EURASIP Journal on Applied Signal Processing 2003:11, 1135–1146

Blind Source Separation Combining Independent Component Analysis and Beamforming

Hiroshi Saruwatari

Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho,
Ikoma, Nara 630-0192, Japan
Email: sawatari@is.aist-nara.ac.jp

Satoshi Kurita

Center for Integrated Acoustic Information Research (CIAIR), Nagoya University, Nagoya 464-8903, Japan

Kazuya Takeda

Center for Integrated Acoustic Information Research (CIAIR), Nagoya University, Nagoya 464-8903, Japan
Email: takeda@nuee.nagoya-u.ac.jp

Fumitada Itakura

Center for Integrated Acoustic Information Research (CIAIR), Nagoya University/CIAIR, Nagoya 464-8903, Japan
Email: itakura@nuee.nagoya-u.ac.jp

Tsuyoki Nishikawa

Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho,
Ikoma, Nara 630-0192, Japan
Email: tsuyo-ni@is.aist-nara.ac.jp

Kiyohiro Shikano

Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho,
Ikoma, Nara 630-0192, Japan
Email: shikano@is.aist-nara.ac.jp

Received 26 November 2002 and in revised form 30 March 2003

We describe a new method of blind source separation (BSS) on a microphone array combining subband independent component analysis (ICA) and beamforming. The proposed array system consists of the following three sections: (1) subband ICA-based BSS section with estimation of the direction of arrival (DOA) of the sound source, (2) null beamforming section based on the estimated DOA, and (3) integration of (1) and (2) based on the algorithm diversity. Using this technique, we can resolve the low-convergence problem through optimization in ICA. To evaluate its effectiveness, signal-separation and speech-recognition experiments are performed under various reverberant conditions. The results of the signal-separation experiments reveal that the noise reduction rate (NRR) of about 18 dB is obtained under the nonreverberant condition, and NRRs of 8 dB and 6 dB are obtained in the case that the reverberation times are 150 milliseconds and 300 milliseconds. These performances are superior to those of both simple ICA-based BSS and simple beamforming method. Also, from the speech-recognition experiments, it is evident that the performance of the proposed method in terms of the word recognition rates is superior to those of the conventional ICA-based BSS method under all reverberant conditions.

Keywords and phrases: blind source separation, microphone array, independent component analysis, beamforming.

1. INTRODUCTION

Source separation for acoustic signals is to estimate original sound source signals from the mixed signals observed in each input channel. This technique is applicable to the realization of noise-robust speech-recognition and high-quality hands-free telecommunication systems. The methods of achieving source separation 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, a method of tracking a formant structure [1], the organization technique for hierarchical perceptual sounds [2], and a method based on auditory scene analysis [3] have been proposed. On the other hand, as multichannel type source separation, the method based on array signal processing, for example, a microphone array system, is one of the most effective techniques [4]. 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 and the adaptive beamformer (ABF) are the conventional and popular microphone arrays currently used for source separation and noise reduction.

For high-quality acquisition of audible signals, several microphone array systems based on the DS array have been implemented since the 1980s. The most successful example was proposed by Flanagan et al. [5] for a speech pickup in auditoriums, in which a two-dimensional array composed of 63 microphones is used with automatic steering to enable detection and location of the desired signal source at any given moment. Recently, many microphone array systems with talker localization have been implemented for hands-free telecommunications or speech recognition [6, 7, 8]. While the DS array has a simple structure, it requires, however, a large number of microphones to achieve high performance, particularly in low-frequency regions. Thus, the degradation of separated signals at low frequencies cannot be avoided in these array systems.

In order to further improve the performance using more efficient methods than the DS array, the ABF has been introduced for acoustic signals analogously to an adaptive array antenna in radar systems [9, 10, 11]. The goal of the adaptive algorithm is to search for optimum directions of the nulls under the specific constraint that the desired signal arriving from the look direction is not significantly distorted. This method can improve the signal-separation performance even with a small array in comparison to that of the DS array. The ABF, however, has the following drawbacks. (1) The look direction for each signal which is 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 of the target signal to avoid any distortion of separated signals. However, in conventional use, we cannot estimate signal breaks in advance. The above-mentioned 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 the 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, that is, information-geometry theory [12, 13]. Blind source separation (BSS) is the approach to estimate 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 [13] and provides us with extended flexibility in which the source-separation procedure requires no training sequences and no a priori information on DOAs of the sound sources. The early contributory works on the BSS have been performed by Cardoso and Jutten [14, 15], where high-order statistics of the signals are used for measuring the independence. Comon [16] has clearly defined the term *independent component analysis* (ICA) and presented an algorithm that measures independence among the source signals. The ICA was later followed by Bell and Sejnowski [17], and was extended to the informax (or the maximum-entropy) algorithm for BSS which is based on a minimization of mutual information of the signals. In recent works on the ICA-based BSS, several methods, in which the complex-valued unmixing matrices are calculated in the frequency domain, have been proposed to deal with the arriving lags among each element of the microphone array system [18, 19, 20, 21]. Since the calculations are carried out at each frequency independently, the following problems arise in these methods: (1) permutation of each sound source, and (2) arbitrariness of each source gain. Various methods to overcome the permutation and scaling problems have been proposed. For example, a priori assumption of similarity among the envelopes of source signal waveforms [19] or interfrequency continuity with respect to the unmixing matrices [18, 20, 21] is necessary to resolve these problems.

In this paper, a new method of BSS on a microphone array using the subband ICA and beamforming is proposed. The proposed array system consists of the following three sections: (1) subband ICA section, (2) null beamforming section, and (3) integration of (1) and (2). First, a new subband ICA is introduced to achieve frequency domain BSS on the microphone array system, where directivity patterns of the array are explicitly used to estimate each DOA of the sound sources [22]. Using this method, we can resolve both permutation and arbitrariness problems simultaneously without the assumption for the source signal waveforms or interfrequency continuity of the unmixing matrices. Next, based on the DOA estimated in the above-mentioned ICA section, we construct a null beamformer in which the directional null is steered to the direction of the undesired sound source, in parallel with the ICA-based BSS. This approach to signal separation has the advantage that there is no difficulty with respect to a low convergence of optimization because the null beamformer is determined by only DOA information without independence between sound sources. Finally, both signal separation procedures are appropriately integrated by the algorithm diversity in the frequency domain [23].

In order to evaluate the effectiveness of the proposed method, both signal-separation and speech-recognition experiments are performed under various reverberant conditions. The results reveal that the performance of the proposed method is superior to that of the conventional ICA-based BSS method [19], and we also show that the proposed method did not cause heavy degradations of the separation

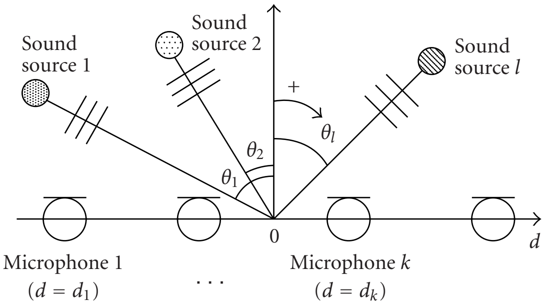


Figure 1: Configuration of a microphone array and signals.

performance compared with those of the previous ICA-based BSS method, particularly when the durations of the observed signals are exceedingly short. In addition, the speech-recognition experiment clarifies that the proposed method is more applicable to the recognition task in multispeaker cases than the conventional BSS.

The rest of this paper is organized as follows. In Sections 2 and 3, the formulation of the general BSS problems and the principle of the proposed method are explained. In Section 4, the signal-separation experiments are described. Following a discussion on the results of the experiments, we give the conclusions in Section 5.

2. SOUND MIXING MODEL OF MICROPHONE ARRAY

In this study, a straight-line array is assumed. The coordinates of the elements are designated as d_k $(k = 1, \ldots, K)$ and the DOAs of multiple sound sources are designated as θ_l $(l = 1, \ldots, L)$ (see Figure 1).

In general, the observed signals in which multiple source signals are mixed linearly are given by the following equation in the frequency domain:

$$\mathbf{X}(f) = \mathbf{A}(f)\mathbf{S}(f), \tag{1}$$

where $\mathbf{X}(f)$ is the observed signal vector, $\mathbf{S}(f)$ is the source signal vector, and $\mathbf{A}(f)$ is the mixing matrix. These are given as

$$\mathbf{X}(f) = \left[X_1(f), \ldots, X_K(f)\right]^{\mathrm{T}}, \tag{2}$$

$$\mathbf{S}(f) = \left[S_1(f), \ldots, S_L(f)\right]^{\mathrm{T}}, \tag{3}$$

$$\mathbf{A}(f) = \begin{bmatrix} A_{11}(f) & \cdots & A_{1L}(f) \\ \vdots & & \vdots \\ A_{K1}(f) & \cdots & A_{KL}(f) \end{bmatrix}. \tag{4}$$

We introduce the model to deal with the arriving lags among each of the elements of the microphone array. In this case, $A_{kl}(f)$ is assumed to be complex valued. Hereafter, for convenience, we only consider the relative lags among each of the elements with respect to the arrival time of the wavefront of each sound source, and neglect the pure delay between the microphone and sound source. Also, $\mathbf{S}(f)$ is identically regarded as the source signals observed at the origin. For example, by neglecting the effect of the room reverberation, we can rewrite the elements in the mixing matrix (4) as the following simple expression:

$$A_{kl}(f) = \exp\left(j2\pi f \tau_{kl}\right), \quad \left(\tau_{kl} \equiv \frac{1}{c} d_k \sin\theta_l\right), \tag{5}$$

where τ_{kl} is the arriving lag with respect to the lth source signal from the direction of θ_l, observed at the kth microphone at the coordinate of d_k. Also, c is the velocity of sound. If the effect of room reverberation is considered, the elements in the mixing matrix $A_{kl}(f)$ are given by more complicated values depending on the room reflections.

3. ALGORITHM

3.1. System overview of the proposed method

This section describes a new BSS method, using a microphone array, and its algorithm. The proposed array system consists of the following three sections (see Figure 2 for the system configuration): (1) subband ICA section for ICA-based BSS and DOA estimation, (2) null beamforming section for efficient reduction of directional interference signals, and (3) integration of (1) and (2) based on the *algorithm diversity* [23], selecting the most appropriate algorithm from (1) and (2) in the frequency domain. The following sections describe each of the procedures in detail.

3.2. Subband ICA section

3.2.1. Estimation on unmixing matrix

In this study, we perform the signal-separation procedure as described below (see Figure 3), where we deal with the case in which the number of sound sources L equals that of microphones K, that is, $K = L$. First, the short-time analysis of the observed signals is conducted by using discrete Fourier transform (DFT) frame by frame. By plotting the spectral values in a frequency bin of one microphone input, frame by frame, we consider them as a time series. The other inputs at the same frequency bin are dealt with in the same manner. Hereafter, we designate the time series as $\mathbf{X}(f, t) = [X_1(f, t), \ldots, X_K(f, t)]^{\mathrm{T}}$. Next, we perform signal separation by using the complex-valued unmixing matrix $\mathbf{W}(f)$ so that the L time series output $\mathbf{Y}(f, t)$ becomes mutually independent; this procedure can be given as

$$\mathbf{Y}(f, t) = \mathbf{W}(f)\mathbf{X}(f, t), \tag{6}$$

where

$$\mathbf{Y}(f, t) = \left[Y_1(f, t), \ldots, Y_L(f, t)\right]^{\mathrm{T}},$$
$$\mathbf{W}(f) = \begin{bmatrix} W_{11}(f) & \cdots & W_{1K}(f) \\ \vdots & & \vdots \\ W_{L1}(f) & \cdots & W_{LK}(f) \end{bmatrix}. \tag{7}$$

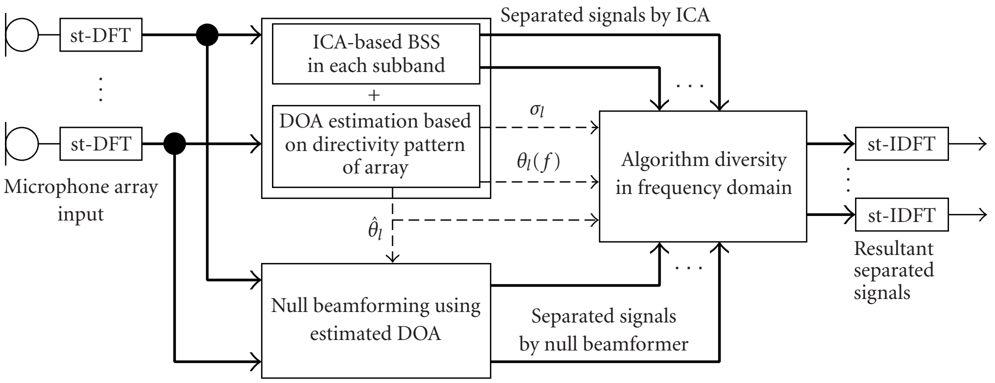

st-DFT
st-DFT
Microphone array input
ICA-based BSS in each subband
+
DOA estimation based on directivity pattern of array
Separated signals by ICA
σ_l
$\theta_l(f)$
$\hat{\theta}_l$
Algorithm diversity in frequency domain
st-IDFT
st-IDFT
Resultant separated signals
Null beamforming using estimated DOA
Separated signals by null beamformer

Figure 2: Configuration of the proposed microphone array system based on subband ICA and beamforming. Here, $\hat{\theta}_l$, $\theta_l(f)$, and σ_l represent estimated DOA of lth sound source, DOA of lth sound source at each frequency f, and deviation with respect to the estimated DOA of lth sound source, respectively. The bold arrows indicate the subband-signal lines. Here "st-DFT" represents the short time DFT.

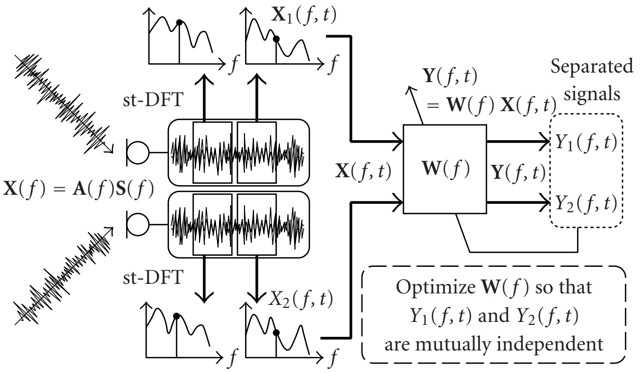


Figure 3: BSS procedure performed in subband ICA section. Here "st-DFT" represents the short time DFT.

We perform this procedure with respect to all frequency bins. Finally, by applying the inverse DFT and the overlap-add technique to the separated time series $\mathbf{Y}(f,t)$, we reconstruct the resultant source signals in the time domain.

Considering the calculation of the unmixing matrix $\mathbf{W}(f)$, we use the optimization algorithm based on the minimization of the Kullback-Leibler divergence; this algorithm has been introduced by Murata and Ikeda for online learning [19] and modified by the authors for offline learning with stable convergence. The optimal $\mathbf{W}(f)$ is obtained by using the following iterative equation:

$$\begin{aligned}\mathbf{W}_{i+1}(f) = \eta\Big[&\operatorname{diag}\left(\langle \mathbf{\Phi}(\mathbf{Y}(f,t))\mathbf{Y}^{\mathrm{H}}(f,t)\rangle_t\right) \\ &- \langle \mathbf{\Phi}(\mathbf{Y}(f,t))\mathbf{Y}^{\mathrm{H}}(f,t)\rangle_t\Big]\left(\mathbf{W}_i^{\mathrm{H}}(f)\right)^{-1} \\ &+ \mathbf{W}_i(f),\end{aligned} \tag{8}$$

where H denotes the Hermitian and $\langle\cdot\rangle_t$ denotes the time-averaging operator, i is used to express the value of the ith step in the iterations, and η is the step size parameter. Also, we define the nonlinear vector function $\mathbf{\Phi}(\cdot)$ as

$$\begin{aligned}\mathbf{\Phi}(\mathbf{Y}(f,t)) &\equiv \left[\Phi(Y_1(f,t)),\ldots,\Phi(Y_L(f,t))\right]^{\mathrm{T}}, \\ \Phi(Y_l(f,t)) &\equiv \left[1+\exp\left(-Y_l^{(\mathrm{R})}(f,t)\right)\right]^{-1} \\ &\quad + j\cdot\left[1+\exp\left(-Y_l^{(\mathrm{I})}(f,t)\right)\right]^{-1},\end{aligned} \tag{9}$$

where $Y_l^{(\mathrm{R})}(f,t)$ and $Y_l^{(\mathrm{I})}(f,t)$ are the real and imaginary parts of $Y_l(f,t)$, respectively.

3.2.2. Source permutation and gain arbitrariness problems and their solutions

This section describes the problems which arise after the signal separation described in Section 3.2.1, and solutions for these problems are newly proposed. Hereafter, we assume a two-channel model without loss of generality, that is, $K = L = 2$.

We assume that the following separation has been completed at frequency bin f:

$$\begin{bmatrix}\hat{S}_1(f,t)\\ \hat{S}_2(f,t)\end{bmatrix} = \begin{bmatrix}W_{11}(f) & W_{12}(f)\\ W_{21}(f) & W_{22}(f)\end{bmatrix}\begin{bmatrix}X_1(f,t)\\ X_2(f,t)\end{bmatrix}, \tag{10}$$

where $\hat{S}_1(f,t)$ and $\hat{S}_2(f,t)$ are the components of the estimated source signals. Since the above calculations are carried out at each frequency bin independently, the following two problems arise (see Figure 4).

Problem 1. The permutation of the source signals $\hat{S}_1(f,t)$ and $\hat{S}_2(f,t)$ arises. That is, the separated signal components can be permuted at every frequency bin, for example, at a frequency bin of $f = f_1$, $\hat{S}_1(f_1,t) = S_1(f_1,t)$, and $\hat{S}_2(f_1,t) = S_2(f_1,t)$, and at another frequency bin of $f = f_2$, $\hat{S}_1(f_2,t) = S_2(f_2,t)$, and $\hat{S}_2(f_2,t) = S_1(f_2,t)$.

Problem 2. The gains of $\hat{S}_1(f,t)$ and $\hat{S}_2(f,t)$ are arbitrary. That is, different gains are obtained at different frequency bins $f = f_1$ and $f = f_2$.

In order to resolve Problems 1 and 2, we focus on the mechanism of the BSS as array signal processing to obtain the separated signals in the acoustical space. For example, from (10), $\hat{S}_1(f,t)$ is given by

$$\hat{S}_1(f,t) = W_{11}(f)X_1(f,t) + W_{12}(f)X_2(f,t). \tag{11}$$

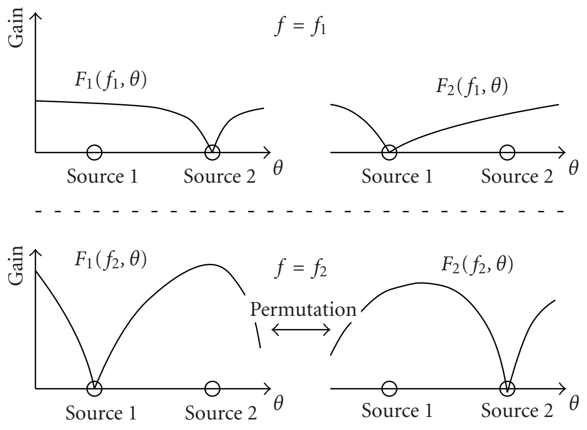


Figure 4: Examples of directivity patterns.

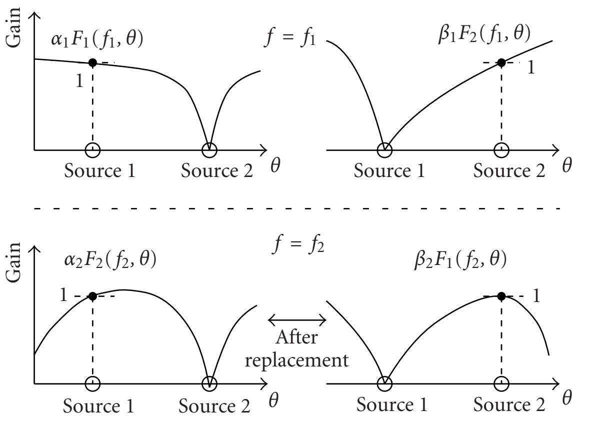

Gain
$\alpha_1 F_1(f_1, \theta)$
$f = f_1$
$\beta_1 F_2(f_1, \theta)$
1
Source 1
Source 2
θ
Gain
$\alpha_2 F_2(f_2, \theta)$
$f = f_2$
$\beta_2 F_1(f_2, \theta)$
After replacement
Source 1
Source 2
θ

Figure 5: Resultant directivity patterns after recovery of permutations and normalization of gains of separated signals.

This equation shows that the resultant output signals are obtained by multiplying the array signals of $X_1(f,t)$ and $X_2(f,t)$ by the weight $W_{lk}(f)$, and then adding them. Thus, from the standpoint of array signal processing, this operation implies that directivity patterns are produced in the array system. Accordingly, we calculate directivity patterns with respect to $W_{lk}(f)$ obtained at every frequency bin. The directivity pattern $F_l(f,\theta)$ is given by [24]

$$F_l(f,\theta) = \sum_{k=1}^{2} W_{lk}(f) \cdot \exp\left[j2\pi f d_k \sin\theta/c\right]. \tag{12}$$

This equation shows that the lth directivity pattern $F_l(f,\theta)$ is produced to extract the lth source signal. Using the directivity pattern $F_l(f,\theta)$, we propose the following procedure to resolve Problems 1 and 2.

Step 1. We plot the directivity patterns in all frequency bins; for example, in the frequency bins of f_1 and f_2, directivity patterns are plotted as shown in Figure 4.

Step 2. In the directivity patterns, directional nulls exist in only two particular directions and these nulls represent DOAs of the sound sources. Accordingly, by obtaining statistics with respect to the directions of nulls at all frequency bins, we can estimate the DOAs of the sound sources. The DOA of the lth sound source, $\hat{\theta}_l$, can be estimated as

$$\hat{\theta}_l = \frac{2}{N}\sum_{m=1}^{N/2} \theta_l(f_m), \tag{13}$$

where N is a total point of DFT and $\theta_l(f_m)$ represents the DOA of the lth sound source at the mth frequency bin. These are given by

$$\begin{aligned}\theta_1(f_m) &= \min\left[\arg\min_{\theta} |F_1(f_m,\theta)|, \arg\min_{\theta} |F_2(f_m,\theta)|\right],\\ \theta_2(f_m) &= \max\left[\arg\min_{\theta} |F_1(f_m,\theta)|, \arg\min_{\theta} |F_2(f_m,\theta)|\right],\end{aligned} \tag{14}$$

where $\min[x,y]$ ($\max[x,y]$) is defined as a function in order to obtain the smaller (larger) value among x and y.

Step 3. From these directivity patterns in all frequency bins, we collect the specific ones in which the directional null is steered to the directions of $\hat{S}_1(f,t)$. Also, we collect the other specific directivity patterns in which the directional null is steered to the directions of $\hat{S}_2(f,t)$. Here, we decide to collect the directivity patterns in which the null is steered to the direction of $\hat{S}_1(f,t)$ ($\hat{S}_2(f,t)$) on the right-(left-)hand side of Figure 5. From this constraint, we replace $F_1(f_2,\theta)$ with $F_2(f_2,\theta)$ at the frequency bin of $f = f_2$. By performing this procedure, we can resolve Problem 1.

Step 4. Problem 2 is resolved by normalizing the directivity patterns according to the gain in each source direction after the classification (see Figure 5). In Figure 5, α_1 and α_2 are the constants which normalize the gain in the direction of $\hat{S}_1(f,t)$, and β_1 and β_2 are the constants which normalize the gain in the direction of $\hat{S}_2(f,t)$.

By applying the above-mentioned modifications, we can finally obtain the unmixing matrix in the ICA section, $\mathbf{W}^{(\mathrm{ICA})}(f)$, as follows:

$$\begin{aligned}\mathbf{W}^{(\mathrm{ICA})}(f_m) &\equiv \begin{bmatrix} W_{11}^{(\mathrm{ICA})}(f_m) & W_{12}^{(\mathrm{ICA})}(f_m)\\ W_{21}^{(\mathrm{ICA})}(f_m) & W_{22}^{(\mathrm{ICA})}(f_m)\end{bmatrix}\\ &= \begin{cases} \begin{bmatrix} 1/F_1(f_m,\hat{\theta}_1) & 0\\ 0 & 1/F_2(f_m,\hat{\theta}_2)\end{bmatrix} \cdot \mathbf{W}(f_m), \\ \text{(without permutation)}, \\ \begin{bmatrix} 0 & 1/F_2(f_m,\hat{\theta}_1)\\ 1/F_1(f_m,\hat{\theta}_2) & 0\end{bmatrix} \cdot \mathbf{W}(f_m), \\ \text{(with permutation)}.\end{cases}\end{aligned} \tag{15}$$

3.3. Beamforming section

In the beamforming section, we can construct an alternative unmixing matrix in parallel, based on the null beamforming technique where the DOA information obtained in the ICA section is used. In the case that the look direction is $\hat{\theta}_1$ and the directional null is steered to $\hat{\theta}_2$, the elements of the unmixing matrix, $W_{1k}^{(\mathrm{BF})}(f_m)$, satisfy the following simultaneous equations:

$$\begin{aligned} F_1(f_m, \hat{\theta}_1) &= \sum_{k=1}^{2} W_{1k}^{(\mathrm{BF})}(f_m) \cdot \exp\left[\frac{j2\pi f_m d_k \sin\hat{\theta}_1}{c}\right] = 1, \\ F_1(f_m, \hat{\theta}_2) &= \sum_{k=1}^{2} W_{1k}^{(\mathrm{BF})}(f_m) \cdot \exp\left[\frac{j2\pi f_m d_k \sin\hat{\theta}_2}{c}\right] = 0. \end{aligned} \tag{16}$$

The solutions of the equations are given by

$$\begin{aligned} W_{11}^{(\mathrm{BF})}(f_m) &= -\exp\left[\frac{-j2\pi f_m d_1 \sin\hat{\theta}_2}{c}\right] \\ &\quad \times \left\{ -\exp\left[\frac{j2\pi f_m d_1 (\sin\hat{\theta}_1 - \sin\hat{\theta}_2)}{c}\right] \right. \\ &\qquad \left. + \exp\left[\frac{j2\pi f_m d_2 (\sin\hat{\theta}_1 - \sin\hat{\theta}_2)}{c}\right] \right\}^{-1}, \\ W_{12}^{(\mathrm{BF})}(f_m) &= \exp\left[\frac{-j2\pi f_m d_2 \sin\hat{\theta}_2}{c}\right] \\ &\quad \times \left\{ -\exp\left[\frac{j2\pi f_m d_1 (\sin\hat{\theta}_1 - \sin\hat{\theta}_2)}{c}\right] \right. \\ &\qquad \left. + \exp\left[\frac{j2\pi f_m d_2 (\sin\hat{\theta}_1 - \sin\hat{\theta}_2)}{c}\right] \right\}^{-1}. \end{aligned} \tag{17}$$

Also in the case that the look direction is $\hat{\theta}_2$ and the directional null is steered to $\hat{\theta}_1$, the elements of the unmixing matrix, $W_{2k}^{(\mathrm{BF})}(f_m)$, satisfy the following simultaneous equations:

$$\begin{aligned} F_2(f_m, \hat{\theta}_2) &= \sum_{k=1}^{2} W_{2k}^{(\mathrm{BF})}(f_m) \cdot \exp\left[\frac{j2\pi f_m d_k \sin\hat{\theta}_2}{c}\right] = 1, \\ F_2(f_m, \hat{\theta}_1) &= \sum_{k=1}^{2} W_{2k}^{(\mathrm{BF})}(f_m) \cdot \exp\left[\frac{j2\pi f_m d_k \sin\hat{\theta}_1}{c}\right] = 0. \end{aligned} \tag{18}$$

The solutions of the equations are given by

$$\begin{aligned} W_{21}^{(\mathrm{BF})}(f_m) &= \exp\left[\frac{-j2\pi f_m d_1 \sin\hat{\theta}_1}{c}\right] \\ &\quad \times \left\{ \exp\left[\frac{j2\pi f_m d_1 (\sin\hat{\theta}_2 - \sin\hat{\theta}_1)}{c}\right] \right. \\ &\qquad \left. - \exp\left[\frac{j2\pi f_m d_2 (\sin\hat{\theta}_2 - \sin\hat{\theta}_1)}{c}\right] \right\}^{-1}, \\ W_{22}^{(\mathrm{BF})}(f_m) &= -\exp\left[\frac{-j2\pi f_m d_2 \sin\hat{\theta}_1}{c}\right] \\ &\quad \times \left\{ \exp\left[\frac{j2\pi f_m d_1 (\sin\hat{\theta}_2 - \sin\hat{\theta}_1)}{c}\right] \right. \\ &\qquad \left. - \exp\left[\frac{j2\pi f_m d_2 (\sin\hat{\theta}_2 - \sin\hat{\theta}_1)}{c}\right] \right\}^{-1}. \end{aligned} \tag{19}$$

These unmixing matrices are approximately optimal for the signal separation when the ideal far-field propagation is only considered and the effect of the room reverberation is negligible. However, these acoustic conditions are oversimplified. In contrast, the optimality cannot hold under reverberant conditions because the signal reduction cannot be achieved by the directional nulls only. This signal-separation approach, however, has the advantage that there is no difficulty with respect to a low-convergence of optimization because the null beamformer is determined by DOA information only without independence between sound sources. The effectiveness of the null beamforming will appear especially when we combine the beamforming and ICA as described in the next section.

3.4. Integration of subband ICA with null beamforming

In order to integrate the subband ICA with null beamforming, we introduce the following strategy for selecting the most suitable unmixing matrix in each frequency bin, that is, algorithm diversity in the frequency domain. If the directional null is steered to the proper estimated DOA of the undesired sound source, we use the unmixing matrix obtained by the subband ICA, $W_{lk}^{(\mathrm{ICA})}(f)$. If the directional null deviates from the estimated DOA, we use the unmixing matrix obtained by the null beamforming, $W_{lk}^{(\mathrm{BF})}(f)$, in preference to that of the subband ICA. The above strategy yields the following algorithm:

$$W_{lk}(f) = \begin{cases} W_{lk}^{(\mathrm{ICA})}(f), & \left(\left|\theta_l(f) - \hat{\theta}_l\right| < h \cdot \sigma_l\right), \\ W_{lk}^{(\mathrm{BF})}(f), & \left(\left|\theta_l(f) - \hat{\theta}_l\right| \geq h \cdot \sigma_l\right), \end{cases} \tag{20}$$

where h is a magnification parameter of the threshold and σ_l represents the deviation with respect to the estimated DOA of the lth sound source; it can be given as

$$\sigma_l = \sqrt{\frac{2}{N} \sum_{m=1}^{N/2} \left(\theta_l(f_m) - \hat{\theta}_l\right)^2}. \tag{21}$$

Using the algorithm with an adequate value of h, we can recover the unmixing matrix trapped on a local minimizer of the optimization procedure in ICA. Also, by changing the parameter h, we can construct various types of array signal processing for BSS, for example, a simple null beamforming with $h = 0$ and a simple ICA-based BSS procedure with $h = \infty$.

By substituting $\mathbf{W}(f)$ after performing the above-mentioned modification for (10) and applying inverse DFT to the outputs $\hat{S}_1(f, t)$ and $\hat{S}_2(f, t)$, we can obtain the source signals correctly.

4. EXPERIMENTS AND RESULTS

Signal-separation experiments are conducted using the sound data convolved with the impulse responses recorded in two environments specified by different reverberation times (RTs). In these experiments, we investigated the performance

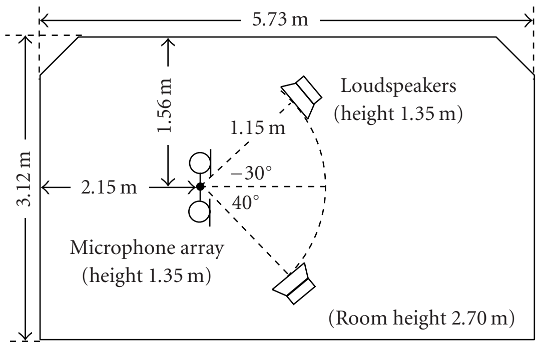


Figure 6: Layout of reverberant room used in experiments.

of separation under different reverberant conditions from two standpoints: an objective evaluation of separated speech quality and a word recognition test.

4.1. Conditions for experiments

A two-element array with the interelement spacing of 4 cm is assumed. We determined this interelement spacing by considering that the spacing should be smaller than half the minimum wavelength to avoid the spatial aliasing effect; it corresponds to 8.5/2 cm in 8 kHz sampling. The speech signals are assumed to arrive from two directions: −30° and 40°. Six sentences spoken by six male and six female speakers selected from the ASJ continuous speech corpus for research [25] are used as the original speech. Using these sentences, we obtain 36 combinations with respect to speakers and source directions. In these experiments, we used the following signals as the source signals: (1) the original speech not convolved with the room impulse responses (only considering the arrival lags among microphones) and (2) the original speech convolved with the room impulse responses recorded in the two environments specified by the different RTs. Hereafter, we designate the experiments using the signals described in (1) as the nonreverberant tests, and those of (2) as the reverberant tests. The impulse responses are recorded in a variable RT room as shown in Figure 6. The RTs of the impulse responses recorded in the room are 150 milliseconds and 300 milliseconds, respectively. These sound data which are artificially convolved with the real impulse responses have the following advantages. (1) We can use the realistic mixture model of two sources neglecting the affection of background noise. (2) Since the mixing condition is explicitly measured, we can easily calculate a reliable objective score to evaluate the separation performance as described in Section 4.2. The analysis conditions of these experiments are summarized in Table 1.

4.2. Objective evaluation score

Noise reduction rate (NRR), defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB, is used as the objective 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

Table 1: Analysis conditions of signal separation.

Sampling frequency	8 kHz
Frame length	32 ms
Frame shift	16 ms
Window	Hamming window
Number of iterations	500
Step size parameter	$\eta = 1.0 \times 10^{-4}$

$$\begin{aligned} \mathrm{NRR} &\equiv \frac{1}{2}\sum_{l=1}^{2}\left(\mathrm{SNR}_l^{(\mathrm{O})} - \mathrm{SNR}_l^{(\mathrm{I})}\right), \\ \mathrm{SNR}_l^{(\mathrm{O})} &= 10\log_{10}\frac{\sum_f \left|H_{ll}(f)S_l(f)\right|^2}{\sum_f \left|H_{ln}(f)S_n(f)\right|^2}, \\ \mathrm{SNR}_l^{(\mathrm{I})} &= 10\log_{10}\frac{\sum_f \left|A_{ll}(f)S_l(f)\right|^2}{\sum_f \left|A_{ln}(f)S_n(f)\right|^2}, \end{aligned} \tag{22}$$

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, $H_{ij}(f)$ is the element in the ith row and the jth column of the matrix $\mathbf{H}(f) = \mathbf{W}(f)\mathbf{A}(f)$, where the mixing matrix $\mathbf{A}(f)$ corresponds to the frequency-domain representation of the room impulse responses described in Section 4.1.

4.3. Alternative method for comparison

In order to perform a comparison with the proposed method, we also performed a BSS experiment using the alternative method proposed by Murata and Ikeda [19] with the modification for offline learning.

Our proposed method is based on the utilization of directivity patterns; in contrast, Murata's method is based on the utilization of $\mathbf{W}^{-1}(f)$ for the normalization of gain and the a priori assumption of similarity among the envelopes of source signal waveforms for the recovery of the source permutation. In this method, the following operations are performed:

$$\begin{aligned} \mathbf{Z}(f,t) &= \left[Z_1(f,t),\ldots,Z_L(f,t)\right]^{\mathrm{T}} = \mathbf{W}(f)\mathbf{X}(f,t), \\ \tilde{\mathbf{S}}_l(f,t) &= \mathbf{W}^{-1}(f)\left[0,\ldots,0,Z_l(f,t),0,\ldots,0\right]^{\mathrm{T}}, \end{aligned} \tag{23}$$

where $\tilde{\mathbf{S}}_l(f,t)$ denotes the component of the lth estimated source signal in the frequency bin of f. By using both $\mathbf{W}(f)$ and $\mathbf{W}^{-1}(f)$, the gain arbitrariness vanishes in the separation procedure. Also, the source permutation can be detected and recovered by measuring the similarity among the envelopes of $\tilde{\mathbf{S}}_l(f,t)$ between the different frequency bins.

4.4. Objective evaluation of separated signal

In order to illustrate the behavior of the proposed array for different values of h, the NRR is shown in Figures 7, 8, and 9. These values are taken as the average of all of the combinations with respect to speakers and source directions.

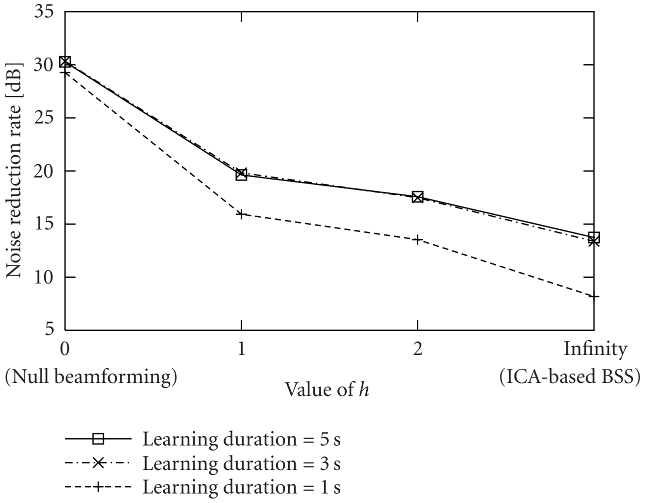


FIGURE 7: Noise reduction rates for different values of threshold parameter h. Reverberation time is 0 milliseconds.

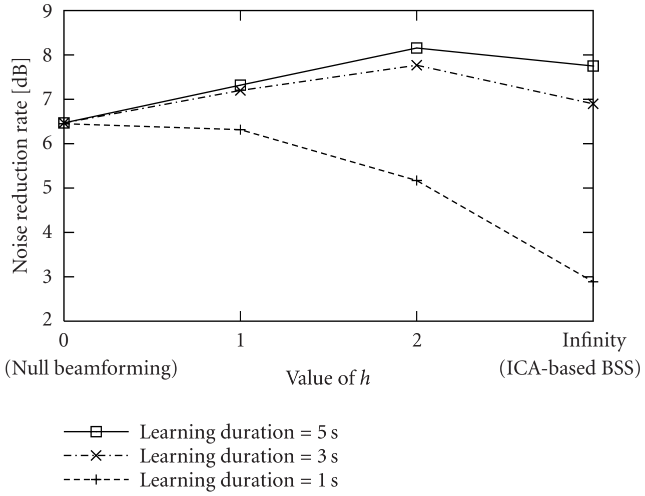


FIGURE 8: Noise reduction rates for different values of threshold parameter h. Reverberation time is 150 milliseconds.

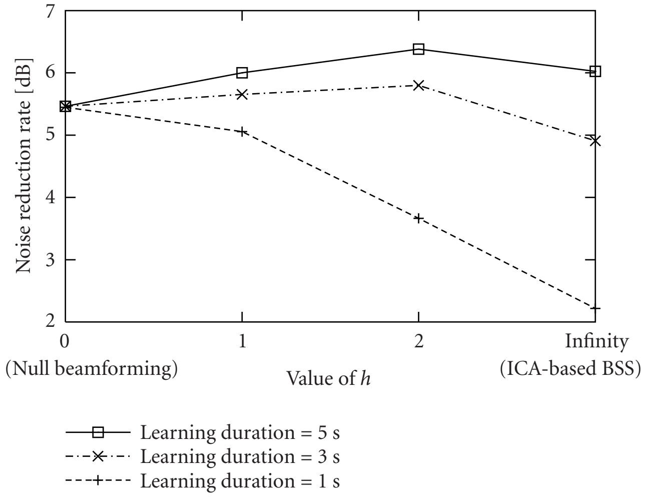


FIGURE 9: Noise reduction rates for different values of threshold parameter h. Reverberation time is 300 milliseconds.

From Figure 7, for the nonreverberant tests, it can be seen that the NRRs monotonically increase as the parameter h decreases, that is, the performance of the null beamformer is superior to that of ICA-based BSS. This indicates that the directions of the sound sources are estimated correctly by the proposed method, and thus the null beamforming technique is more suitable for the separation of directional sound sources under nonreverberant condition.

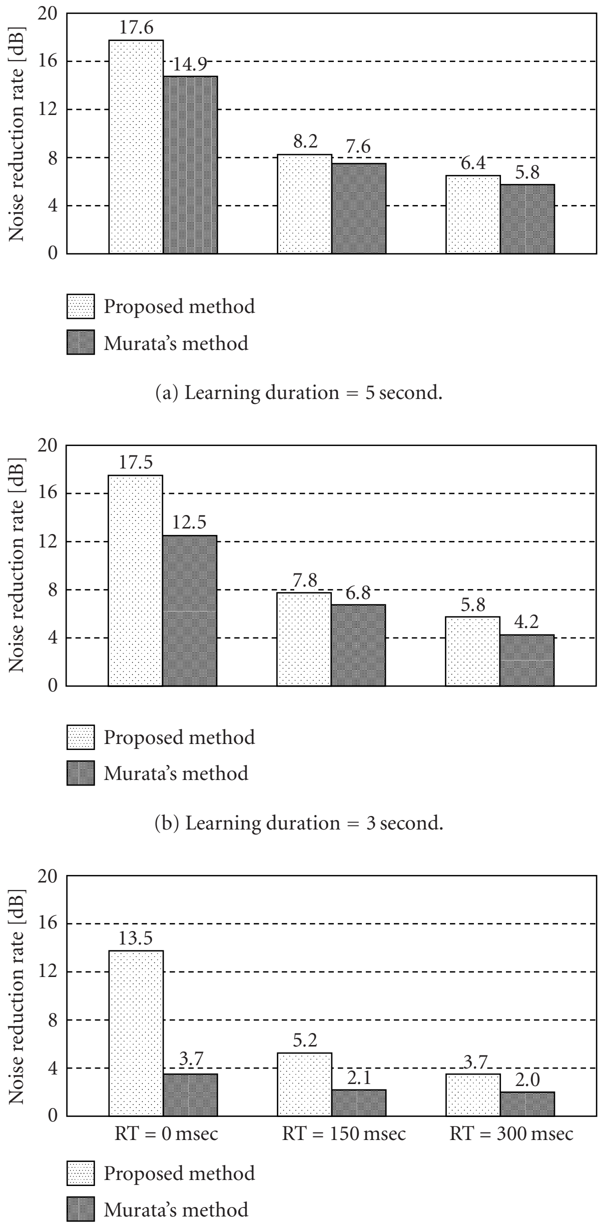
In contrast, from Figures 8 and 9, for the reverberant tests, it is shown that the NRR monotonically increases as the parameter h decreases in the case that the observed signals of 1 second duration are used to learn the unmixing matrix, and we can obtain the optimum performances by setting the appropriate value of h, for example, $h = 2$, in the case that the learning durations are 3 seconds and 5 seconds. We can summarize from these results that the proposed combination algorithm of ICA and null beamforming is effective for the signal separation, particularly under the reverberant conditions.

In order to perform a comparison with the conventional BSS method, we also perform the same BSS experiments using Murata's method as described in Section 4.3. Figure 10a shows the results obtained using the proposed method and Murata's method where the observed signals of 5 second duration are used to learn the unmixing matrix, Figure 10b shows those of 3 second duration, and Figure 10c shows those of 1 second duration. In these experiments, the parameter h in the proposed method is set to be 2.

From Figure 10, in both nonreverberant and reverberant tests, it can be seen that the BSS performances obtained by using the proposed method are the same as or superior to those of Murata's conventional method. In particular, from Figure 10c, it is evident that the NRRs of Murata's method degrade markedly in the case that the learning duration is 1 second; however, there are no significant degradations in the case of the proposed method compared with those of Murata's method. By looking at the similarity, for example, *frequency-averaged cosine distance* defined by

$$\frac{2}{N}\sum_{m=1}^{N/2}\frac{\left|\left\langle Y_1(f_m,t)Y_2(f_m,t)^*\right\rangle_t\right|}{\left\langle\left|Y_1(f_m,t)\right|^2\right\rangle_t^{1/2}\left\langle\left|Y_2(f_m,t)\right|^2\right\rangle_t^{1/2}}, \tag{24}$$

among the source signals of different lengths, we can summarize the main reasons for the degradations in Murata's method as follows (see Figure 11). (1) The envelopes of the original source speech become more similar to each other as the duration of the speech shortens. (2) The separated signals' envelopes at the same frequency are similar to each other since the inaccurate unmixing matrix is estimated to have many components of crosstalk. Therefore, the recovery of the permutation tends to fail in Murata's method. In contrast, our method did not fail to recover the source



(a) Learning duration = 5 second.

(b) Learning duration = 3 second.

(c) Learning duration = 1 second.

Figure 10: Comparison of noise reduction rates obtained by the proposed method ($h = 2$) and Murata's method in the case that the learning duration for ICA is (a) 5 seconds, (b) 3 seconds, and (c) 1 second.

permutation because we did not use any informations of signal waveforms, but rather used only the directivity patterns.

4.5. Word recognition test

The HMM continuous speech recognition (CSR) experiment is performed in a speaker-dependent manner. For the CSR experiment, 10 sentences spoken by one speaker are used as test data, and the monophone HMM model is trained using 140 phonetically balanced sentences. Both test and training sets are selected from the ASJ continuous speech corpus for research. The remaining conditions are summarized in Table 2.

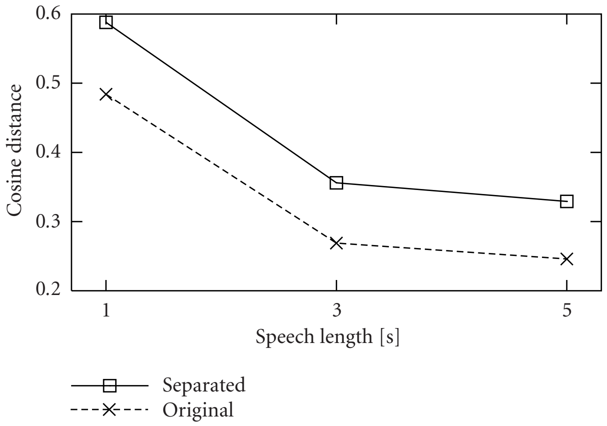


Figure 11: Cosine distances for different speech lengths. These values are the average of all of the frequency bins.

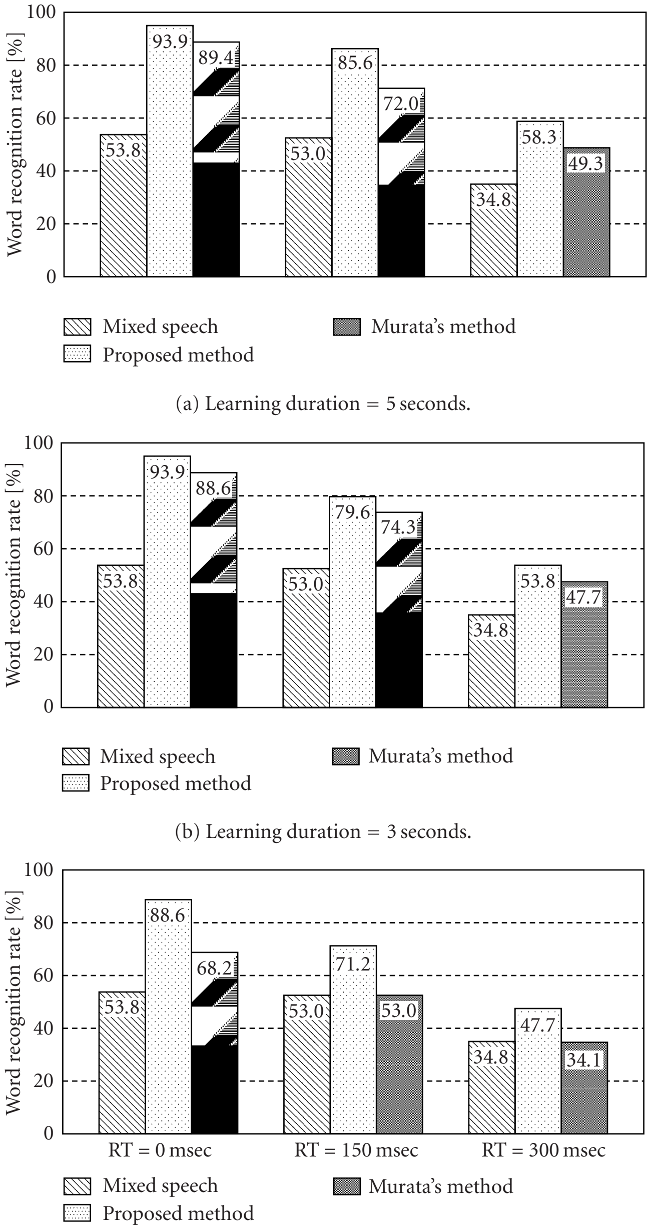
Table 2: Analysis conditions for CSR experiments.

Frame length	25 ms
Frame shift	10 ms
Window	Hamming window
Feature vector	12th order MFCC [26] + 12th order Δ MFCC + 12th order ΔΔ MFCC + ΔPOWER + ΔΔ POWER
Number of states	5
Vocabulary	68

Figure 12 shows the results in terms of word recognition rates under different reverberant conditions. Compared with the results of Murata's BSS method, it is evident that the improvements of the proposed method are superior to those of the conventional ICA-based BSS method under all conditions with respect to both reverberation and learning duration. These results indicate that the proposed method is applicable to the speech-recognition system, particularly when confronted with interfering speech signals.

5. CONCLUSION

In this paper, a new BSS method using subband ICA and beamforming was described. In order to evaluate its effectiveness, signal-separation and speech-recognition experiments were performed under various reverberant conditions. The signal-separation experiments with observed signals of sufficient duration reveal that the NRR of about 18 dB is obtained under the nonreverberant condition, and NRRs of 8 dB and 6 dB are obtained in the case that the RTs are 150 milliseconds and 300 milliseconds, respectively. These performances were superior to those of both simple ICA-based BSS and simple



(a) Learning duration = 5 seconds.

(b) Learning duration = 3 seconds.

(c) Learning duration = 1 seconds.

FIGURE 12: Comparison of word recognition rates obtained by the proposed method ($h = 2$) and Murata's method in the case that the learning duration for ICA is (a) 5 seconds, (b) 3 seconds, and (c) 1 second.

beamforming technique. Also, it was evident that the NRRs of Murata's ICA-based BSS method degrade markedly in the case that the learning duration is 1 second; however, there are no significant degradations in the case of the proposed method. From the speech-recognition experiments, compared with the results of Murata's BSS method, it was evident that the improvements of the proposed method are superior to those of Murata's BSS method under all conditions with respect to both reverberation and learning duration. These results indicate that the proposed method is applicable to the speech-recognition system, particularly when confronted with interfering speech signals.

In this paper, we mainly showed that the utilization of beamforming in ICA can improve the separation performance. As for the other application of beamforming to ICA, we have already presented a method [27] in which we are particularly concerned with the acceleration of convergence speed in the ICA learning. These results show the explicit evidence for the effectiveness of beamforming used in ICA framework; however, further study and development on the alternative combination technique between ICA and beamforming is an open problem.

ACKNOWLEDGMENT

This work was partly supported by a Grant in Aid for COE Research no. 11CE2005 and CREST (Core Research for Evolutional Science and Technology) in Japan.

REFERENCES

[1] T. W. Parsons, "Separation of speech from interfering speech by means of harmonic selection," *Journal of the Acoustical Society of America*, vol. 60, no. 4, pp. 911–918, 1976.

[2] K. Kashino, K. Nakadai, T. Kinoshita, and H. Tanaka, "Organization of hierarchical perceptual sounds," in *Proc. 14th International Joint Conference on Artificial Intelligence*, vol. 1, pp. 158–164, Montreal, Quebec, Canada, August 1995.

[3] M. Unoki and M. Akagi, "A method of signal extraction from noisy signal based on auditory scene analysis," *Speech Communication*, vol. 27, no. 3, pp. 261–279, 1999.

[4] G. W. Elko, "Microphone array systems for hands-free telecommunication," *Speech Communication*, vol. 20, no. 3-4, pp. 229–240, 1996.

[5] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1508–1518, 1985.

[6] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 187–190, Munich, Germany, April 1997.

[7] K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura, and J. Kojima, "A microphone array system for speech recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 215–218, Munich, Germany, April 1997.

[8] M. Omologo, M. Matassoni, P. Svaizer, and D. Giuliani, "Microphone array based speech recognition with different talker-array positions," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 227–230, Munich, Germany, April 1997.

[9] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.

[10] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.

[11] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 6, pp. 1391–1400, 1986.

[12] T.-W. Lee, *Independent Component Analysis: Theory and Applications*, Kluwer Academic Publishers, Boston, Mass, USA, 1998.

[13] S. Haykin, *Unsupervised Adaptive Filtering*, John Wiley & Sons, New York, NY, USA, 2000.

[14] J. F. Cardoso, "Eigenstructure of the 4th-order cumulant tensor with application to the blind source separation problem," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 2109–2112, Glasgow, Scotland, UK, May 1989.

[15] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.

[16] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[17] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[18] V. Capdevielle, C. Serviere, and J. Lacoume, "Blind separation of wide-band sources in the frequency domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 2080–2083, Detroit, Mich, USA, May 1995.

[19] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," in *Proc. International Symposium on Nonlinear Theory and Its Application*, vol. 3, pp. 923–926, Le Regent, Crans-Montana, Switzerland, September 1998.

[20] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, 1998.

[21] 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, 2000.

[22] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, pp. 3140–3143, Istanbul, Turkey, June 2000.

[23] Y. Karasawa, T. Sekiguchi, and T. Inoue, "The software antenna: a new concept of kaleidoscopic antenna in multimedia radio and mobile computing era," *IEICE Transaction on Communications*, vol. E80-B, no. 8, pp. 1214–1217, 1997.

[24] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

[25] T. Kobayashi, S. Itabashi, S. Hayashi, and T. Takezawa, "ASJ continuous speech corpus for research," *Journal of the Acoustical Society of Japan*, vol. 48, no. 12, pp. 888–893, 1992.

[26] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[27] H. Saruwatari, T. Kawamura, and K. Shikano, "Blind source separation based on fast-convergence algorithm using ICA and beamforming," in *Proc. IEEE/EURASIP International Workshop on Acoustic Echo and Noise Control*, pp. 119–122, Darmstadt, Germany, September 2001.

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

Satoshi Kurita received the B.E. and M.E. degrees in electrical engineering from Nagoya University in 1998 and 2000, respectively. His research interests include array signal processing and blind source separation. He is a member of the IEICE and the Acoustical Society of Japan.

Kazuya Takeda was born in Sendai, Japan in 1960. He received the B.E.E., M.E.E., and D.Eng. degrees, all from Nagoya University, in 1983, 1985, and 1994, respectively. In 1986, he joined ATR (Advanced Telecommunication Research Laboratories), where he was involved in the two major projects of speech database construction and speech synthesis system development. In 1989, he moved to KDD R & D Laboratories and participated in a project for constructing voice-activated telephone extension system. Since 1995, he has been working for Nagoya University. He is a leader of speech recognition group in CIAIR (Center for Integrated Acoustic Information Research).

Fumitada Itakura received the undergraduate and graduate degrees from Nagoya University in 1963 and 1965, respectively. In 1968, he joined the NTT's Musashino Electrical Communication Laboratory, Tokyo. He completed his D.Eng. degree in speech analysis and synthesis based on a statistical method in 1972. He worked at the Acoustics Research Department of Bell Labs under James Flanagan from 1973 to 1975. From 1975 to 1981, he researched problems in speech analysis and synthesis based on the line spectrum pair (LSP) method. In 1981, he was appointed as Chief of the Speech and Acoustics Research Section at NTT. He left this position in 1984 to take a professorship in communications theory and signal processing at Nagoya University. His awards include the IEEE ASSP 1975 Senior Award, an award from Japan's Ministry of Science and Technology in 1977, the IEEE 1986 Morris N. Liebmann Award (with B. S. Atal), the IEEE Signal Processing 1996 Society Award, the IEEE Third Millennium Medal, the IEICE 2002 Distinguished Achievement and Contributions Award, and the 2003 Purple Ribbon Medal from Japanese Government. He is a Fellow of the IEEE and the Institute of Electronics, Information and Communication Engineers of Japan, and a member of the Acoustical Society of Japan.

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

Kiyohiro Shikano received the B.S., M.S., and Ph.D. degrees in electrical engineering from Nagoya University in 1970, 1972, and 1980, respectively. He is currently a Professor at Nara Institute of Science and Technology (NAIST), where he is directing Speech and Acoustics Laboratory. His major research areas are speech recognition, multimodal dialog system, speech enhancement, adaptive microphone array, and acoustic field reproduction. From 1972 to 1993, he had been working at NTT Laboratories, where he had been engaged in speech recognition research. During 1986–1990, he was Head of Speech Processing Department at ATR Interpreting Telephony Research Laboratories. During 1984–1986, he was a Visiting Scientist in Carnegie Mellon University. He received the Yonezawa Prize from IEICE in 1975, the Signal Processing Society 1990 Senior Award from IEEE in 1991, the Technical Development Award from ASJ in 1994, IPSJ Yamashita SIG Research Award in 2000, and Paper Award from the Virtual Reality Society of Japan in 2001. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan (IEICE), Information Processing Society of Japan, the Acoustical Society of Japan (ASJ), Japan VR Society, the Institute of Electrical and Electronics, Engineers (IEEE), and International Speech Communication Society.