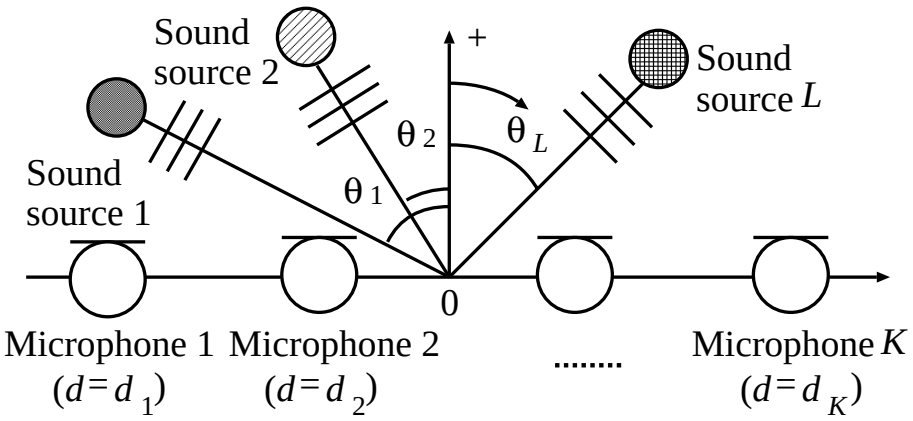


Figure 5. Configuration of microphone array and signals.

among the elements of the microphone array and the room reverberations. Also, K is the number of array elements (microphones), L is the number of multiple sound sources, d_k $(k = 1 \cdots K)$ denotes the coordinates of the elements, and θ_l $(l = 1 \cdots L)$ denotes DOAs for the l-th source $s_l(t)$. $\boldsymbol{A}(z)$ is the z-transform of the mixing filter $\boldsymbol{a}(\tau)$ $(\tau = 0, \cdots, P-1)$; these are given as

$$\boldsymbol{A}(z) = \sum_{\tau=0}^{P-1} \boldsymbol{a}(\tau) z^{-\tau}, \tag{5}$$

$$= \begin{bmatrix} \sum_{\tau=0}^{P-1} a_{11}(\tau) z^{-\tau} & \cdots & \sum_{\tau=0}^{P-1} a_{1L}(\tau) z^{-\tau} \\ \vdots & \ddots & \vdots \\ \sum_{\tau=0}^{P-1} a_{K1}(\tau) z^{-\tau} & \cdots & \sum_{\tau=0}^{P-1} a_{KL}(\tau) z^{-\tau} \end{bmatrix}, \tag{6}$$

where z^{-1} is used as the unit-delay operator for convenience, i.e., $z^{-\tau} \cdot x(t) = x(t-\tau)$.

We can simplify the convolutive mixture (Eq. (1)) down to the following simultaneous mixtures by the frequency transform;

$$\boldsymbol{X}(f) = \boldsymbol{A}(f)\boldsymbol{S}(f), \tag{7}$$

where $\boldsymbol{X}(f)$ is the observed signal vector which is discrete Fourier transform (DFT) of $\boldsymbol{x}(t)$, $\boldsymbol{S}(f)$ is the source signal vector which is DFT of $\boldsymbol{s}(t)$, and $\boldsymbol{A}(f)$

is the mixing matrix which is DFT of $\boldsymbol{a}(\tau)$; these are given as

$$\begin{aligned} \boldsymbol{X}(f) &= [X_1(f), \cdots, X_K(f)]^{\mathrm{T}}, & (8) \\ \boldsymbol{S}(f) &= [S_1(f), \cdots, S_L(f)]^{\mathrm{T}}, & (9) \\ \boldsymbol{A}(f) &= \begin{bmatrix} A_{11}(f) & \cdots & A_{1L}(f) \\ \vdots & \ddots & \vdots \\ A_{K1}(f) & \cdots & A_{KL}(f) \end{bmatrix}, & (10) \end{aligned}$$

where f is arbitrary frequency bin. In this case, $\boldsymbol{A}(f)$ is the mixing matrix which is assumed to be complex-valued because we introduce a model to deal the room reverberations. Generally speaking, the reverberation time (RT) (the time interval in which the decay level drops down by 60 dB) of a typical small room is about 300 ms and large halls have reverberation times between 700 ms and 2.0 s [26]. Also, reverberation times of conference rooms are about 800 ms and Tatami rooms have reverberation times between 400 ms and 600 ms [27]. For instance, we show the impulse responses recorded in a real room with the reverberation time of 300 ms (see Fig. 6).

Figure 7 shows the general sound mixture procedure in a real acoustic environment. In this environment, there are two speakers and we use a two-element array. Figure 8 (a) and (b) show the source waveforms of two speakers. In real acoustic (reverberant) environments, multiple source signals are convoluted with room impulse responses which include reverberant components and reflection components and the interference signals are mixed. The observed signal recorded by the microphone in this situation is shown in Fig. 9.

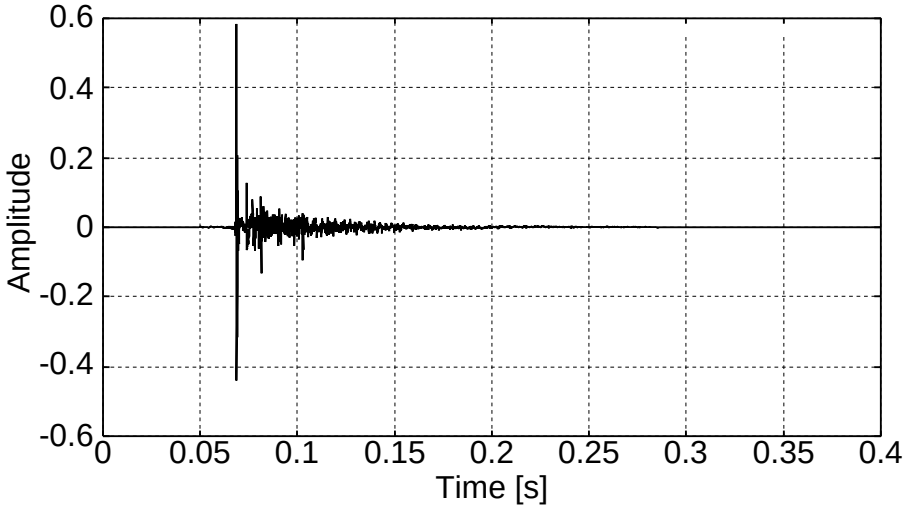


Figure 6. Impulse response recorded in a typical room with the reverberation time of 300 ms.

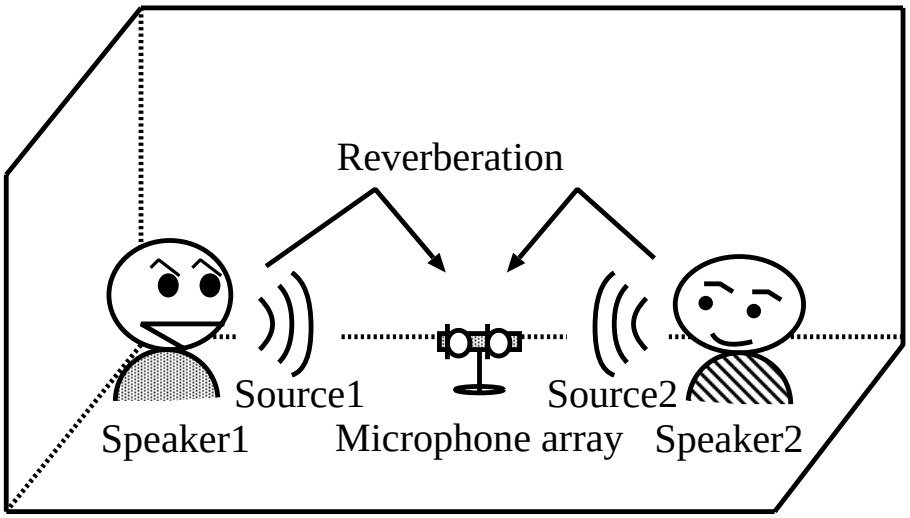


Figure 7. General sound mixture procedure in a real acoustic environment.

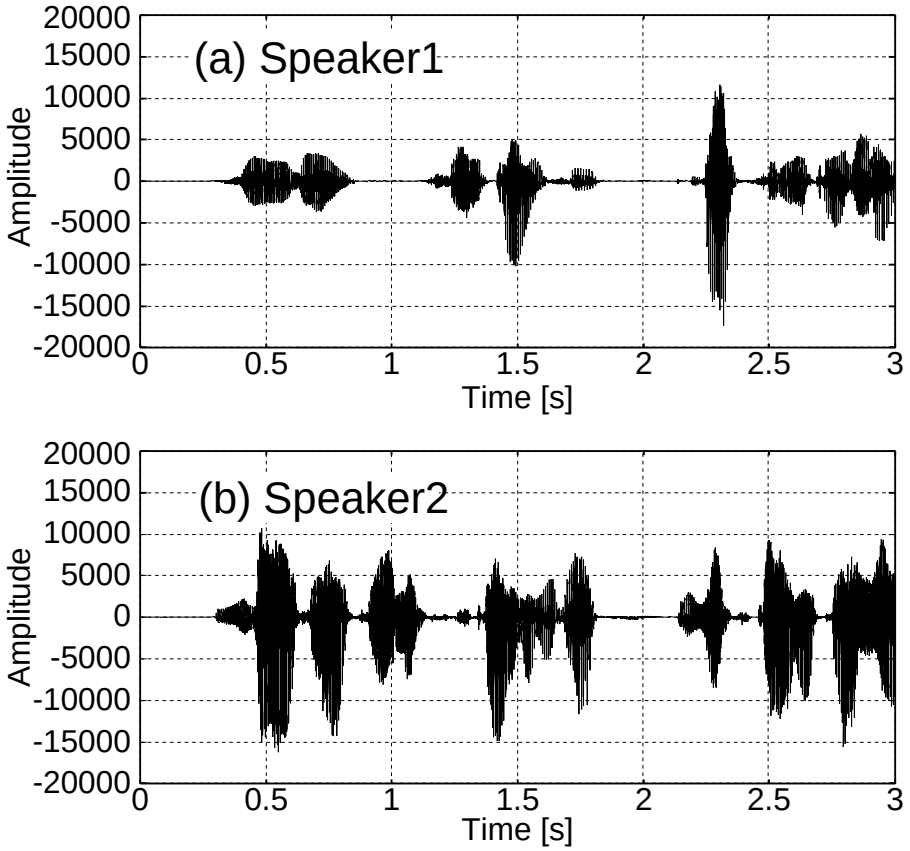


Figure 8. Waveforms of source signals of two speakers.

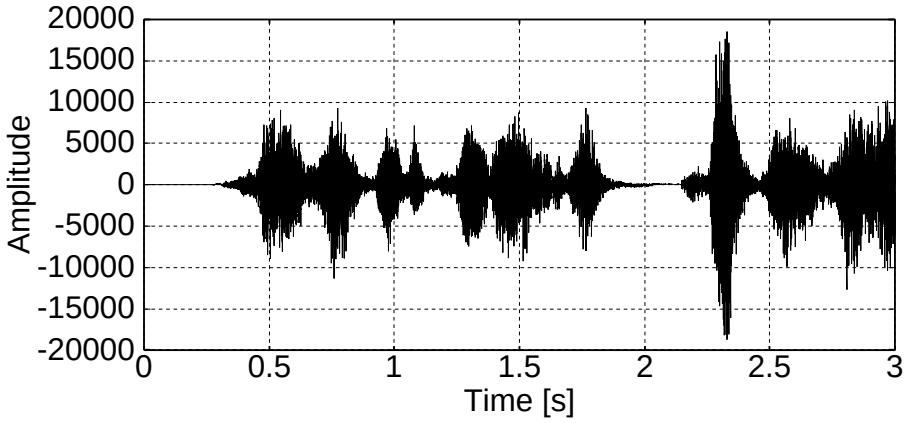


Figure 9. Waveform of observed signal recorded by a microphone.

2.3 FDICA-Based BSS

2.3.1 Calculation of Separation Matrices

We investigate the procedure to optimize the separation matrices based on the minimization of the KLD. In this section, we describe the procedure for real-valued time series in Sect. 2.3.2 preliminarily, and then extend it to the complex-valued case in Sect. 2.3.3.

2.3.2 Minimization of KLD [17]

In this study, we can obtain the separated signals $\hat{\boldsymbol{s}}(t)$ by multiplying the observed signal vector $\hat{\boldsymbol{x}}(t)$;

$$\hat{\boldsymbol{y}}(t) = \boldsymbol{w}\hat{\boldsymbol{x}}(t), \tag{11}$$

where $\hat{\boldsymbol{y}}(t)$ is the separated signal vector, $\boldsymbol{w}$ is the separation matrix formed by real-valued coefficients, $\boldsymbol{x}(t)$ is the observed signal vector obtained by the instantaneous mixture model;

$$\hat{\boldsymbol{x}}(t) = \boldsymbol{a}\boldsymbol{s}(t). \tag{12}$$

where $\boldsymbol{s}(t)$ is the source signal vector (Eq. (3)) and $\boldsymbol{a}$ is the mixing matrix formed by real-valued coefficients. These are given as

$$\hat{\boldsymbol{x}}(t) = [\hat{x}_1(t), \cdots, \hat{x}_K(t)]^{\mathrm{T}}, \tag{13}$$

$$\hat{\boldsymbol{y}}(t) = [\hat{y}_1(t), \cdots, \hat{y}_L(t)]^{\mathrm{T}}, \tag{14}$$

$$\boldsymbol{a} = \begin{bmatrix} a_{11} & \cdots & a_{1K} \\ \vdots & \ddots & \vdots \\ a_{L1} & \cdots & a_{LK} \end{bmatrix}, \tag{15}$$

$$\boldsymbol{w} = \begin{bmatrix} w_{11} & \cdots & w_{1L} \\ \vdots & \ddots & \vdots \\ w_{K1} & \cdots & w_{KL} \end{bmatrix}. \tag{16}$$

Under the assumption that each $s_l(t)$ ($l = 1, \cdots, L$) is stationary and a non-Gaussian process, the separation matrix $\boldsymbol{w}$ is optimized so that $\hat{y}_l(t)$ is mutually independent.

In the case that the joint probability density function (p.d.f) $p(\hat{\boldsymbol{y}}(t))$ of $\hat{\boldsymbol{y}}(t)$ is given by

$$p(\hat{\boldsymbol{y}}(t)) = p(y_1(t), \cdots, y_L(t)). \tag{17}$$

When $\hat{\boldsymbol{y}}(t)$ is mutually independent, the following equation holds:

$$p(\hat{\boldsymbol{y}}(t)) = \prod_{l=1}^{L} q(\hat{y}_l(t)), \tag{18}$$

where $q(\hat{y}_l(t))$ represents the marginal p.d.f of $\hat{y}_l(t)$. The KLD of the joint p.d.f from the product of marginal p.d.f is given as

$$KL(\boldsymbol{w}) = \int p(\hat{\boldsymbol{y}}(t)) \log \frac{p(\hat{\boldsymbol{y}}(t))}{\prod_{l=1}^{L} q(\hat{y}_l(t))} d\hat{\boldsymbol{y}}(t). \tag{19}$$

As we assume that the sources are non-Gaussian, $KL(\boldsymbol{w})$ vanishes if and only if reconstructed signals $\hat{y}_l(t)$ are mutually independent. Thus, we estimate $\boldsymbol{w}$ by minimizing $KL(\boldsymbol{w})$. The partial differentiation of $KL(\boldsymbol{w})$ by the separation matrix $\boldsymbol{w}$ is given by the following equation:

$$\begin{aligned}
\frac{\partial KL(\boldsymbol{w})}{\partial \boldsymbol{w}} &= -\boldsymbol{w}^{-\mathrm{T}} + \int p(\hat{\boldsymbol{x}}(t)) \boldsymbol{\varphi}(\hat{\boldsymbol{y}}(t)) \hat{\boldsymbol{x}}(t)^{\mathrm{T}} d\hat{\boldsymbol{x}}(t), & (20) \\
\boldsymbol{\varphi}(\hat{\boldsymbol{y}}(t)) &= -\left(\frac{\partial \log p(\hat{y}_1(t))}{\partial \hat{y}_1(t)}, \cdots, \frac{\partial \log p(\hat{y}_L(t))}{\partial \hat{y}_L(t)} \right)^{\mathrm{T}} \\
&= -\left(\frac{1}{p(\hat{y}_1(t))} \frac{\partial p(\hat{y}_1(t))}{\partial \hat{y}_1(t)}, \cdots, \frac{1}{p(\hat{y}_L(t))} \frac{\partial p(\hat{y}_L(t))}{\partial \hat{y}_L(t)} \right)^{\mathrm{T}}, & (21)
\end{aligned}$$

thus, using the following gradient,

$$\begin{aligned}
\Delta \boldsymbol{w} &\propto -\frac{\partial KL(\boldsymbol{w})}{\partial \boldsymbol{w}} \\
&= \left[\boldsymbol{w}^{-\mathrm{T}} - \mathrm{E}[\boldsymbol{\varphi}(\hat{\boldsymbol{y}}(t)) \hat{\boldsymbol{x}}(t)^{\mathrm{T}}] \right] \\
&= \left[\boldsymbol{I} - \mathrm{E}[\boldsymbol{\varphi}(\hat{\boldsymbol{y}}(t)) \hat{\boldsymbol{y}}(t)^{\mathrm{T}}] \right] \boldsymbol{w}^{-\mathrm{T}}, & (22)
\end{aligned}$$

we can obtain the optimal $\boldsymbol{w}$ based on the steepest decent method. $E[\cdot]$ denotes the expectation operator. Practically, this expectation is replaced as the

time average. Thus we optimize the $\boldsymbol{w}$ by using the following off-line iterative equations:

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \alpha \left[\boldsymbol{I} - \langle \boldsymbol{\varphi}(\hat{\boldsymbol{y}}(t)) \hat{\boldsymbol{y}}(t)^{\mathrm{T}} \rangle_t \right] \boldsymbol{w}_i^{-\mathrm{T}}, \tag{23}$$

where $\langle \cdot \rangle_t$ denotes the time-averaging operator, i is used to express the value of the i-th step in the iteration, and α is the step size parameter.

Here, the important problem which remains with this approach is how to define the function $\boldsymbol{\varphi}(\hat{\boldsymbol{y}}(t))$. In general, in the case that the amplitude distributions of the source signals such as speech are super-Gaussian (kurtosis ≤ 0), the l-th element of $\boldsymbol{\varphi}(\hat{\boldsymbol{y}}(t))$ is approximated by the following sigmoid function with respect to $\hat{y}_l(t)$ [16, 17];

$$\varphi(\hat{y}_l(t)) = \frac{1}{1 + \exp(-\hat{y}_l(t))}. \tag{24}$$

2.3.3 Extension to Complex-Valued Signal

In the case of the instantaneous model, we can estimate the separation matrix $\boldsymbol{w}$ easily using the iterative equation (Eq. (23)). However, the model in a real acoustic environment with reverberation is a convolutive mixture and we cannot estimate the separation matrix $\boldsymbol{w}$ by Eq. (23). Therefore we simplify the convolutive mixture down to the simultaneous mixtures of the complex-valued coefficients by the frequency transform and we extend the iterative equation (Eq. (23)) into the complex-valued case.

In this procedure, first, the short-time analysis of observed signals is conducted by frame-by-frame DFT. By plotting the spectral values in a frequency bin of each microphone input frame by frame, we consider them as time series. Hereafter, we designate the narrow-band time series as

$$\boldsymbol{X}(f, m) = [X_1(f, m), \cdots, X_K(f, m)]^{\mathrm{T}}, \tag{25}$$

where m is the short-time analysis frame. We apply ICA for these narrow-band observed signals and perform this procedure with respect to all frequency bins. This procedure is called FDICA (see Fig. 10). The signal separation using the complex-valued separation matrix $\boldsymbol{W}(f)$ is performed by the following equation:

$$\boldsymbol{Y}(f, m) \;\; = \;\; \boldsymbol{W}(f) \boldsymbol{X}(f, m), \tag{26}$$

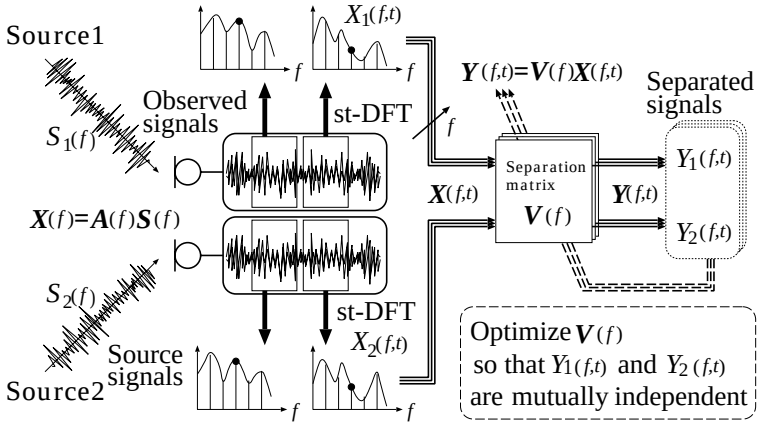


Figure 10. Blind source separation procedure performed in FDICA.

where $\boldsymbol{Y}(f,m)$ is the separated signal vector; these are given as

$$\boldsymbol{Y}(f,m) = [Y_1(f,m),\cdots,Y_L(f,m)]^{\mathrm{T}}, \tag{27}$$

$$\boldsymbol{W}(f) = \begin{bmatrix} W_{11}(f) & \cdots & W_{1K}(f) \\ \vdots & \ddots & \vdots \\ W_{L1}(f) & \cdots & W_{LK}(f) \end{bmatrix}. \tag{28}$$

Furthermore, the iterative equation (23) and $\boldsymbol{\varphi}(\boldsymbol{y}(t))$ in the Eq. (23) are replaced as

$$\boldsymbol{W}_{i+1}(f) = \alpha\left[\boldsymbol{I} - \left\langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,m))\boldsymbol{Y}(f,m)^{\mathrm{H}}\right\rangle_m\right]\boldsymbol{W}_i(f)^{-\mathrm{H}} + \boldsymbol{W}_i(f), \tag{29}$$

$$\boldsymbol{\Phi}\Big(\boldsymbol{Y}(f,m)\Big) = [\Phi(Y_1(f,m)),\cdots,\Phi(Y_L(f,m))]^{\mathrm{T}}, \tag{30}$$

$$\begin{aligned}\Phi\Big(Y_l(f,m)\Big) &= \Phi\Big(\mathrm{Re}[Y_l(f,m)]\Big) + j\cdot\Phi\Big(\mathrm{Im}[Y_l(f,m)]\Big) \\ &= \frac{1}{1+\exp(-\mathrm{Re}[Y_l(f,m)])} + j\frac{1}{1+\exp(-\mathrm{Im}[Y_l(f,m)])},\end{aligned} \tag{31}$$

where $^{-\mathrm{H}}$ denotes the inverse of the Hermitian transposition. Also, $\mathrm{Re}[Y_l(f,m)]$ and $\mathrm{Im}[Y_l(f,m)]$ are the real and the imaginary parts of $Y_l(f,m)$, respectively. In the recent work, the following alternative approximation is mainly used;

$$\Phi(Z_l(f,m)) = \tanh(\mathrm{Re}[Z_l(f,m)]) + j\cdot\tanh(\mathrm{Im}[Z_l(f,m)]). \tag{32}$$

This nonlinear function vanishes the bias component which occurs during the calculation of the time-averaging operation.

In the case in which the signals are real-valued, if w_i in the iterative equation (23) converges in the correct value, the following equation holds:

$$\langle \boldsymbol{\varphi}(\boldsymbol{y}(t))\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t = \boldsymbol{I}, \tag{33}$$

that is,

$$\langle \varphi(y_p(t))y_q(t)\rangle_t = \delta_{pq}, \tag{34}$$

where δ_{pq} is the Kronecker's delta. Based on the relationship in Eq. (34), we consider the average of the matrices $\langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,m))\ \boldsymbol{Y}(f,m)^{\mathrm{H}}\rangle_m$ in Eq. (30). The element of p-th row and q-th column in the matrix can be written as

$$\begin{aligned}
&\Big\langle \Big[\boldsymbol{\Phi}(\boldsymbol{Y}(f,m))\Big]_p \Big[\boldsymbol{Y}(f,m)^{\mathrm{H}}\Big]_q \Big\rangle_m \\
&= \Big\langle \{\Phi\left(\mathrm{Re}[Y_p(f,m)]\right) + j\cdot\Phi\left(\mathrm{Im}[Y_p(f,m)]\right)\} \\
&\quad \cdot \{\mathrm{Re}[Y_q(f,m)] - j\cdot\mathrm{Im}[Y_q(f,m)]\}\Big\rangle_m \\
&= \Big[\langle\Phi\left(\mathrm{Re}[Y_p(f,m)]\right)\mathrm{Re}[Y_q(f,m)]\rangle_m \\
&\quad + \langle\Phi\left(\mathrm{Im}[Y_p(f,m)]\right)\mathrm{Im}[Y_q(f,m)]\rangle_m\Big] \\
&\quad + j\cdot\Big[\langle\Phi\left(\mathrm{Re}[Y_p(f,m)]\right)\mathrm{Im}[Y_q(f,m)]\rangle_m \\
&\quad - \langle\Phi\left(\mathrm{Im}[Y_p(f,m)]\right)\mathrm{Re}[Y_q(f,m)]\rangle_m\Big],
\end{aligned} \tag{35}$$

where $[\cdot]_p$ is the p-th element of argument vector. In the case that $p \neq q$ holds, from the assumption that signals are mutually independent, each term on the right-hand side of Eq. (35) converges in 0. On the other hand, in the case that $p = q$ holds, from the Eq. (23), the first and second terms of the real parts converge in 1, respectively. However, in the imaginary part, we don't give a particular constraint because we don't care whether $\mathrm{Re}[Y_p(f,m)]$ and $\mathrm{Im}[Y_p(f,m)]$ are mutually independent. For the realization of such convergences, Murata et al. proposed a new constraint in which $\boldsymbol{I}$ in the right-hand side of the Eq. (33) is replaced as $\mathrm{diag}\left(\langle\boldsymbol{\Phi}(\boldsymbol{Y}(f,m))\boldsymbol{Y}(f,m)^{\mathrm{H}}\right)$ [17], where $\mathrm{diag}(\cdot)$ is the operation for setting every off-diagonal element of matrix as zero. From this, the condition for the diagonal elements becomes soft. In this improvement, it occurs the problem

of the gain arbitrariness, because the arbitrariness remains in the diagonal elements. However, since all of the off-diagonal elements converge in 0, the mutual independence of $Y_p(f,m)$ and $Y_q(f,m)$ can be achieved in the technique. Since the gain arbitrariness can be solved by the method using the directivity pattern of the separation matrix [28] and the method using the inverse matrix of the separation matrix [17], we introduce the above technique. Thus, we obtain the following equation:

$$\begin{aligned}\boldsymbol{W}_{i+1}(f) \\ = \alpha \Big[\mathrm{diag}\Big(\Big\langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,m))\boldsymbol{Y}(f,m)^{\mathrm{H}}\Big\rangle_m\Big) - \Big\langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,m))\boldsymbol{Y}(f,m)^{\mathrm{H}}\Big\rangle_m\Big] \\ \cdot \boldsymbol{W}_i(f)^{-\mathrm{H}} + \boldsymbol{W}_i(f). \end{aligned} \tag{36}$$

To achieve a fast convergence and a stable learning, Murata et al. proposed following FDICA [17] based on natural gradient [29]:

$$\begin{aligned}\boldsymbol{W}_{i+1}(f) \\ = \alpha \Big[\mathrm{diag}\Big(\Big\langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,m))\boldsymbol{Y}(f,m)^{\mathrm{H}}\Big\rangle_m\Big) - \Big\langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,m))\boldsymbol{Y}(f,m)^{\mathrm{H}}\Big\rangle_m\Big] \\ \cdot \boldsymbol{W}_i(f)^{\mathrm{H}} + \boldsymbol{W}_i(f). \end{aligned} \tag{37}$$

In the recent FDICA work, Sawada et al. has been proposed the alternative nonlinear function based on polar coordinate to resolve the problem of the convergence point [20]. To achieve a fast-convergence, BSS methods combining FDICA and beamforming has been proposed by Saruwatari et al. [19, 30, 31]. In the conventional method, a specific case of two sources and two microphones is assumed [19]. Therefore this algorithm cannot be applied to the source-separation problem of multiple sources and multiple microphones (more than 2 sources with more than 2 microphones). To resolve this problem, we propose a extended algorithm in which ICA and beamforming are combined for the blind separation of multiple sources (see Appendix A).

2.4 TDICA-Based BSS

2.4.1 Calculation of Separation Filter Matrices

Figure 11 shows the procedure of TDICA. The separated signals $\boldsymbol{y}(t)$ from TDICA can be given as

$$\boldsymbol{y}(t) = \sum_{\tau=0}^{Q-1} \boldsymbol{w}(\tau)\boldsymbol{x}(t-\tau) = \boldsymbol{W}(z) \cdot \boldsymbol{x}(t), \tag{38}$$

where $\boldsymbol{w}(\tau)$ is the separation filter matrix for TDICA, $\boldsymbol{W}(z)$ is the z-transform of the separation filter coefficient $\boldsymbol{w}(\tau)$ $(\tau = 0, \cdots, Q-1)$, and Q is the length of the separation filter of TDICA; these are given as

$$\boldsymbol{y}(t) = [y_1(t), \cdots, y_L(t)]^{\mathrm{T}}, \tag{39}$$

$$\boldsymbol{w}(\tau) = \begin{bmatrix} w_{11}(\tau) & \cdots & w_{1K}(\tau) \\ \vdots & \ddots & \vdots \\ w_{L1}(\tau) & \cdots & w_{LK}(\tau) \end{bmatrix}, \tag{40}$$

$$\boldsymbol{W}(z) = \sum_{\tau=0}^{Q-1} \boldsymbol{w}(\tau) z^{-\tau}, \tag{41}$$

$$= \begin{bmatrix} \sum_{\tau=0}^{Q-1} w_{11}(\tau) z^{-\tau} & \cdots & \sum_{\tau=0}^{Q-1} w_{1K}(\tau) z^{-\tau} \\ \vdots & \ddots & \vdots \\ \sum_{\tau=0}^{Q-1} w_{L1}(\tau) z^{-\tau} & \cdots & \sum_{\tau=0}^{Q-1} w_{LK}(\tau) z^{-\tau} \end{bmatrix}. \tag{42}$$

We investigate the procedures of two types of TDICA to optimize the separation filter matrices. In Sect. 2.4.2, we describe the procedure of TDICA based on simultaneous decorrelation of nonstationary signal. In Sect. 2.4.3, we describe the procedure of TDICA based on the minimization of the KLD for convolutive mixtures.

2.4.2 Simultaneous Decorrelation of Nonstationary Signal [21, 22]

In this section, we introduce the TDICA based on the simultaneous decorrelation of nonstationary signal. We separate the sources by minimizing the nonnegative

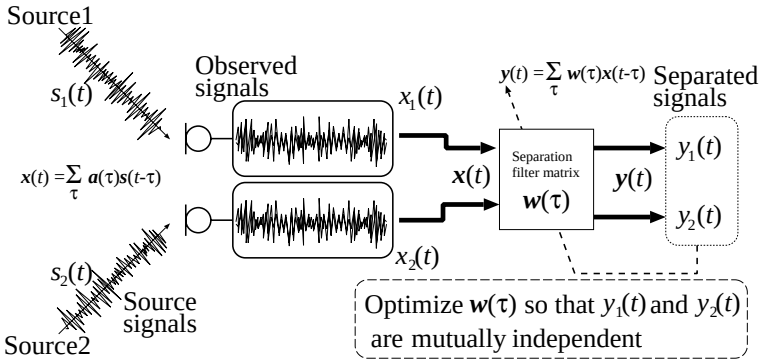


Figure 11. Blind source separation procedure performed in TDICA.

cost function which takes the minimum value only when the second-order cross-correlation becomes zero if the source signals are nonstationary. The cost function can be given as [21]

$$Q(\boldsymbol{w}(\tau)) \;=\; \frac{1}{2B}\sum_{b=1}^{B}\log\left(\frac{\det \operatorname{diag}\boldsymbol{R}_y^{(b)}(0)}{\det \boldsymbol{R}_y^{(b)}(0)}\right), \tag{43}$$

where B is the number of local analysis blocks. $\boldsymbol{R}_y^{(b)}(\tau)$ is the correlation matrix of the separated signals:

$$\boldsymbol{R}_y^{(b)}(\tau) = \langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau)^{\mathrm{T}}\rangle_t^{(b)}, \tag{44}$$

where $\langle\cdot\rangle_t^{(b)}$ denotes the time-averaging operator for the b-th local analysis block, $\boldsymbol{y}(t)$ is the output signal vector. Equation (43) becomes zero only when $y_p(t)$ and $y_q(t)$ are uncorrelated for all of the local analysis blocks. The iterative equation of the separation filter $\boldsymbol{w}(\tau)$ to minimize the cost function $Q(\boldsymbol{w}(\tau))$ is given as (hereafter we designate the iterative equation as "SD-TDICA") [21]:

[SD-TDICA]

$$\begin{aligned}\boldsymbol{w}_{i+1}(\tau) &= \boldsymbol{w}_i(\tau) + \beta\Delta\boldsymbol{w}_i(\tau)\\ &= \boldsymbol{w}_i(\tau) + \frac{\beta}{B}\sum_{b=1}^{B}\sum_{d=0}^{Q-1}\Big\{\boldsymbol{I}\\ &\quad -\Big(\operatorname{diag}\langle\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\Big)^{-1}\langle\boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)}\Big\}\boldsymbol{w}_i(d),\end{aligned} \tag{45}$$

where β is the step-size parameter. This derivation of learning rule Eq. (45) by Kawamoto includes *mathematical error* [32]) and we will propose a TDICA algorithm based on the correct derivation in Sect. 3.3.1.

2.4.3 Minimization of KLD [29]

Amari proposed the TDICA algorithm which optimizes the separation filter by minimizing the KLD between the joint probability density function and the marginal probability density function of the separated signals [29]. The KLD is given by

$$KLD(\boldsymbol{w}(\tau)) = \int p(\boldsymbol{y}(t)) \log \frac{p(\boldsymbol{y}(t))}{\prod_{l=1}^{L} \prod_{t=0}^{T-1} q(y_l(t))} d\boldsymbol{y}(t), \tag{46}$$

where $p(\cdot)$ is the joint probability density function, $q(\cdot)$ is the marginal probability density function, and T is the length of the separated signals. The iterative equation of the separation filter $\boldsymbol{w}^{(\mathrm{H})}(\tau)$ to minimize the KLD is given as (hereafter we designate the iterative equation as "H-TDICA"):
[H-TDICA]

$$\boldsymbol{w}_{i+1}^{(\mathrm{H})}(\tau) = \boldsymbol{w}_{i}^{(\mathrm{H})}(\tau) + \beta \sum_{d=0}^{Q-1} \Big\{ \boldsymbol{I}\delta(\tau - d) - \langle \boldsymbol{\phi}(\boldsymbol{y}(t)) \boldsymbol{y}(t - \tau + d)^{\mathrm{T}} \rangle_t \Big\} \boldsymbol{w}_{i}^{(\mathrm{H})}(d), \tag{47}$$

where $\boldsymbol{I}$ is the identity matrix and $\delta(\tau)$ is delta function, where $\delta(0) = 1$ and $\delta(\tau) = 0$ $(\tau \neq 0)$. Also, we define the nonlinear vector function $\boldsymbol{\phi}(\cdot)$ as

$$\boldsymbol{\phi}(\boldsymbol{y}(t)) \equiv [\tanh(y_1(t)), \cdots, \tanh(y_L(t))]^{\mathrm{T}}. \tag{48}$$

The H-TDICA forces the separated signals to have the characteristic that their higher-order autocorrelation is $\delta(\tau)$, i.e., the signals are temporally decorrelated. This performance might have a negative influence on the source separation. In order to solve the problem, Choi proposed a modified TDICA algorithm with a nonholonomic constraint [40]. In this algorithm, the constraint for the diagonal component of $\{\cdot\}$ part in Eq. (47), i.e., the higher-order autocorrelation of separated signals, is set to be arbitrary. The iterative equation of the separation

filter $\boldsymbol{w}^{(\mathrm{NH})}(\tau)$ is given as (hereafter we designate the iterative equation as "NH-TDICA"):
[NH-TDICA]

$$\begin{aligned}\boldsymbol{w}_{i+1}^{(\mathrm{NH})}(\tau) &= \boldsymbol{w}_{i}^{(\mathrm{NH})}(\tau) + \beta \sum_{d=0}^{Q-1} \Big\{ \mathrm{diag}\Big(\langle \boldsymbol{\phi}(\boldsymbol{y}(t)) \boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big) \\ &\quad - \langle \boldsymbol{\phi}(\boldsymbol{y}(t)) \boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big\} \boldsymbol{w}_{i}^{(\mathrm{NH})}(d). \end{aligned} \tag{49}$$

2.5 Initial Value for ICA

As for the initial value of ICA, various coefficients have been applied. Major initial values are listed as follows:

1. FIR-filter which elements are random values [33]

2. FIR-filter only though the straight pass [22]

$$\boldsymbol{w}(\tau) = \begin{cases} \delta\left(\tau - \dfrac{Q}{2}\right) & \text{(diagonal components)} \\ 0 & \text{(off-diagonal components)} \end{cases}, \tag{50}$$

   where $\delta(0) = 1$ and $\delta(\tau) = 0, (\tau \neq 0)$. Also, Q is the length of the separation filter.

3. DS array [28, 30]

4. NBF [34, 35]
   In this case that the look direction is θ_1 and the directional null is steered to θ_2, the elements of the matrix for signal separation are given as

$$\begin{aligned} W_{11}(f) &= -\exp\left[-j2\pi(f f_s/N) d_1 \sin\theta_2/c\right] \\ &\quad \cdot \Big\{ -\exp\left[j2\pi(f f_s/N) d_1 (\sin\theta_1 - \sin\theta_2)/c\right] \\ &\quad + \exp\left[j2\pi(f f_s/N) d_2 (\sin\theta_1 - \sin\theta_2)/c\right] \Big\}^{-1}, \end{aligned} \tag{51}$$

$$\begin{aligned} W_{12}(f) &= \exp\left[-j2\pi(f f_s/N) d_2 \sin\theta_2/c\right] \\ &\quad \cdot \Big\{ -\exp\left[j2\pi(f f_s/N) d_1 (\sin\theta_1 - \sin\theta_2)/c\right] \\ &\quad + \exp\left[j2\pi(f f_s/N) d_2 (\sin\theta_1 - \sin\theta_2)/c\right] \Big\}^{-1}, \end{aligned} \tag{52}$$

where where c is velocity of sound, f_s is sampling frequency and N is a DFT size. Also, in the case that the look direction is θ_2 and the directional null is steered to θ_1, the elements of the matrix are given as

$$\begin{aligned} W_{21}(f) &= \exp[-j2\pi(ff_s/N)d_1\sin\theta_1/c] \\ &\quad \cdot\Big\{\exp[j2\pi(ff_s/N)d_1(\sin\theta_2-\sin\theta_1)/c] \\ &\quad -\exp[j2\pi(ff_s/N)d_2(\sin\theta_2-\sin\theta_1)/c]\Big\}^{-1}, \qquad (53) \\ W_{22}(f) &= -\exp[-j2\pi(ff_s/N)d_2\sin\theta_1/c] \\ &\quad \cdot\Big\{\exp[j2\pi(ff_s/N)d_1(\sin\theta_2-\sin\theta_1)/c] \\ &\quad -\exp[j2\pi(ff_s/N)d_2(\sin\theta_2-\sin\theta_1)/c]\Big\}^{-1}. \qquad (54) \end{aligned}$$

2.6 Conclusion

In this section, the sound mixing model of the microphone array is explained. Next, we introduced two types of ICA for the convolutive mixture. The first one is FDICA based on the minimization of KLD and the other one is TDICA. We also described two types of TDICA, i.e., TDICA based on simultaneous decorrelation of nonstationary signal and TDICA based on the minimization of KLD. Moreover, the advantages and disadvantages in FDICA and TDICA are explained.

3. Experimental Analyses of FDICA and TDICA

3.1 Introduction

FDICA and TDICA have the disadvantages as shown in the Sect. 3.2.4 and 3.3.4. In this section, we verify these problems by source-separation experiments under the real acoustic (reverberant) conditions and we newly describe the experimental analyses of FDICA and TDICA. In Sect. 3.2, we describe the experimental analysis of FDICA. In Sect. 3.3, we describe the experimental analysis of TDICA.

3.2 Experimental Analysis of FDICA

The performances of traditional noise reduction methods and dereverberation methods are improves as the filter length (the number of subbands) is increased. We speculate that the source-separation performance is also improved as the number of subbands is increased in FDICA. Then, we investigate the relationship between the separation performance and the number of subbands.

3.2.1 Experimental Setup

A two-element array with the interelement spacing of 4.0 cm is assumed. The speech signals are assumed to arrive from two directions, -30° and 40°. (direction normal to the array is set to be 0°). The distance between the microphone array and the loudspeakers is 1.15 m (see Fig. 12). Two sentences spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research [44] are used as the original speech samples. The sampling frequency is 8 kHz and the length of speech is limited to within 3 seconds. Using these sentences, we obtain 12 combinations with respect to speakers and source directions. In these experiments, we use the following signals as the source signals: the original speech convolved with the impulse responses specified by the reverberation times of 300 ms. 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 and neglect the effect of background noise, and (2) since the mixing condition is explicitly measured, we can easily calculate a reliable objective score for evaluating the separation performance as described

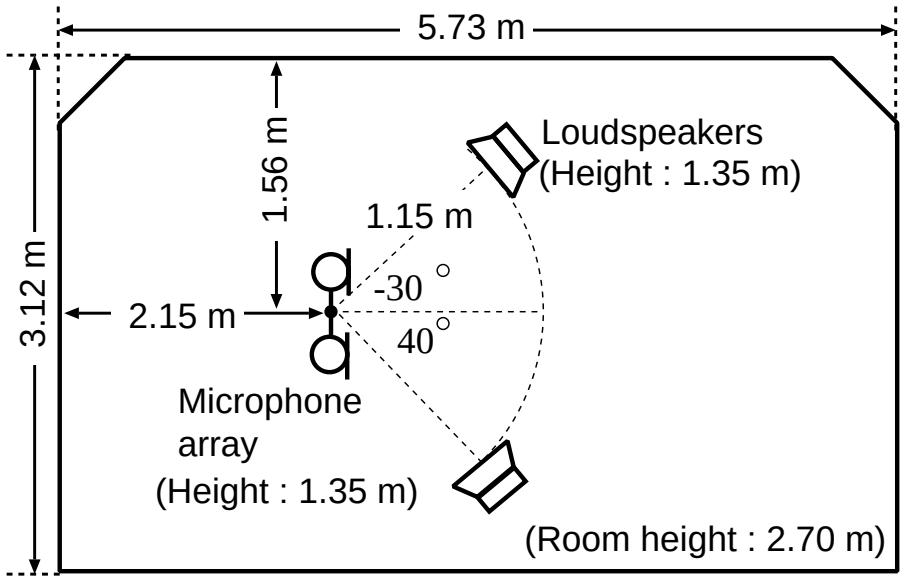


Figure 12. Layout of reverberant room used in experiments.

in Sect. 3.2.2.

3.2.2 Objective Evaluation Score; Noise Reduction Rate

Noise reduction is a necessary task for achieving a noise-robust hands-free speech recognition and a high-quality hands-free telecommunication system. To evaluate the degree of the noise reduction, we introduce the *Noise reduction rate* (NRR). NRR defined as the output SNR in dB minus input SNR in dB, is used as the objective evaluation score in this experiment. The SNRs are calculated under the assumption that the speech signal of the undesired speaker is regarded as noise. The NRR is defined as

$$\mathrm{NRR} \equiv \frac{1}{L}\sum_{l=1}^{L}\left(\mathrm{SNR}_l^{(\mathrm{O})} - \mathrm{SNR}_l^{(\mathrm{I})}\right), \tag{55}$$

$$\mathrm{SNR}_l^{(\mathrm{O})} = 10\log_{10}\frac{\sum_f |H_{ll}(f)S_l(f)|^2}{\sum_f |H_{ln}(f)S_n(f)|^2}, \tag{56}$$

$$\mathrm{SNR}_l^{(\mathrm{I})} = 10\log_{10}\frac{\sum_f |A_{ll}(f)S_l(f)|^2}{\sum_f |A_{ln}(f)S_n(f)|^2}, \tag{57}$$

Table 1. Analysis conditions of FDICA

Number of Subbbands	32, 64, 128, 256, 512, 1024, 2048, 4096 [points]
	(4, 8, 16, 32, 64, 128, 256, 512 [ms])
Frame Shift	16 [points] (2 [ms])
Window	Hamming window
Number of Iterations	30
Step Size Parameter *alpha*	1.0×10^{-5}

where $\mathrm{SNR}_l^{(\mathrm{O})}$ and $\mathrm{SNR}_l^{(\mathrm{I})}$ are the output SNR and the input SNR, respectively, and $l \neq n$. Also, $S_l(f)$ is the frequency-domain representation of the source signal, $s_l(t)$, $H_{ij}(f)$ is the element in the i th row and the j th column of the matrix $\boldsymbol{H}(f) = \boldsymbol{W}(f)\boldsymbol{A}(f)$ where $\boldsymbol{A}(f)$ is the mixing matrix which corresponds to the frequency-domain representation of the room impulse responses described in Sect. 2.2, and $\boldsymbol{W}(f)$ is the frequency-domain representation of the separation filter matrix of ICA, $\boldsymbol{w}(\tau)$.

3.2.3 Relation between Separation Performance and Number of Subbands in FDICA

In order to confirm the independence problem of narrow-band signals in FDICA ((F3) described in Sect. 3.2.4), we carried out the preliminary experiment under the analysis conditions shown in Table 1. As for the initial value of $\boldsymbol{W}(f)$, we apply the NBF [31] Eqs. (51) to (54). In this experiments, we apply the NBF in which the null steered toward $\pm 60^\circ$.

Figure 13 shows the NRR results for different numbers of frequency bins in FDICA. As shown in Fig. 13, the NRR of FDICA obviously degrades when the number of frequency bins becomes too large, and the separation performance is saturated before reaching a sufficient performance. This is because we transform the fullband signals into the narrow-band signals and the independence assumption collapses in each frequency bin, particularly when the number of frequency bins is large.

In order to confirm the fact, we newly define the following objective measure

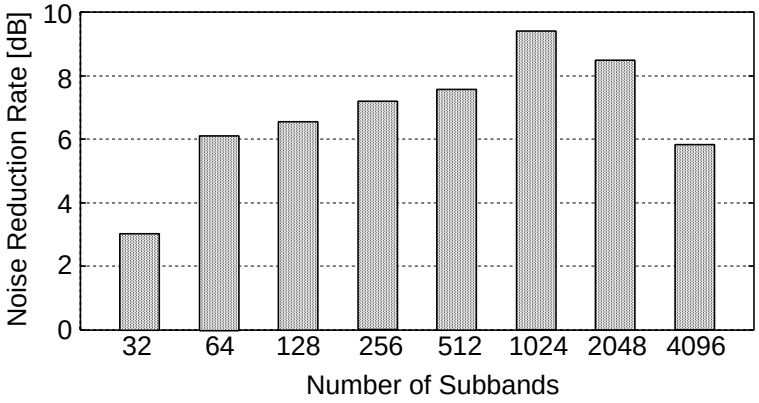


Figure 13. Relation between separation performances and the number of subbands in conventional FDICA.

to quantify an independence, and investigate the relation between the number of frequency bins and the independence among subband signals.

$$J = \left\langle \left\| \mathrm{diag}\Big(\langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,m)) \boldsymbol{Y}^{\mathrm{H}}(f,m) \rangle_m \Big) - \langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,m)) \boldsymbol{Y}^{\mathrm{H}}(f,m) \rangle_m \right\| \right\rangle_f, \quad (58)$$

where $\| \cdot \|$ is frobenius norm of matrix. This measure J is a part of the iterative equation (37) and has no dimension. Therefore the absolute value of J is meaningless itself, however, the relative value between the different numbers of frequency bins is important. If narrow-band signals become mutually independent, the measure J becomes zero. Also we can consider that the independence of subband signals is high when J is small. In order to evaluate the independence of real narrow-band speech signals, we carried out the experiment in which the input signal, $\boldsymbol{Y}(f,m)$, in Eq. (58) is regarded as the perfectly separated sources, i.e., original speech samples. Figure 14 shows the relation between the number of frequency bins and the value of J which corresponds to the independence of subband signals. Figure 14 shows that the independence decreases as the number of frequency bins increases, especially when the number of frequency bins is large.

Next, to compare the correlation among source signals, we compare the waveform of the narrow-band signals of sources. Figures 15 and 16 are the narrow-band signals of the source 1 and the source 2 analyzed in the conditions in which the number of subbands is set to be 32 points and 2048 points, respectively. In these

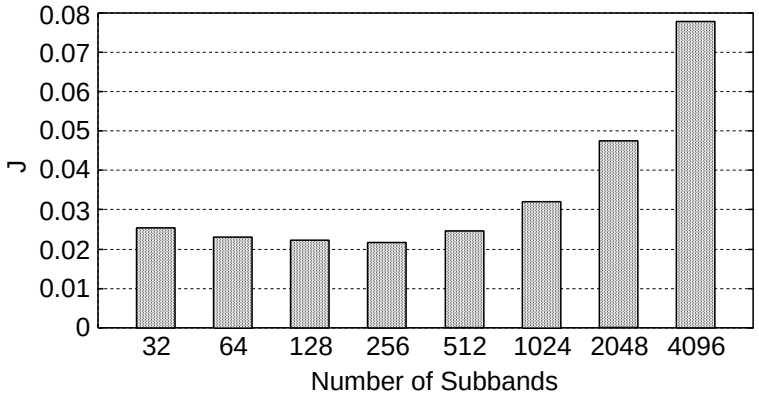


Figure 14. Relation between the number of subbands and the value of J defined by Eq. (58), which corresponds to the independence of subband signals.

figures, (a) and (b) are the real part and the imaginary part of the narrow-band signal of the source 1, respectively. (c) and (d) are the real part and the imaginary part of the narrow-band signal of the source 2, respectively. Compared Fig. 15 with 16, we can also confirm the following characteristic: (1) the correlation among narrow-band signals is low when the number of subbabds becomes small, (2) the correlation among narrow-band signals is high when the number of subbabds becomes large.

Above-mentioned experimental results clarify the disadvantage that the separation performance is saturated in FDICA because we transform the fullband signals into the narrow-band signals. We should lengthen the separation filter (or FFT length for analysis) when we confront with a long reverberation. In this case, however, the independence of subband signals decreases. Thus, there is a trade-off relation among the independence of subband signals and robustness against reverberation as shown in Figure 17. On the basis of these results, we should cascade another signal processing analysis, e.g., TDICA, with FDICA to obtain the further separation performances.

3.2.4 Advantages and Disadvantages of FDICA

We can conclude that FDICA has the following advantages and disadvantages.

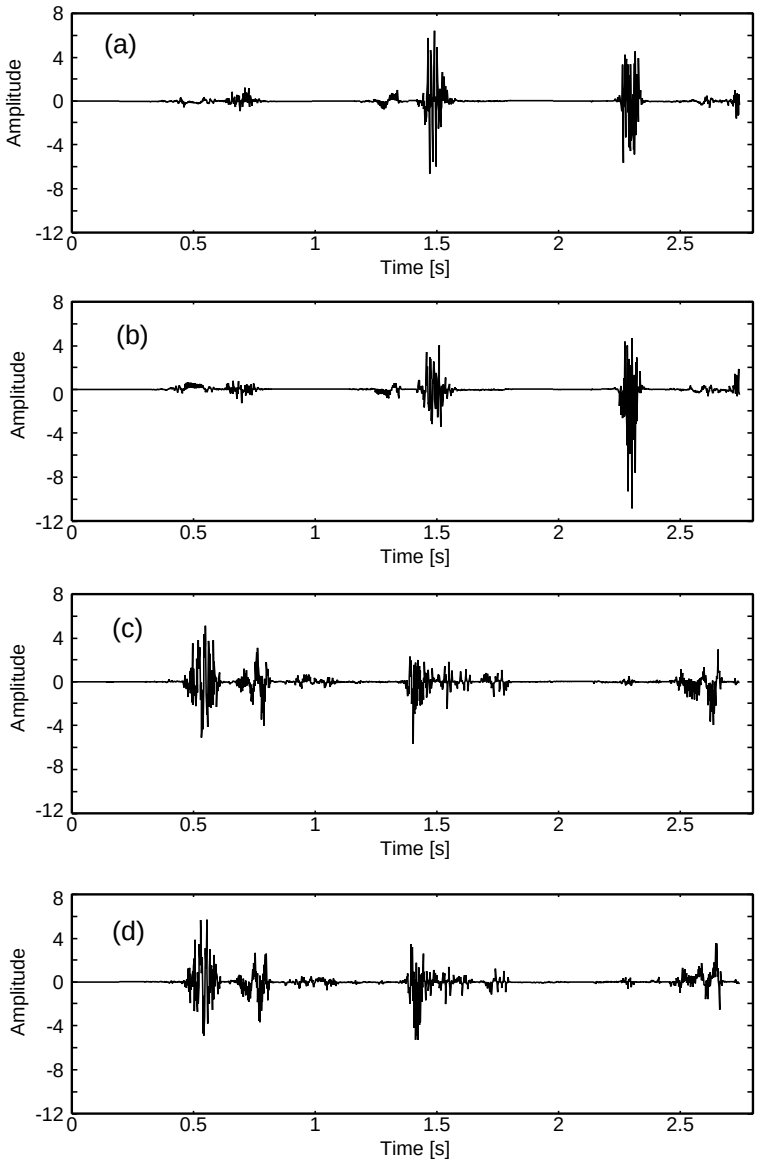


Figure 15. Narrow-band signals of source signals analyzed under the conditions in which the number of subbands is set to be 32 points and 1 kHz. (a) and (b) are the real part and the imaginary part of the narrow-band signal of the source 1, respectively. (c) and (d) are the real part and the imaginary part of the narrow-band signal of the source 2, respectively.

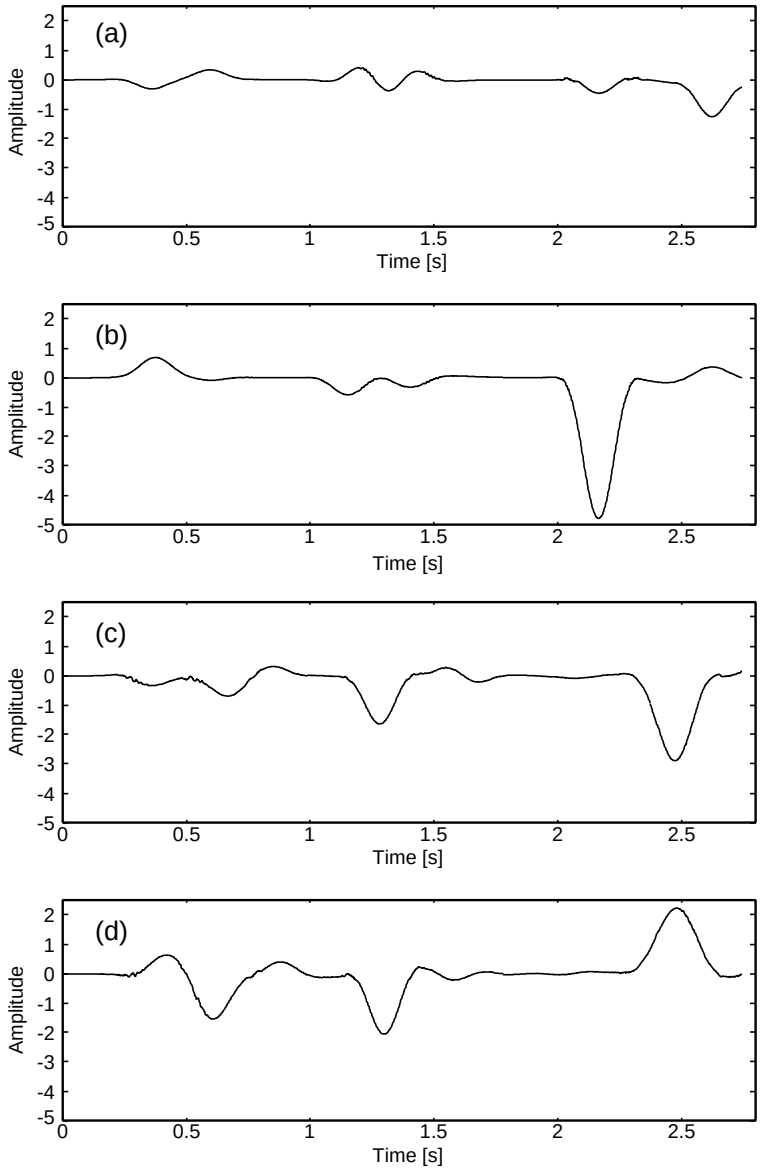


Figure 16. Narrow-band signals of source signals analyzed under the conditions in which the number of subbands is set to be 2048 points and 1 kHz. (a) and (b) are the real part and the imaginary part of the narrow-band signal of the source 1, respectively. (c) and (d) are the real part and the imaginary part of the narrow-band signal of the source 2, respectively.

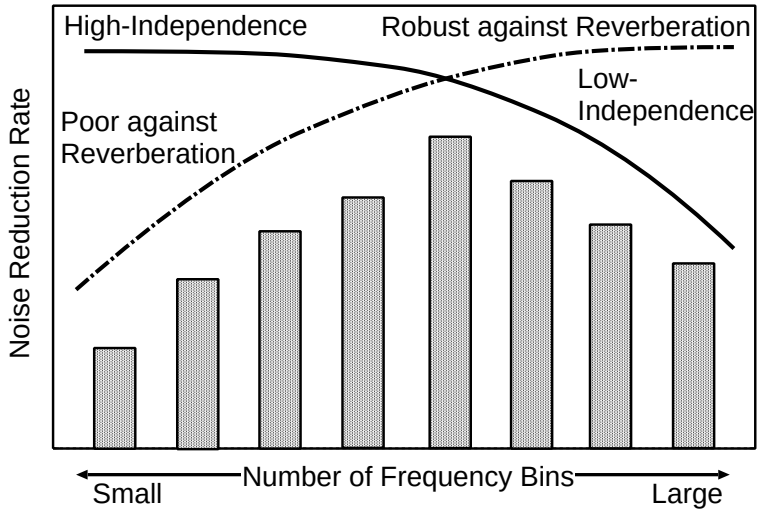


Figure 17. Trade-off relation between the independence of subband signals and robustness against reverberation.

Advantages:

(F1) We can simplify the convolutive mixture down to simultaneous mixtures by the frequency transform.

(F2) It is easy to converge the separation filter in iterative ICA learning with high stability.

Disadvantages:

(F3) The separation performance is saturated before reaching a sufficient performance because the independence assumption collapses in each narrow-band [25] (see Sect. 3.2.3).

(F4) Permutation among source signals and indeterminacy of each source gain in each subband.

As for disadvantage (F4), various solutions have already been proposed [17, 28, 36, 37, 38]. However, the collapse of the independence assumption, (F3), is a serious and inherent problem, and this prevents us from applying FDICA in a real acoustic environment with a long reverberation.

3.3 Experimental Analysis of TDICA

First, we drive the TDICA algorithm based on the simultaneous decorrelation of nonstationary signal. The cost function is defined as Eq. (43) (this cost function has been already proposed by Kawamoto et al. [21], but their derivation of learning rule Eq. (45) includes *mathematical error* [32]) and we proposed a TDICA algorithm based on the correct derivation.

We speculate that the source-separation performance is also improved as the number of subbands is increased in TDICA. Then, we investigate the relationship between the separation performance and the filter length. Since H-TDICA (see Eq. (47)) causes the distortion effect by whitening effect, we do not investigate the separation performance by H-TDICA and we compare SD-TDICA 1 (see Eq. (63)) , SD-TDICA 2 (see Eq. (64)), and NH-TDICA (see Eq. (49)).

3.3.1 TDICA Based on Simultaneous Decorrelation of Nonstationary Signal and Its Extension

The optimal separation filter is found by minimizing the cost function Q. In order to achieve the minimization, we consider the following natural gradient [29]:

$$\Delta \boldsymbol{w}(\tau) = -\frac{\partial Q(\boldsymbol{w}(\tau))}{\partial \boldsymbol{w}(\tau)} \boldsymbol{W}(z^{-1})^{\mathrm{T}} \boldsymbol{W}(z), \tag{59}$$

where

$$\boldsymbol{W}(z^{-1}) \;=\; \sum_{\tau=0}^{Q-1} \boldsymbol{w}(\tau) z^{\tau}. \tag{60}$$

The standard gradient, $\partial Q(\boldsymbol{w}(\tau))/\partial \boldsymbol{w}(\tau)$, on the right-hand side in Eq. (59) is rewritten as

$$\frac{\partial Q(\boldsymbol{w}(\tau))}{\partial \boldsymbol{w}(\tau)} = \frac{1}{B} \sum_{b=1}^{B} \Big\{ \Big(\mathrm{diag} \boldsymbol{R}_y^{(b)}(0) \Big)^{-1} \boldsymbol{R}_y^{(b)}(\tau) - \Big(\boldsymbol{R}_y^{(b)}(0) \Big)^{-1} \boldsymbol{R}_y^{(b)}(\tau) \Big\} \boldsymbol{W}(z^{-1})^{-\mathrm{T}}, \tag{61}$$

where $^{-\mathrm{T}}$ represents transpose of inverse matrix. The derivation of Eq. (61) is given by Appendix B. Substituting Eq. (61) into Eq. (59) and using the relationship Eq. (41), $\Delta \boldsymbol{w}(\tau)$ is obtained the following equation (the derivation is given

by Appendix C);

$$\begin{aligned}\Delta \boldsymbol{w}(\tau) &= \frac{1}{B}\sum_{b=1}^{B}\sum_{d=0}^{Q-1}\Big\{\Big(\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\Big)^{-1}\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)} \\ &\quad -\Big(\mathrm{diag}\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\Big)^{-1}\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)}\Big\}\boldsymbol{w}(d). \qquad (62)\end{aligned}$$

From Eq. (62), the iterative equation of the separation filter (hereafter we designate the iterative equation as "SD-TDICA 1"):
[SD-TDICA 1]

$$\begin{aligned}\boldsymbol{w}_{i+1}(\tau) &= \boldsymbol{w}_i(\tau) + \beta\Delta\boldsymbol{w}_i(\tau) \\ &= \boldsymbol{w}_i(\tau) + \frac{\beta}{B}\sum_{b=1}^{B}\sum_{d=0}^{Q-1}\Big\{\Big(\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\Big)^{-1}\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)} \\ &\quad -\Big(\mathrm{diag}\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\Big)^{-1}\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)}\Big\}\boldsymbol{w}_i(d), \qquad (63)\end{aligned}$$

where β is the step-size parameter. In this equation, when d is larger than τ, we calculate the future correlation between the separated signals. On the other hand, when τ is larger than d, we calculate the past correlation. Therefore we can achieve the iterative leaning for a both-side filter, and we can treat the mixing condition even including the non-minimum phase systems [39].

Since the Eq. (63) evaluates only off-diagonal of $\boldsymbol{R}_y^{(b)}(0)$, we confirmed that the iterative equation of Eq. (63) could not achieve a superior separation performance under the reverberant condition (see Sect. 3.3.2). Namely, the source separation is not achieved by only using nonstationarity of signals. Therefore we use not only nonstationarity of signals but also time-delayed decorrelation approach. We expand Eq. (63) to the following equation to evaluate the off-diagonal of $\boldsymbol{R}_y^{(b)}(\tau)$ for all time delays τ (hereafter we designate the iterative equation as "SD-TDICA 2"):
[SD-TDICA 2]

$$\begin{aligned}\boldsymbol{w}_{i+1}(\tau) &= \boldsymbol{w}_i(\tau) + \frac{\beta}{B}\sum_{b=1}^{B}\sum_{d=0}^{Q-1}\Big\{\Big(\mathrm{diag}\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\Big)^{-1}\mathrm{diag}\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)} \\ &\quad -\Big(\mathrm{diag}\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\Big)^{-1}\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)}\Big\}\boldsymbol{w}_i(d). \qquad (64)\end{aligned}$$

We confirmed that the separation performance was improved by using both nonstationarity and time-delayed decorrelation approach (see Sect. 3.3.2).

Table 2. Analysis conditions of SD-TDICA 1 and SD-TDICA 2

Filter Length Q	16, 32, 64, 128, 256, 512, 1024, 2048 [taps]
	(2, 4, 8, 16, 32, 64, 128, 256 [ms])
Local Analysis Block B	1, 2, 3, 4, 5, 6, 8, 10
	(3.0, 1.5, 1.0, 0.8, 0.6, 0.5, 0.4, 0.3 [s])
Maximum Iterations	500
Step Size Parameter β	1 / Q

3.3.2 Fundamental Limitation of SD-TDICA 1 Based on Simultaneous Decorrelation of Nonstationary Signal

To evaluate the effectiveness of the simultaneous decorrelation of nonstationary signal, we compare the source-separation performances of SD-TDICA 1 (see Eq. (63)) and SD-TDICA 2 (see Eq. (64)) under reverberant conditions.

We carried out the experiments under the condition as shown in Fig. 12. The analysis conditions of these experiments are shown in Table 2. As for the initial value of $\boldsymbol{w}(\tau)$, we apply the filter matrix Eq. (50). As for the local analysis block for SD-TDICA1 and SD-TDICA2, we divided the signals equally into B parts ($B = 1 \sim 10$). We chose the optimal B and number of iterations for each combination of speaker because the convergence is different for every combination.

Figure 18 (a) and (b) show the NRR results in the SD-TDICA 1 and SD-TDICA 2. Figure 18 (a) shows that SD-TDICA 1 can not achieve a signal separation under the reverberant condition. Comparing SD-TDICA 1 with SD-TDICA 2 in Fig. 18, we confirm that SD-TDICA 2 can achieve a superior separation performance to SD-TDICA 1. These results show that it is necessary to evaluate correlations of different times to achieve a superior performance.

3.3.3 Relation between Separation Performance and Filter Length in TDICA

We carried out the experiments using SD-TDICA 1 (see Eq. (64)) and NH-TDICA (see Eq. (49)), to evaluate the contribution of these TDICAs for improving the separation performances under reverberant conditions. The analysis conditions

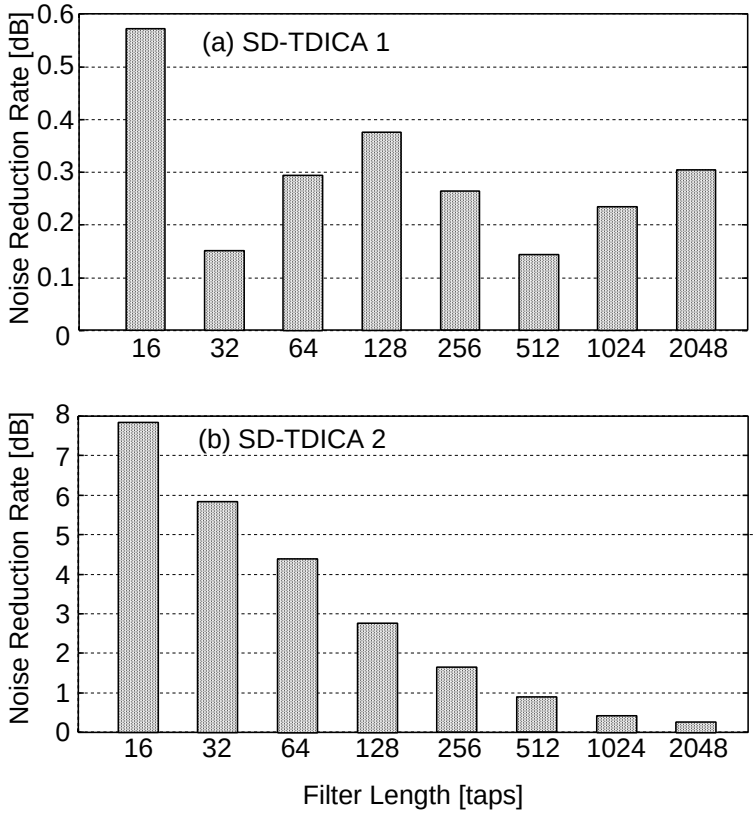


Figure 18. Separation performances in (a) SD-TDICA 1, and (b) SD-TDICA 2.

of these experiments are as shown in Table 3. As for the initial value of $\boldsymbol{w}(\tau)$, we apply the filter matrices Eq. (50) and the inverse DFT of NBF Eq. (51) $\sim$ (54). We chose the optimal B and number of iterations for each combination of speaker because the convergence is different for every combination.

Figure 19 shows the NRR results in the SD-TDICA 2 and NH-TDICA for different filter lengths. These results reveal that both TDICAs can not achieve a signal separation under the reverberant condition compared with conventional FDICA (see Fig. 13). This reason is that (1) the iterative rule for FIR-filter learning is complicated and (2) the convergence degrades under reverberant conditions. Therefore we can conclude that the conventional TDICA cannot achieve a superior separation performance because of the problems (1) and (2).

Table 3. Analysis conditions of SD-TDICA 2 and NH-TDICA

Filter Length Q	32, 64, 128, 256, 512, 1024, 2048, 4096 [taps]
	(4, 8, 16, 32, 64, 128, 256 [ms])
Local Analysis Block B	1, 2, 3, 4, 5, 6, 8, 10
(SD-TDICA 2)	(3.0, 1.5, 1.0, 0.8, 0.6, 0.5, 0.4, 0.3 [s])
Maximum Iterations	500
Step Size Parameter β	$5.0 \times 10^{-1} \sim 1.0 \times 10^{-7}$

The separation performances of NH-TDICA and SD-TDICA 2 using the optimal B are not so different (see Fig. 19) [45, 46]. In SD-TDICA 2, we utilize the nonstationarity of the signals. The separation performance is not monotonic as the parameter B is changed because the nonstationarity for every source signal is different [47]. However, it is difficult and impractical to estimate the optimal B because the quantification of nonstationarity for source signal is the difficult task. Therefore, NH-TDICA is more feasible than SD-TDICA 2 because NH-TDICA does not require the estimation of the optimal B.

The estimation of the separation filter becomes complicated when we use the initial value far from the optimal solution, i.e, Eq. (50) compared with the initial value Eq. (51) $\sim$ (54). This phenomenon is distinguished especially when the filter length is increased. From this results, the use of the effective initial value is important to achieve a superior separation performance in BSS. On the basis of these results, we should perform useful preprocessing, e.g., FDICA and NBF before TDICA to obtain the further separation performances.

3.3.4 Advantages and Disadvantages of TDICA

We can conclude that TDICA has the following advantages and disadvantages.

Advantages:

(T1) We can treat the fullband speech signals where the independence assumption of sources usually holds.

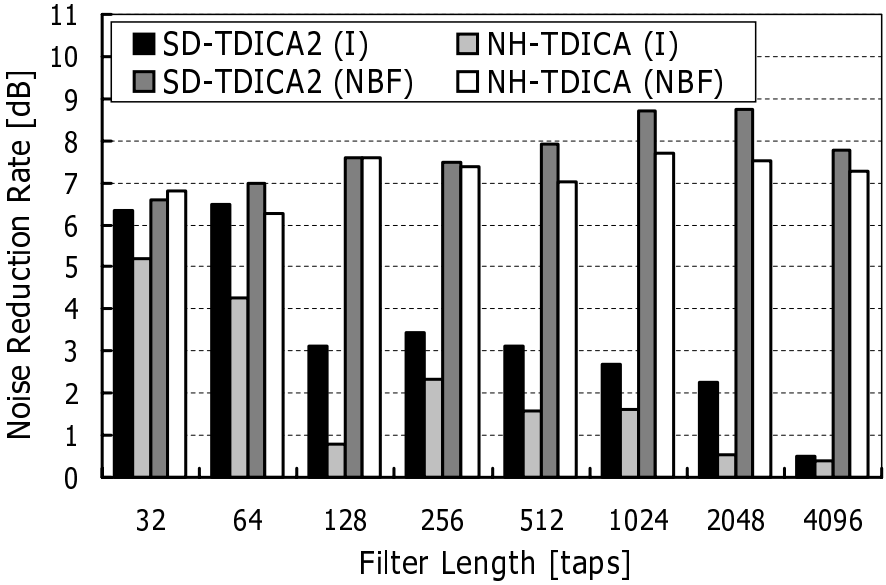


Figure 19. Relation between separation performance and filter length in SD-TDICA 2 and NH-TDICA. "I" and "NBF" denote that the initial value for TDICA are Eq. (50) and Eq. (51) $\sim$ (54), respectively.

(T2) High-convergence possibility near the optimal point.

Disadvantages:

(T3) The iterative rule for FIR-filter learning is complicated.

(T4) The convergence degrades under reverberant conditions.

It is known that TDICA works only in the case of mixtures with a short-tap FIR filter, i.e., less than 100 taps. Also, TDICA fails to separate source signals under real acoustic environments because of disadvantages (T3) and (T4).

3.4 Conclusion

In this section, we performed the experimental analyses of FDICA and TDICA and we described the experimental analyses of both ICAs under the real acoustic conditions.

First, the results of the signal separation experiment with FDICA reveals that the separation performance of FDICA obviously degrades when the number of subbands becomes too large, and is saturated before reaching a sufficient performance. We can conclude that this is because the independence assumption of the narrow-band signals collapses.

Secondly, the results of the signal separation experiment with TDICA reveals that the separation performance of TDICA is not sufficient compared with FDICA. We can conclude that this is because the iterative learning rule becomes more complicated as the reverberation increases. Also, we confirmed that not only the simultaneous decorrelation of nonstationary signals but also the utilization of correlations of different times is required to achieve a superior performance.

4. MSICA-Based BSS

4.1 Introduction

In the above section, we described some specific disadvantages and the applicable limitations of the conventional ICAs. In this section, to resolve the problems of the conventional ICAs, we propose MSICA-based BSS.

In Sect. 4.2, we describe the procedure of the proposed MSICA. Next, in Sect. 4.3, we carry out source-separation experiments under reverberant conditions and we describe the effectiveness of the proposed MSICA. Finally, in Sect. 4.4, we compare the source-separation performance by the proposed MSICA with those by the conventional ICAs.

4.2 Motivation and Strategy

As described in Sect. 3.2 and 3.3, the conventional ICA methods have some disadvantages. However, note that the advantages and disadvantages of FDICA and TDICA are mutually complementary (see Fig. 20), i.e., (F3) can be resolved by (T1) and (T2), and (T3) and (T4) can be resolved by (F1) and (F2). Hence, in order to resolve the disadvantages, we propose a new algorithm, MSICA, in which FDICA and TDICA are combined (see Fig. 21).

MSICA is conducted with the following steps. In the first stage, we perform FDICA to separate the source signals to some extent with the high-stability advantages of FDICA, (F1) and (F2). The output signals $\boldsymbol{z}(t) = [z_1(t), \cdots, z_L(t)]^{\mathrm{T}}$ from FDICA can be given as

$$\boldsymbol{z}(t) = \sum_{\tau=0}^{Q-1} \boldsymbol{v}(\tau)\boldsymbol{x}(t-\tau), \tag{65}$$

where $\boldsymbol{v}(\tau)$ is the separation filter matrix for FDICA , Q is the length of the separation filter of FDICA, and we optimize $\boldsymbol{v}(\tau)$ by Eq. (37);

$$\begin{aligned}\boldsymbol{V}_{i+1}(f) &\\ = \alpha \Big[\mathrm{diag}\Big(\Big\langle \boldsymbol{\Phi}(\boldsymbol{Z}(f,m))\boldsymbol{Z}(f,m)^{\mathrm{H}}\Big\rangle_m\Big) - \Big\langle \boldsymbol{\Phi}(\boldsymbol{Z}(f,m))\boldsymbol{Z}(f,m)^{\mathrm{H}}\Big\rangle_m\Big] &\\ \cdot \boldsymbol{V}_i(f)^{\mathrm{H}} + \boldsymbol{V}_i(f), &\end{aligned} \tag{66}$$

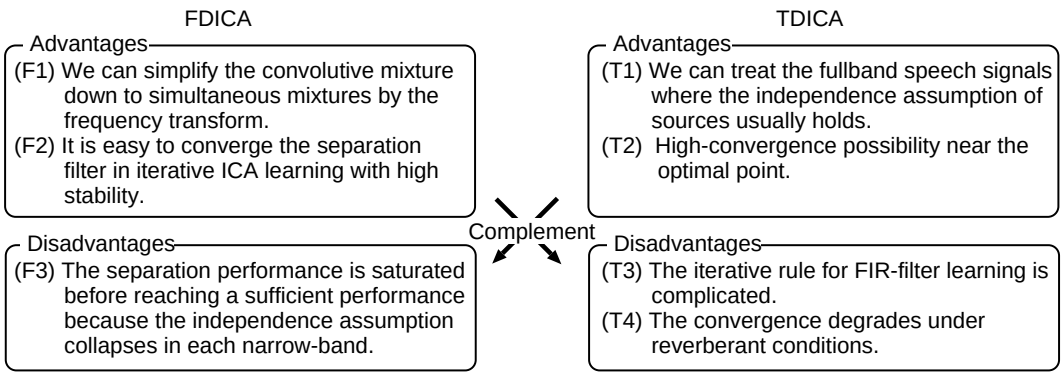


Figure 20. Complementary relation between the advantages and the disadvantages of FDICA and TDICA.

where $\boldsymbol{V}(f)$ is the DFT of $\boldsymbol{v}(\tau)$ and $\boldsymbol{Z}(f,m)$ is the narrow-band signals for time-domain signals $\boldsymbol{z}(t)$. In the second stage, we regard the output signals $\boldsymbol{z}(t)$ of FDICA as the input signals for TDICA, and we remove the residual crosstalk components of FDICA by using TDICA. The output signals $\boldsymbol{y}(t) = [y_1(t), \cdots, y_L(t)]^{\mathrm{T}}$ of TDICA can be given as

$$\boldsymbol{y}(t) = \sum_{\tau=0}^{R-1} \boldsymbol{w}(\tau)\boldsymbol{z}(t-\tau), \tag{67}$$

where $\boldsymbol{w}(\tau)$ is the separation filter matrix for TDICA and R is the length of the separation filter of TDICA. In this procedure, we optimize $\boldsymbol{w}(\tau)$ by Eq. (49);

$$\begin{aligned} \boldsymbol{w}_{i+1}(\tau) &= \boldsymbol{w}_i(\tau) + \beta \sum_{d=0}^{R-1} \Big\{ \mathrm{diag}\Big(\langle \boldsymbol{\phi}(\boldsymbol{y}(t)) \boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big) \\ &\quad - \langle \boldsymbol{\phi}(\boldsymbol{y}(t)) \boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big\} \boldsymbol{w}_i(d). \end{aligned} \tag{68}$$

Finally, we regard the output signals of TDICA as the resultant separated signals. MSICA can achieve a high stability and a superior separation performance to that of conventional FDICA and TDICA.

4.3 Effectiveness for Cascading TDICA

We carried out the experiments using MSICA to evaluate the contribution of increments of separation-filter length for improving the separation performances

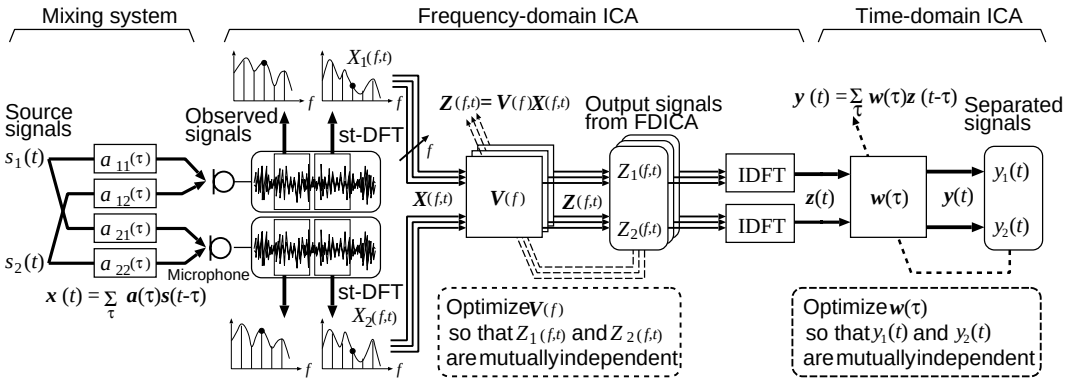


Figure 21. Blind source separation procedure performed in MSICA.

under reverberant conditions. The experimental condition is the same as that given in Sect. 3.2.1. The analysis conditions of these experiments are shown in Table 4. As for the initial value of the separation filter of FDICA is the NBF in which the null steered toward $\pm 60°$. Figures 22 shows the NRR results in the MSICA for different filter lengths.

As shown in Fig. 19, when we use the initial value Eq. (50) and the long separation filter, the separation performance of the TDICA degrades. This implies that the estimation of the separation filter becomes complicated when we use the initial value far from the optimal solution. This phenomenon is distinguished especially when the filter length is increased. On the other hand, in Fig. 22, the separation performance of MSICA is improved when the filter length is longer. This reveals that the TDICA part in MSICA can separate the source signals even with the reverberation components, and the TDICA is still useful near the optimal point.

Table 4. Analysis conditions of MSICA

FDICA part	
Number of Subbbands Q	1024 points
Frame Shift	16 points
Window	Hamming window
Number of Iterations	30
Step Size Parameter α	1.0×10^{-5}
TDICA part	
Filter Length R	16, 32, 64, 128, 256, 512, 1024, 2048 taps
Maximum Iterations	500
Step Size Parameter β	$5.0 \times 10^{-1} \sim 1.0 \times 10^{-7}$

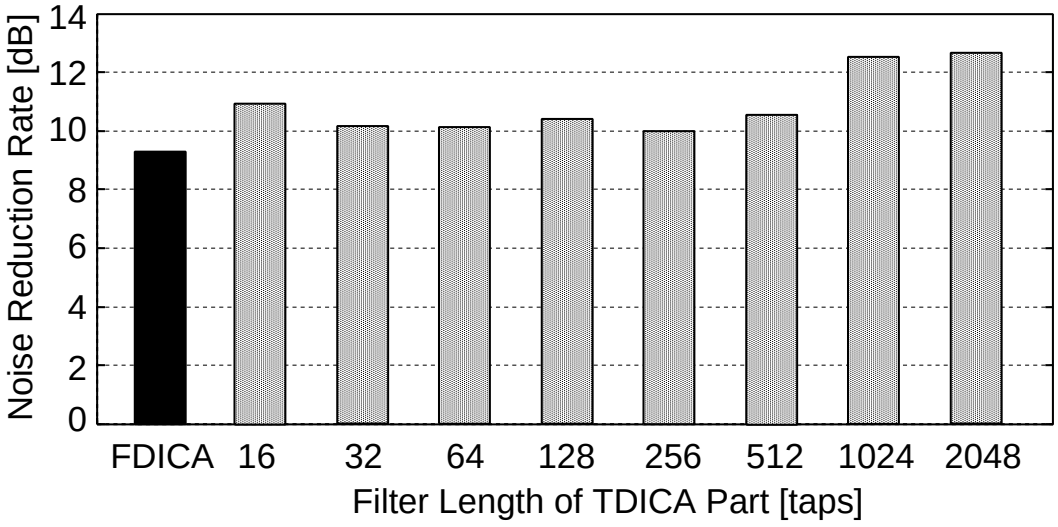


Figure 22. Relation between the separation performance and filter length in TDICA part in MSICA.

Table 5. Analysis condition of TDICA

Filter Length R	1024 taps
Number of Iterations	500
Step Size Parameter β	2.0×10^{-6}

Table 6. Analysis condition of FDICA

Number of Subbbands	1024 points
Frame Shift	16 points
Window	Hamming window
Number of Iterations	30
Step Size Parameter α	1.0×10^{-5}

4.4 Comparison between Conventional ICAs and MSICA

We compared the performance of the proposed MSICA with those of the conventional ICAs under the reverberant condition. The experimental condition is the same as that given in Sect. 3.2.1. As for TDICA, FDICA, MSICA, the analysis conditions are shown in Tables 5, 6, 7. As for the initial value of the separation filter of FDICA is the NBF in which the null steered toward $\pm 60^\circ$.

Figure 23 shows the NRRs of the conventional FDICA, TDICA, and MSICA. In this figure, we separately plot the NRRs for different combination of speakers, and the averages of their NRRs. The results reveal that the separation performances of the proposed MSICA are superior to those of the conventional FDICA and TDICA. Specifically, compared with the conventional ICA, the proposed method can improve the NRR by about 2.7 dB over that of FDICA and by about 4.3 dB over that of TDICA, for an average of 12 combinations.

As described in Sect. 3.2, the FDICA in this study showed the saturation of NRR when we used the 1024-subband analysis. As described in Sect. 3.3, the simple TDICA could not separate the source signals accurately under the reverberant condition. These findings indicate the practical limitations of the separation per-

Table 7. Analysis condition of MSICA

FDICA part	
Number of Subbbands	1024 points
Frame Shift	16 points
Window	Hamming window
Number of Iterations	30
Step Size Parameter α	1.0×10^{-5}
TDICA part	
Filter Length R	2048 taps
Number of Iterations	500
Step Size Parameter β	1.0×10^{-6}

formances of conventional ICA-based BSS methods. From the results of Fig. 23, however, we can confirm that the proposed MSICA can inherently remove these limitations, and is effective for improving the separation performance and convergence under reverberant conditions.

4.5 Discussion on Combination Order in MSICA

As described in the previous section, the combination of FDICA and TDICA can contribute to the improvement of separation. In this combination, the advantage (F2) of FDICA is useful in the initial step of separation procedure and the advantage (T2) of TDICA is also useful in the later step. Therefore we use FDICA as the first-stage ICA and TDICA as the second-stage ICA. In order to confirm the availability of this combination order, we compare the proposed combination (hereafter we designate this combination as "MSICA1") with the combination in which TDICA is used in the first stage and FDICA is used in the second stage (hereafter we designate this swapped combination as "MSICA-SWAP").

The experiment of MSICA-SWAP was carried out in the following manner. As for TDICA part in MSICA-SWAP, the number of iterations is 400 and and the filter length is 10 taps. As for FDICA part in MSICA-SWAP, the analysis conditions are the same as those given in Table 6. Figure 24 shows the comparison

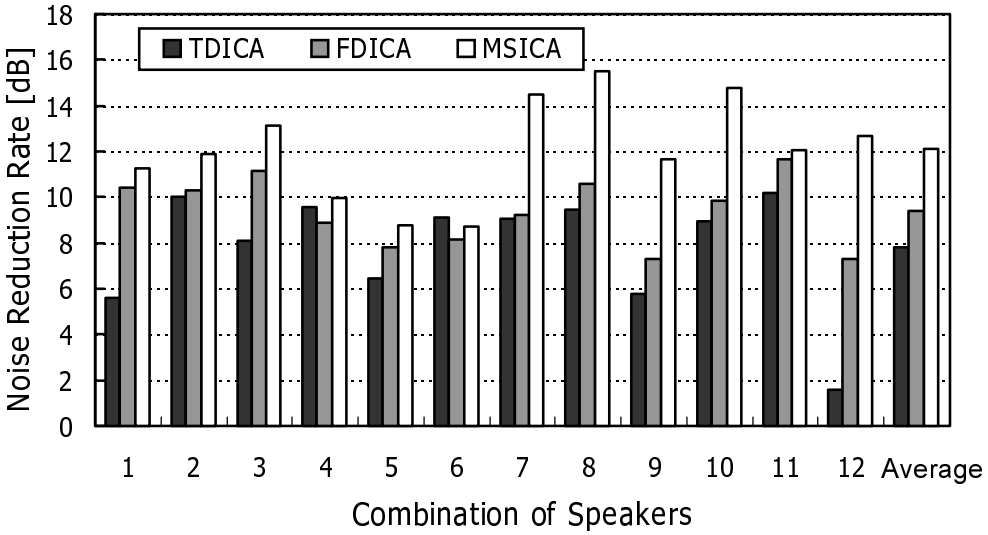


Figure 23. Comparison of noise reduction rates obtained by MSICA, conventional FDICA and TDICA.

of noise reduction rates obtained by simple TDICA, simple FDICA not using the beamforming technique [19], proposed MSICA, and MSICA-SWAP. As the result, the NRR of 7.5 dB is obtained in MSICA-SWAP and this performance is better than that of simple TDICA but is poorer than that of the original MSICA and simple FDICA. In MSICA-SWAP, the separation performance is still improved by using FDICA in the second stage, however, the separation performance is saturated because of the disadvantage (F3) of FDICA. MSICA-SWAP can not achieve the separation performance of 9.4 dB which corresponds to NRR of simple FDICA. This reason is that FDICA in this section uses the beamforming technique and the directivity pattern of the array which provide a good initial value of the separation matrix to improve the convergence [19], however, such kind of information is no longer valid in the combination order of MSICA-SWAP because we can not know the effective positions of the array elements after the first-stage TDICA and can not depict the directivity pattern. Thus the separation performance of MSICA-SWAP is almost equal to that of a raw FDICA without the beamforming technique (from [19] we can see the NRR

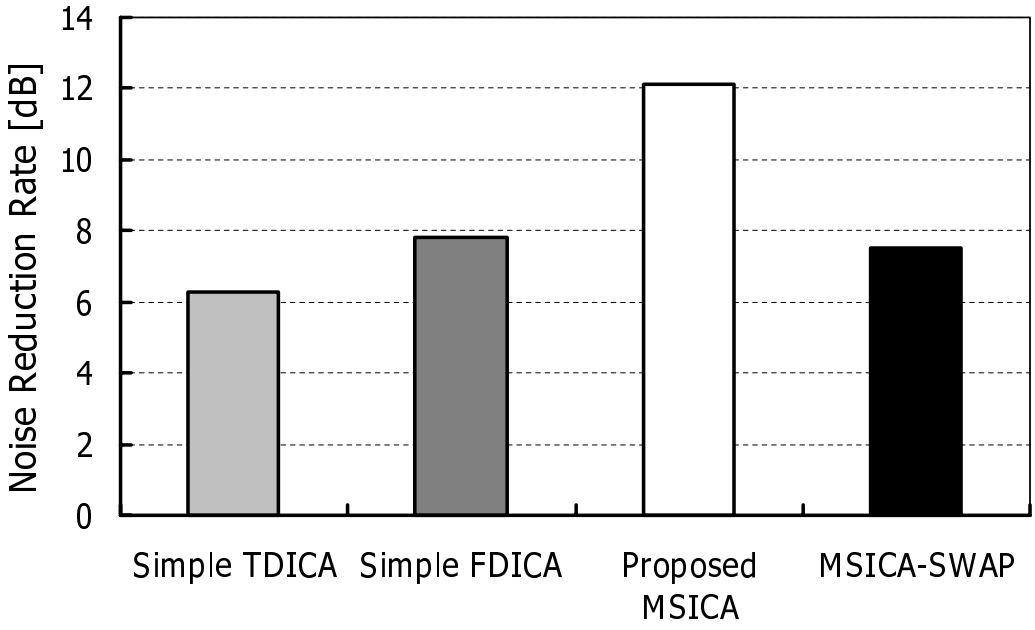


Figure 24. Comparison of noise reduction rates obtained by simple TDICA, simple FDICA, proposed MSICA, and MSICA-SWAP.

of about 7.5 dB at the 30-iteration point). This fact indicates that the swapped combination order of MSICA-SWAP has no contribution to the improvement of the separation performance, and the proposed combination order of the original MSICA (FDICA in the first stage and TDICA in the second stage) is essential.

4.6 Application of MSICA to Speech Recognition in Room Environment

4.6.1 Experimental Conditions

One of the applications of BSS is a hands-free speech recognition system. In this section, we provide a experimental evaluation with a large vocabulary continuous speech recognition task. Figures 25 and 26 show the layouts of the reverberant

Table 8. Analysis conditions of MSICA

FDICA part	
Number of Subbbands Q	2048 points
Frame Shift	128 points
Number of Iterations	100
TDICA part	
Filter Length R	4096 taps
Maximum Iterations	200 (interference speech)
	500 (PC noise)

room (RT=200 ms) used in the experiment. A two-element array with interelement spacing of 2.1 cm is used. The target speech arrive from the direction 30°; a loudspeaker is placed on the right-hand side (30°) and the distance between the loudspeaker and the microphone array is 1.15 m. We consider the following two situations where several noises are added: (1) interference speech of female selected from ASJ database [44] with 0 dB SNR which is placed on the left-hand side (−50°) (see Fig. 25), (2) a tower-type personal computer (PC) with 10 dB SNR which is placed on the left-hand side (−60°) (see Fig. 26). The analysis conditions of these experiments are shown in Table 8. Table 9 shows the experimental conditions for speech recognition.

4.6.2 Experimental Results

In order to valuate the speech recognition performance, we adopt the Word Accuracy (WA) and the Word Correct (WC) as an evaluation score. WA and WC are given by

$$\mathrm{WA}[\%] = \frac{W - S - D - I}{W} \times 100, \tag{69}$$

$$\mathrm{WC}[\%] = \frac{W - S - D}{W} \times 100, \tag{70}$$

where W is the total number of words in the test speech, S is the number of substitution errors, D is the number of deletion errors, and I is the number of insertion errors. We average each WA and WC obtained from 200 speech in total.

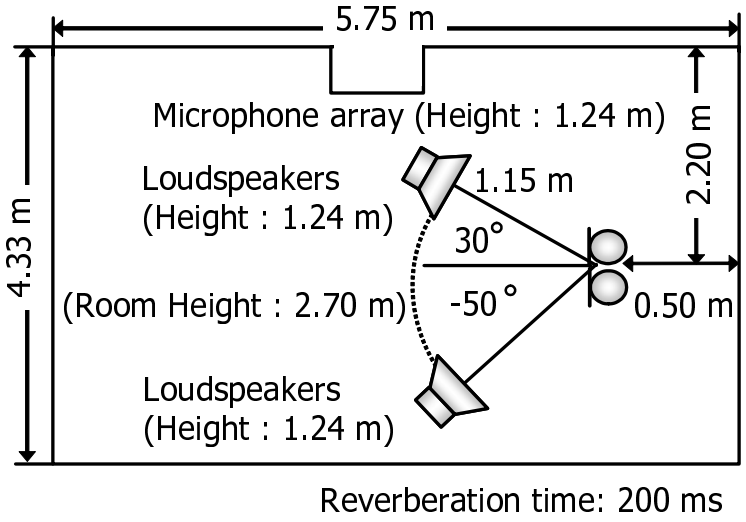


Figure 25. Layout of reverberant room used in real recording experiment. We use the interference speech as the noise signal.

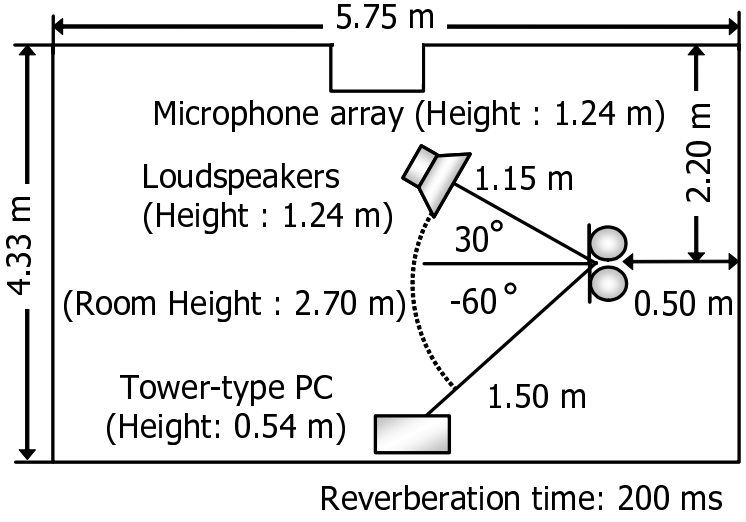


Figure 26. Layout of reverberant room used in real recording experiment. We use the personal computer as the noise signal.

Table 9. Experimental conditions for speech recognition

Database	JNAS [48],306 speakers (150 sentences / 1 speaker)
Task	20-k newspaper dictation
Acoustic model	phonetic tied mixture (PTM) [49] (clean model)
Number of training speakers	260 speakers (150 sentences / 1 speaker)
Number of testing speakers	46 speakers (200 sentences)
Decoder	JULIUS ver.3.4.2 [50]
Sampling frequency	16 kHz
Frame size	20 ms (400 sample)

Figure 27 and 28 show the results in terms of word accuracy and word correct under different noise conditions. In these figures, the black bars represent the speech recognition results for the observed signals at the single microphone, the gray bars represent the results by conventional FDICA, the white bars represent the results of the proposed MSICA, and the black line represent the results by a observed signal without noise components which correspond the upper limit in this evaluation, respectively. The word accuracy in the experiments with the interference speech is far more inferior compared with that of using the PC noise. This phenomenon occurs due to the fact that the insertion errors are increased because the interference noise is a speech signal.

The improvements of word accuracy and word correct can be found in Fig. 28 in both FDICA and MSICA compared with the results using a single microphone. Regarding the reduction of the interference speech, we can confirm that MSICA outperformed FDICA about 35 % for word accuracy and 10 % for word correct from FDICA. As for the reduction of the PC noise, there are no obvious improvements in the proposed MSICA compared with FDICA, but also no deterioration; this means that the proposed MSICA has no serious side-effects.

In summary, these results indicate that the proposed MSICA is applicable to the hands-free speech recognition system, particularly when confronted with the

Table 10. Experimental conditions for speech recognition

Task	69 isolated word recognition with network grammar
Acoustic model	diphone HMM by single Gaussian mixture (speaker-independent)
Number of testing speakers	23 speakers (69 sentences / 1speaker) 17 male and 5 female
Decoder	VORERO Ver.4.3 [51]
Sampling frequency	11 kHz
Processing for noise-robust	(1) continuous spectral subtraction [52] (2) normalized least mean square error with frame-wise voice activity detection [53] (3) exact cepstrum mean normalization [52]

interference speech.

4.7 Application of MSICA to Speech Recognition in Car Environment

4.7.1 Experimental Conditions

One promising application of BSS is a navigation system in a car environment, where many kinds of noises, e.g., interfering speech from the assistant seat, engine noise and air-conditioner noise exist. The objective of this section is to provide a experimental evaluation of applicabilities of BSS in car environments.

Table 10 shows the experimental conditions for speech recognition. A two-element array with the interelement spacing of 4 cm is used to record the sounds in a real car environment as shown in Fig. 29. The target signal is set to a driver's speech, and the interference noise to be reduced is (a) assistant speech or (b) air-conditioner noise. As for the background noise, we consider the following situations where several diffuse noises are added: (1) engine noise at idle (**idling**), (2) engine noise and road noise from the car tires at a speed of 60 km/h with the

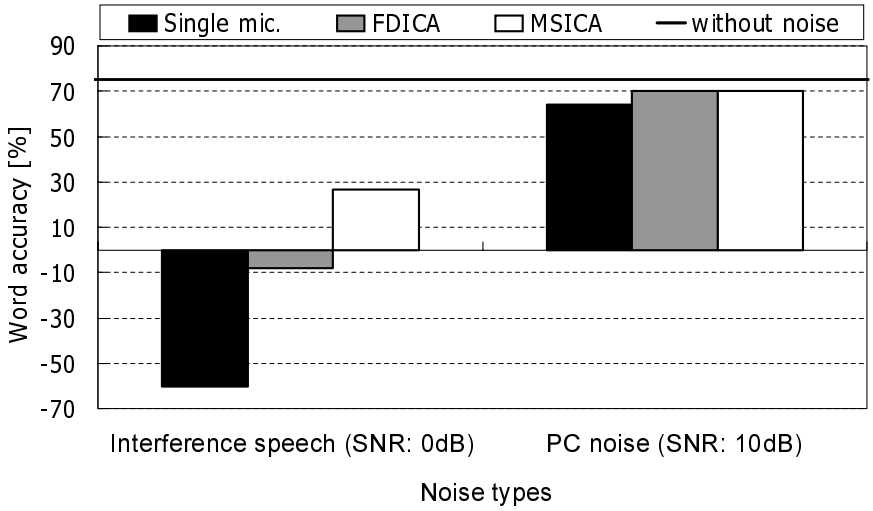


Figure 27. Comparison of word accuracy obtained by a single microphone, conventional FDICA, the proposed MSICA, and a observed signal without noise components (upper limit) under the condition that the speech signal or PC noise interferes the target speech.

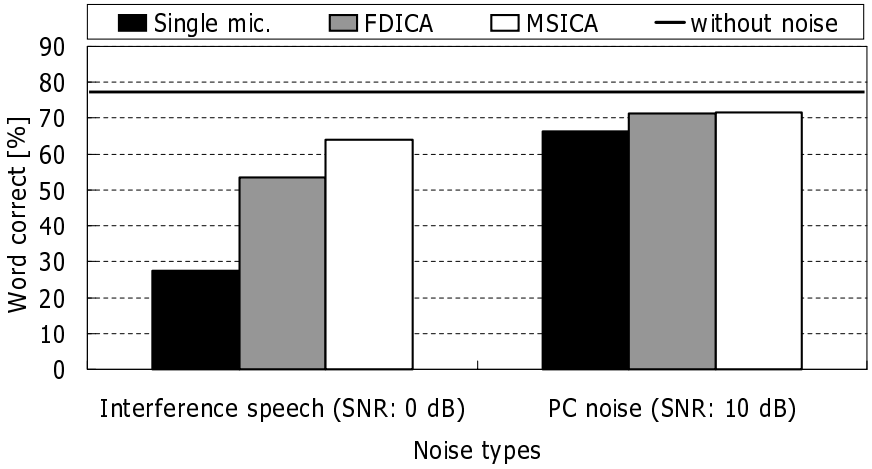


Figure 28. Comparison of word correct obtained by a single microphone, conventional FDICA, the proposed MSICA, and a observed signal without noise components (upper limit) under the condition that the speech signal or PC noise interferes the target speech.

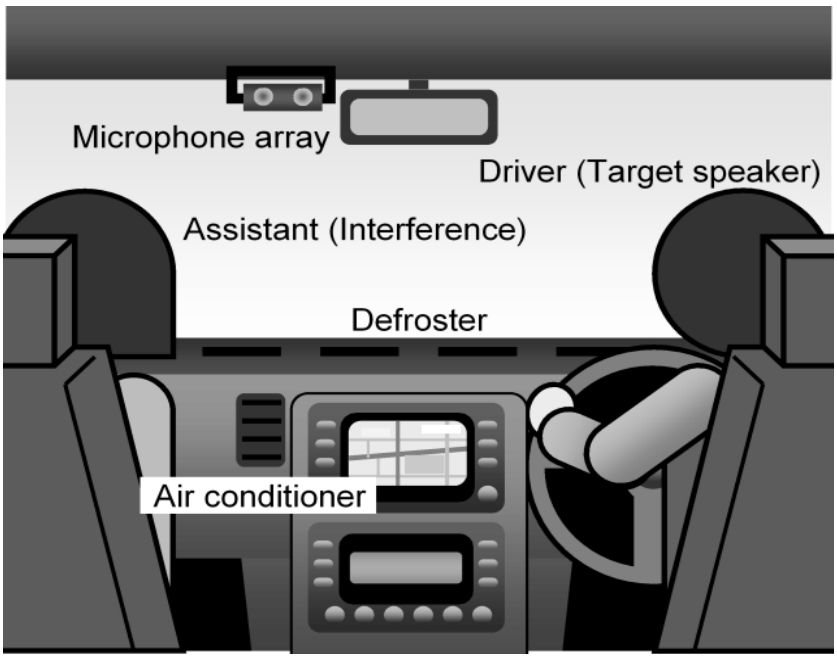


Figure 29. Layout of a real car environment used in experiments.

car window close (**60km/h(C)**), (3) engine noise and road noise from the car tires at a speed of 60 km/h with the car window open (**60km/h(O)**), and (4) engine noise and road noise from the car tires at a speed of 100 km/h with the car window close (**100km/h(C)**). The analysis conditions of these experiments are shown in Table 11.

4.7.2 Experimental Results

Figures 30 and 31 show the results in terms of word accuracy under different noise conditions. In these figures, the black bars represent the speech recognition results for the observed signals at the single microphone, the gray bars represent the results by FDICA, and the white bars represent the results by MSICA, respectively.

The remarkable improvements of word accuracy can be found in Fig. 30 in both FDICA and MSICA compared with the results using a single microphone. Regarding the reduction of the assistant speech, we can confirm an MSICA's

Table 11. Analysis conditions of MSICA

FDICA part	
Number of Subbbands Q	256 points
Frame Shift	32 points
Number of Iterations	100
TDICA part	
Filter Length R	512 taps
Maximum Iterations	100

slight outperformance from FDICA in all situations for the background noise. As for the reduction of the air-conditioner noise, there are no obvious improvements in the proposed MSICA compared with FDICA, but also no deterioration; this means that the proposed MSICA has no serious side-effects. Figure 32 shows the results of the assistant speech reduction in the case that a defroster noise is further added into the background noises. We can see the same tendency as in Fig. 30.

In summary, these results indicate that the proposed MSICA is applicable to the speech recognition system, particularly when confronted with the assistant speech.

4.8 Conclusion

In this section, we propose a new algorithm for BSS, in which FDICA and TDICA are combined to achieve a superior source-separation performance under reverberant conditions. Also, we provide a comparison results for the separation performance of FDICA, TDICA, and the proposed method under the real acoustic condition.

The results of the signal separation experiment with the proposed method reveals that the separation performance and the speech recognition performance of the proposed algorithm are superior to that of conventional ICA-based BSS methods, and the combination of FDICA and TDICA is inherently effective for improving the separation performance. Specifically, the proposed method can

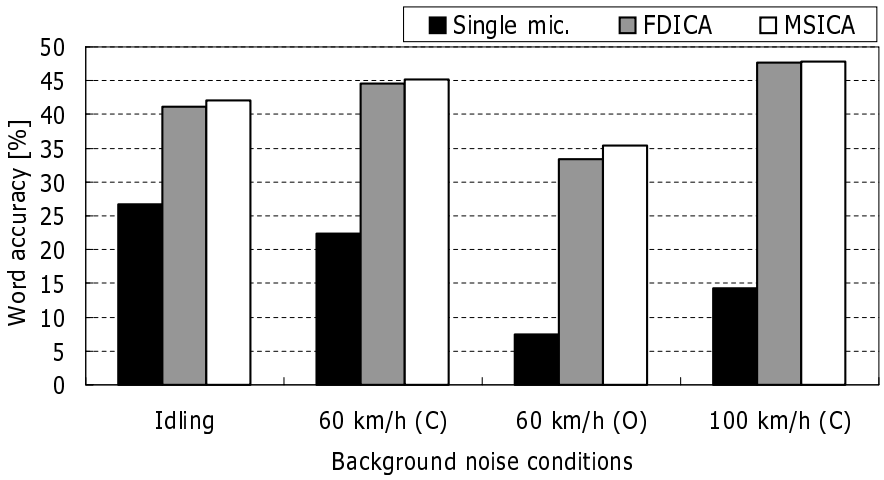


Figure 30. Comparison of word accuracy rates obtained by a single microphone, conventional FDICA and the proposed MSICA under the condition that the assistant speech interferes the target driver's speech.

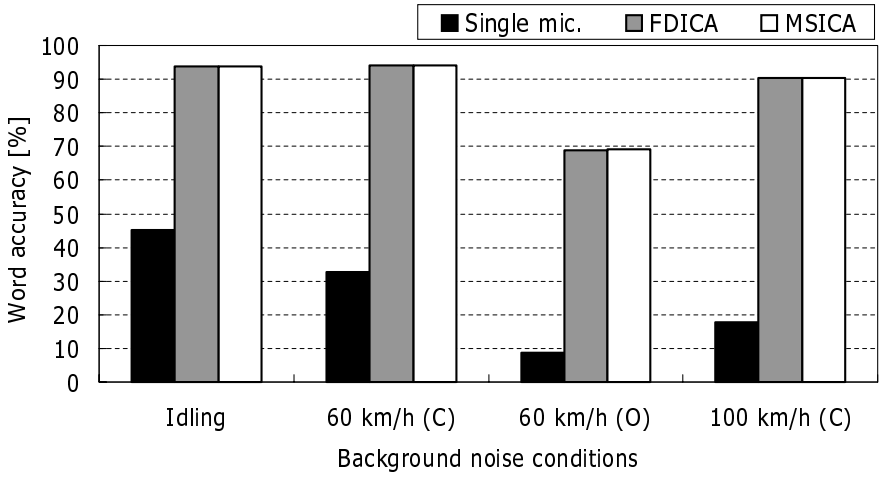


Figure 31. Comparison of word accuracy rates obtained by a single microphone, conventional FDICA and the proposed MSICA under the condition that the air-conditioner interferes the target driver's speech.

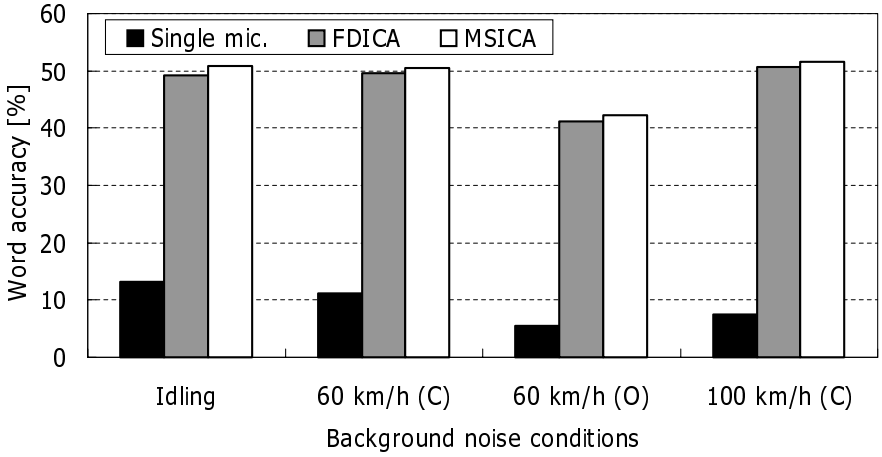


Figure 32. Comparison of word accuracy rates obtained by a single microphone, conventional FDICA and the proposed MSICA under the condition that the assistant speech interferes the target driver's speech with defroster noise.

improve the SNR by about 2.7 dB over that of FDICA and by about 4.3 dB over that of TDICA, for an average of 12 speaker-combinations.

5. Overdetermined BSS on MSICA Using Subarray Processing

5.1 Introduction

In order to improve the separation performance, we have proposed MSICA (see Sect. 4), in FDICA [17, 18, 19] and time-domain ICA (TDICA) [11, 22, 29, 40] are cascaded (see Fig. 33). In the original MSICA research, the specific mixing model is assumed where the number of microphones is equal to that of sources. However, additional microphones are required to achieve an improved separation performance because of the existence of the reflection and the reverberation component. In this section, we set the number of microphones to be larger than that of sources and we extend the conventional MSICA into a new MSICA method using a large microphones. We point out that the following problems arise in the simple extension of MSICA: (1) the permutation problem [17, 38] in FDICA part becomes very complicated, and (2) the solution of FDICA is likely to be trapped within a trivial solution as described in Sect. .

In this section, as a robust method against these problems, we propose a new MSICA method using subarray processing, where the number of each subarray's microphones is set to be equal to that of the sources, and the outputs of FDICA performed in every subarray are weighted to be inserted into TDICA.

The rest of this section is organized as follows. In Sect. , the simple extensions of the original MSICA algorithm and their problems are explained. In Sect. , the proposed MSICA using the subarray processing is described in detail. In Sect. , from the signal-separation experiments, the problems in the simply extended MSICA are described and the superiority of the proposed subarray technique over the conventional method is shown.

5.2 Simple Extension of Conventional MSICA

In the original MSICA, the specific mixing model is assumed, where the number of microphones is equal to that of sources (see Fig. 33). However, additional microphones are required to achieve an improved separation performance because of the reflection and the reverberation component. Thus, we should set the

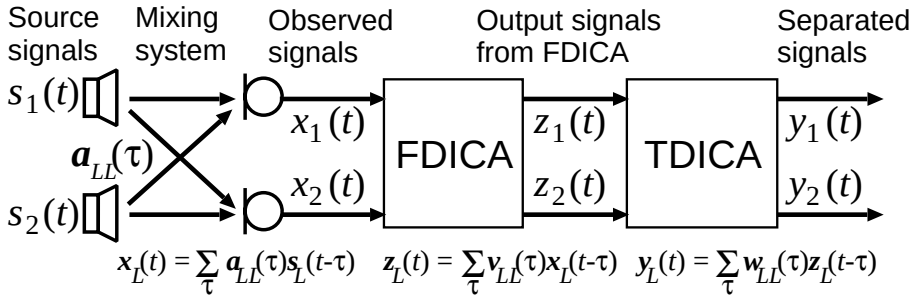


Figure 33. BSS procedure performed in original MSICA.

number of microphones to be larger than that of sources (i.e., $K > L$), and we extend the original MSICA into a new MSICA method by using a large number of microphones. First, as the simple extension of MSICA, we consider the following two methods in the specific case of $K > L$.

[**Method 1**]

Figure 34 shows the BSS procedure performed in Method 1-based MSICA. In this method, the K output signals are obtained from FDICA and L separated signals are obtained from TDICA:

$$\boldsymbol{z}_K(t) = \sum_{\tau=0}^{Q-1} \boldsymbol{v}_{KK}(\tau)\boldsymbol{x}_K(t-\tau), \tag{71}$$

$$\boldsymbol{y}_L(t) = \sum_{\tau=0}^{R-1} \boldsymbol{w}_{LK}(\tau)\boldsymbol{z}_K(t-\tau). \tag{72}$$

There is a permutation problem [17] of sources in every frequency bin in FDICA. By using recently proposed techniques [28, 36, 37, 38], we can easily solve the problem only in the case of $K = L$. However, **(P1)** the permutation problem in FDICA becomes very complicated as the number of microphones is increased. Also, **(P2)** the discrimination of the output signals corresponding to the true sources is needed because there exist $K - L$ imaginary outputs. Therefore Method 1 is not applicable to separating sources in the real environment.

[**Method 2**]

Figure 35 shows the BSS procedure performed in Method 2-based MSICA. In this method, the L output signals are obtained from FDICA and the L separated

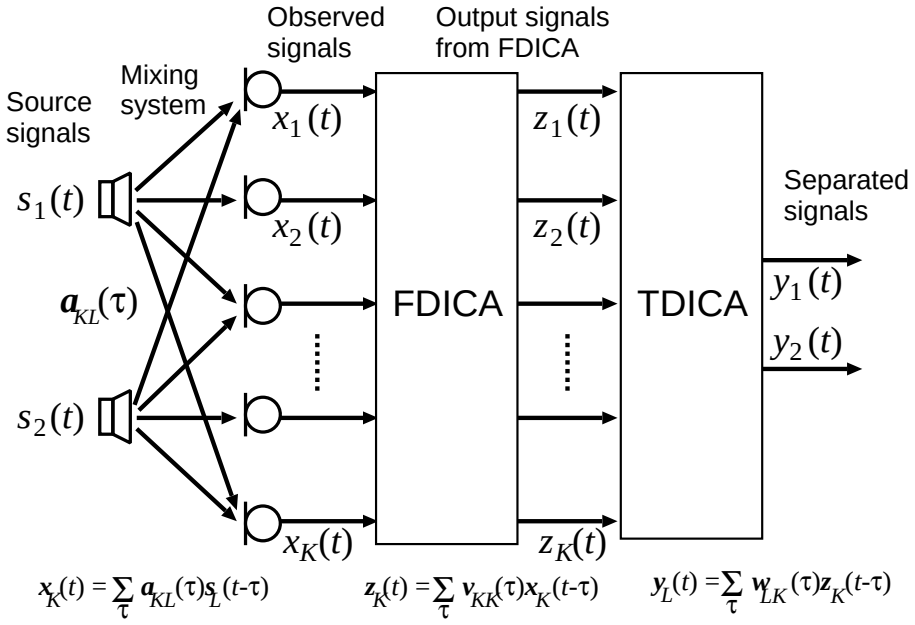


Figure 34. BSS procedure performed in Method 1-based MSICA.

signals are obtained from TDICA:

$$\boldsymbol{z}_L(t) = \sum_{\tau=0}^{Q-1} \boldsymbol{v}_{LK}(\tau)\boldsymbol{x}_K(t-\tau), \tag{73}$$

$$\boldsymbol{y}_L(t) = \sum_{\tau=0}^{R-1} \boldsymbol{w}_{LL}(\tau)\boldsymbol{z}_L(t-\tau). \tag{74}$$

There still exist some problems as follows. **(P3)** In the iterative learning of FDICA, the solution is likely to be trapped within a trivial solution as described in Sect. 5.3.2. **(P4)** We cannot utilize all the information of the observed signals at K microphones in TDICA because the number of the input signals for TDICA is decreased to L by FDICA.

Due to these problems, a new extension algorithm of MSICA which is not affected by **(P1)**–**(P4)** is desired to achieve a superior separation performance. Therefore, in the next section we propose a new BSS algorithm based on the extended MSICA using subarray processing.

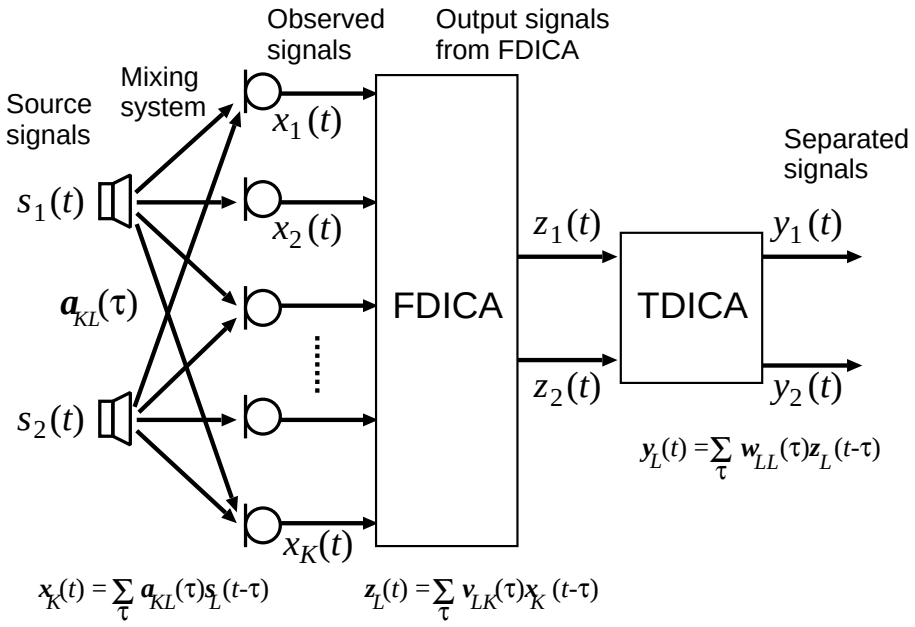


Figure 35. BSS procedure performed in Method 2-based MSICA.

5.3 Simulation Experiments Using Simply Extended MSICA Based on Method 2

5.3.1 Experimental Setup

A 14-element array with the interelement spacing of 2.83 cm is assumed. The speech signals are assumed to arrive from two directions, $-40°$ and $20°$ (direction normal to the array is set to be $0°$). The distance between the microphone array and the loudspeakers is 2.0 m (see Fig. 36). Two sentences spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research [44] are used as the original speech samples. The sampling frequency is 8 kHz and the length of speech is limited to within 3 seconds. Using these sentences, we obtain 12 combinations with respect to speakers and source directions. In these experiments, we use the following signals as the source signals: the original speech convolved with the impulse responses specified by the reverberation times of 300 ms. We use the impulse responses recorded in a real room selected from the

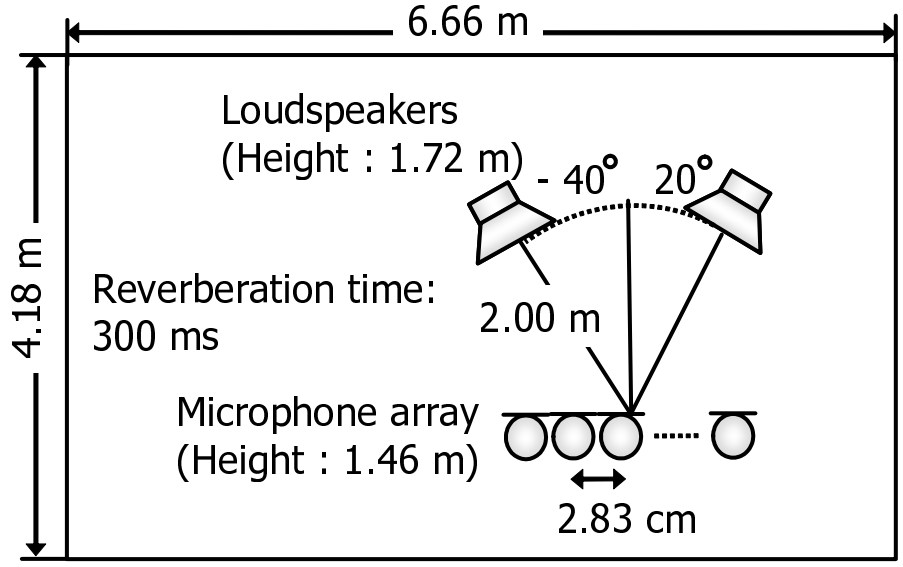


Figure 36. Layout of reverberant room [27] used in experiments.

Real World Computing Partnership (RWCP) sound scene database [27]. These sound data are artificially convolved with the real impulse responses. In order to evaluate the performance, we used the NRR defined in Sect. 3.2.2.

5.3.2 Problems in Simply Extended Method 2-Based MSICA

As the analysis conditions, the filter length of FDICA is 1024 taps and the initial value of FDICA is the DS beamformer in which the beam steered toward $\pm 60°$. Also, the number of iterations of FDICA is 150.

In order to visually evaluate the convergence by FDICA of Method 2, we plot the directivity pattern of the separation filter $\boldsymbol{v}_{LK}(\tau)$ provided by FDICA of Method 2 (Eq. (73)). Figure 37 shows the directivity pattern for a different number of microphones ($K = 2$ or 12), where "Filter 1" is extracting source 1, and "Filter 2" is extracting source 2. In Fig. 37 (a), the directional nulls of the separation filters given by FDICA steer in the direction of interference when two microphones are used. However, in Fig. 37 (b) where 12 microphones are used, the nulls of separation filter 2 steer not only in the direction of interference but

also in the target speech direction. Therefore, the output signal from separation filter 2 becomes a zero signal.

In FDICA, the separation filters are updated so that the output signals are mutually independent and the separated signal from FDICA can be generally given as

$$Z_l(f,m) = c_l(f)S_l(f,m), \tag{75}$$

where $S_l(f,m)$ is the source signal in the time-frequency domain and $c_l(f)$ is the arbitrary complex-valued coefficient. The coefficient $c_l(f)$ is not determined because we evaluate only the independence between the output signals in FDICA. The coefficient $c_1(f)$ in Fig. 37(b) becomes approximately zero and the output signal from filter 1 becomes the zero signal. The speech signal and the zero signal are mutually independent and consequently, the independence assumption holds. However, needless to say, this solution is trivial with respect to the separation of source signals. This phenomenon occurs due to the fact that the degree of freedom of the separation filter becomes high when we use many microphones. We can conclude that the separation filter with a low degree of freedom is desirable in FDICA. This is the motivation behind proposing the extended MSICA using subarray processing in which the number of each subarray's microphones is equal to that of sources.

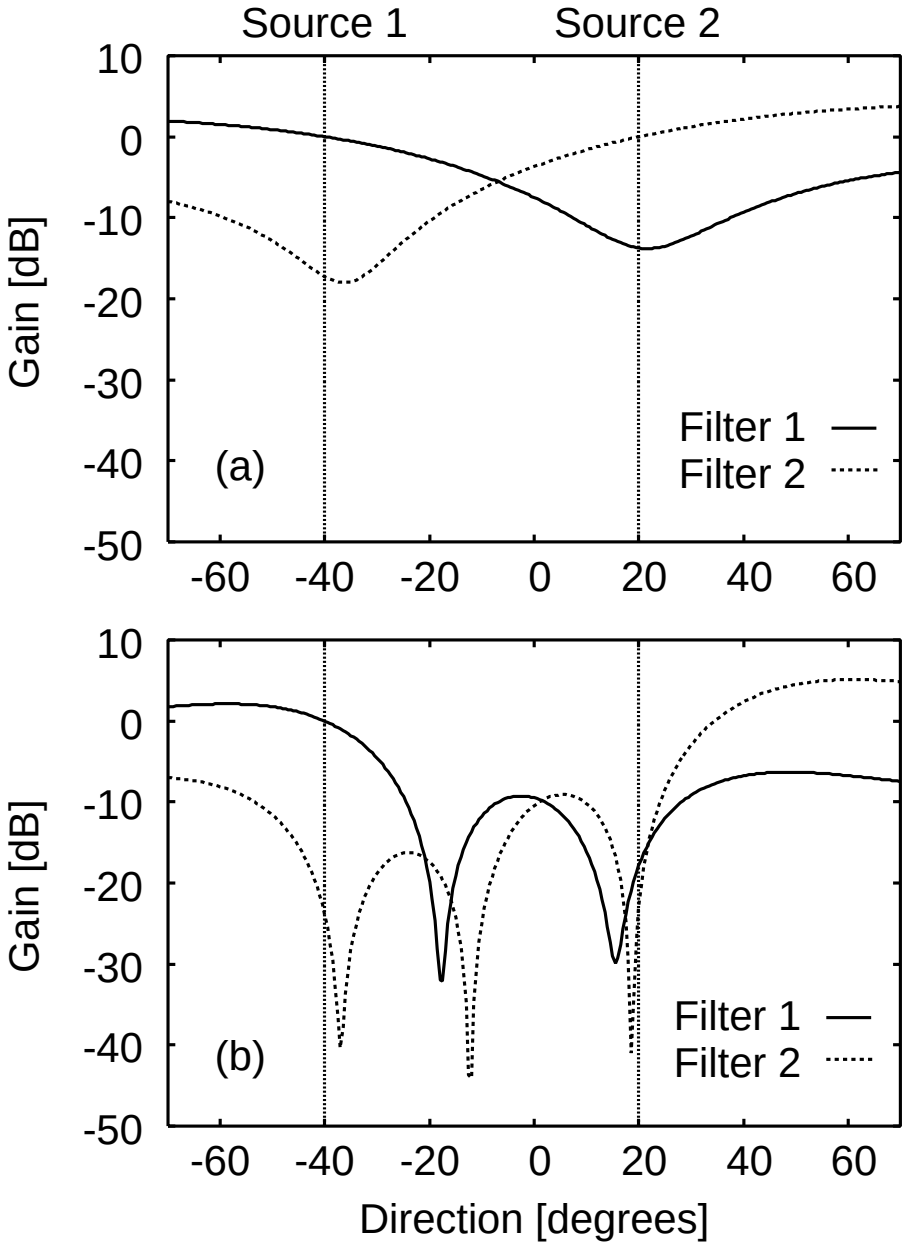


Figure 37. Directivity patterns in 1812.5 Hz of the separation filters provided by FDICAs of Method 2 (Eq. (73)) by using (a) two microphones and (b) 12 microphones. The number of sources is two.

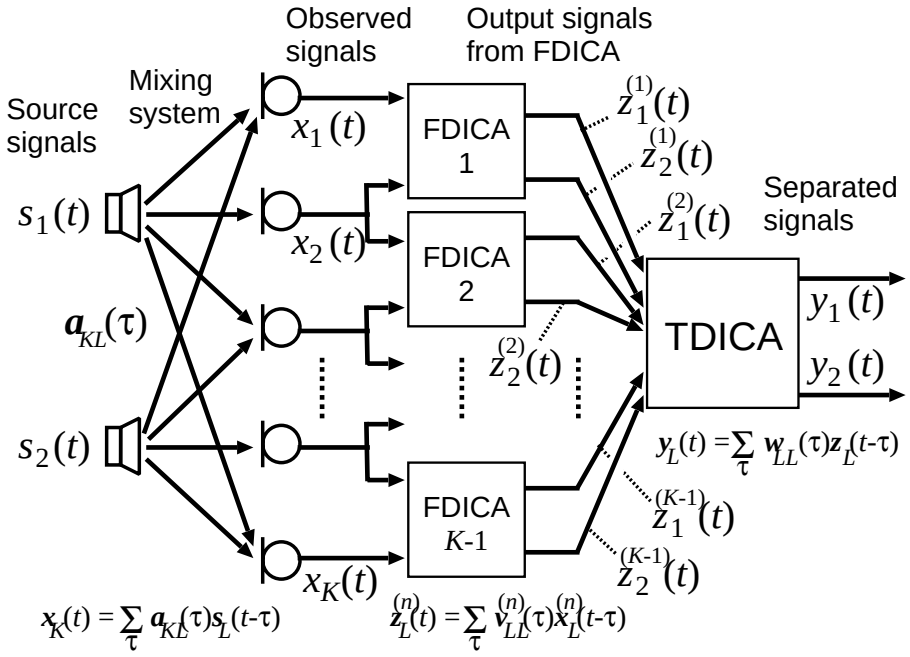


Figure 38. BSS procedure performed in the proposed MSICA using subarray processing.

5.4 Proposed MSICA Using Subarray Processing

5.4.1 Source Separation Algorithm

In the proposed extended MSICA, we regard the K observed signals as combinations of the $L(< K)$ observed signals, and we regard this combination as a subarray (see Fig. 38). First, we divide the whole inputs into $K-1$ subarrays, and we perform FDICA in every subarray. The output signals $\boldsymbol{z}_L^{(n)}(t)$,

$$\boldsymbol{z}_L^{(n)}(t) = [z_1^{(n)}(t), \cdots, z_L^{(n)}(t)]^{\mathrm{T}}, \tag{76}$$

from FDICA in the n-th subarray can be given as

$$\boldsymbol{z}_L^{(n)}(t) = \sum_{\tau=0}^{R-1} \boldsymbol{v}_{LL}^{(n)}(\tau)\boldsymbol{x}_L^{(n)}(t-\tau), \tag{77}$$

where $\boldsymbol{v}_{LL}^{(n)}(\tau)$ is the separation filter matrix of FDICA in the n-th subarray and

$$\boldsymbol{x}_L^{(n)}(t) = [x_n(t), x_{n+1}(t), \cdots, x_{n+L-1}(t)]^{\mathrm{T}}. \tag{78}$$

which consists the adjoining microphones. In this study, we construct the subarray consisting of microphones in which the spatial aliasing effect does not arise. As the FDICA algorithm for optimization of the separation filter $\boldsymbol{v}_{LL}^{(n)}(\tau)$, we introduce the fast-convergence FDICA proposed by one of the authors [19]. In the FDICA, the optimal $\boldsymbol{v}_{LL}^{(n)}(\tau)$ is obtained by the following iterative equation (37) [17] :

$$\begin{aligned}\boldsymbol{V}_{LL}^{(n)}(f)_{[i+1]} &= \alpha\Big[\mathrm{diag}\Big(\langle\boldsymbol{\Phi}(\boldsymbol{Z}_L^{(n)}(f,m))\boldsymbol{Z}_L^{(n)}(f,m)^{\mathrm{H}}\rangle_m\Big) \\ &\quad -\langle\boldsymbol{\Phi}(\boldsymbol{Z}_L^{(n)}(f,m))\boldsymbol{Z}_L^{(n)}(f,m)^{\mathrm{H}}\rangle_m\Big]\boldsymbol{V}_{LL}^{(n)}(f)_{[i]} + \boldsymbol{V}_{LL}^{(n)}(f)_{[i]}, \end{aligned}\tag{79}$$

where $\boldsymbol{V}_{LL}^{(n)}(f)$ is a Fourier transform result of $\boldsymbol{v}_{LL}^{(n)}(\tau)$ and $\boldsymbol{Z}_L^{(n)}(f,m)$ is the narrow-band output signal in the time-frequency domain. Also, f is frequency, m is the analysis frame of short-time DFT, $\langle\cdot\rangle_m$ denotes the frame-averaging operator. We define the nonlinear vector function $\boldsymbol{\Phi}(\cdot)$ as Eq. (32).

Next, we regard all output signals from FDICA in $K-1$ subarrays as the input signals for TDICA, and we remove the residual crosstalk components from FDICAs. The resultant separated signals $\boldsymbol{y}_L^{(n)}(t)$ can be given as

$$\boldsymbol{y}_L(t) = \sum_{\tau=0}^{R-1} \boldsymbol{w}_{LL\cdot(K-1)}(\tau)\boldsymbol{z}_{L\times K-1}(t-\tau), \tag{80}$$

where $\boldsymbol{w}_{LL\cdot(K-1)}(\tau)$ is the $LL\cdot(K-1)$ separation filter matrix and

$$\begin{aligned}\boldsymbol{z}_{L\cdot(K-1)}(t) &= [z_1^{(1)}(t), z_1^{(2)}(t), \cdots, z_1^{(K-1)}(t), \\ &\quad z_2^{(1)}(t), z_2^{(2)}(t), \cdots, z_2^{(K-1)}(t), \cdots, \\ &\quad z_L^{(1)}(t), z_L^{(2)}(t), \cdots, z_L^{(K-1)}(t)]^{\mathrm{T}}. \end{aligned}\tag{81}$$

In the TDICA, the optimal $\boldsymbol{w}_{LL\cdot(K-1)}(\tau)$ is obtained by the following iterative equation (49) [40]:

$$\begin{aligned}\boldsymbol{w}_{LL\cdot(K-1)}(\tau)_{[i+1]} &= \beta\sum_{d=0}^{R-1}\Big\{\mathrm{diag}\Big(\langle\boldsymbol{\phi}(\boldsymbol{y}_L(t))\boldsymbol{y}_L(t-\tau+d)^{\mathrm{T}}\rangle_t\Big) \\ &\quad -\langle\boldsymbol{\phi}(\boldsymbol{y}_L(t))\boldsymbol{y}_L(t-\tau+d)^{\mathrm{T}}\rangle_t\Big\}\boldsymbol{w}_{LL\cdot(K-1)}(d)_{[i]} \end{aligned}$$

$$+\boldsymbol{w}_{LL\cdot(K-1)}(\tau)_{[i]}. \tag{82}$$

We can easily solve the permutation problem by using the conventional methods [28, 36, 38] because the number of microphones is equal to that of sources in every subarray. Also, the discrimination of the output signals corresponding to the true sources is not required because the number of output signals from FDICA is equal to that of sources, i.e., there are no imaginary outputs. The separation filter of FDICA is likely to converge on the optimal point, particularly in the case of $K = L$ (see Sect. 5.3.2). Therefore, in the proposed MSICA, the problems **(P1)**–**(P3)** described in Sect. 5.2 do not arise. In addition, we can utilize the information of all the element of the microphone array in the TDICA because we use the output signals from FDICA in all subarrays with the information from all microphones. Therefore, **(P4)** is also solved by the proposed MSICA.

5.4.2 Initial Value for TDICA Part in Proposed MSICA

As the initial value of the TDICA part in the proposed MSICA, we introduce the following coefficient:

$$\boldsymbol{w}_{LL\times K-1}(\tau)_{[0]} = \begin{cases} \left[\dfrac{c_{k-(l-1)\times K-1}^{-\gamma}}{\sum_{n=1}^{K-1} c_n^{-\gamma}} \cdot \mathrm{IDFT}[\exp(-j\omega d_{lk})]\right]_{lk} \\ \qquad\qquad \text{if } (l-1)\times K-1 < k \le l\times K-1, \\ [0]_{lk} \qquad\quad \text{otherwise,} \end{cases} \tag{83}$$

$$c_n = \sum_{\tau=-T}^{T} \Big\{ |\langle \phi(z_i^{(n)}(t)) z_j^{(n)}(t-\tau)\rangle_t| + |\langle \phi(z_j^{(n)}(t)) z_i^{(n)}(t-\tau)\rangle_t| \Big\}, \tag{84}$$

where $[\cdot]_{lk}$ denotes the matrix in which the lk-th element is $[\cdot]$, IDFT$[\cdot]$ denotes an inverse DFT of $\cdot$, T is the length of the output signals from FDICA, ω is an angular frequency, and d_{lk} is the phase delay of input signals for TDICA so that the correlation between the input signal ${z_l}^{(i)}$ and ${z_l}^{(j)}$ is maximum. Also, γ is the enhancement parameter to weight with the correlation c_n. c_n corresponds to the Frobenius norm of the update term $\{\cdot\}$ in the TDICA algorithm given by Eq. (49), and we estimate the degree of the separation performance by using this value. We introduce this filter (Eq. (83)) as the initial value of the TDICA part in MSICA. If $\gamma = 0$ in Eq. (83), this filter corresponds to a conventional

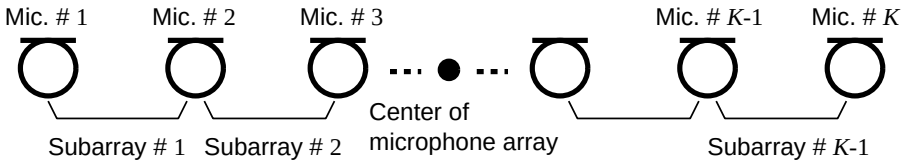


Figure 39. Configuration of the microphone array and the subarray used in experiments.

delay-and-sum beamformer. On the other hand, highly separated output signals from specific FDICAs are strongly weighted as the γ is increased. We compare the separation performances of the initial value and the proposed MSICA by changing γ and the number of microphones.

5.5 Simulation Experiments Using Subarray Processing

5.5.1 Separation Results of FDICA and Conventional MSICA in Each Subarray

The experimental condition is the same as that given in Sect. 5.3.1. Figure 39 shows the numbers of the microphone array and the subarray used in the experiments. We determined the subarrays so that the spatial aliasing effect does not arise. That is, the interelement spacing should be smaller than half of the minimum wavelength to avoid the spatial aliasing effect. In this experiment, this spacing is 8.5/2 cm because the sampling frequency is 8 kHz. The interelement spacing is 2.83 cm in this experimental condition and we used the adjoining microphones for constructing a subarray. As the analysis conditions, the filter length of FDICA is 1024 taps and the initial value of FDICA is the null beamformer in which the null steered toward $\pm 60^\circ$. The separation filter length of the TDICA part in MSICA is 2048 taps. Also, the number of iterations of FDICA is 150 and that of TDICA is 500.

Figure 40 shows the NRR results of FDICA and the conventional MSICA for different subarrays. For example, "2+3 (2)" denotes the experimental result in subarray #2 which consists of microphones #2 and #3 (see Fig. 39). These

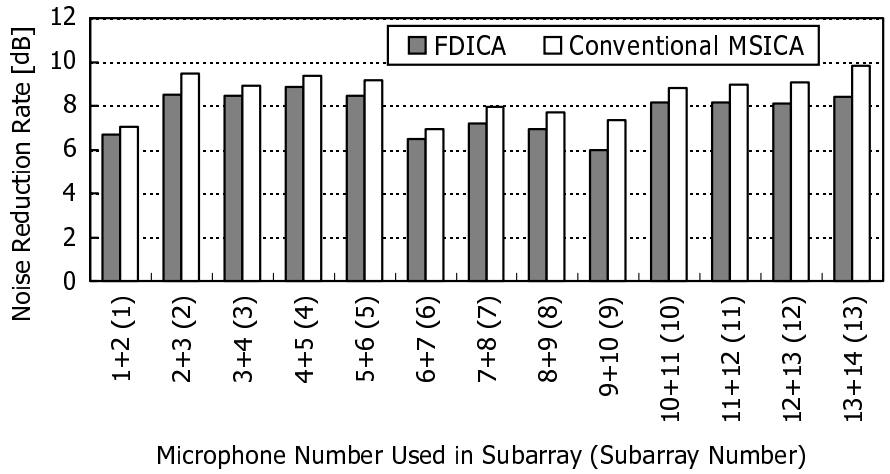


Figure 40. Comparison of the source-separation performance by FDICA and conventional MSICA in every subarray. For microphone number and subarray number see Fig. 39.

separation performances are averaged for 12 combinations of speakers. From Fig. 40, we can confirm that the source-separation performances in each subarray are disperse. We speculate the reason as being that there are differences in the standing wave condition, the reflection component, and reverberant component at each microphone. The blind determination of the subarray which can achieve a superior separation performance is a difficult problem. Also, we must perform the conventional MSICA in all subarrays and huge amounts of calculations are required. Therefore, it is unreasonable to perform the original MSICA in each subarray.

5.5.2 Separation Results of Proposed MSICA for Different Initial Values in TDICA Part

In the proposed MSICA using subarray processing, the microphones which are selected symmetrically with respect to the array center (see the black circle in Fig. 39) are used. For example, the "four-element array" consists of microphones #5, #6, #7, and #8.

Figures 41 and 42 show the NRR results of the initial value and the proposed

MSICA for different γ and numbers of microphones. From Fig. 41, the separation performances of the initial value for the proposed MSICA are improved as γ is increased in all microphones. Therefore, the weighting equation (Eq. (83)) with the input signals for TDICA works effectively. The final separation performance is improved as the number of microphones is increased (see Fig. 42). However, the separation performances of the proposed MSICA which are improvements from the initial values using different γ are not very different in all microphones. We can conclude that the proposed MSICA does not depend on the initial value in the TDICA part and we can achieve a superior separation performance by using the information from many microphones.

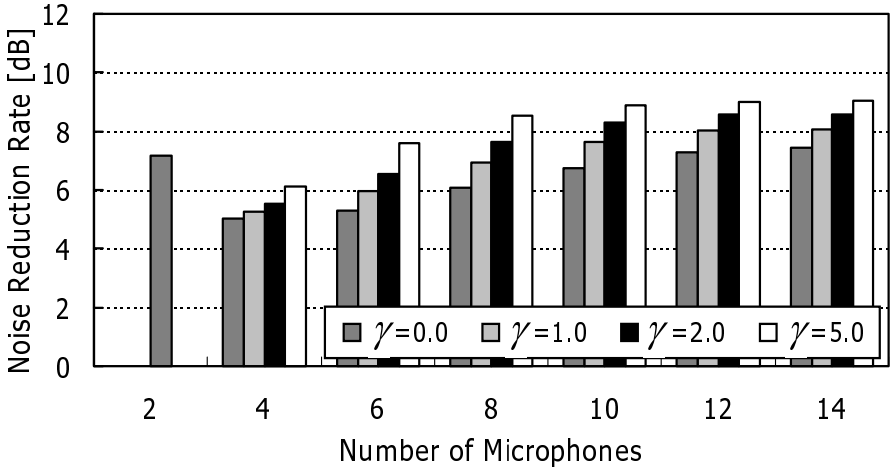


Figure 41. Comparison of the initial values in the TDICA part of the proposed MSICA for different γ and numbers of microphones.

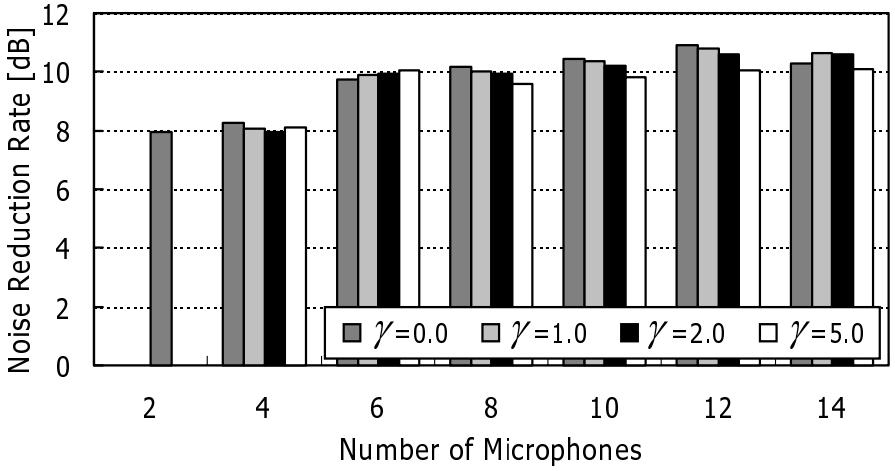


Figure 42. Comparison of the proposed MSICA for different γ and numbers of microphones.

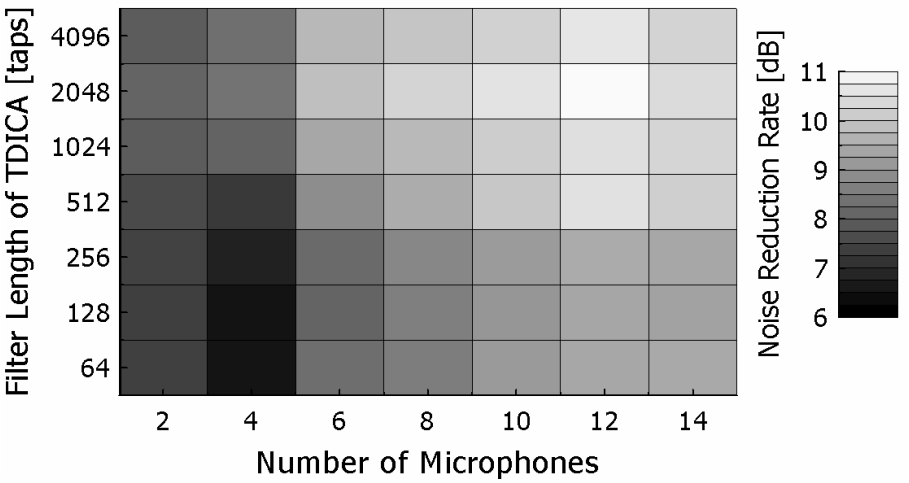


Figure 43. Relationship between the source-separation performance and the number of microphones or filter length of TDICA part.

5.5.3 Relationship between Separation Performance and the Number of Microphones or Filter Length

Figure 43 shows the NRR results of the proposed extended MSICA for different numbers of microphones or filter lengths of the TDICA part. In Fig. 43, the horizontal axis shows the number of microphones, the vertical axis shows the filter length and the tone shows the separation performance.

We investigate the relationship between the source-separation performance and the number of microphones or the filter length. On observing the horizontal axis in Fig. 43 it is seen that the separation performance is improved as the number of microphones is increased. Moreover, on observing the vertical axis we note that the separation performance is also improved as the filter length is increased. These results show the same tendencies as those for the conventional microphone array processing, e.g., in terms of delay and sum beamformer. However, in the proposed MSICA, huge amounts of calculations are required. The increase in the number of microphones corresponds to an increase in the number of FDICAs. Therefore, as a future work, we should propose the MSICA with an effective subarray structure.

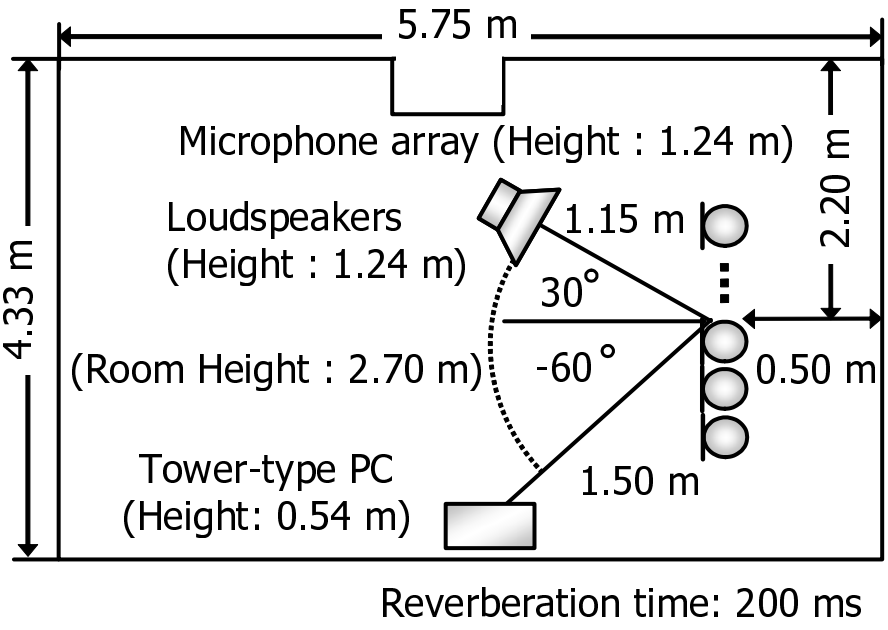


Figure 44. Layout of reverberant room used in real recording experiment.

5.6 Illustrative Experiment with Real Recordings

5.6.1 Conditions for Experiment

In this section, the BSS experiment is performed using actual devices in a real acoustic environment. The experiment was carried out in an ordinary room, which has the RT of 200 ms, as shown in Fig. 44. A 14-element array with interelement spacing of 2.1 cm is used. The source signals arrive from two directions, $-60°$ and $30°$; a loudspeaker is placed on the right-hand side ($30°$) to sound a target female speech, and a tower-type personal computer (PC) is placed on the left-hand side ($-60°$) as an interference (noise) sound generator.

As the analysis conditions for these experiments, the sampling frequency is 16 kHz, the filter length of FDICA and TDICA are 2048 taps and the initial value of FDICA is the null beamformer in which the null steered toward $-30°$ and $50°$. Also, the number of iterations of FDICA is 200, that of TDICA is 200, and $\gamma = 0.0$.

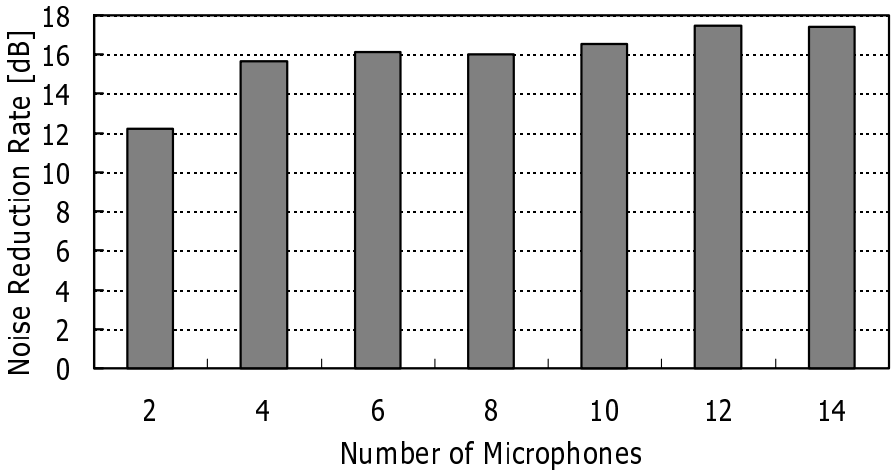


Figure 45. Noise reduction rates for different numbers of microphones under real recording condition. RT is 200 ms, and the background noise level is 37 dB(A).

The level of background noise, which is not the PC noise but an ambient noise, and the target speech level measured at the array origin, were 37 dB(A) and 54 dB(A), respectively. The levels of the target speech and the PC noise are almost even. It also should be mentioned that all of the experimental apparatus may include possible sensor noise, environment noise, and/or nonlinear error which is produced in, for example, amplifiers.

5.6.2 Results

Figure 45 shows NRR results in the proposed MSICA. Note that we only depict the NRR with regard to the target speech in this figure because we consider the PC noise as an uninteresting and hence undesired sound source. The results reveal that the separation performance is also improved as the number of microphones is increased, like the simulation results in Sect. 5.5. This indicates encouraging evidence for the feasibility of the proposed algorithm for real-world applications such as a robust hands-free speech communication system and a hands-free telecommunication system.

5.7 Conclusion

In this section, we proposed a MSICA, by setting the number of microphones to be larger than that of sources to achieve an improved separation performance. In the FDICA part in the simple extension of MSICA, the use of additional microphones led to alternative problems: the solution is likely to be trapped within a trivial solution and the permutation problem in FDICA becomes very complicated. In order to solve these problems, we proposed a new extended MSICA using subarray processing, where the number of microphones and that of sources are set to be the same in every subarray. The experimental results obtained under real acoustic environmental conditions reveal that the separation performance of the proposed MSICA is improved as the number of microphones is increased.

6. Stable and Low-Distortion Algorithm Based on Blind Separation of Temporally Correlated Acoustic Signals

6.1 Introduction

In order to achieve a superior separation performance, we have proposed an efficient BSS algorithm called MSICA, in which FDICA and TDICA are combined. In this method, first, FDICA can find an approximate solution to separate the sources to a certain extent, and finally TDICA can remove the residual crosstalk components from FDICA. Therefore, the improvement of TDICA is a primary issue because the quality of resultant separated signals is determined by TDICA. In this section, we discuss the stability of the TDICA algorithm, and newly propose two stable algorithms for temporally correlated signals, e.g., speech signals. First, the following points are explicitly noted: (1) The stability of learning in conventional TDICA with a holonomic constraint [29] is highly acceptable. The *stability of learning* used in this thesis is defined as, "The separation performance is improved monotonically and the solution of ICA converges in the optimal point or the local minimum point and stays of this point when we use small step-size parameter which is not affected by the divergence or the vibration.". However, the method cannot work well for speech signals due to the deconvolution property; i.e., the separated speech is harmfully distorted by the whitening process. (2) To decrease the whitening effect, TDICA with a nonholonomic constraint has been proposed [40]. This method, however, includes the inherent drawback that the stability of learning cannot be guaranteed.

In order to resolve these problems, The method to compensate the sound qualities has been proposed by Murata [17]. In this method, the inverse matrix of the separation matrix is used for the compensation. However, the stability of the inverse matrix is not guaranteed (see Sect. 6.4.3). Also, the affection of the circular convolution causes serious deterioration when we use this method [54]. As a method without the inverse matrix, Matsuoka et al. have proposed a ICA based on the Minimal Distortion Principle in which the ICA's outputs should be the single components in observed signals at a specific microphone point [55].

However, there is a problem that the cost function for the constraint of the ICA's outputs does not become zero even in the optimal point. The blind determination of the optimal point is difficult because of this problem. Takatani et al. have proposed a ICA based on single-input multiple-output (SIMO) model for solving this problem on the cost function [56, 57, 58]. The ICA's outputs are constrained to the SIMO model at specific microphone points. However, a huge number of calculations are required because we have to estimate all SIMO-model signals at all microphone points in this method. Abe et al. have proposed a ICA based on multiple-input single-output (MISO) model [59]. The ICA's outputs are constrained to the MISO model at a specific microphone point. However, in this method, the optimization of the separation filter becomes complicated because we determined not only step-size parameter for ICA but also the additional parameter for the minimization of the cost function to constrain separates source signals.

In order to solve both problems simultaneously, we propose two stable and low-distortion algorithms for the two cases, i.e., (Case 1) the number of microphones is equal to that of sources, and (Case 2) the number of microphones to be larger than that of sources. For the Case 1, we propose the novel approach in which the linear predictors estimated from the roughly separated source signals by FDICA are inserted before TDICA with a holonomic constraint as a prewhitening processing (after TDICA, the dewhitening is also performed). The stability of the learning in TDICA can be guaranteed by the holonomic constraint, and it is still possible to separate the temporally correlated signals because the pre/dewhitening processing prevents the ICA from performing the decorrelation. For the Case 2, to avoid the distortions, we estimate the distortion components by TDICA with the holonomic constraint and we compensate the sound qualities by using the estimated components. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the proposed compensation method prevents the ICA from performing the decorrelation. The experimental results under a reverberant condition reveal that the proposed algorithm results in the higher stability and higher separation performance than the conventional MSICA.

The rest of this section is organized as follows. In Sect. and , conventional

MSICA algorithms and their problems are explained. In Sect. and , the proposed ICA is described in detail. In Sect. and , the signal-separation experiments are described and the results are compared with those of the conventional methods.

6.2 Conventional MSICA and Their Problems

In this section, we above the procedure of conventional MSICAs (see Fig. 46) and their problems. The separated signals of the conventional MSICA can be given as

$$\boldsymbol{y}(t) = \sum_{\tau=0}^{R-1} \boldsymbol{w}(\tau)\boldsymbol{z}(t-\tau), \tag{85}$$

where $\boldsymbol{y}(t)$ is the resultant separated signal vector of MSICA and $\boldsymbol{z}(t)$ is the input signal vector for the TDICA part in MSICA (i.e., the output signals from FDICA). Also, $\boldsymbol{w}(\tau)$ is the separation filter matrix of TDICA. In this procedure, we optimize $\boldsymbol{w}(\tau)$ so that the separated signals are mutually independent. The selection of TDICA is an important issue because the quality of resultant separated signals is determined by TDICA. We have following two choices for TDICA algorithms; TDICA with a holonomic constraint [29] and TDICA with a nonholonomic constraint [40].

In the TDICA with a holonomic constraint (see Eq. 47), the separation filter is optimize by following iterative equation;
[**H-TDICA**]

$$\boldsymbol{w}_{i+1}^{(\mathrm{H})}(\tau) \;=\; \boldsymbol{w}_{i}^{(\mathrm{H})}(\tau) + \beta \sum_{d=0}^{R-1} \Big\{ \boldsymbol{I}\delta(\tau-d) - \langle \boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big\} \boldsymbol{w}_{i}^{(\mathrm{H})}(d). \tag{86}$$

Here, we note the convergence point of H-TDICA. In the convergence point, the term inside the braces $\{\cdot\}$ becomes zero matrix. This results to the convergence point of off-diagonal elements when the higher-order cross-correlation becomes zero. However this is also the convergence point of the diagonal elements in which the higher-order autocorrelation becomes delta function. This diagonal convergence means that the separated signal from H-TDICA is harmfully distorted by whitening effect and this is the disadvantage of H-TDICA. Thus, this method

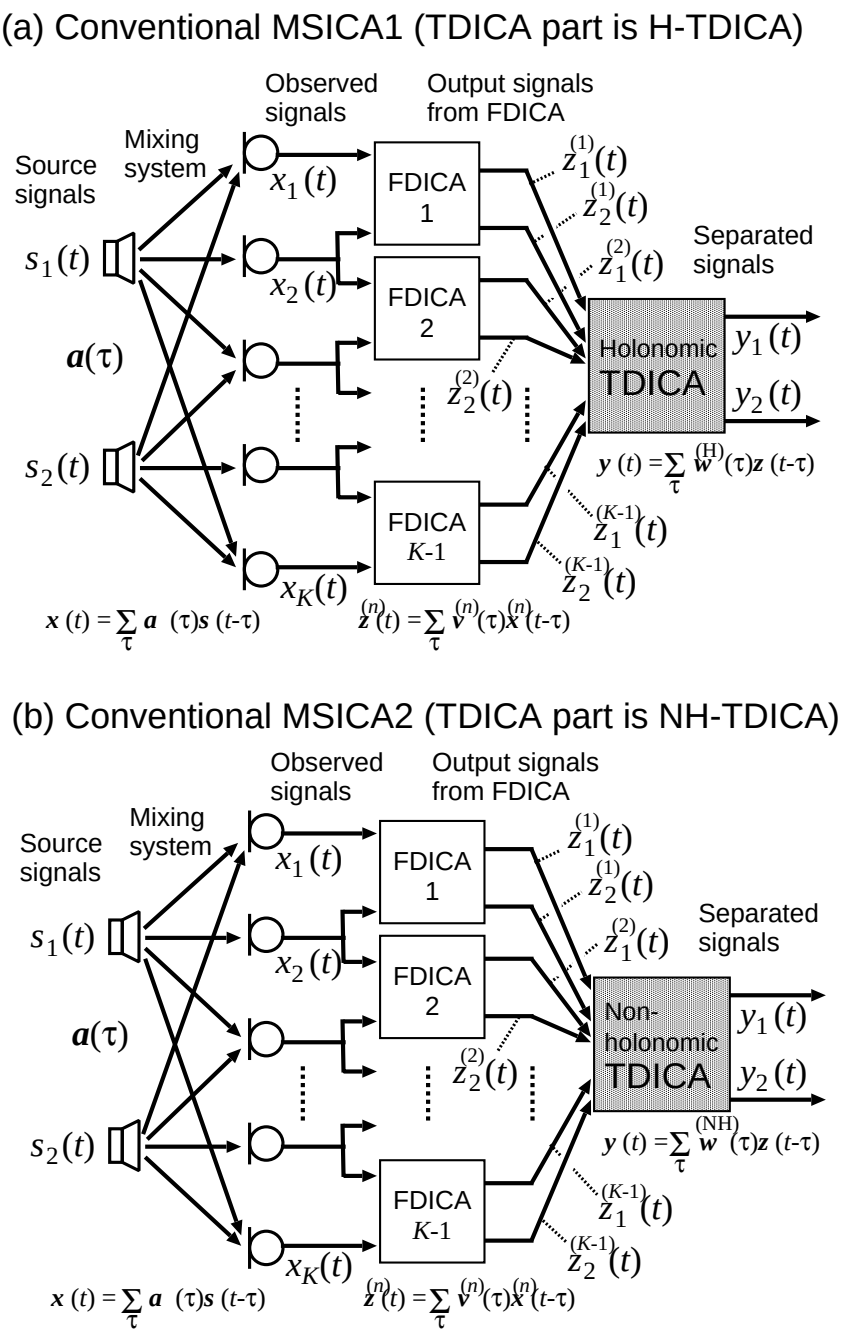


Figure 46. Blind source separation procedures performed in (a) conventional MSICA 1 (TDICA part is H-TDICA) and (b) conventional MSICA 2 (TDICA part is NH-TDICA) .

cannot work well for speech signals due to the deconvolution property. In order to solve the problem of the H-TDICA, Choi proposed a modified TDICA algorithm with a nonholonomic constraint [40]. In the TDICA with a nonholonomic constraint (see Eq. 49), the separation filter is optimize by following iterative equation;

[**NH-TDICA**]

$$\begin{aligned}\boldsymbol{w}_{i+1}^{(\mathrm{NH})}(\tau) &= \boldsymbol{w}_{i}^{(\mathrm{NH})}(\tau) + \beta \sum_{d=0}^{R-1} \Big\{ \mathrm{diag}\Big(\langle \boldsymbol{\phi}(\boldsymbol{y}(t)) \boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big) \\ &\quad - \langle \boldsymbol{\phi}(\boldsymbol{y}(t)) \boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big\} \boldsymbol{w}_{i}^{(\mathrm{NH})}(d). \end{aligned} \tag{87}$$

In this algorithm, Choi changed the first term inside the braces $\{\cdot\}$, $\boldsymbol{I}\delta(\tau - d)$, to be only diagonal components of the second term inside the braces $\{\cdot\}$. By changing this term, the convergence point of diagonal elements becomes an arbitrary value. This arbitrary value will give us freedom to minimize whitening effect, i.e., NH-TDICA is applicable to speech signals. This method, however, includes the inherent drawback that the stability of learning cannot be guaranteed only if we use small step-size parameter which is not affected by the divergence or the vibration. That is, the convergence in an optimal point or a local minimum point and the halt at a balanced point cannot be guaranteed because this algorithm is made artificially without the mathematical derivation from H-TDICA.

The advantage and disadvantage of conventional TDICAs can be summarized as follows. (1) The stability of learning in H-TDICA is satisfactory because H-TDICA is based on direct and correct differentiation from KLD. However, the method cannot work well for speech signals due to the deconvolution property; i.e., the separated speech is harmfully distorted by the whitening process. (2) On the other hand, NH-TDICA possibly performs no deconvolution, i.e., NH-TDICA is applicable to speech signals. This method, however, includes the inherent drawback that the stability of learning cannot be guaranteed as described in Sect.6.4.3. Thus, the separation of temporally correlated signals such as speech cannot be achieved only using the conventional TDICAs.

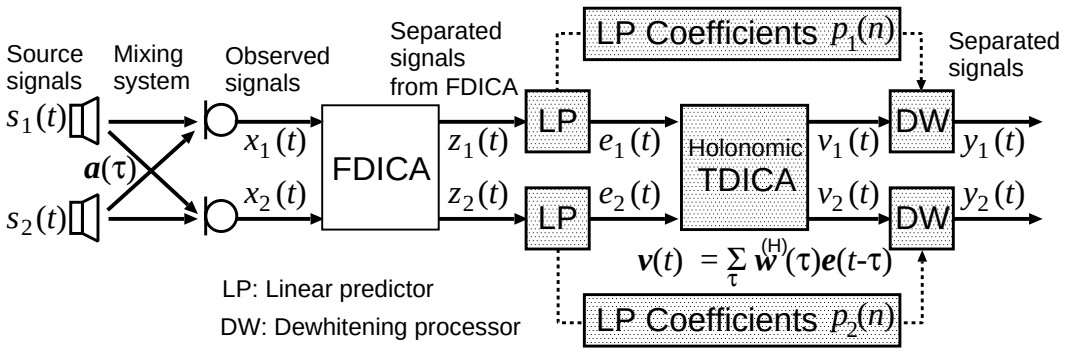


Figure 47. BSS procedure performed in the proposed algorithm combining MSICA and linear prediction. In this system, the stability of the learning in TDICA can be guaranteed by the holonomic constraint, and it is still possible to separate the temporally correlated signals because the pre/dewhitening processing prevents the ICA from performing the decorrelation.

6.3 Proposed Algorithm Combining MSICA and Linear Prediction

This section describes a new stable algorithm combining the linear prediction technique with an original MSICA (see Fig. 46) for the case where the number of microphones is equal to that of sources. In the proposed algorithm, the linear predictors estimated from the roughly separated source signals by FDICA are inserted before TDICA with a holonomic constraint as a prewhitening processing (see Fig. 47). After TDICA, the dewhitening is also performed. The stability of the learning in TDICA can be guaranteed by the holonomic constraint, and it is still possible to separate the temporally correlated signals because the pre/dewhitening processing prevents the ICA from performing the decorrelation. The detailed process using the proposed algorithm is as follows.

[STEP 1. FDICA]

First, we perform FDICA to separate the source signals to some extent, where we apply the iterative equation (37). For example, the typical separation performance in FDICA is about 8 to 10 dB under the condition that the reverberation time is 300 ms [19, 22, 35]. Also, the mel cepstral distortion [63] between the ob-

served signal with the single source component at the microphone and the output signals from FDICA is about 2 to 3 dB [60]. The separation filter of FDICA has spectrally flat characteristics in the direction of each sound source [19, 61]. From this, we can estimate the approximate spectra of the sources blindly.

[STEP 2. Prewhitening by Linear Prediction]

In the linear prediction, the auto-regressive model of the generation process of the output signals $\boldsymbol{z}(t)$ from FDICA is given as

$$z_l(t) = -\sum_{d=1}^{D} p_l(n) z_l(t-n) + e_l(t) \quad (l = 1, \cdots, L), \tag{88}$$

where $p_l(n)$ is a linear prediction coefficient for the l-th input signal, $e_l(t)$ is the input signal of this model, and D is the order of the linear prediction coefficient. The linear prediction coefficient is obtained by calculating the following Yule-Walker's simultaneous equations:

$$\begin{bmatrix} r_l(0) & \cdots & r_l(N-1) \\ \vdots & \ddots & \vdots \\ r_l(N-1) & \cdots & r_l(0) \end{bmatrix} \begin{bmatrix} p_l(1) \\ \vdots \\ r_l(D) \end{bmatrix} = - \begin{bmatrix} r_l(1) \\ \vdots \\ r_l(D) \end{bmatrix}, \tag{89}$$

where $r_l(\tau)$ is the autocorrelation of $z_l(t)$, i.e.,

$$r_l(\tau) = \langle z_l(t) z_l(t-\tau) \rangle_t. \tag{90}$$

Solving Eq. (89) basically involves the inversion of the matrix on the left-hand side. However, this can be simplified because the matrix is Toeplitz, i.e., all elements on each superdiagonal are equal and all elements on each subdiagonal are equal. Based on this property, we can easily and efficiently determine $p_l(n)$ by using Levinson-Durbun's recursive algorithm [62].

The whitened signal $e_l(t)$ is obtained by convolving the linear prediction coefficient $p_l(n)$ with $z_l(t)$ as

$$e_l(t) = \sum_{n=0}^{D} p_l(n) z_l(t-n). \tag{91}$$

[STEP 3. TDICA with Holonomic Constraint]

H-TDICA is performed with whitened signals. The output signals of H-TDICA can be given as

$$\boldsymbol{o}(t) = \sum_{\tau=0}^{R-1} \boldsymbol{w}^{(\mathrm{H})}(\tau)\boldsymbol{e}(t-\tau), \tag{92}$$

where $\boldsymbol{o}(t) = [o_1(t), \cdots, o_L(t)]^{\mathrm{T}}$ is the separated signal vector of H-TDICA, and $\boldsymbol{e}(t) = [e_1(t), \cdots, e_L(t)]^{\mathrm{T}}$ is the input signal vector whitened by the linear prediction for the H-TDICA part in MSICA. We optimize $\boldsymbol{w}^{(\mathrm{H})}(\tau)$ by H-TDICA (see Eq. (47)):

$$\begin{aligned} \boldsymbol{w}_{i+1}^{(\mathrm{H})}(\tau) &= \boldsymbol{w}_i^{(\mathrm{H})}(\tau) + \beta \sum_{d=0}^{R-1} \Big\{ \boldsymbol{I}\delta(\tau-d) \\ &\quad - \langle \boldsymbol{\phi}(\boldsymbol{o}(t))\boldsymbol{o}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big\} \boldsymbol{w}_i^{(\mathrm{H})}(d). \end{aligned} \tag{93}$$

[STEP 4. Dewhitening]
The dewhitening process is performed by using the linear prediction coefficients $p_l(n)$ obtained in STEP 2. The resultant separated signals $y_l(t)$ can be obtained by the following IIR filtering:

$$y_l(t) = -\sum_{n=1}^{D} p_l(n) y_l(t-n) + v_l(t) \quad (l = 1, \cdots, L). \tag{94}$$

Note that the stability of the filtering is guaranteed because $p_l(n)$ is calculated from Levinson-Durbun's algorithm [62].

6.4 Experiments and Results in Case 1

In this section, we compare the proposed algorithm combining MSICA and linear prediction with the conventional MSICAs under the condition where the number of microphones K is equal to that of sources L and $K = L = 2$.

6.4.1 Postprocessing for Spectral Compensation

In order to compare the various ICAs fairly, we perform postprocessing for the spectral compensation of the separated signals in this experiment. This processing

is based on the utilization of the inverse of the separation filter matrix for the normalization of gain [17]. In this method, the following operation is performed:

$$\tilde{Y}_l(f) = \Big[\boldsymbol{W}(f)^{-1}[\,0,\,\cdots,\,0\,,\,Y_l(f),\,0,\,\cdots,\,0\,]^{\mathrm{T}}\Big]_l, \tag{95}$$

where $Y_l(f)$ denotes the frequency-domain component of the l-th estimated source signal by the TDICA part in MSICA, $\tilde{Y}_l(f)$ denotes the l-th resultant separated source signal in the frequency domain after the spectral compensation ($[\cdot]_l$ denotes the l-th element of the argument), and $\boldsymbol{W}(f)$ is the frequency-domain representation of the separation filter matrix in the TDICA part, $\boldsymbol{w}^{(\mathrm{H})}(\tau)$ or $\boldsymbol{w}^{(\mathrm{NH})}(\tau)$. After the operation, we can obtain the spectrally compensated output signals in the time domain by applying the inverse DFT. In this experiment, we use the DFT and the inverse DFT of 2^{18} points.

By using $\boldsymbol{W}(f)^{-1}$, the gain arbitrariness vanishes in the separation procedure. However, this procedure often fails and yields harmful results for signal reconstruction, particularly when the condition number of $\boldsymbol{W}(f)$ is large because the invertibility of $\boldsymbol{W}(f)$ cannot be guaranteed.

6.4.2 Objective Evaluation Score; Mel Cepstral Distortion

Aside from noise reduction, low-distortion is a necessary task to achieve a noise-robust hands-free speech recognition and a high-quality hands-free telecommunication system. To evaluate the degree of the distortion of the separated signal, we introduce the 16-order *Mel cepstral distortion* (MelCD) [63]. The MelCD for l-th source signal is defined as

$$\mathrm{MelCD_l} \equiv \frac{20}{\ln 10}\sqrt{2\sum_{i=1}^{16}\Big(m_l^{(\mathrm{ref})}(i) - m_l^{(\mathrm{target})}(i)\Big)}, \tag{96}$$

where $m_l^{(\mathrm{ref})}(i)$ is the i-th mel cepstral coefficient for the observed signal with the single source component at the microphone and $m_l^{(\mathrm{target})}(i)$ is that for the output signal from ICA.

6.4.3 Experimental Results and Discussion

In this study, we compare the following MSICAs:

MSICA1: FDICA is followed by NH-TDICA,

MSICA2: FDICA is followed by NH-TDICA with spectral compensation,

MSICA3: FDICA is followed by H-TDICA,

MSICA4: FDICA is followed by H-TDICA with spectral compensation, and

MSICA5: FDICA is followed by the proposed method combining H-TDICA and linear prediction.

The experimental condition is the same as that given in Sect. 3.2.1. The analysis conditions of FDICA in these experiments are shown in Table 6. The length of the separation filters of TDICAs, $\boldsymbol{w}^{(\mathrm{H})}(\tau)$ or $\boldsymbol{w}^{(\mathrm{NH})}(\tau)$, is 2048. In the proposed algorithm, the order N in the linear predictor is 1024.

Figures 48, 49, and 50 show the NRR results of MSICA1–MSICA5 for different iteration points. These values were averages of all of the combinations with respect to speakers and source directions. The step-size parameters are chosen independently for each of the NH-TDICA, H-TDICA, and the proposed algorithm so that the NRR scores at the early iterations are almost the same in Figs. 48, 49, and 50. From these results, the following are revealed. (1) In the conventional MSICA1 and MSICA2 in which the NH-TDICA is used, the behavior of the NRR is not monotonic and there are remarkably consistent deteriorations, even when the step-size parameter is changed. (2) In the proposed algorithm, MSICA5, there are no deteriorations of NRRs. Therefore, the separation performances are almost completely retained during all of the iterations.

Regarding the separation performance of MSICA3 and MSICA4 in which the H-TDICA is used, the following are revealed. (1) The separation performance of MSICA3 is obviously superior to that of the proposed MSICA5. (2) However, its effective separation performance, i.e., the performance of MSICA4, is inferior to that of MSICA5. We speculate that the *specious* performance in MSICA3 is due to the exceeding emphasis of high-frequency components by the whitening effect of H-TDICA. Figure 51 shows the MelCD between the observed signal with the single source component at the microphone and the output signals from (a) conventional MSICA3 or (b) proposed MSICA5. Also, Fig. 52 shows the typical long-time averaged spectra of the separated signals obtained by MSICA3 and

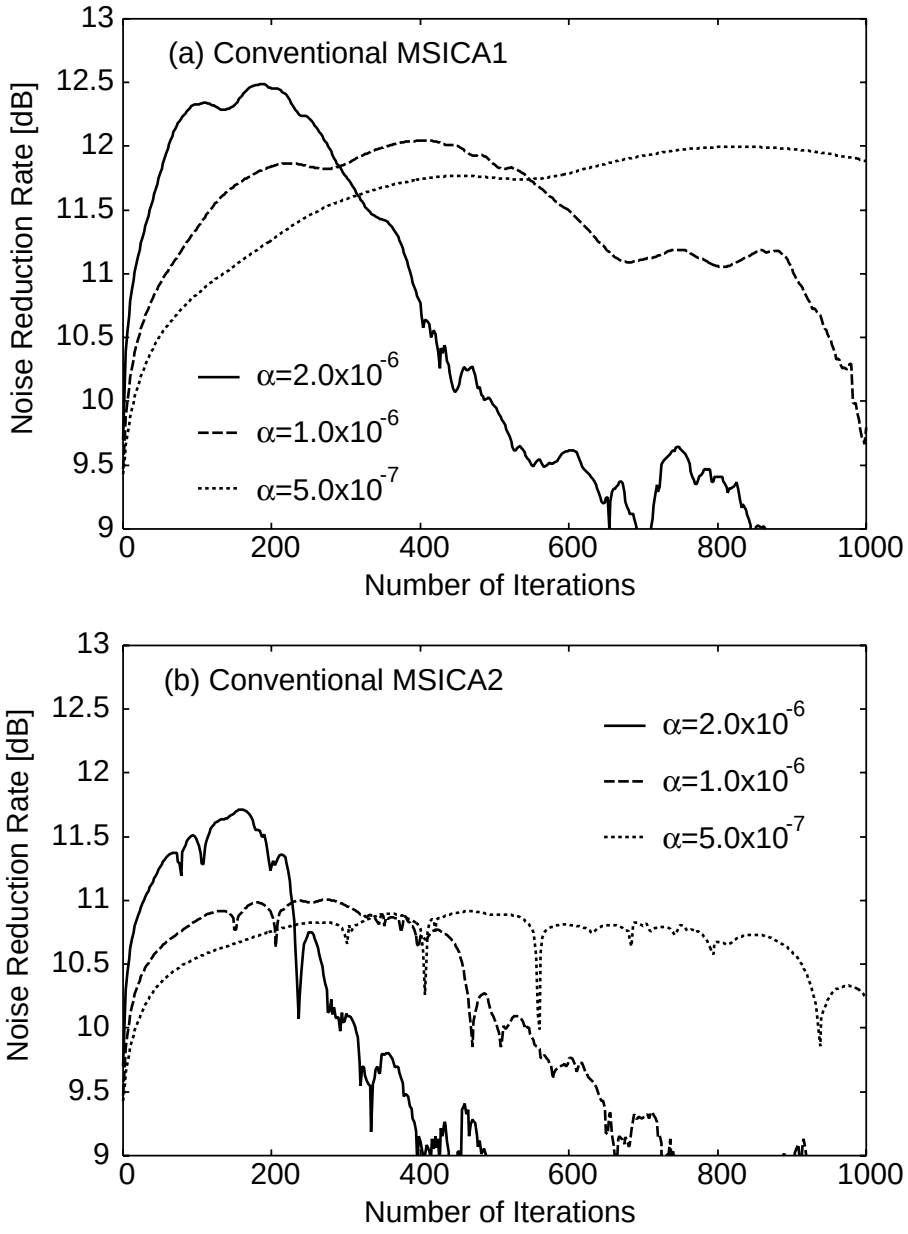


Figure 48. Comparison of the noise reduction rates in (a) conventional MSICA1: FDICA is followed by NH-TDICA and (b) conventional MSICA2: FDICA is followed by NH-TDICA with spectral compensation.

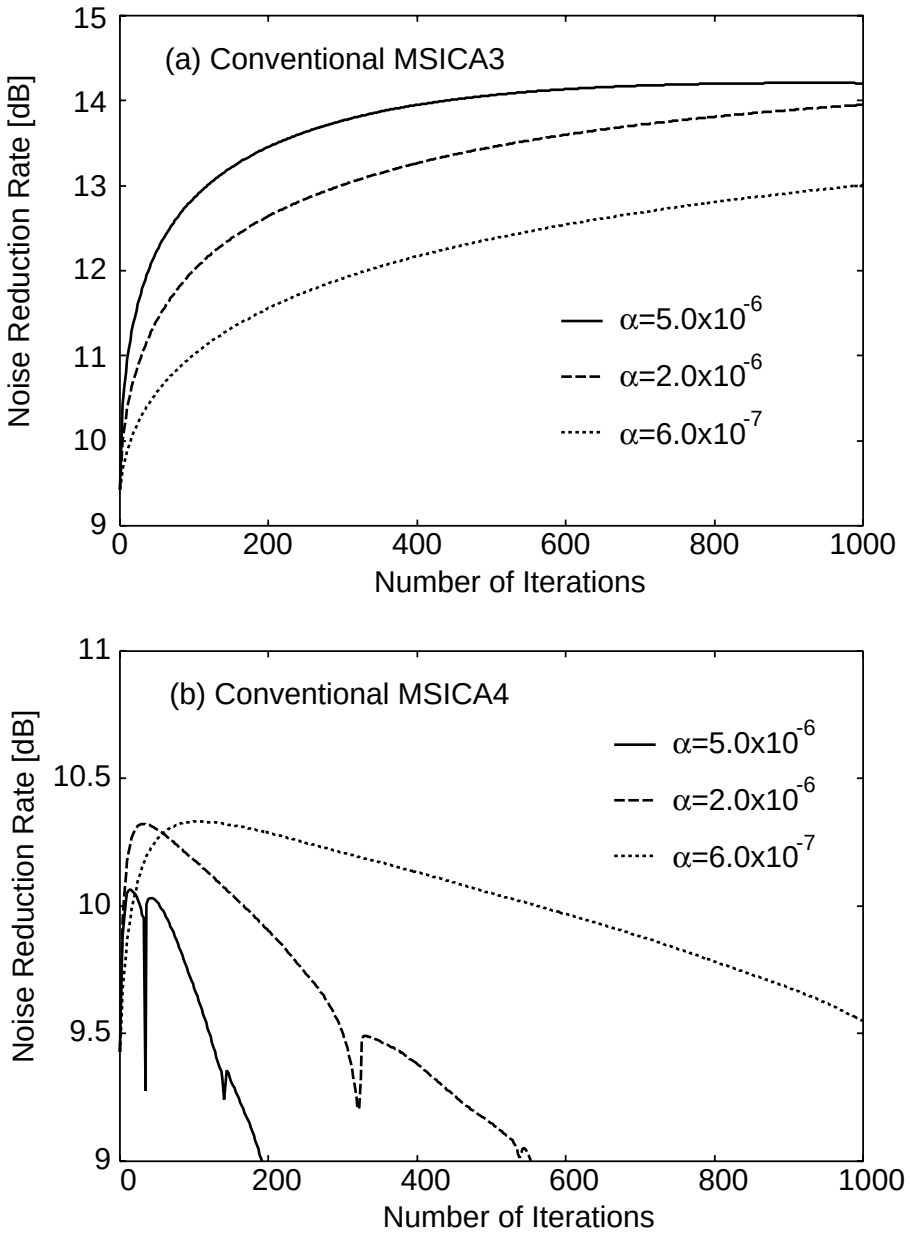


Figure 49. Comparison of the noise reduction rates in (a) conventional MSICA1: FDICA is followed by H-TDICA and (b) conventional MSICA2: FDICA is followed by H-TDICA with spectral compensation.

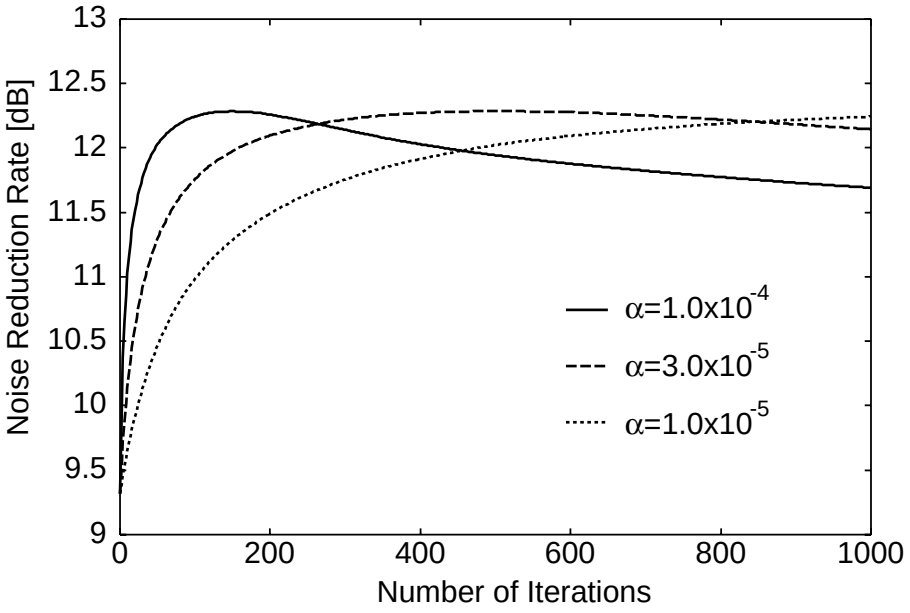


Figure 50. The noise reduction rates in (e) proposed MSICA5: FDICA is followed by the proposed method combining H-TDICA and linear prediction.

MSICA4. From these results, we can confirm the spectral distortion in MSICA3. In general, the separation in the high-frequency region is easier than that in the low-frequency region [64, 65] because the reverberation is shorter as the frequency increases [26]. Thus, MSICA3 gains the improvement of the NRR only in the high-frequency region, and consequently we can conclude that MSICA3 is useless for separating the speech signals from the practical viewpoint.

In order to confirm the convergence of each MSICA learning, we evaluate the frobenius norms of $\{\cdot\}$ parts on the right-hand side in Eqs. (49) and (47), which are defined as

$$\begin{aligned} FN^{(\mathrm{NH})} &= \frac{1}{Q^2}\sum_{\tau=0}^{Q-1}\sum_{d=0}^{Q-1}\Big\|\mathrm{diag}\Big(\langle\boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t\Big) \\ &\quad -\langle\boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t\Big\|, \qquad (97) \\ FN^{(\mathrm{H})} &= \frac{1}{Q^2}\sum_{\tau=0}^{Q-1}\sum_{d=0}^{Q-1}\Big\|\boldsymbol{I}\delta(\tau-d)-\langle\boldsymbol{\phi}(\boldsymbol{v}(t))\boldsymbol{v}(t-\tau+d)^{\mathrm{T}}\rangle_t\Big\|. \qquad (98) \end{aligned}$$

Figures 53(a) and (b) show $FN^{(\mathrm{NH})}$ of the conventional MSICA1 and $FN^{(\mathrm{H})}$ of the proposed MSICA5. These scores correspond to the stability of the iterative learning; it should be monotonically decreased. As shown in these figures, the conventional ICA loses its stability under the nonholonomic constraint. However, the proposed method can converge in every situation and consequently, we can conclude that the proposed algorithm is effective for improving the stability of the learning.

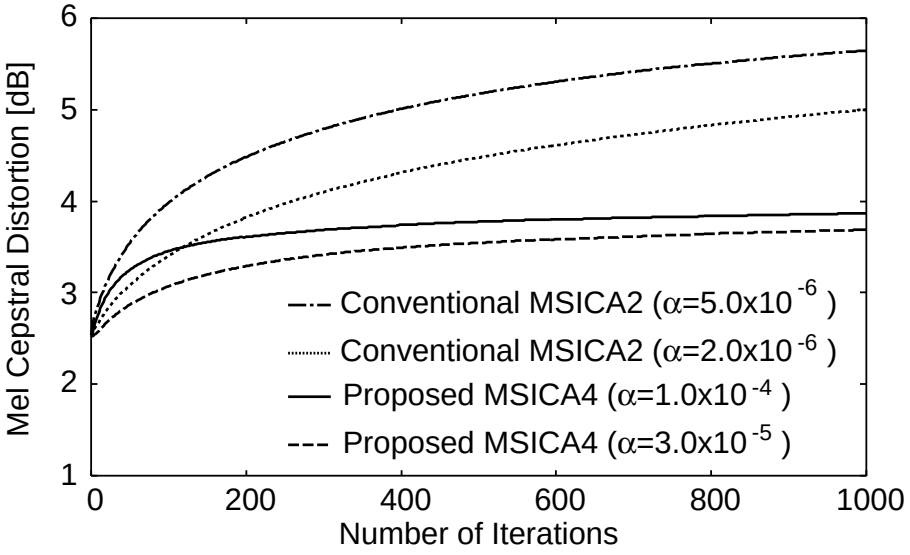


Figure 51. Comparison of the mel cepstral distortions between the observed signal with the single source component at the microphone and the output signals from (a) conventional MSICA3 or (b) proposed MSICA5.

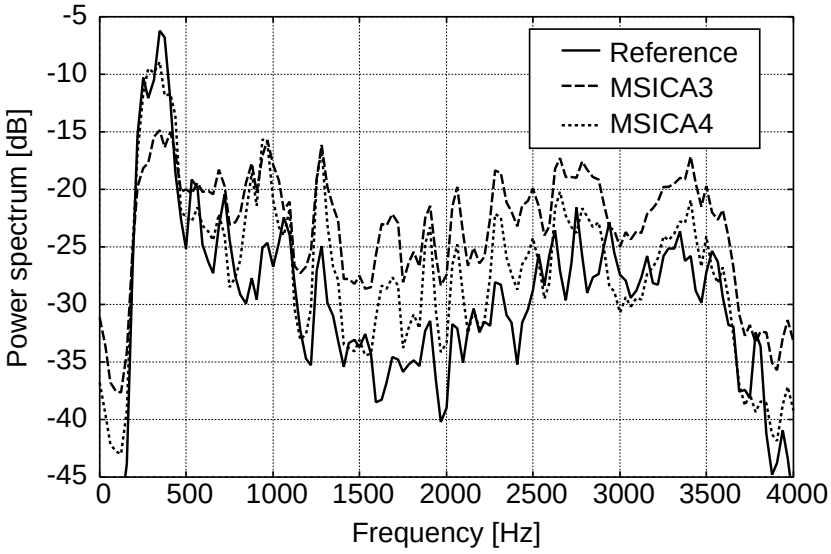


Figure 52. Comparison of the power spectra in conventional MSICA3: FDICA is followed by H-TDICA, conventional MSICA4: FDICA is followed by H-TDICA with spectral compensation, and reference signal ($s_1(t)$ component recorded at the microphone).

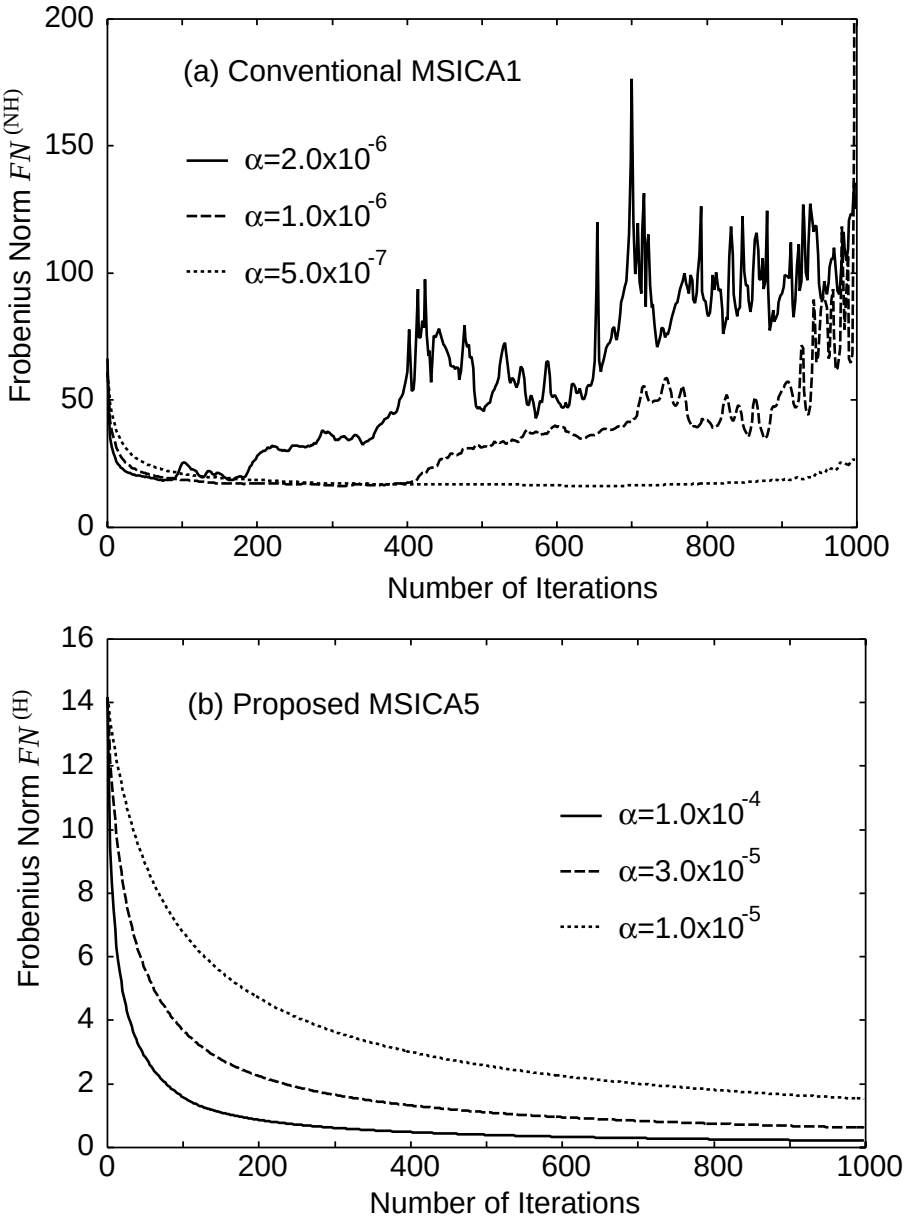


Figure 53. Comparison of frobenius norms Eqs. (97), (98) in (a) conventional MSICA1: FDICA is followed by NH-TDICA and (b) proposed MSICA5: FDICA is followed by the proposed method combining H-TDICA and linear prediction.

6.5 Proposed Stable and Low-Distortion Algorithm for Overdetermined BSS

6.5.1 Problems in Original Stable and Low-Distortion Algorithm

To solve the problems for stability and the distortion of conventional TDICA, we have already proposed the algorithm combining MSICA and linear prediction in Sect. 6.3. However this method does not apply to the model, where the number of microphones larger than that of the sources because the compensation to the sound qualities of separated signals become difficult. Thus, we should propose a new algorithm, in which the source-separation of temporally correlated signals such as speech is free of distortion caused by the decorrelation effect in the model where the number of microphones larger than that of the sources.

6.5.2 Proposed Stable and Low-Distortion MSICA for Case 2

This section describes a new algorithm with stable and low-distortion property based on MSICA using subarray processing. Figures 54 (a) and (b) show the procedure performed in the proposed stable and low-distortion MSICA using subarray processing. The proposed method consists of two steps, i.e., the iterative learning process of H-TDICA (see Fig. 54 (a)) and the compensation process (see Fig. 54 (b)). In the iterative learning of the proposed algorithm, we perform H-TDICA. However the separated signals are distorted by the decorrelation effect as described in Sect. 6.4.3. Therefore we estimate the the components which contribute to the distortion and we compensate the sound qualities of the separated signals.

First, we explicate the mechanism of H-TDICA algorithm. The iterative learning of H-TDICA Eq. (47) can be decomposed into the components which contribute to the separation and decorrelation given; these are given as

$$\begin{aligned}
\boldsymbol{w}_{i+1}^{(\mathrm{H})}(\tau) &= \boldsymbol{w}_{i}^{(\mathrm{H})}(\tau) + \beta \sum_{d=0}^{R-1} \Big\{ \boldsymbol{I}\delta(\tau - d) - \langle \boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big\} \boldsymbol{w}_{i}^{(\mathrm{H})}(d) \\
&= \boldsymbol{w}_{0}^{(\mathrm{H})}(\tau) + \sum_{j=0}^{i} \Big[\beta \sum_{d=0}^{R-1} \Big\{ \boldsymbol{I}\delta(\tau - d) - \langle \boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}} \rangle_t \Big\} \boldsymbol{w}_{j}^{(\mathrm{H})}(d) \Big] \\
&= \boldsymbol{w}_{0}^{(\mathrm{H})}(\tau)
\end{aligned}$$

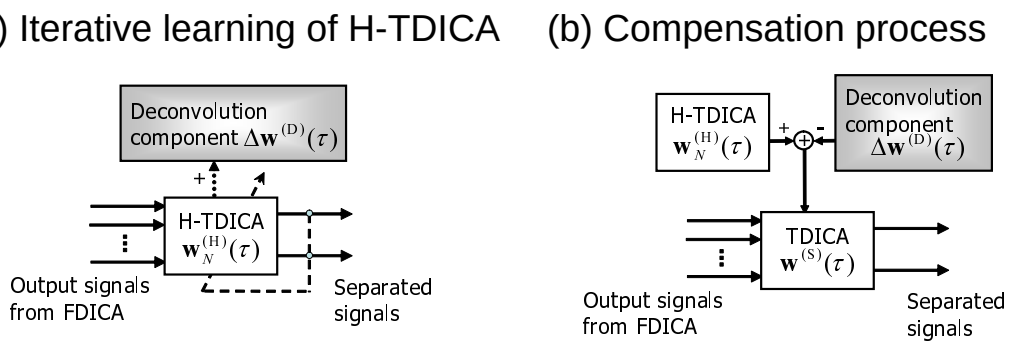


Figure 54. Procedure performed in the proposed stable and low-distortion MSICA using subarray processing. The proposed method consists of following two steps, i.e., (a) the iterative learning process of H-TDICA and (b) the compensation process.

$$
\begin{aligned}
&+\sum_{j=0}^{i}\Big[\beta\sum_{d=0}^{R-1}\Big\{\boldsymbol{I}\delta(\tau-d)-\mathrm{diag}(\langle\boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t)\Big\}\boldsymbol{w}_j^{(\mathrm{H})}(d)\\
&-\beta\sum_{d=0}^{R-1}\Big\{\text{off-diag}(\langle\boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t)\Big\}\boldsymbol{w}_j^{(\mathrm{H})}(d)\Big]\\
=\;&\boldsymbol{w}_0^{(\mathrm{H})}(\tau)+\sum_{j=0}^{i}\Big(\Delta\boldsymbol{w}_j^{(\mathrm{D})}(\tau)+\Delta\boldsymbol{w}_j^{(\mathrm{S})}(\tau)\Big), \qquad (99)
\end{aligned}
$$

$$
\Delta\boldsymbol{w}_j^{(\mathrm{D})}(\tau) = \beta\sum_{d=0}^{R-1}\Big\{\boldsymbol{I}\delta(\tau-d)-\mathrm{diag}(\langle\boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t)\Big\}\boldsymbol{w}_j^{(\mathrm{H})}(d), \qquad (100)
$$

$$
\Delta\boldsymbol{w}_j^{(\mathrm{S})}(\tau) = -\beta\sum_{d=0}^{R-1}\Big\{\text{off-diag}(\langle\boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t)\Big\}\boldsymbol{w}_j^{(\mathrm{H})}(d), \qquad (101)
$$

where $\boldsymbol{w}_0^{(\mathrm{H})}(\tau)$ is the initial filter for the iterative learning, $\boldsymbol{w}_j^{(\mathrm{D})}(\tau)$ is the component which contributes to the decorrelation, and $\boldsymbol{w}_j^{(\mathrm{S})}(\tau)$ is the component which contributes to the separation. The iterative learning of NH-TDICA is the algorithm in which $\boldsymbol{w}_j^{(\mathrm{D})}(\tau)=\boldsymbol{0}$. This modification yields the source separation without the decorrelation effect, but, the stability in the iterative learning degrades shown in Sect. 6. We note that the only source separation is performed by the nonholonomic constraint; indeed it is experimentally proved that $\boldsymbol{w}_j^{(\mathrm{S})}(\tau)$ is the component which contribute to the separation. Also, we can understand

that $\boldsymbol{w}_j^{(\mathrm{D})}(\tau)$ is the component which contribute to the decorrelation. Therefore we estimate the distortion components by using $\boldsymbol{w}_j^{(\mathrm{D})}(\tau)$ and we compensate the sound quality of the separated signal. The desired optimized separation filter $\boldsymbol{w}^{(\mathrm{S})}(\tau)$ is obtained by subtracting the component to the distortion from $\boldsymbol{w}_N^{(\mathrm{H})}(\tau)$:

$$\boldsymbol{w}^{(\mathrm{S})}(\tau) \;=\; \boldsymbol{w}_N^{(\mathrm{H})}(\tau) - \sum_{i=0}^{N-1} \Delta \boldsymbol{w}_i^{(\mathrm{D})}(\tau), \tag{102}$$

where N is the number of iterations for TDICA.

The compensation in the proposed method is achieved by recovering the higher-order autocorrelation of the separated signal to be that of the input signals for TDICA. If we apply the conventional simple TDICA, the higher-order autocorrelation of the separated signal approximates to that of the mixed signal. The distortion occurs because the autocorrelation of the mixed signal and that of the desired source signal at the microphone point are completely different. Therefore the application of this algorithm to the conventional simple TDICA is not effective. On the other hand, in the application of this algorithm to MSICA, the input signals for TDICA is the output signals from FDICA. The typical separation performance in FDICA using subarray processing is about 8 to 10 dB under the condition that the reverberation time is 300 ms as shown in Sect. 5.5.2 [35] and the MelCD between the observed signal with the single source component at the microphone and the output signals from FDICA using subarray processing is about 2 to 3 dB [66]. From this, we can infer that the autocorrelation of the output signal from FDICA which corresponds to the input signal for TDICA part in MSICA and that of the desired source signal at the microphone point are similar. Therefore it is possible that the autocorrelation of the separated signal approximates to that of the output signal from FDICA even in the case where the number of microphones to be larger than that of sources, and we can compensate the distortion effectively in the proposed method

6.5.3 Experiments and Results in Case 2

In this section, we compare the proposed MSICA shown in Sect. 6.5.2 with the conventional MSICAs Eqs. (86) and (87) (see Fig. 46) under the condition where the number of microphones K to be larger than that of sources L, i.e., $L > K$.

The experimental condition is the same as that given in Sect. 5.3.1. The analysis conditions of FDICA in these experiments are shown in Table 6. The filter length in FDICA is 1024 taps and the filter length in TDICA is 2048 taps.

In this study, we compare the following MSICAs:

Conventional MSICA1: FDICA is followed by NH-TDICA,

Conventional MSICA2: FDICA is followed by H-TDICA,

Proposed MSICA: FDICA is followed by the proposed TDICA algorithm,

Figures 55 (a) and (b) show the NRR results of the conventional MSICA1, MSICA2, and the proposed MSICA for different iteration points. Figures 56 (a) and (b) show the MelCD between the observed signal with the single source component at the microphone and the output signals from (a) the conventional MSICA1 and the proposed MSICA or (b) the conventional MSICA2.

These values were averages of all of the combinations with respect to speakers and source directions. From these results, the following are revealed. First, in the conventional MSICA1 in which the NH-TDICA is used, the behavior of the NRR is not monotonic (see Fig. 55 (a)) and there is remarkably consistent deterioration. The MelCD of the conventional MSICA1 in the initial step of the iterative learning is superior, but, the MelCD degrades as the number of iterations is increased (see Fig. 56 (a)). Secondly, regarding the separation performance of the conventional MSICA2 in which the H-TDICA is used, the separation performance of the conventional MSICA2 is obviously superior to that of the proposed MSICA (see Fig. 55 (b)). However, the MelCD degrades as the number of iterations is increased (see Fig. 56 (b)). We speculate that the *specious* performance in MSICA2 is due to the exceeding emphasis of high-frequency components by the whitening effect of H-TDICA. MSICA2 gains the improvement of the NRR only in the high-frequency region as shown in Sect. 6.4.3 , and consequently we can conclude that MSICA2 is ineffective in separating the speech signals from the practical viewpoint.

On the other hand, in the proposed MSICA, there is no deterioration of NRR (see Fig. 56 (a)). Therefore, the separation performances are almost completely retained during all of the iterations and the proposed MSICA is effective for

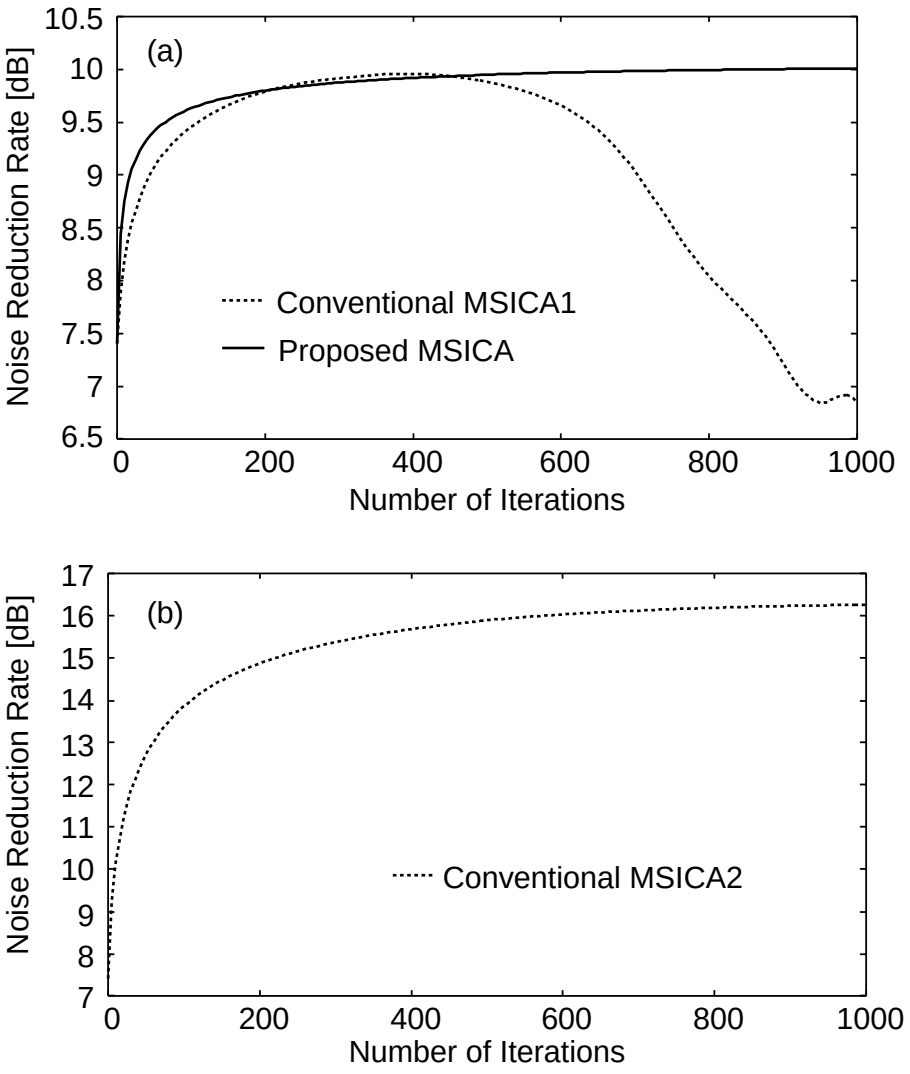


Figure 55. Comparison of the noise reduction rates in (a) conventional MSICA1 and proposed MSICA, and (b) conventional MSICA2.

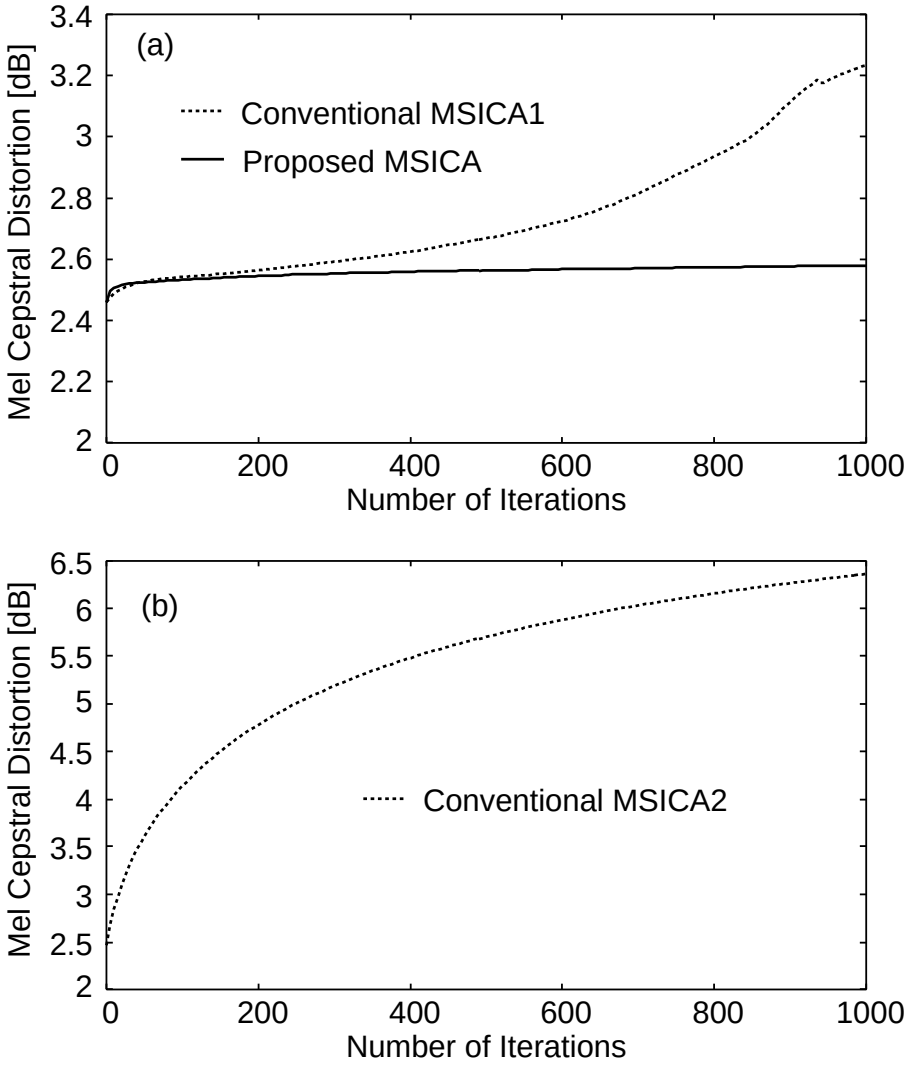


Figure 56. Comparison of the mel cepstral distortion in (a) conventional MSICA1 and proposed MSICA, and (b) conventional MSICA2.

the stability in the iterative learning. Also, there are almost no degradation of the MelCD in the proposed MSICA than that of the conventional MSICA2. From these results, we can conclude that the proposed algorithm is effective for improving the stability of the learning.

6.6 Conclusion

We newly proposed a stable and low-distortion algorithm combining MSICA and linear prediction for BSS in the case where the number of microphones is equal to that of sources. In the proposed algorithm, the linear predictors estimated from the roughly separated signals by FDICA are inserted before the holonomic TDICA as a prewhitening processing, and the dewhitening is performed after TDICA. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the pre/dewhitening processing prevents the decorrelation. Moreover, we proposed a new algorithm with a stability and low-distortion for overdetermined BSS based on MSICA using subarray processing in the case where the number of microphones to be larger than that of sources. In the proposed algorithm, to solve the problem of the stability, we perform TDICA with the holonomic constraint. Also, to avoid the distortions, we estimate the distortion components by TDICA with the holonomic constraint and we compensate the sound qualities by using the estimated components. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the proposed compensation method prevents the distortion. The experimental results under a reverberant condition revealed that the proposed algorithm provides the higher stability and the higher separation performance, compared with the conventional MSICA including H-TDICA or NH-TDICA.

7. Conclusion

7.1 Summary of the Thesis

We addressed to the blind source separation (BSS) to realize a high-quality hands-free speech recognition system and a hands-free telecommunication system. BSS is an approach for estimating original source signals only from the information of the mixed signals observed in each input channel. Many BSS methods based on independent component analysis (ICA) have been proposed for the acoustic signal separation. However, the performances of these methods degrade particularly seriously under extreme reverberant conditions.

In order to improve the separation performance, we proposed two novel BSS algorithms, i.e., (1) BSS based on multistage ICA (MSICA), in which frequency-domain ICA (FDICA) and time-domain ICA (TDICA) are cascaded and (2) Overdetermined BSS based on MSICA using subarray processing. Also, in order to achieve a stability in the iterative learning of ICA and the separated signal with low-distortion, we proposed two novel BSS algorithms, i.e., (1) BSS combining MSICA and linear prediction and (2) Overdetermined BSS based on MSICA using the new compensation method.

In Sect. 2, first, the sound mixing model of the microphone array is explained. Next, we introduced two types of ICA for the convolutive mixture, i.e., FDICA and TDICA. Moreover, the advantages and disadvantages in FDICA and TDICA was explained.

In Sect. 3, we investigated the disadvantages of FDICA and TDICA and we newly described the applicable limitations of both ICAs under the real acoustic conditions. First, the results of the signal separation experiment with FDICA revealed that the separation performance of FDICA obviously degrades when the number of subbands become too large, and was saturated before reaching a sufficient performance. We can conclude that this is because the independence assumption of the narrow-band signals collapses. Secondly, the results of the signal separation experiment with TDICA revealed that the separation performance of TDICA was not sufficient compared with FDICA. We can conclude that this is because the iterative learning rule become more complicated as the reverberation increases.

In Sect. 4, we proposed a new algorithm for BSS, in which FDICA and TDICA were combined to achieve a superior source-separation performance under reverberant conditions. Also, we provided a comparison results for the separation performance of FDICA, TDICA, and the proposed method under the same acoustic condition. The results of the signal separation experiment with the proposed method revealed that the separation performance and the speech recognition of the proposed algorithm were superior to that of conventional ICA-based BSS methods, and the combination of FDICA and TDICA was inherently effective for improving the separation performance.

In Sect. 5, we proposed a MSICA, by setting the number of microphones to be larger than that of sources to achieve an improved separation performance. In the FDICA part in the simple extension of MSICA, the use of additional microphones led to alternative problems: the solution was likely to be trapped within a trivial solution and the permutation problem in FDICA become very complicated. In order to solve these problems, we proposed a new extended MSICA using subarray processing, where the number of microphones and that of sources were set to be the same in every subarray. The experimental results obtained under real acoustic environmental conditions revealed that the separation performance of the proposed MSICA was improved as the number of microphones is increased.

In Sect. 6, we newly proposed a stable and low-distortion algorithm combining MSICA and linear prediction for BSS in the case where the number of microphones was equal to that of sources. In the proposed algorithm, the linear predictors estimated from the roughly separated signals by FDICA were inserted before the holonomic TDICA as a prewhitening processing, and the dewhitening was performed after TDICA. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the pre/dewhitening processing prevents the decorrelation. Moreover, we proposed a novel algorithm with a stability and low-distortion for overdetermined BSS based on MSICA using subarray processing in the case where the number of microphones to be larger than that of sources. In the proposed algorithm, to solve the problem of the stability, we performed TDICA with the holonomic constraint. Also, to avoid the distortions, we estimated the distortion components by TDICA with the holonomic constraint and we compensated the sound qualities by using the estimated

components. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the proposed compensation method prevents the distortion. The experimental results under a reverberant condition revealed that the proposed algorithm provides the higher stability and the higher separation performance, compared with the conventional MSICA including holonomic TDICA or nonholonomic TDICA.

In summary, we confirmed that the proposed MSICA-based BSS and overdetermined BSS based on MSICA using subarray processing are effective for improving the source-separation performance. In addition, the proposed BSS combining MSICA and linear prediction and overdetermined BSS based on MSICA using the novel compensation method are also effective for achieving a stable learning and low-distortion.

7.2 Future Work

Although we have improved the source-separation performance, the stability in the learning of ICA, and the distortion of the separated signal, a number of problems still remain to be solved.

Reduction of the degree of the calculation for ICA

In the application of FDICA, we can achieve a real-time processing except that about 3 s is required for the adaptation of the initial separation filter. MSICA could improve the source separation performance compared with the conventional ICAs. However, the degree of the calculation in the proposed method is more expensive compared with that of the conventional FDICA (about one minute). We have to reduce the degree of calculation by improving the convergence of MSICA and optimizing the combination of FDICA and TDICA. In recent work, Mukai et al. propose the BSS using ICA and residual crosstalk subtraction [69]. In the future work, the combination with another noise reduction technique and speech enhancement technique, e.g., spectral subtraction [67, 68], is required to assist ICA and reduce the degree of the calculation.

Dereverberation

For achieving a superior speech recognition performance, not only the separation of the user's speech but also the reduction of the reverberant components is an important task because the speech recognition performance is degraded as the

reverberation time is lengthened. Our research members have been proposing a new blind source separation and deconvolution of MIMO-FIR system with colored sound inputs using SIMO-model-based ICA [70]. However, in the case of mixtures with a long-tap FIR filter, the performance of the blind source separation and deconvolution obviously degrades. Also, various type of dereverberation methods and improvements of acoustic model in the speech recognition system have proposed [71, 72, 73]. A new approach combining ICA and this dereverberation method is desired to achieve superior speech recognition performance.

Discrimination of the user's speech

In the real applications of the hands-free speech recognition and a hands-free telecommunication system, we have to discriminate the user's speech from separated signals. However, the discrimination of the user using only speech information is very difficult. In the future, we will require a novel discrimination method using not only the speech information but also the image information captured by the image sensor is required.

Estimation of the number of source signals

In general, the number of sound sources is about 3 in typical environments that a hands-free system is applied. Therefore we only use the 3-elements array in the application of BSS. However the improvement of the separation performance is desired by combining with the estimation method of the number of sound sources [41, 42, 43] because of the effect for the convergence point mentioned in Sect.5.3.2.

Acknowledgements

I would like to express my deepest appreciation to Professor Kiyohiro Shikano of Nara Institute of Science and Technology, my thesis adviser, for his constant guidance and encouragement through my master's course and doctral course.

I would also like to express my gratitude to Professor Kenji Sugimoto of Nara Institute of Science and Technology for his invaluable comments to the thesis.

I would especially like to express my sincere gratefulness to Associate Professor Hiroshi Saruwatari of Nara Institute of Science and Technology, for his continuous support and valuable advice through the master's course and the doctral course. The core of this work originated with his pioneering ideas in blind source separation. This work could not have been accomplished without his direction. I could learn many lessons from his attitude toward study. I have always been happy to carry out research with him.

I would like to thank Assistant Professor Akinobu Lee and Assistant Professor Hiromichi Kawanami of Nara Institute of Science and Technology, 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 thank Dr. Takanobu Nishiura, who is currently Associate Professor at Ritsumeikan University, Dr. Yosuke Tatekura, who is currently Assistant Professor at Shizuoka University, Dr. Tomoki Toda, who is currently a Research Fellow of the Japan Society for the Promotion of Schience in Graduate School of Engineering, Nagoya Institute of Technology, Dr. Ryuichi Nisimura, who is currently Assistant Professor at Wakayama University, for providing thoughtful advice and discussions on my research. I owe a great deal to Mrs. Noriko Abe and Mrs. Kyoko Yoshida, secretary of Speech and Acoustics Laboratory, for their support in the laboratory.

I would sincerely like to thank Dr. Shoji Makino, Executive Manager at Media Information Laboratory of NTT Communication and Science Laboratories, Miss Shoko Araki, Mr. Ryo Mukai, Dr. Hiroshi Sawada, of NTT Communication and Science Laboratories, for providing livily and fruitful discussions about blind source separation and microphone array processing. I would also like Dr. Shigeru Katagiri, Deputy Directer of NTT Communication and Science Laboratories, Dr.

Shoji Makino, and Miss Shoko Araki, for their support when I was a student intern at NTT Communication and Science Laboratories. I would sincerely like to thank Dr. Atsunobu Kaminuma, NISSAN MOTOR CO., LTD., for fruitful discussions about speech enhancement and blind source separation in the car environments.

I would like to express my gratitude to Professor Noboru Nakasako and Professor Hisanao Ogura of Kinki University, for their support, guidance, and having recommended that I enter Nara Institute of Science and Technollogy.

I would like to thank Mr. Toshiya Kawamura (who is currently a Researcher at DENSO Corporation), Mr. Tomoya Takatani, doctoral candidate of Nara Institute of Science and Technology, Mr. Yoichi Hinamoto, doctoral candidate of Kyoto University, Mr. Hiroaki Yamajo and Mr. Hiroshi Abe, former master's cource of Nara Institute of Science and Technology, Mr. Satoshi Ukai, Mr. Yasuaki Ohashi, Miss Sachiko Obara, and Mr. Masayuki Shimada, master's cource of Nara Institute of Science and Technology, and Mr. Robert Aichner, doctoral candidate of University of Erlangen-Nuremberg, for their fruitful discussions about blind source separation and speech enhancement. I would also like to thank Mr. Randy Gomez, doctoral candidate of Nara Institute of Science and Technology, for his English support.

I am indebted to many Researchers and Professors. I would especially like to express my gratitude to Dr. Futoshi Asano, Group Reader at Media Interaction Group of National Institute of Advanced Industrial Science and Technology, and Dr. Mitsuru Kawamoto, Associate Professor at Shimane University, for their valuable advice and discussions. I would like to thank Mr. Akira Baba, a Researcher at Matsushita Electric Works, Ltd., Dr. Manabu Otsuka, a Researcher at DENSO Corporation, and Mr. Daisuke Saitoh, a Researcher at NISSAN MOTOR CO., LTD., for providing fruitful discussions.

I would like to acknowledge my friends for their support. I would especially like to thank Mr. Miichi Yamada, Mr. Takashi Uchida, Mr. Hidekazu Kamiyanagida, Mr. Yuu Nagata, Mr. Goshu Nagino, Mr. Keisuke Noma, Mr. Tsuyoshi Masuda, Miss Kanako Matsunami, Mr. Koichi Mino, Mr. Yuichiro Mera, Mr. Katsuyuki Sawai, and Sahoko Hata.

Finally, I would like to acknowledge my family for their support.

References

[1] B. H. Juang and F. K. Soong, "Hands-free telecommunications," *Proc. International Conference on Hands-Free Speech Communication*, pp.5–10, April 2001.

[2] J. S. Lim, *Speech enhancement*, Prentice-Hall, Inc., New Jersey, 1994.

[3] T. W. Parsons, "Separation of speech from interfering speech by means of harmonic selection," *J. Acoust. Soc. Am.*, Vol.60, pp.911–918, 1976.

[4] K. Kashino, K. Nakadai, T. Kinoshita, and H. Tanaka, "Organization of hierarchical perceptual sounds," *Proc. 14 th Int. Conf. Artificial Intelligence*, Vol.1, pp.158–164, 1995.

[5] M. Unoki and M. Akagi, "A method of signal extraction from noisy signal based on auditory scene analysis," *Speech Communication*, Vol.27, pp.261–279, 1999.

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

[7] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *Acoustical Society of America*, Vol.78, pp.1508–1518, Nov. 1985.

[8] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," Proceedings of IEEE , Vol. 60, No. 8, pp.926–935, 1972.

[9] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Transactions on Audio Processing, Vol. 30, No. 1, pp.27–34, 1982.

[10] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," IEEE Transactions on Speech and Audio Processing, Vol. 34, No. 6, pp.1391–1400, 1986.

[11] T. W. Lee, *Independent component analysis*, Kluwer academic publishers, 1998.

[12] S. Haykin, *Unsupervised adaptive filtering*, John Wiley & Sons, Inc., 2000.

[13] 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.

[14] 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.

[15] P. Common, "Independent component analysis, a new concept?," *Signal Processing*, Vol.36, pp.287–314, 1994.

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

[17] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. of 1998 International Symposium on Nonlinear Theory and Its Application (NOLTA98)*, pp.923–926, Sep. 1998.

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

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

[20] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," *IEICE Trans. Fundamentals*, Vol.E86-A, No.3, pp.590–595, March 2003.

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

[22] 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, pp.846–858, April, 2003.

[23] M. Kawamoto and Y. Inouye, "Generalized deflation algorithms for the blind source-factor separation of MIMO-FIR channels," *Proc. International Symposium on ICA and BSS*, pp.561–566, April 2003.

[24] S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Subband-based blind source separation for convolutive mixtures of speech," IEEE Transactions on Speech and Audio Processing, (accepted).

[25] S. Araki, S. Makino, R. Mukai, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," IEEE Transactions on Speech and Audio Processing, Vol. 11, No. 2, pp.109–116, March 2003.

[26] H. Kuttruff, *Room acoustics (Fourth Ed.)*, Spon Press, London, 2000.

[27] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," *Proc. International Conference on Language Resources and Evaluation*, pp.965–968, June 2000.
http://tosa.mri.co.jp/sounddb/indexe.htm

[28] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. ICASSP2000*, Vol.5, pp.3140–3143, June 2000.

[29] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," *Proc. SPAWC97*, pp.101–104, April 1997.

[30] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, T. Nishikawa, and K. Shikano, "Blind source separation combining ICA and beamforming," EURASIP Journal on Applied Signal Processing, Vol.2003, No.11, pp.1135-1146, 2003.

[31] H. Saruwatari, T. Kawamura, and K. Shikano, H. Saruwatari, T. Kawamura, T. Nishikawa, and K. Shikano, "Fast-convergence algorithm for blind source separation based on array signal processing," *IEICE Trans. Fundamentals*, Vol.E86-A, No.3, pp.286–291, March 2003.

[32] M. Kawamoto and Y. Inoue, "Blind deconvolution of MIMO-FIR systems with colored inputs using," *IEICE Trans. Fundamentals*, Vol.E86-A, No.3, pp.597–604, March 2003.

[33] L. Parra and C. V. Alvino, "Geometric source separation: merging convolutive source separation with geometric beamforming," *IEEE Trans. Speech & Audio Process.*, Vol.10, No.6, pp.352–362, 2002.

[34] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures," EURASIP Journal on Applied Signal Processing, Vol.2003, No.11, pp.1157–1166, 2003.

[35] T. Nishikawa, H. Abe, H. Saruwatari, K. Shikano, and A. Kaminuma, "Overdetermined blind separation for real convolutive mixtures of speech based on multistage ICA using subarray processing," IEICE Trans. Fundamentals, Vol.E87-A, No.8, pp.1924–1932, Aug. 2004.

[36] 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.

[37] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, N. Kitawaki, "A combined approach of array processing and independent component analysis for blind separation of acoustic signals," *Proc. ICASSP2001*, pp.2729–2732, May 2001.

[38] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," IEEE Transactions on Speech and Audio Processing, Vol. 12, No. 5, pp.530–538, Sep. 2004.

[39] M. Kawamoto, A. K. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Blind signal separation for convolved nonstationary signals," *Electronics and Communications in Japan*, Part 3, Vol. 84, No.2, 2001.

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

[41] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol.33, No.2, pp.387–392, April 1985.

[42] K. Yamamoto, F. Asano, W. F. G. Rooijen, E. Y. L. Ling, T. Yamada, and N. Kitawaki, "Estimation of the number of sound sources using support vector machines and its application to sound source separation," *Proc. ICASSP2003*, pp.485–488, April 2003.

[43] H. Sawada, S. Winter, R. Mukai, S. Araki, and S. Makino, "Estimating the number of sources for frequency-domain blind source separation," *Proc. International Symposium on ICA and BSS*, pp.610–617, Sep. 2004.

[44] 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).

[45] T. Nishikawa, T. Takatani, H. Saruwatari, K. Shikano, S. Araki, and S. Makino, "Comparison of time-domain ICA methods based on minimization of KL divergence and simultaneous decorrelation of nonstationary signal," *The 2002 Autumn Meeting of the ASJ*, pp.545–546, Sep. 2002 (in Japanese).

[46] T. Nishikawa, H. Saruwatari, K. Shikano, S. Araki, and S. Makino, "Multistage ICA for blind source separation of real acoustic convolutive mixture," *Proc. International Symposium on ICA and BSS*, pp.523–528, April 2003.

[47] T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation using frequency-domain ICA and time-domain ICA in real acoustic environment," *The 2001 Autumn Meeting of the ASJ*, pp.625–626, Oct. 2001 (in Japanese).

[48] K. Itou, M. Yamamoto, K. Takeda, T .Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, "JNAS : Japanese speech corpus for large vocabulary continuous speech recognition research," *J. Acoust. Soc. Jpn. (E)*, Vol.20, No.3, pp.199–206, 1999.

[49] A. Lee, T. Kawahara, K. Takeda, K. Shikano, "A new phonetic tied-mixture model for efficient decoding," *Proc. ICASSP2000*, Vol.III, pp.1269–1272, 2000.

[50] A. Lee, T. Kawahara, K. Shikano, "Julius – An open source real-time large vocabulary recognition engine," *Proc. EUROSPEECH2001*, pp.1691–1694, 2001.

[51] M. Shozakai, "Development of automatic speech recognition middleware VORERO for embedded appliances," *The 2004 Spring Meeting of the ASJ*, pp.31–32, March 2004 (in Japanese).
http://www.vorero.com

[52] M. Shozakai, S. Nakamura, and K. Shikano, "An evaluation of speech enhancement approach E-CMN/CSS for speech recognition in car environments," IEICE Trans., Vol.J81-DII, No.1, pp.1–9, Jan. 1998 (in Japanese).

[53] M. Shozakai, S. Nakamura, and K. Shikano, "A robust speech recognition using adaptive filter based on frame-wise voise activity detection in car environments," IEICE Trans., Vol.J81-DII, No.6, pp.1074–1083, June 1998 (in Japanese).

[54] H. Sawada, R. Mukai, S. Ryhove, S. Araki, and S. Makino, "Spectral smoothing for frequency-domain blind source separation," *Proc. International Workshop on Acoustic Echo and Noise Control*, pp.311–314, Sep. 2003.

[55] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. Int. Symp. on ICA and BSS (ICA2001)*, pp.722–727, Dec. 2001.

[56] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fidelity blind separation of acoustic signals using SIMO-model-based indepen-

dent component analysis," IEICE Trans. Fundamentals, Vol.E87-A, No.8, pp.2063–2072, Aug. 2004.

[57] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fildeity blind separation of acoustic signals using SIMO-model-based ICA with information-geometric learning," *Proc. International Workshop on Acoustic Echo and Noise Control*, pp.251–254, Sep. 2003.

[58] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison between SIMO-ICA with least squares criterion and SIMO-ICA with information-geometric learning," *Proc. International Congress on Acoustics*, Vol.I, pp.329–332, April 2004.

[59] T. Nishikawa, H. Abe, H. Saruwatari, and K. Shikano, "Overdetermined blind separation of acoustic signals based on MISO-constrained frequency-domain ICA," *Proc. International Congress on Acoustics*, Vol.IV, pp.3143–3146, April 2004.

[60] T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable learning algorithm for low-distortion blind separation of real speech mixture combining multistage ICA and linear prediction," ISCA tutorial and research workshop on non-linear speech processing, pp.31–34, May 2003.

[61] T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable learning algorithm for blind separation of temporally correlated acoustic signals combining multi-stage ICA and Linear Prediction," *IEICE Trans. Fundamentals*, Vol.E86-A, No.8, pp.2028–2036, Aug. 2003.

[62] A. Papoulis and S. U. Pillai, *Probability, random variables and stochastic processes (Fourth Ed.)*, McGraw-Hill Series in Electrical and Computer Engineering, New York, 2002.

[63] S. Furui, *Digital speech processing, synthesis, and recognition (Second Ed.)*, Signal Processing and Communications Series in Marcel Dekker, Inc., New York, 2000.

[64] T. Nishikawa, T. Kawamura, H. Saruwatari, and K. Shikano, "Overdetermined source separation with blind beamformer," *The 2000 Autumn Meeting of the ASJ*, pp.447–448, Sep. 2000 (in Japanese).

[65] R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time domain ICA blind source separation of non-stationary convolved signals by utilizing geometric beamforming," *Proc. IEEE International Workshop on Neural Networks for Signal Processing*, pp.445–454, Sep. 2002.

[66] T. Nishikawa, H. Saruwatari, K. Shikano, and A. Kaminuma, "Stable and low-distortion algorithm based on overdetermined blind separation for convolutive mixtures of speech," *Proc. International Symposium on ICA and BSS*, pp.881–888, Sep. 2004.

[67] S. F. Boll, "Supression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Speech and Audio Processing, Vol. 27, No. 2, pp.113–120, 1979.

[68] M. Mizumachi and M. Akagi, "Noise reduction by paired microphones using spectral subtraction," *Proc. ICASSP '98*, pp.1001–1004, 1998.

[69] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Blind source separation for moving speech signals using blockwise ICA and residual crosstalk subtraction," IEICE Trans. Fundamentals, Vol. E87-A, No. 8, pp. 1941–1948, Aug. 2004.

[70] H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa, and K. Shikano, "Blind separation and deconvolution of MIMO-FIR system with colored sound inputs using SIMO-model-based ICA," 2003 IEEE Workshop on Statistical Signal Processing (SSP2003), pp.421–424, Sep. 2003.

[71] A. Baba, A. Lee, H. Saruwatari, and K. Shikano, "Speech recognition by reverberation adapted acoustic models," *The 2002 Autumn Meeting of the ASJ*, pp.27–28, Sep. 2002 (in Japanese).

[72] T. Takiguchi and M. Nishimura, "Acoustic model adaptation using first order prediction for revereberant speech ," *Proc. ICASSP2004*, pp.869–872, May 2004.

[73] A. Baba, D. Matsumoto, A. Lee, H. Saruwatari, and K. Shikano, "Recognition of speech with dereverberation by spectrum subtraction in home environment," *The 2004 Autumn Meeting of the ASJ*, pp.9–10, Sep. 2004 (in Japanese).

[74] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Norwell, MA: Kluwer Academic Publishers, 1998.

[75] D. H. Johnson and D. E. Dudgeon, *Array signal processing: concepts and techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[76] 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.

[77] M. Athans, "The matrix minimum principle," Information Control, Vol. 11, pp.592–606, 1967.

[78] S. Kodama, *Matrix theory for systems and control*, Corona publishing CO., LTD., p.30, p.46, 1978 (in Japanese).

Appendix

A. Fast-Convergence FDICA for More Than Two Sources Combining ICA and Beamforming

A.1 Introduction

In this section, we address to the complex-valued ICA i.e., FDICA [17, 18, 19, 20, 28, 30, 31]. The FDICA-based BSS approach seems to be a very flexible and effective technique for the source separation, but it has an inherent disadvantage in that there is difficulty with the slow convergence of the optimization in ICA [30].

Saruwatari et al. have solved this problem in a specific case of two sources and two microphones by introducing the fast-convergence algorithm combining ICA and beamforming [19]. However, this algorithm cannot be extended to the source-separation problem of multiple sources and multiple microphones (more than 2 sources with more than 2 microphones). To resolve this problem, in this section, we describe a new extended algorithm [19] in which ICA and beamforming are combined for the blind separation of multiple sources. The proposed method consists of the following three parts: (a) frequency-domain ICA with estimation of the DOA of the sound source using a Lloyd clustering algorithm [74], (b) null beamforming based on the estimated DOA, and (c) integration of (a) and (b) based on the algorithm diversity in both iteration and frequency domain. The temporal utilization of null beamforming through ICA iterations can realize fast- and high-convergence optimization. The results of the signal separation experiments reveal that the signal separation performance of the proposed algorithm is superior to that of the conventional ICA-based BSS method, and the utilization of null beamforming in ICA is effective for improving the separation performance and convergence, even under reverberant conditions.

A.2 Proposed Algorithm

A.2.1 Motivation and Strategy

The conventional FDICA method inherently has a significant disadvantage which is due to slow and poor convergence through nonlinear optimization in ICA, particularly when introducing a poor initial setting of the separation matrix.

Meanwhile, we has recently provided an insight into the close relationship between ICA and the fixed null beamformer [34]. It is reported that, after the filter update has been completed, ICA with the small number of sensors (e.g., $K = 2$) often provides directional nulls against the undesired source signals, unlike the traditional DS array which enhances the target signal via the directional lobe. Indeed, the null beamforming is approximately optimal for the signal separation when the effect of the room reverberation is negligible, but this optimality cannot hold under reverberant conditions because the exact signal reduction cannot be achieved by using only the directional nulls. The null-beamforming approach, however, still has the advantage that there is no difficulty with respect to the slow convergence of optimization because the null beamformer is determined by using only DOA information without independence between sound sources.

The above-mentioned findings motivate us to combine FDICA and null beamforming. That is, a specific separation matrix which is designed on the basis of null beamforming can assist FDICA in the convergence and yield a good initial value of $\boldsymbol{W}(f)$ with regard to an advance removal of the direct sound of the interference. In this section, we propose an algorithm based on the temporal alternation of learning between FDICA and null beamforming; the separation matrix $\boldsymbol{W}(f)$ obtained through FDICA is temporally substituted by the matrix based on null beamforming for a temporal initialization or acceleration of the iterative optimization.

It is worth noting that even in the proposed algorithm, DOA information for each source is needed before the construction of the null beamformer, similarly to other beamforming techniques. However, this DOA estimation was considered as a tough problem under common BSS tasks where the number of sources, L, equals that of sensors, K. For instance, the traditional high-resolution DOA estimator, e.g., MUSIC and minimum variance methods [75] cannot be applied because these

methods require the condition that $K > L$. To achieve the DOA estimation blindly in the case of K=L, we introduce a new combination in which the DOA estimation follows one-time FDICA iteration and can be performed by using the separation matrix obtained from FDICA. This DOA estimation method is mainly based on our earlier finding that the directional null is steered to the DOA of the suppressed source in FDICA. Consequently, we can approximately estimate the DOAs only to find the null directions in the directivity patterns obtained from FDICA. Although the proposed combination approach partly includes heuristics with no guarantee of mathematically exact convergence, the effectiveness will be experimentally discussed in Sect. A.3. The proposed algorithm is conducted by the following steps with respect to all frequency bins in parallel (see Fig. 57).

A.2.2 Procedure of Proposed Algorithm

[Step 1: Initialization] Set the initial $\boldsymbol{W}_i(f)$, i.e., $\boldsymbol{W}_0(f)$, to a conventional DS array, where the subscript i is set to be 0.
[Step 2: 1-time ICA iteration] Optimize $\boldsymbol{W}_i(f)$ using the one-time FDICA iteration (see Eq. (37)), where the superscript "(ICA)" is used to express the fact that the separation matrix is obtained by ICA, whereas $\boldsymbol{W}_i(f)$ originated from either ICA or null beamforming, as described in **step 5**.
[Step 3: DOA estimation] Estimate DOAs $\boldsymbol{\theta} = \{\theta_1, \cdots \theta_L\}$ of the sound sources by utilizing the directivity pattern of the array system. The directivity pattern for the l th output is designated by $F_l(f, \theta)$, which is generally obtained by the multiplication of array weights and a *steering vector* as [75]

$$[F_1(f,\theta), \cdots, F_L(f,\theta)]^{\mathrm{T}} = \boldsymbol{W}^{(\mathrm{ICA})}(f)\boldsymbol{e}(f,\theta), \tag{103}$$

where $\boldsymbol{e}(f, \theta)$ is the steering vector which is defined by

$$\begin{aligned} \boldsymbol{e}(f,\theta) \;=\; & [\exp[j2\pi(ff_s/N)d_1 \sin\theta/c], \\ & \cdots, \exp[j2\pi(ff_s/N)d_K \sin\theta/c]\,]^{\mathrm{T}}, \end{aligned} \tag{104}$$

where c is velocity of sound, f_s is sampling frequency and N is a DFT size.

In the case of $K = L = 2$, directional nulls in the directivity patterns exist in only two particular directions, and thus we can heuristically drop the directions θ_l into two categories "large (max)" or "small (min)" [19]. This procedure is very

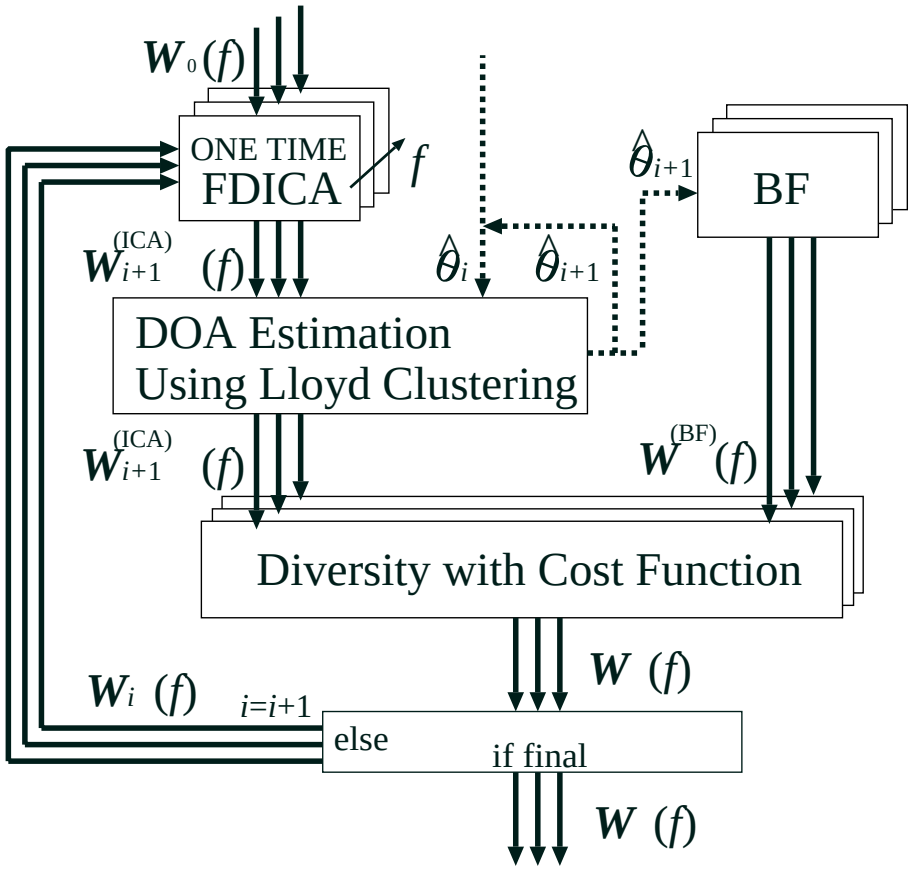


Figure 57. Proposed algorithm combining FDICA and beamforming.

simple and has the benefit of the low computational cost, but obviously the rule cannot be available in $K = L > 2$. To overcome the problem, we newly introduce an extended DOA estimation algorithm based on a directional null clustering technique, which can work even in the general case of $K = L > 2$.

In the l th directivity pattern $F_l(f, \theta)$ at the f th frequency bin, at most $L-1$ directional nulls can be found. We define the set of DOAs corresponding to the directional nulls as $\Theta^{(l)}(f)$; which is given by

$$\begin{aligned} \Theta^{(l)}(f) \;=\; \Big\{\theta|\; [F_l(f,\theta) - F_l(f,\theta-\Delta\theta)] \le 0; \\ [F_l(f,\theta+\Delta\theta) - F_l(f,\theta)] > 0\Big\}, \end{aligned} \tag{105}$$

where $\Delta\theta$ is a positive small value, and $\{\theta|\; A;\; B\}$ represents a set of θ which satisfies the conditions A and B simultaneously.

$\Theta^{(l)}(f)$ is evidently a good candidate of source directions. To estimate the DOAs of sources, we classify $\Theta^{(l)}(f)$ with all f and l into L categories, and then regard the centroids as the estimated DOAs. This classification can be carried out by using a Lloyd clustering algorithm [74] as follows.

(a) Make the whole set of $\Theta^{(l)}(f)$ to be classified, as

$$\Theta \;=\; \Big\{\theta_1, \theta_2, \cdots, \theta_Q\Big\} = \sum_{l=1}^{L}\sum_{f=1}^{N/2} \Theta^{(l)}(f), \tag{106}$$

where Q is the total number of detected directional nulls and at most $Q = (L-1)\cdot L \cdot N/2$.

(b) Set initial L centroids $\boldsymbol{\theta}^{(\mathrm{C})} = \{\theta_1^{(\mathrm{C})}, \cdots \theta_L^{(\mathrm{C})}\}$ to the DOAs estimated in the previous ICA iteration, i.e., $\boldsymbol{\theta}^{(\mathrm{C})} = \hat{\boldsymbol{\theta}}_i$. In the first iteration ($i = 0$), we set $\boldsymbol{\theta}^{(\mathrm{C})}$ to an arbitrary value.

(c) Set $L-1$ partitions $\theta_p^{(\mathrm{P})}$ as $\theta_p^{(\mathrm{P})} = (\theta_p^{(\mathrm{C})} + \theta_{p+1}^{(\mathrm{C})})/2$ where $p = 1, \cdots, L-1$. Also, the terminal partitions $\theta_0^{(\mathrm{P})}$ and $\theta_L^{(\mathrm{P})}$ are fixed at -90 and 90, respectively, throughout the algorithm.

(d) Given the partitions, calculate the L centroids $\theta_l^{(\mathrm{C})}$ ($l = 1, \cdots, L$) as

$$\theta_l^{(\mathrm{C})} \;=\; \frac{1}{Q_l}\Big\{\sum_{\theta_{l-1}^{(\mathrm{P})} \le \theta_q < \theta_l^{(\mathrm{P})}} \theta_q\Big\}, \tag{107}$$

where Q_l denotes the number of θ_q under $\theta_{l-1}^{(\mathrm{P})} \le \theta_q < \theta_l^{(\mathrm{P})}$.

(e) Go back to (c) for updating the new partitions by using the new centroids, and repeat the loop in (c)∼(e) with an appropriate number of iterations. The final centroids $\boldsymbol{\theta}^{(\mathrm{C})}$ are regarded as the resultant estimated DOAs $\hat{\boldsymbol{\theta}}_{i+1}$ in the $(i+1)$ th iteration, i.e., $\hat{\boldsymbol{\theta}}_{i+1} = \boldsymbol{\theta}^{(\mathrm{C})}$.

[Step 4: Beamforming] Construct an alternative matrix for signal separation, $\boldsymbol{W}^{(\mathrm{BF})}(f)$, based on the null-beamforming technique where the DOA information obtained in the ICA section is used. The separation matrix $\boldsymbol{W}^{(\mathrm{BF})}(f)$ can be obtained as

$$\boldsymbol{W}^{(\mathrm{BF})}(f) = \left[\boldsymbol{e}(f,\hat{\theta}_1), \boldsymbol{e}(f,\hat{\theta}_2), \cdots, \boldsymbol{e}(f,\hat{\theta}_L)\right]^{-1}. \tag{108}$$

[Step 5: Diversity using cost function] In order to integrate the FDICA with null beamforming, we introduce the following strategy for selecting the most suitable separation matrix in each frequency bin and at each iteration point, i.e., algorithm diversity in both iteration and frequency domain. As a cost function for achieving the diversity, we introduce a *coherence function* among L separated signals:

$$\begin{aligned} & C(\boldsymbol{W}(f)) \\ & \quad = \frac{1}{{}_L\mathrm{C}_2} \sum_{l=2}^{L} \sum_{l'<l} \frac{|\langle Y_{l'}(f,t) Y_l(f,t)^* \rangle_t|}{\sqrt{\langle\, |Y_{l'}(f,t)|^2 \,\rangle_t \langle\, |Y_l(f,t)|^2 \,\rangle_t}}, \end{aligned} \tag{109}$$

where $Y_l(f,t)$ and $Y_{l'}(f,t)$ are the separated signals. We calculate the estimated coherence function once for $\boldsymbol{W}(f) = \boldsymbol{W}^{(\mathrm{ICA})}(f)$ and once for $\boldsymbol{W}(f) = \boldsymbol{W}^{(\mathrm{BF})}(f)$; these are written as $C(\boldsymbol{W}^{(\mathrm{ICA})}(f))$ and $C(\boldsymbol{W}^{(\mathrm{BF})}(f))$, respectively. In fact, the coherence function cannot indicate the exact independence between sources, unlike ICA. However, we use this function to assess the source independence approximately because of the feasible advantage that the coherence function does not include any nonlinear calculations which often entail large computational complexity.

If the expected separation performance of beamforming is superior to that of ICA, the following condition holds, $C(\boldsymbol{W}^{(\mathrm{ICA})}(f)) > C(\boldsymbol{W}^{(\mathrm{BF})}(f))$; otherwise, $C(\boldsymbol{W}^{(\mathrm{ICA})}(f)) \leq C(\boldsymbol{W}^{(\mathrm{BF})}(f))$. Thus, an observation of the conditions yields the following algorithm:

$$\boldsymbol{W}_{i+1}(f)$$

$$= \begin{cases} \boldsymbol{W}^{(\mathrm{ICA})}(f), & (C(\boldsymbol{W}^{(\mathrm{ICA})}(f)) \leq C(\boldsymbol{W}^{(\mathrm{BF})}(f))) \\ \boldsymbol{W}^{(\mathrm{BF})}(f), & (C(\boldsymbol{W}^{(\mathrm{ICA})}(f)) > C(\boldsymbol{W}^{(\mathrm{BF})}(f))). \end{cases} \tag{110}$$

If the $(i+1)$ th iteration is the final iteration, go to **step 6**; otherwise, go back to **step 2** and repeat the ICA iteration, inserting $\boldsymbol{W}_{i+1}(f)$ as given by (110) into $\boldsymbol{W}_i(f)$ in (37) with an increment of i.

[Step 6: Ordering and scaling] Using the DOA information obtained in **step 3**, we can detect and correct the source permutation and the gain inconsistency [28]. From the directivity patterns in all frequency bins, we approximate the degree of the noise reduction by the differences between the gain at the direction of the target and those of the interferences. By comparing the degree of the estimated noise reduction, we can resolve the permutation problem. The gain inconsistency problem is resolved by normalizing the directivity patterns according to the gain in each source direction after the classification.

A.3 Experiments and Results

A.3.1 Experimental Setup

The source separation experiment with $K = L = 3$ is conducted. In this thesis, we assume that the number of sound sources is known in advance. Regarding the estimation of the number of sound sources, many methods are available, e.g., [41, 42, 43]. In order to generate the room impulse responses, we use the image method [76] assuming the artificial room as shown in Fig. 58, where the reverberation time is set to be 300 ms. A three-element array with interelement spacing of 4 cm is used. Three sound sources are placed at three directions, $-60°$, $0°$, and $70°$ to sound the speech signals. Two sentences spoken by two male and two female speakers are used as the original speech samples and the sampling frequency is 8 kHz. Using these sentences, we obtain 12 combinations with respect to speakers and source directions. We use the DS-array-based initial value $\boldsymbol{W}_0(f)$ which steers the look directions to $-90°$, $20°$, and $90°$. The frame length is 128 ms and the frame shift is 2 ms. The step-size parameter α is 2.0×10^{-6}. In order to evaluate the performance, we used the NRR as described in Sect. 3.2.2.

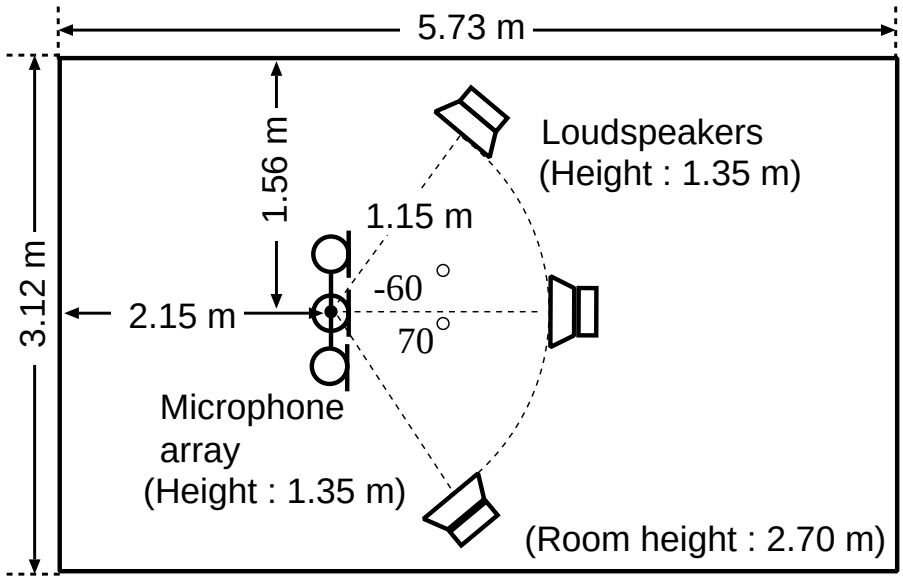


Figure 58. Layout of reverberant room used in image method.

A.3.2 DOA Estimation Results

Figure 59 shows the DOA estimation results (averaged DOAs for 12 combinations) for different number of loops in the Lloyd clustering algorithm (see Sect. A.2, Step 3 (e)). We compared four patterns, in which the number of Lloyd loops is 1, 3, 5, and 10 times. These results reveal that the performances of the DOA estimation using 5 and 10 Lloyd loops are equivalent. From these results, the Lloyd clustering converges at 5 times. In the next source-separation experiment, we set the number of loops for the Lloyd clustering algorithm to be 5 times.

A.3.3 Source-Separation Result

Figure 60 shows the NRR results (averaged NRR for 12 combinations) of the proposed method and the conventional BSS. This figure contains the following three curves.

Proposed Method : Our proposed BSS method described in Section A.2.

Conventional ICA : The conventional FDICA-based BSS method described

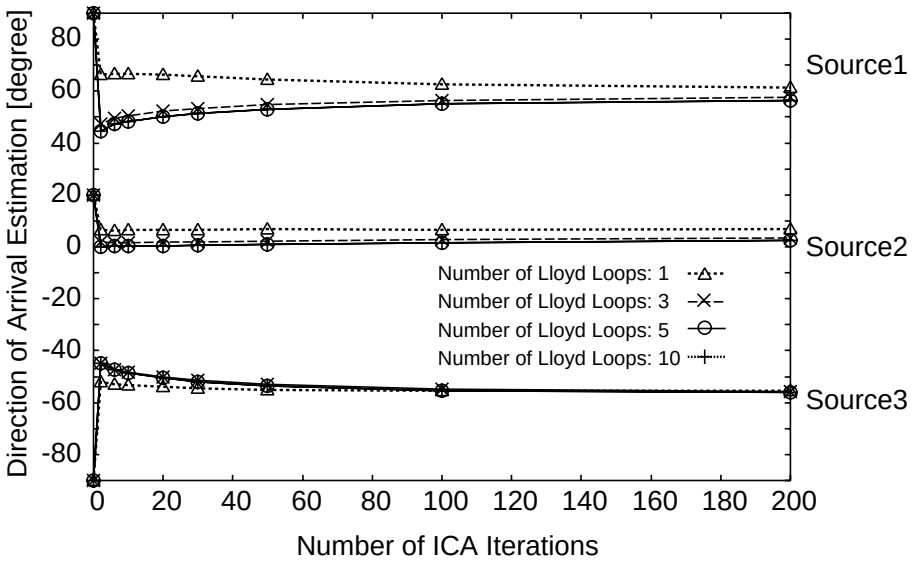


Figure 59. Results of DOAs estimated by the Lloyd clustering algorithm for different number of loops in the proposed method.

in Section 2.3. This also corresponds to the special case that ICA is always (wrongly) chosen in **step 5** of the proposed algorithm, i.e., always $\boldsymbol{W}_{i+1}(f) = \boldsymbol{W}^{(\mathrm{ICA})}(f)$ in (110).

Null Beamformer : The iteratively optimized null beamformer which corresponds to the special case that the null beamformer is always (wrongly) chosen in **step 5** of the proposed algorithm, i.e., always $\boldsymbol{W}_{i+1}(f) = \boldsymbol{W}^{(\mathrm{BF})}(f)$ in (110).

These results reveal that the proposed method obviously outperforms both the conventional FDICA-based BSS and null beamformer at every iteration point. Thus, it can be asserted that the proposed method is feasible for the case of $K = L = 3$ as well as $K = L = 2$ [19].

In addition, Figure 61 shows the example of alternation results between ICA and null beamforming through iterative optimization by the proposed algorithm. As shown in Fig. 61, the proposed algorithm can function automatically as follows.

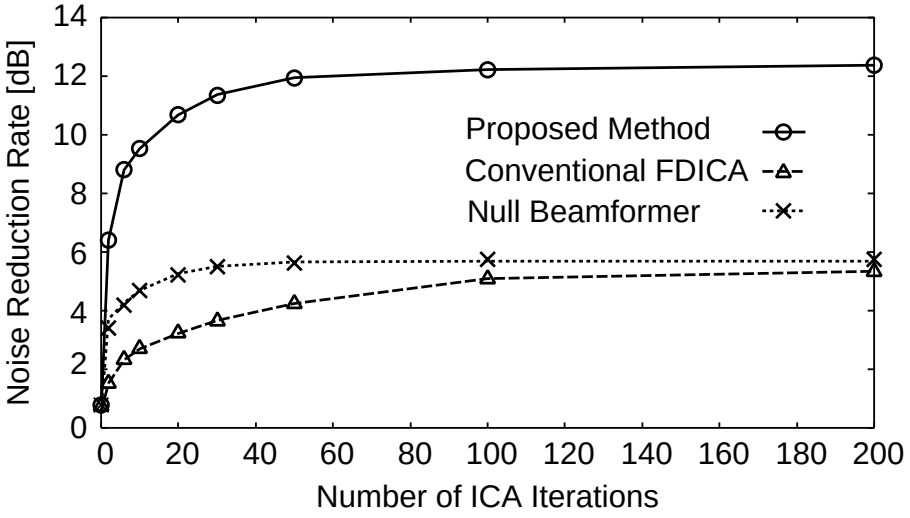


Figure 60. Noise reduction rates for different iteration points in proposed method, conventional ICA, and iteratively optimized null beamformer.

- Null beamforming is used for the acceleration of learning early in the iterations because $\boldsymbol{W}^{(\mathrm{BF})}(f)$ is a rough approximation of the separation matrix.
- ICA is used after the early part of the iterations because it can update the separation matrix more accurately.
- The separation matrix obtained by ICA is substituted by the matrix based on null beamforming through all iteration points at particular frequency bins where the independence between the sources is low.

From these results, although null beamforming is not suitable for signal separation under the condition that direct sounds and their reflections exist, we can confirm that the temporal utilization of null beamforming for algorithm diversity through ICA iterations is effective for improving the separation performance and convergence.

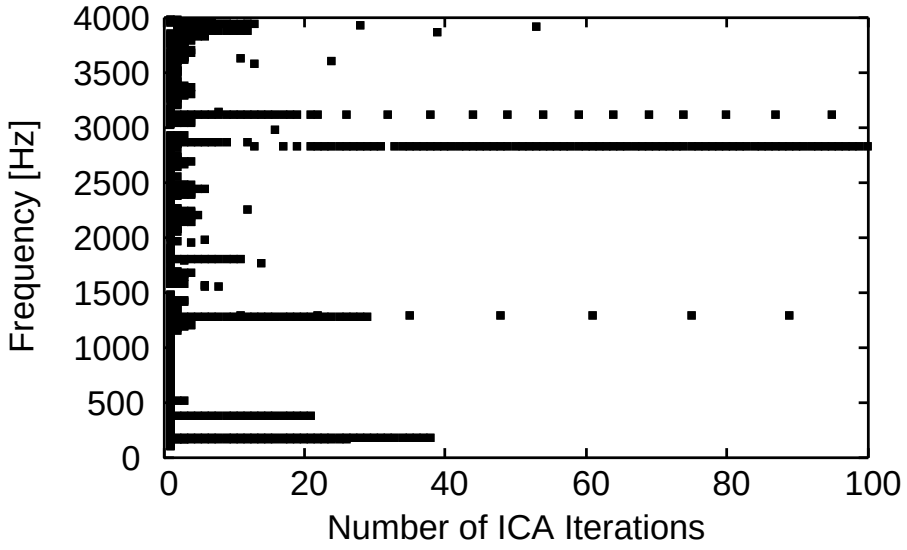


Figure 61. Result of alternation between ICA and null beamforming through iterative optimization by the proposed algorithm. The symbol "-" indicates that the null beamforming is used at the iteration point and frequency bin.

A.4 Conclusion

In this section, we described a fast- and high-convergence blind separation algorithm for multiple source signals where null beamforming is temporally used for algorithm diversity through ICA iterations. The simulation results of the signal separation experiments reveal that the signal separation performance of the proposed algorithm is superior to that of the conventional FDICA-based BSS method, and the utilization of null beamforming in ICA is effective for improving the separation performance and convergence, even under reverberant conditions.

B. Derivation of Eq. (61)

The cost function Eq. (43) is

$$\begin{aligned}\frac{\partial Q(\boldsymbol{w}(\tau))}{\partial \boldsymbol{w}(\tau)} &= \frac{1}{2B}\sum_{b=1}^{B}\frac{\partial}{\partial \boldsymbol{w}(\tau)}\log\left(\frac{\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\det \boldsymbol{R}_y^{(b)}(0)}\right)\\ &= \frac{1}{2B}\sum_{b=1}^{B}\Big\{\frac{\partial}{\partial \boldsymbol{w}(\tau)}\log(\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0))\\ &\quad -\frac{\partial}{\partial \boldsymbol{w}(\tau)}\log(\det \boldsymbol{R}_y^{(b)}(0))\Big\}. \end{aligned} \tag{111}$$

In this thesis, we give the derivation of the case in which the number of the source signals and that of the observed signals are 2, i.e., the separation filter matrix $\boldsymbol{w}(\tau)$ is 2×2 matrix.

The partial differentiation $\partial \log(\det \boldsymbol{R}_y^{(b)}(0))/\partial \boldsymbol{w}(\tau)$ is given by the following equation (the differentiation of matrix were refered from [77, 78]).

$$\begin{aligned}\frac{\partial \log(\det \boldsymbol{R}_y^{(b)}(0))}{\partial \boldsymbol{w}(\tau)} &= \frac{\partial \log(\det \boldsymbol{R}_y^{(b)}(0))}{\partial \det \boldsymbol{R}_y^{(b)}(0)}\cdot\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial \boldsymbol{w}(\tau)}\\ &= \left(\det \boldsymbol{R}_y^{(b)}(0)\right)^{-1}\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial \boldsymbol{w}(\tau)}. \end{aligned} \tag{112}$$

Here, $\boldsymbol{R}_y^{(b)}(0)$ is rewritten as the following equation:

$$\begin{aligned}\boldsymbol{R}_y^{(b)}(0) &= \langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\\ &= \langle (\boldsymbol{W}(z)\boldsymbol{x}(t))(\boldsymbol{W}(z)\boldsymbol{x}(t))^{\mathrm{T}}\rangle_t^{(b)}\\ &= \langle \boldsymbol{W}(z)\boldsymbol{x}(t)\boldsymbol{x}(t)^{\mathrm{T}}\boldsymbol{W}(z)^{\mathrm{T}}\rangle_t^{(b)}\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{W}(z^{-1})\langle \boldsymbol{x}(t)\boldsymbol{x}(t)^{\mathrm{T}}\rangle_t^{(b)}\boldsymbol{W}(z^{-1})^{\mathrm{T}} \\
&= \boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)\boldsymbol{W}(z^{-1})^{\mathrm{T}}, \qquad (113)
\end{aligned}
$$

where

$$
\boldsymbol{R}_x^{(b)}(0) = \langle \boldsymbol{x}(t)\boldsymbol{x}(t)^{\mathrm{T}}\rangle_t^{(b)}. \qquad (114)
$$

Also, the elements of $\boldsymbol{R}_y^{(b)}(0)$ are given as the following equation by using the elements of $\boldsymbol{R}_x^{(b)}(0)$ and $\boldsymbol{W}(z^{-1})$:

$$
\begin{aligned}
[\boldsymbol{R}_y^{(b)}(0)]_{11} &= [\boldsymbol{W}(z^{-1})]_{11}^2[\boldsymbol{R}_x^{(b)}(0)]_{11} + [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{12} \\
&\quad \cdot\Big([\boldsymbol{R}_x^{(b)}(0)]_{21} + [\boldsymbol{R}_x^{(b)}(0)]_{12}\Big) + [\boldsymbol{W}(z^{-1})]_{12}^2[\boldsymbol{R}_x^{(b)}(0)]_{22}, \qquad (115) \\
[\boldsymbol{R}_y^{(b)}(0)]_{12} &= [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{21} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}, \qquad (116) \\
[\boldsymbol{R}_y^{(b)}(0)]_{21} &= [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{21} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}, \qquad (117) \\
[\boldsymbol{R}_y^{(b)}(0)]_{22} &= [\boldsymbol{W}(z^{-1})]_{21}^2[\boldsymbol{R}_x^{(b)}(0)]_{11} + [\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{W}(z^{-1})]_{22} \\
&\quad \cdot\Big([\boldsymbol{R}_x^{(b)}(0)]_{21} + [\boldsymbol{R}_x^{(b)}(0)]_{12}\Big) + [\boldsymbol{W}(z^{-1})]_{22}^2[\boldsymbol{R}_x^{(b)}(0)]_{22}, \qquad (118)
\end{aligned}
$$

where $[\cdot]_{ij}$ denotes the ij-th element of the argument. The partial differentiation $\partial \det \boldsymbol{R}_y^{(b)}(0)/\partial \boldsymbol{w}(\tau)$ in Eq. (112) is expanded as the following equation by substituting Eq. (113):

$$
\begin{aligned}
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial \boldsymbol{w}(\tau)} &= \frac{\partial}{\partial \boldsymbol{w}(\tau)}\Big([\boldsymbol{R}_y^{(b)}(0)]_{11}[\boldsymbol{R}_y^{(b)}(0)]_{22} - [\boldsymbol{R}_y^{(b)}(0)]_{12}[\boldsymbol{R}_y^{(b)}(0)]_{21}\Big) \\
&= \Big\{[\boldsymbol{W}(z^{-1})]_{11}^2[\boldsymbol{R}_x^{(b)}(0)]_{11} + [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{12} \\
&\quad \cdot\Big([\boldsymbol{R}_x^{(b)}(0)]_{21} + [\boldsymbol{R}_x^{(b)}(0)]_{12}\Big) + [\boldsymbol{W}(z^{-1})]_{12}^2[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\} \\
&\quad \cdot\Big\{[\boldsymbol{W}(z^{-1})]_{21}^2[\boldsymbol{R}_x^{(b)}(0)]_{11} + [\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{W}(z^{-1})]_{22} \\
&\quad \cdot\Big([\boldsymbol{R}_x^{(b)}(0)]_{21} + [\boldsymbol{R}_x^{(b)}(0)]_{12}\Big) + [\boldsymbol{W}(z^{-1})]_{22}^2[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\} \\
&\quad - \Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{21}
\end{aligned}
$$

$$
\begin{aligned}
&+ [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&+ [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\} \\
&\cdot\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&+ [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{21} \\
&+ [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&+ [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\}. \qquad (119)
\end{aligned}
$$

Hereafter, we resolve Eq. (119) into the partial differentiation for each element of $\boldsymbol{w}(\tau)$ as

$$
\begin{aligned}
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{11}} &= z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{22}\Big\{2[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}\Big([\boldsymbol{R}_x^{(b)}(0)]_{21} + [\boldsymbol{R}_x^{(b)}(0)]_{12}\Big)\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{21}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{12}\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{12}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{21}\Big\}\Big) \\
&= 2z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{22}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{21}\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{12}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{21}\Big\}\Big), \qquad (120)
\end{aligned}
$$

where we used the partial differentiation as

$$
\frac{\partial [\boldsymbol{W}(z^{-1})]_{ij}}{\partial [\boldsymbol{w}(\tau)]_{ij}} = \frac{\partial}{\partial [\boldsymbol{w}(\tau)]_{ij}}\Big(\sum_{\tau'=0}^{Q-1}[\boldsymbol{w}(\tau')]_{ij}z^{\tau'}\Big) = z^{\tau}, \qquad (121)
$$

and the following relation defined from the symmetry of correlation matrix,

$$
\begin{aligned}
[\boldsymbol{R}_x^{(b)}(0)]_{12} &= [\boldsymbol{R}_x^{(b)}(0)]_{21}, \qquad (122) \\
[\boldsymbol{R}_y^{(b)}(0)]_{12} &= [\boldsymbol{R}_y^{(b)}(0)]_{21}. \qquad (123)
\end{aligned}
$$

By calculating the other elements, we obtain

$$
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{12}} = z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{22}\Big\{2[\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{22}
$$

$$
\begin{aligned}
&\quad + [\boldsymbol{W}(z^{-1})]_{11}\Big([\boldsymbol{R}_x^{(b)}(0)]_{21} + [\boldsymbol{R}_x^{(b)}(0)]_{12}\Big)\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{21}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{21} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{12}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\}\Big) \\
&= 2z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{22}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{12}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\}\Big), \qquad (124)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{21}} &= z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{11}\Big\{2[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{22}\Big([\boldsymbol{R}_x^{(b)}(0)]_{21} + [\boldsymbol{R}_x^{(b)}(0)]_{12}\Big)\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{21}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{21}\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{12}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{12}\Big\}\Big) \\
&= 2z^{\tau}\Big(-[\boldsymbol{R}_y^{(b)}(0)]_{21}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{21}\Big\} \\
&\quad + [\boldsymbol{R}_y^{(b)}(0)]_{11}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{21}\Big\}\Big), \qquad (125)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{22}} &= z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{11}\Big\{2[\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{21}\Big([\boldsymbol{R}_x^{(b)}(0)]_{21} + [\boldsymbol{R}_x^{(b)}(0)]_{12}\Big)\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{21}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\} \\
&\quad - [\boldsymbol{R}_y^{(b)}(0)]_{12}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{21} \\
&\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\}\Big) \\
&= 2z^{\tau}\Big(-[\boldsymbol{R}_y^{(b)}(0)]_{21}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{12}
\end{aligned}
$$

$$
\begin{aligned}
&+ [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\} \\
&+ [\boldsymbol{R}_y^{(b)}(0)]_{11}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&+ [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\}\Big). \qquad (126)
\end{aligned}
$$

Here, the elements of $\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)$ are given as the following equation by using the elements of $\boldsymbol{R}_x^{(b)}(0)$ and $\boldsymbol{W}(z^{-1})$:

$$
\begin{aligned}
[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{11} &= [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{11} + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{21}, \qquad (127)\\
[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{12} &= [\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{12} + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{22}, \qquad (128)\\
[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{21} &= [\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11} + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{21}, \qquad (129)\\
[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{22} &= [\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{12} + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}. \qquad (130)
\end{aligned}
$$

Substituting Eqs. (127)–(130) into Eqs (120)–(126), the following equations are obtained:

$$
\begin{aligned}
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{11}} &= 2z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{22}[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad -[\boldsymbol{R}_y^{(b)}(0)]_{12}[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{21}\Big), \qquad (131)\\
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{12}} &= 2z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{22}[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&\quad -[\boldsymbol{R}_y^{(b)}(0)]_{12}[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{22}\Big), \qquad (132)\\
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{21}} &= 2z^{\tau}\Big(-[\boldsymbol{R}_y^{(b)}(0)]_{21}[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{11} \\
&\quad +[\boldsymbol{R}_y^{(b)}(0)]_{11}[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{21}\Big), \qquad (133)\\
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{22}} &= 2z^{\tau}\Big(-[\boldsymbol{R}_y^{(b)}(0)]_{21}[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{12} \\
&\quad +[\boldsymbol{R}_y^{(b)}(0)]_{11}[\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{22}\Big). \qquad (134)
\end{aligned}
$$

Thus, Eq. (119) is rewritten as the following matrix form:

$$
\begin{aligned}
\frac{\partial \det \boldsymbol{R}_y^{(b)}(0)}{\partial \boldsymbol{w}(\tau)} &= 2z^{\tau}\begin{bmatrix} [\boldsymbol{R}_y^{(b)}(0)]_{22} & -[\boldsymbol{R}_y^{(b)}(0)]_{12} \\ -[\boldsymbol{R}_y^{(b)}(0)]_{21} & [\boldsymbol{R}_y^{(b)}(0)]_{11} \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} [\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{11} & [\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{12} \\ [\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{21} & [\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{22} \end{bmatrix}
\end{aligned}
$$

$$= 2z^{\tau}\mathrm{adj}\Big[\boldsymbol{R}_y^{(b)}(0)\Big]\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0), \tag{135}$$

where adj[·] is adjoint matrix. By substituting Eq. (135) into Eq. (112), $\partial \log(\det \boldsymbol{R}_y^{(b)}(0))/\partial \boldsymbol{w}(\tau)$ is rewritten as

$$\begin{aligned}\frac{\partial \log(\det \boldsymbol{R}_y^{(b)}(0))}{\partial \boldsymbol{w}(\tau)} &= 2z^{\tau}\Big(\det \boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\mathrm{adj}\Big[\boldsymbol{R}_y^{(b)}(0)\Big]\cdot \boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0) \\ &= 2z^{\tau}\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0) \\ &\quad \cdot \boldsymbol{W}(z^{-1})^{\mathrm{T}}\boldsymbol{W}(z^{-1})^{-\mathrm{T}}. \end{aligned} \tag{136}$$

By using relation Eq. (113), Eq. (136) is expanded as

$$\begin{aligned}\frac{\partial \log(\det \boldsymbol{R}_y^{(b)}(0))}{\partial \boldsymbol{w}(\tau)} &= 2\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1} z^{\tau}\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\boldsymbol{W}(z^{-1})^{-\mathrm{T}} \\ &= 2\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\langle z^{-\tau}\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}\boldsymbol{W}(z^{-1})^{-\mathrm{T}} \\ &= 2\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\langle \boldsymbol{y}(t)\boldsymbol{y}(t-n)^{\mathrm{T}}\rangle_t^{(b)}\boldsymbol{W}(z^{-1})^{-\mathrm{T}} \\ &= 2\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{W}(z^{-1})^{-\mathrm{T}}. \end{aligned} \tag{137}$$

On the other hand, $\partial \log(\det \mathrm{diag}\boldsymbol{R}_{\mathrm{y}}^{(\mathrm{b})}(0))/\partial \boldsymbol{w}(\tau)$ is also derived as the same manner in the above derivation:

$$\begin{aligned}\frac{\partial \log(\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0))}{\partial \boldsymbol{w}(\tau)} &= \frac{\partial \log(\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0))}{\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}\cdot\frac{\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\partial \boldsymbol{w}(\tau)} \\ &= \Big(\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\frac{\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\partial \boldsymbol{w}(\tau)}. \end{aligned} \tag{138}$$

The elements of $\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)/\partial \boldsymbol{w}(\tau)$ are obtained by

$$\begin{aligned}\frac{\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{11}} &= 2z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{22}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{11} \\ &\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{21}\Big\}\Big), \end{aligned} \tag{139}$$

$$\begin{aligned}\frac{\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{12}} &= 2z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{22}\Big\{[\boldsymbol{W}(z^{-1})]_{11}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\ &\quad + [\boldsymbol{W}(z^{-1})]_{12}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\}\Big), \end{aligned} \tag{140}$$

$$\frac{\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{21}} = 2z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{11}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{11}$$

$$+ [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{21}\Big\}\Big), \tag{141}$$

$$\begin{aligned}\frac{\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\partial [\boldsymbol{w}(\tau)]_{22}} &= 2z^{\tau}\Big([\boldsymbol{R}_y^{(b)}(0)]_{11}\Big\{[\boldsymbol{W}(z^{-1})]_{21}[\boldsymbol{R}_x^{(b)}(0)]_{12} \\ &\quad + [\boldsymbol{W}(z^{-1})]_{22}[\boldsymbol{R}_x^{(b)}(0)]_{22}\Big\}\Big). \end{aligned} \tag{142}$$

Thus, $\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)/\partial \boldsymbol{w}(\tau)$ is given by the following matrix form:

$$\begin{aligned}\frac{\partial \det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\partial \boldsymbol{w}(\tau)} &= 2z^{\tau}\begin{bmatrix} [\boldsymbol{R}_y^{(b)}(0)]_{22} & 0 \\ 0 & [\boldsymbol{R}_y^{(b)}(0)]_{11} \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} [\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{11} & [\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{12} \\ [\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{21} & [\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0)]_{22} \end{bmatrix} \\ &= 2z^{\tau}\mathrm{adj}\Big[\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big]\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0). \end{aligned} \tag{143}$$

$\partial \log(\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0))/\partial \boldsymbol{w}(\tau)$ is rewritten as following equation by substituting Eq. (143) into Eq. (138):

$$\begin{aligned}\frac{\partial \log(\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0))}{\partial \boldsymbol{w}(\tau)} &= 2z^{\tau}\Big(\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\mathrm{adj}\Big[\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big] \\ &\quad \cdot \boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0) \\ &= 2z^{\tau}\Big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{W}(z^{-1})\boldsymbol{R}_x^{(b)}(0) \\ &\quad \cdot \boldsymbol{W}(z^{-1})^{\mathrm{T}}\boldsymbol{W}(z^{-1})^{-\mathrm{T}} \\ &= 2\Big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{W}(z^{-1})^{-\mathrm{T}}. \end{aligned} \tag{144}$$

As a result, $\partial Q(\boldsymbol{w}(\tau))/\partial \boldsymbol{w}(\tau)$ is obtained by following equation by substituting Eqs. (137), (144) into Eq. (111):

$$\begin{aligned}\frac{\partial Q(\boldsymbol{w}(\tau))}{\partial \boldsymbol{w}(\tau)} &= \frac{1}{2B}\sum_{b=1}^{B}\Big\{2\Big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{W}(z^{-1})^{-\mathrm{T}} \\ &\quad - 2\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{W}(z^{-1})^{-\mathrm{T}}\Big\} \\ &= \frac{1}{B}\sum_{b=1}^{B}\Big\{\Big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau) \\ &\quad - \Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau)\Big\}\boldsymbol{W}(z^{-1})^{-\mathrm{T}}. \end{aligned} \tag{145}$$

C. Derivation of Eq. (62)

Substituting Eq. (61) into Eq. (59) we obtaine the following natural gradient:

$$\begin{aligned}\Delta \boldsymbol{w}(\tau) &= -\frac{1}{B}\sum_{b=1}^{B}\Big\{\Big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau) - \Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau)\Big\}\boldsymbol{W}(z) \\ &= \frac{1}{B}\sum_{b=1}^{B}\Big\{\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{W}(z) \\ &\quad -\Big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{W}(z)\Big\}. \qquad (146)\end{aligned}$$

The convolution operation $\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{H}(z)$ is defined as

$$\begin{aligned}\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{H}(z) &\equiv \sum_{d=0}^{Q-1}\boldsymbol{R}_y^{(b)}(\tau)\boldsymbol{h}(d)z^{-d} \\ &= \sum_{d=0}^{Q-1}\boldsymbol{R}_y^{(b)}(\tau-d)\boldsymbol{h}(d) \\ &= \sum_{d=0}^{Q-1}\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)}\boldsymbol{h}(d), \qquad (147)\end{aligned}$$

where $\boldsymbol{H}$(z) is the z-transform of an arbitrary FIR-filter matrix $\boldsymbol{h}(\tau)$. Using the definition Eq. (147) and Eq. (44) , Equation (146) is rewritten as

$$\begin{aligned}\Delta \boldsymbol{w}(\tau) &= \frac{1}{B}\sum_{b=1}^{B}\Big\{\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\Big(\sum_{d=0}^{Q-1}\Big\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\Big\rangle_t^{(b)}\boldsymbol{w}(d)\Big) \\ &\quad -\Big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\Big(\sum_{d=0}^{Q-1}\Big\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\Big\rangle_t^{(b)}\boldsymbol{w}(d)\Big)\Big\} \\ &= \frac{1}{B}\sum_{b=1}^{B}\sum_{d=0}^{Q-1}\Big\{\Big(\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\Big\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\Big\rangle_t^{(b)} \\ &\quad -\Big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\Big)^{-1}\Big\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\Big\rangle_t^{(b)}\Big\}\boldsymbol{w}(d) \\ &= \frac{1}{B}\sum_{b=1}^{B}\sum_{d=0}^{Q-1}\Big\{\Big(\Big\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\Big\rangle_t^{(b)}\Big)^{-1}\Big\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\Big\rangle_t^{(b)} \\ &\quad -\Big(\mathrm{diag}\Big\langle \boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\Big\rangle_t^{(b)}\Big)^{-1}\Big\langle \boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\Big\rangle_t^{(b)}\Big\}\boldsymbol{w}(d). \qquad (148)\end{aligned}$$

List of Publications

Journal Papers

1. 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, pp.846–858, April 2003.

2. T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable learning 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, Aug. 2003.

3. T. Nishikawa, H. Abe, H. Saruwatari, K. Shikano, and A. Kaminuma, "Overdetermined blind separation for real convolutive mixtures of speech based on multistage ICA using subarray processing," IEICE Trans. Fundamentals, Vol.E87-A, No.8, pp.1924–1932, Aug. 2004.

4. H. Saruwatari, T. Kawamura, T. Nishikawa, and K. Shikano, "Fast-convergence algorithm for blind source separation based on array signal processing", IEICE Trans. Fundamentals, Vol.E86-A, No.3, pp.286–291 March 2003.

5. S. Araki, S. Makino, R. Mukai, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," IEEE Transactions on Speech and Audio Processing, Vol. 11, No. 2, pp.109–116, March 2003.

6. H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, T. Nishikawa, and K. Shikano, "Blind source separation combining ICA and beamforming," EURASIP Journal on Applied Signal Processing, Vol.2003, No.11, pp.1135–1146, 2003.

7. S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures," EURASIP Journal on Applied Signal Processing, Vol.2003, No.11, pp.1157–1166, 2003.

8. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fidelity blind separation of acoustic signals using SIMO-model-based independent component analysis," IEICE Trans. Fundamentals, Vol.E87-A, No.8, pp.2063–2072, Aug. 2004.

9. H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on fast-convergence algorithm combining ICA and beamforming," IEEE Transactions on Speech and Audio Processing (accepted).

10. T. Takatani, S. Ukai, T. Nishikawa, H. Saruwatari, and K. Shikano, "A self-generator method for initial filters of SIMO-ICA applied to blind separation of binaural sound mixtures," IEICE Trans. Fundamentals, (accepted).

International Conferences

1. T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation based on multi-stage ICA using frequency-domain ICA and time-domain ICA," The International Conference on Fundamentals of Electronics, Communications and Computer Sciences (ICFS2002), Vol.1, pp.7–12, March 2002.

2. T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation based on multi-stage ICA combining frequency-domain ICA and time-domain ICA," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2002), pp.2938–2941, May 2002.

3. T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of time-domain ICA, frequency-domain ICA and multistage ICA," The 2002 European Signal Processing Conference (EUSIPCO2002), Vol.II, pp.15–18, Sep. 2002.

4. T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable learning algorithm for blind separation of temporally correlated signals combining multistage ICA and linear prediction," 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp.327–342, April 2003.

5. T. Nishikawa, H. Saruwatari, K. Shikano, S. Araki, and S. Makino, "Multistage ICA for blind source separation of real acoustic convolutive mixture," 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp.523–528, April 2003.

6. T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable learning algorithm for low-distortion blind separation of real speech mixture combining multistage ICA and linear prediction," ISCA tutorial and research workshop on nonlinear speech processing (NOLISP2003), pp.31–34, May 2003.

7. T. Nishikawa, H. Abe, H. Saruwatari, and K. Shikano, "Overdetermined blind source separation of real acoustic sounds based on multistage ICA using subarray processing," 2003 International Symposium on Signal Processing and Information Technology (ISSPIT2003), Dec. 2003.

8. T. Nishikawa, H. Abe, H. Saruwatari, and K. Shikano, "Overdetermined blind separation of acoustic signals based on MISO-constrained frequency-domain ICA," 18th International Congress on Acoustics (ICA2004), Vol.IV, pp.3143–3146, April 2004.

9. T. Nishikawa, H. Abe, H. Saruwatari, and K. Shikano, "Overdetermined blind separation for convolutive mixtures of speech based on multistage ICA using subarray processing," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2004), Vol.I, pp.225–228, May 2004.

10. T. Nishikawa, H. Saruwatari, K. Shikano, and A. Kaminuma, "Stable and low-distortion algorithm based on overdetermined blind separation for convolutive mixtures of speech," 5th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2004), pp.881–888, Sep. 2004.

11. T. Nishikawa, H. Saruwatari, K. Shikano, "Fast-convergence blind separation of more than two sources combining ICA and beamforming," International Workshop on Nonlinear Signal and Image Processing (NSIP2005), May 2005, (accepted).

12. S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Limitation of frequency domain blind source separation for convolutive mixture of speech," IEEE International Workshop on Hands-Free Speech Communication (HSC2001), pp.91–94, April 2001.

13. S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2001), pp.2737–2740, May 2001.

14. S. Araki, S. Makino, R. Mukai, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolved mixture of speech," 3rd International Conference on Independent Component Analysis and Blind Signal Separation (ICA2001), pp.132–137, Dec. 2001.

15. Y. Hinamoto, T. Nishikawa, H. Saruwatari, S. Araki, S. Makino, and R. Mukai, "Equivalence between frequency domain blind source separation and adaptive beamforming," The International Conference on Fundamentals of Electronics, Communications and Computer Sciences (ICFS2002), Vol.1, pp.13–18, March 2002.

16. S. Araki, S. Makino, R. Mukai, Y. Hinamoto, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2002), pp.1899–1902, May 2002.

17. R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming," 2002 IEEE International Workshop on Neural Networks for Signal Processing (NNSP2002), Sep. 2002.

18. S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Blind source separation for convolutive mixtures of speech using subband processing," Second International Workshop on Spectral Methods and Multirate Signal Processing (SMMSP2002), pp.195–202, Sep. 2002.

19. S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Subband based blind source separation with appropriate processing for each frequency band," 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp.499–504, April 2003.

20. H. Saruwatari, T. Takatani, H. Yamajo, T. Nishikawa, and K. Shikano, "Blind separation and deconvolution for real convolutive mixture of temporally correlated acoustic signals using SIMO-model-based ICA," 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp.549–554, April 2003.

21. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "SIMO-model-based independent component analysis for high-fidelity blind separation of acoustic signals," 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp.993–998, April 2003.

22. S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Blind source separation for convolutive mixtures of speech," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2003), pp.509–512, April 2003.

23. T. Takatani, T. Nishikawa, and H. Saruwatari, "Blind source separation based on binaural ICA," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2003), pp.321–324, April 2003.

24. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fidelity blind separation for convolutive mixture of acoustic signals using SIMO-model-based independent component analysis," Seventh International Symposium on Signal Processing and its Applications (ISSPA2003), July 2003.

25. H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa, and K. Shikano, "Parallel structured independent component analysis for SIMO-model-based blind separation and deconvolution of convolutive speech mixture," 2003 International Joint Conference on Neural Networks (IJCNN2003), pp.714–719, July 2003.

26. H. Yamajo, H. Saruwatari, T. Takatani, T. Nishikawa, and K. Shikano, "Blind separation and deconvolution for convolutive mixture of speech using SIMO-model-based ICA and multichannel inverse filtering," 8th European Conference on Speech Communication and Technology (Eurospeech2003), pp.I-537–540, Sep. 2003.

27. H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa, and K. Shikano, "Blind separation and deconvolution of MIMO system driven by colored inputs using SIMO-model-based ICA with information-geometric learning," IEEE Neural Network for Signal Processing Workshop 2003 (NNSP2003), pp.379–388, Sep. 2003.

28. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fildeity blind separation of acoustic signals using SIMO-model-based ICA with information-geometric learning," 2003 International Workshop on Acoustic Echo and Noise Control (IWAENC2003), pp.251–254, Sep. 2003.

29. H. Yamajo, H. Saruwatari, T. Takatani, T. Nishikawa, and K. Shikano, "Evaluation of blind separation and deconvolution for convolutive speech mixture using SIMO-model-based ICA," 2003 International Workshop on Acoustic Echo and Noise Control (IWAENC2003), pp.299–302 , Sep. 2003.

30. H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa, and K. Shikano, "Blind separation and deconvolution of MIMO-FIR system with colored sound inputs using SIMO-model-based ICA," 2003 IEEE Workshop on Statistical Signal Processing (SSP2003), pp.421–424, Sep. 2003.

31. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "SIMO-model-based ICA with information-geometric learning for high-fidelity blind source separation," 2003 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2003), pp.108–113, Dec. 2003.

32. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison between SIMO-ICA with least squares criterion and SIMO-ICA with information-geometric learning," 18th International Congress on Acoustics (ICA2004), Vol.I, pp.329–332, April 2004.

33. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind separation of binaural sound mixtures using SIMO-model-based independent component analysis," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2004), Vol.IV, pp.113–116, May 2004.

34. H. Yamajo, H. Saruwatari, T. Takatani, T. Nishikawa, and K. Shikano, "Evaluation of blind separation and deconvolution for binaural-sound mixtures using SIMO-model-based ICA," The 2004 European Signal Processing Conference (EUSIPCO2004), pp.1709–1712, Sep. 2004.

35. S. Ukai, T. Takatani, T. Nishikawa, and H. Saruwatari, "Blind source separation combining SIMO-model-based ICA and adaptive beamforming," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2005), March 2005 (accepted).

36. Y. Ohashi, K. Sakamoto, T. Nishikawa, H. Saruwatari, and K. Shikano, "Hands-free speech recognition using spatial subtraction array," International Workshop on Hands-Free Speech Communication and Microphone Arrays, March 2005 (accepted).

37. Y. Ohashi, T. Nishikawa, H. Saruwatari, A. Lee, and K. Shikano, "Noise robust speech recognition based on spatial subtraction array," International Workshop on Nonlinear Signal and Image Processing (NSIP2005), May 2005, (accepted).

38. T. Takatani, S. Ukai, T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind Decomposition of Binaural Mixed Signals Using High-convergence algorithm combining SIMO-ICA and DOA estimation," International Workshop on Nonlinear Signal and Image Processing (NSIP2005), May 2005, (accepted).

Invited Talks

1. H. Saruwatari, T. Nishikawa, S. Araki, and S. Makino, "Blind source separation based on multi-stage ICA combining time-domain ICA and frequency-

domain ICA," 2001 Annual Conference of Japanese Neural Network Society, pp.99–100, Sep. 2001 (in Japanese).

2. H. Saruwatari, T. Nishikawa, and K. Shikano, "Blind source separation of acoustic signals based on multistage independent component analysis," Summer Meeting of Acoustical Society of Korea, pp.9–14, Aug. 2002.

3. H. Saruwatari, T. Nishikawa, T. Takatani, and K. Shikano, "Recent advance in acoustic blind source separation," 2005 IEICE General Conference of The Institute of Electronics, Information and Communication Engineers, March 2005 (in Japanese).

4. H. Saruwatari, K. Sawai, T. Nishikawa, A. Lee, K. Shikano, A. Kaminuma, M. Sakata, and D. Saito, "(tentative) Speech enhancement based on blind source separation in car environments," International Workshop on Real-world Multimedia Corpora in Mobile Environment, April 2005.

5. H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa, K. 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. T. Nishikawa, S. Araki, S. Makino, and H. Saruwatari, "Optimization on the number of subbands in frequency-domain blind source separation," IEICE Technical Report, EA2000-95, pp.53–59, Jan. 2001 (in Japanese).

2. T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation using frequency-domain ICA and time-domain ICA," IEICE Technical Report, EA2001-35, pp.57–64, Aug. 2001 (in Japanese).

3. T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of blind source separation methods based on time-domain ICA using nonstationarity and multistage ICA," IEICE Technical Report, EA2001-112, pp.45–52, Jan. 2002.

4. T. Nishikawa, H. Saruwatari, and K. Shikano, "stable learning algorithm for blind separation of temporally-correlated signals using multistage ICA and linear prediction," IEICE Technical Report, EA2002-109, pp.25–30, Jan. 2003.

5. T. Nishikawa, H. Abe, H. Saruwatari, K. Shikano, and A. Kaminuma, "Overdetermined blind separation of speech signal based on multistage ICA using subarray processing," IEICE Technical Report, EA2003-141, pp.7–12, Jan. 2004.

6. T. Nishikawa, H. Saruwatari, and K. Shikano, "Fast-convergence algorithm combining ICA and beamforming for blind separation of more than two sources," IEICE Technical Report, EA2004-92, pp.13–18, Nov. 2004.

7. T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable learning and low-distortion algorithm for overdetermined blind separation of temporally correlated signals," IEICE Technical Report, EA2005-120, pp.19–24, Jan. 2005.

8. Y. Hinamoto, T. Nishikawa, H. Saruwatari, S. Araki, S. Makino, and R. Mukai, "Equivalence between frequency domain blind source separation and adaptive beamformer," IEICE Technical Report, EA2001-84, pp.75–82, Nov. 2001 (in Japanese).

9. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fidelity blind source separation using SIMO model based ICA," IEICE Technical Report, EA2002-108, pp.19–24, Jan. 2003.

10. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "SIMO-model-based ICA with infomax constraint for high-fidelity blind source separation," IEICE Technical Report, EA2003-12, pp.25–30, April 2003.

11. H. Yamajo, H. Saruwatari, T. Takatani, T. Nishikawa, and K. Shikano, "Blind separation and deconvolution of convolutive speech mixture using SIMO-model-based ICA and multichannel inverse filtering," IEICE Technical Report, EA2003-11, pp.19–24, April 2003.

12. H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa, and K. Shikano, "Blind separation and deconvolution of convolutive speech mixture using SIMO-model-based ICA with information-geometric learning," IEICE Technical Report, EA2003-46, pp.25–32, June 2003.

13. H. Abe, T. Nishikawa, H. Saruwatari, and K. Shikano, "Overdetermined blind source separation based on MISO-constrained frequency-domain ICA," IEICE Technical Report, EA2003-160, pp.31–36, March 2004 (in Japanese).

14. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-convergence algorithm combining SIMO-ICA and DOA estimation for blind binaural signal separation," IEICE Technical Report, EA2004-43, pp.31–36, Aug. 2004.

15. S. Ukai, T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Evaluation of blind source separation combining SIMO-model-based ICA and adaptive beamforming," IEICE Technical Report, SIS2004-48, pp.33–38, Dec. 2004.

16. Y. Ohashi, T. Nishikawa, H. Saruwatari, and K. shikano, "Hands-free speech recognition using spatial subtraction array and known noise superimposition," IEICE Technical Report, EA2005-123, pp.37–42, Jan. 2005 (in Japanese).

17. S. Obara, T. Nishikawa, H. Saruwatari, and K. Shikano, "Selection of optimal null beamformer based on kurtosis criterion," IEICE Technical Report, EA2005-122, pp.31–36, Jan. 2005 (in Japanese).

Meetings

1. T. Nishikawa, T. Kawamura, H. Saruwatari, and K. Shikano, "Overdetermined source separation with blind beamformer," *The 2000 Autumn Meeting of the ASJ*, pp.447–448, Sep. 2000 (in Japanese).

2. T. Nishikawa, S. Araki, S. Makino, and H. Saruwatari, "Optimization on the number of subbands in blind source separation with subband ICA," *The*

2001 Spring Meeting of the ASJ, pp.569–570, March 2001 (in Japanese).

3. T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation using frequency-domain ICA and time-domain ICA in real acoustic environment," *The 2001 Autumn Meeting of the ASJ*, pp.625–626, Oct. 2001 (in Japanese).

4. T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of blind source separation methods based on time-domain ICA, frequency-domain ICA, and multistage ICA," *The 2002 Spring Meeting of the ASJ*, pp.617–618, March 2002 (in Japanese).

5. T. Nishikawa, T. Takatani, H. Saruwatari, K. Shikano, S. Araki, and S. Makino, "Comparison of time-domain ICA methods based on minimization of KL divergence and simultaneous decorrelation of nonstationary signal," *The 2002 Autumn Meeting of the ASJ*, pp.545–546, Sep. 2002 (in Japanese).

6. T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable learning algorithm for blind separation of acoustic signals combining multistage ICA and linear prediction," *The 2003 Spring Meeting of the ASJ*, pp.675–676, March 2003 (in Japanese).

7. T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation based on multistage ICA using multi-element array," *The 2003 Autumn Meeting of the ASJ*, pp.543–544, Sep. 2003 (in Japanese).

8. T. Nishikawa, H. Saruwatari, and K. Shikano, "Stable and low-distortion algorithm for overdetermined blind source separation," *The 2004 Autumn Meeting of the ASJ*, pp.739–740, Sep. 2004 (in Japanese).

9. T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind separation of more than two sources based on fast-convergence algorithm combining ICA and beamforming," *The 2005 Spring Meeting of the ASJ*, March 2005 (in Japanese).

10. T. Nishikawa, H. Saruwatari, A. Lee, K. Shikano, D. Saitoh, and A. Kaminuma, "Speech recognition using multistage independent component analysis in real environments," 2005 IEICE General Conference of The Institute of

Electronics, Information and Communication Engineers, March 2005 (in Japanese).

11. S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Limitation of frequency domain blind source separation for convolutive mixture of speech," *The 2001 Spring Meeting of the ASJ*, pp.567–568, March 2001 (in Japanese).

12. S. Araki, R. Aichner, S. Makino, T. Nishikawa, and H. Saruwatari, "Blind source separation using SSB subband ," *The 2002 Spring Meeting of the ASJ*, pp.619–620, March 2002 (in Japanese).

13. S. Araki, R. Aichner, S. Makino, T. Nishikawa, and H. Saruwatari, "Time-domain blind source separation with utilization of null beamforming," *The 2002 Autumn Meeting of the ASJ*, pp.543–544, Sep. 2002 (in Japanese).

14. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Speech enhancement using blind source separation based on minimal distortion principle," *The 2002 Autumn Meeting of the ASJ*, pp.547–548, Sep. 2002 (in Japanese).

15. H. Yamajo, T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation and deconvolution using SIMO-model-based ICA and blind multichannel inverse filtering," *The 2003 Spring Meeting of the ASJ*, pp.673–674, March 2003 (in Japanese).

16. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation using SIMO-model-based ICA," *The 2003 Spring Meeting of the ASJ*, pp.677–678, March 2003 (in Japanese).

17. S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Subband based blind source separation with appropriate processing for each frequency band," *The 2003 Spring Meeting of the ASJ*, pp.781–782, March 2003 (in Japanese).

18. H. Abe, T. Nishikawa, H. Saruwatari, and K. Shikano, "Frequency domain blind source separation using multiple microphones," *The 2003 Autumn Meeting of the ASJ*, pp.541–542, Sep. 2003 (in Japanese).

19. T. Takatani, H. Yamajo, T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation using SIMO-model-based ICA with information-geometric learning," *The 2003 Autumn Meeting of the ASJ*, pp.537–538, Sep. 2003 (in Japanese).

20. H. Yamajo, H. Saruwatari, T. Takatani, T. Nishikawa, and K. Shikano, "Evaluation of blind source separation and deconvolution using SIMO-model-based ICA wiht information-geometric learning," *The 2003 Autumn Meeting of the ASJ*, pp.539–540, Sep. 2003 (in Japanese).

21. T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation for binaural mixed signals using SIMO-model-based independent component analysis", *The 2004 Spring Meeting of the ASJ*, pp.497–498, March 2004 (in Japanese).

22. H. Yamajo, H. Saruwatari, T. Takatani, T. Nishikawa, and K. Shikano, "Separation and equalization for colored signal mixtures with HRTF using two-stage blind separation and deconvolution," *The 2004 Spring Meeting of the ASJ*, pp.545–546, March 2004 (in Japanese).

23. T. Takatani, S. Ukai, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-convergence algorithm using SIMO-ICA and DOA estimation for blind binaural signal separation," *The 2004 Autumn Meeting of the ASJ*, pp.747–748, Sep. 2004 (in Japanese).

24. S. Obara, T. Nishikawa, H. Saruwatari, and K. Shikano, "A study on kurtosis-based selection of beamformer," *The 2004 Autumn Meeting of the ASJ*, pp.751–752, Sep. 2004 (in Japanese).

25. D. Saitoh, A. Kaminuma, H. Saruwatari, A. Lee, and T. Nishikawa, "Speech recognition in car environment using FDICA with speech sub-band passing filter," *The 2004 Autumn Meeting of the ASJ*, pp.755–756, Sep. 2004 (in Japanese).

26. Y. Ohashi, T. Nishikawa, H. Saruwatari, A. Lee, and K. Shikano, "Improvement of hands-free speech recognition using spatial subtraction array," *The 2005 Spring Meeting of the ASJ*, March 2005 (in Japanese).

27. S. Obara, T. Nishikawa, H. Saruwatari, and K. Shikano, "Evaluation of optimal null beamformer selection based on kurtosis criterion," *The 2005 Spring Meeting of the ASJ*, March 2005 (in Japanese).

28. T. Takatani, S. Ukai, T. Nishikawa, H. Saruwatari, and K. Shikano, "Evaluation of SIMO-ICA with self-generator for initial filter," *The 2005 Spring Meeting of the ASJ*, March 2005 (in Japanese).

29. S. Ukai, T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "Evaluation of blind source separation using SIMO-model-signal extraction and adaptive beamforming," *The 2005 Spring Meeting of the ASJ*, March 2005 (in Japanese).

Other Meetings

1. T. Nishikawa, "Hands-free speech recognition system using speech interface based on blind source separation," 2003 NAIST COE International Symposium - Ubiquitous Networked Media Computing -, pp.61–62, Oct. 2003.

Master's Thesis

1. T. Nishikawa, "Blind source separation based on multistage ICA combining frequency-domain ICA and time-domain ICA," *Master's thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology*, NAIST-IS-MT0051076, Feb. 2002 (in Japanese).

Award

1. TELECOM System Technology Award of the Telecommunications Advancement Foundation

2. TELECOM System Technology Student Award of the Telecommunications Advancement Foundation

3. C&C Young Best Paper Award of the Foundation for C&C Promotion

4. Funai Information Technology Award for Young Researchers of the Funai Foundation for Information Technology