

Frequency-Domain Blind Source Separation of Many Speech Signals Using Near-Field and Far-Field Models

Ryo Mukai, Hiroshi Sawada, Shoko Araki, and Shoji Makino

NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai, Seika-Cho, Soraku-Gun, Kyoto 619-0237, Japan

Received 19 December 2005; Revised 26 April 2006; Accepted 11 June 2006

We discuss the frequency-domain blind source separation (BSS) of convolutive mixtures when the number of source signals is large, and the potential source locations are omnidirectional. The most critical problem related to the frequency-domain BSS is the permutation problem, and geometric information is helpful as regards solving it. In this paper, we propose a method for obtaining proper geometric information with which to solve the permutation problem when the number of source signals is large and some of the signals come from the same or a similar direction. First, we describe a method for estimating the absolute DOA by using relative DOAs obtained by the solution provided by independent component analysis (ICA) and the far-field model. Next, we propose a method for estimating the spheres on which source signals exist by using ICA solution and the near-field model. We also address another problem with regard to frequency-domain BSS that arises from the circularity of discrete-frequency representation. We discuss the characteristics of the problem and present a solution for solving it. Experimental results using eight microphones in a room show that the proposed method can separate a mixture of six speech signals arriving from various directions, even when two of them come from the same direction.

Copyright © 2006 Ryo Mukai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Blind source separation (BSS) [1, 2] is a technique for estimating original source signals using only observed mixtures. The BSS of audio signals has a wide range of applications including speech enhancement [3] for speech recognition, hands-free telecommunication systems, and high-quality hearing aids. Independent component analysis (ICA) [4–7] is one of the main statistical methods used for BSS. It is theoretically possible to solve the BSS problem with a large number of sources by ICA, if we assume that the number of sensors is equal to or greater than the number of source signals. However, there are many practical difficulties.

In most realistic audio applications, the signals are mixed in a convolutive manner with reverberations, and the separation system that we have to estimate is a matrix of filters, not just a matrix of scalars. Although many studies have been undertaken on BSS in a reverberant environment [8], most of them have assumed two source signals arriving from different directions, and only a few studies have dealt with more than two source signals.

There are two major approaches to solving the convolutive BSS problem. The first is the time-domain approach, where ICA is applied directly to the convolutive mixture

model [1, 9, 10, 12, 13]. Matsuoka et al. [11] have shown that time-domain ICA can solve the convolutive BSS problem of eight sources with eight microphones in a real environment. Unfortunately, the time-domain approach incurs considerable computational cost, and it is difficult to obtain a solution in a practical time.

The other approach is frequency-domain BSS, where ICA is applied to multiple instantaneous mixtures in the frequency domain [14–24]. This approach takes much less computation time than time-domain BSS. However, it poses another problem in that we need to align the output signal order for every frequency bin so that a separated signal in the time domain contains frequency components from one source signal. This problem is known as the permutation problem.

Many methods have been proposed for solving the permutation problem, and the use of geometric information, such as beam patterns [17, 19, 20], direction of arrival (DOA), and source locations [14], is an effective approach. We have proposed a robust method that combines the DOA-based method [17, 19] and the correlation-based method [18], which almost completely solves the problem for two-source cases [22]. However it is insufficient when the number of signals is large or when the signals come from the same

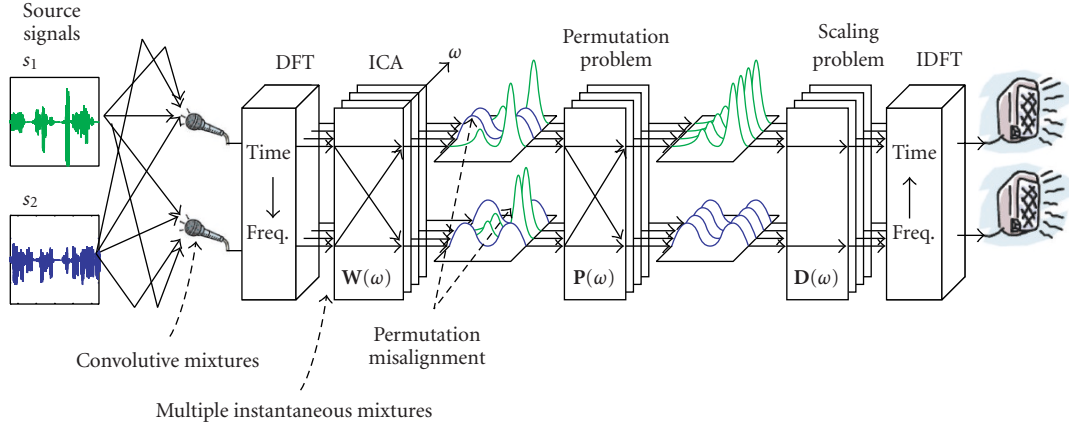


FIGURE 1: Flow of frequency-domain BSS ($N = M = 2$).

or similar direction. In this paper, we propose a method for obtaining proper geometric information for solving the permutation problem in such cases.

There is another problem with regard to the frequency-domain approach. Frequency-domain BSS is influenced by the circularity of the discrete-frequency representation. This causes a problem when we convert separation matrices in the frequency domain into separation filters in the time domain [25, 26]. This problem is not well known since it is not serious in a two-source case but it becomes serious as the number of sources increases. We also discuss the characteristics and the reason for this problem and present a solution based on spectral smoothing.

This paper is an extended version of our conference papers [23–25], whose contents are partially summarized in our survey articles [27, 28]. In this paper, we describe problems of sensitivity and ambiguity regarding DOA estimation in detail. We also carry out detailed experiments to examine the effectiveness of the spectral smoothing and the scaling adjustment when the number of source signals is large.

This paper is organized as follows. In Section 2, we review frequency-domain BSS and its inherent problems of permutation and scaling. In Section 3, we propose a method for localizing source signals by using the ICA solution with near-field and far-field models. The geometric information obtained with our method is useful for solving the permutation problem. In Section 4, we discuss the problem of the circularity, which becomes crucial when the number of source signals is large, and propose a solution. The experimental results and discussions are presented in Section 5. Section 6 concludes this paper.

2. FREQUENCY-DOMAIN BSS

When N source signals are $s_1(t), \dots, s_N(t)$ and the signals observed by M sensors are $x_1(t), \dots, x_M(t)$, the mixing model can be described by the following equation:

$$x_j(t) = \sum_{i=1}^N \sum_l h_{ji}(l) s_i(t-l), \quad (1)$$

where $h_{ji}(l)$ is the impulse response from source i to sensor j . We assume that the number of sources N is known or can be estimated in some way (e.g., by [20]), and the number of sensors M is equal to or greater than N ($N \leq M$). The separation system typically consists of a set of FIR filters $w_{kj}(l)$ of length L designed to produce N separated signals $y_1(t), \dots, y_N(t)$, and it is described as

$$y_k(t) = \sum_{j=1}^M \sum_{l=0}^{L-1} w_{kj}(l) x_j(t-l). \quad (2)$$

Figure 1 shows the flow of BSS in the frequency domain. Each convolutive mixture in the time domain is converted into multiple instantaneous mixtures in the frequency domain. Therefore, we can apply an ordinary ICA algorithm [7] in the frequency domain to solve a BSS problem in a reverberant environment. Using a short-time discrete Fourier transform (DFT), the mixing model is approximated as

$$\mathbf{x}(f, m) = \mathbf{H}(f) \mathbf{s}(f, m), \quad (3)$$

where f denotes a frequency, m is a frame index, $\mathbf{s}(f, m) = [s_1(f, m), \dots, s_N(f, m)]^T$ is a vector of the source signals in the frequency bin f , $\mathbf{x}(f, m) = [x_1(f, m), \dots, x_M(f, m)]^T$ is a vector of the observed signals, and $\mathbf{H}(f)$ is a matrix consisting of the frequency responses $H_{ji}(f)$ from source i to sensor j . The separation process can be formulated in each frequency bin as

$$\mathbf{y}(f, m) = \mathbf{W}(f) \mathbf{x}(f, m), \quad (4)$$

where $\mathbf{y}(f, m) = [y_1(f, m), \dots, y_N(f, m)]^T$ is a vector of the separated signals, and $\mathbf{W}(f)$ represents the separation matrix. $\mathbf{W}(f)$ is determined so that the elements of $\mathbf{y}(f, m)$ become mutually independent for each f .

In the experiments shown in Section 5, we calculated \mathbf{W} by using a complex-valued version of FastICA [7, 30] and improved it further by using InfoMax [5] combined with the natural gradient [31] whose nonlinear function is based on the polar coordinate [32].

2.1. Permutation and scaling problems

The ICA solution suffers permutation and scaling ambiguities. This is due to the fact that if $\mathbf{W}(f)$ is a solution, then $\mathbf{D}(f)\mathbf{P}(f)\mathbf{W}(f)$ is also a solution, where $\mathbf{D}(f)$ is a diagonal complex-valued scaling matrix, and $\mathbf{P}(f)$ is an arbitrary permutation matrix. Before constructing output signals in the time domain, we have to align the permutation so that each channel contains frequency components from one source signal.

The scaling ambiguity causes a filtering effect in the time domain. We have to determine $\mathbf{D}(f)$ so that the output signals become natural based on certain criteria. There is a simple and reasonable solution for the scaling problem:

$$\mathbf{D}(f) = \text{diag} \{ [\mathbf{P}(f)\mathbf{W}(f)]^{-1} \}, \quad (5)$$

which is obtained by the minimal distortion principle (MDP) [9] or the projection back method [18], and we can use it. By using this solution, the output signal y_i becomes an estimation of the reverberant version of source s_i measured at sensor i . On the other hand, the permutation problem is complicated, especially when the number of source signals is large, since the number of possible permutations increases to the factorial of N .

2.2. Solutions for permutation problem

There are various methods for solving the permutation problem. Geometric information, such as beam patterns [17, 19, 20], direction of arrival (DOA), and source locations [14], is useful for solving the problem. This approach is robust, however, it is not precise since the estimation of the geometric information fails in some frequency bins, especially in lower frequency bins. Another approach is based on the interfrequency correlations of output signal envelopes [18]. However, the correlation-based method is not robust since a misalignment at one frequency bin causes consecutive misalignments.

We have proposed a robust and precise method by combining the DOA-based method and the correlation-based method, which almost completely solves the permutation problem for two sources that come from different directions [22]. However the DOA-based method fails in the first stage when the signals come from the same or similar directions. Even if the signals come from different directions, when the number of signals is large or the source locations are omnidirectional, there are problems of sensitivity and ambiguity regarding DOA estimation, which are described later. In such cases, we have to rely on the correlation-based method, which is unstable. In the next section, we propose a method for obtaining proper geometric information for solving the permutation problem in such cases. The first method is to unify relative DOAs obtained by ICA solution. The second method is to estimate spheres on which source signals exist by using the ICA solution and near-field model.

3. SOURCE LOCALIZATION BY ICA

As Comon has suggested in [4], a two-stage procedure, consisting of ICA and using the knowledge of the array manifold, is useful for source localization. However, a simple comparison of the ICA solution with the propagation model does not yield proper information because of the scaling ambiguity in the ICA solution. This is the major difference from source localization using blind identification [14], where the mixing system is estimated directly.

This section presents a new source localization method that involves the ICA solution. The information about the source locations can be used to solve the permutation problem.

3.1. Invariant in ICA solution

The frequency response matrix $\mathbf{H}(f)$ is closely related to the locations of the sources and sensors. If a separation matrix $\mathbf{W}(f)$ is calculated successfully and it extracts source signals with a scaling ambiguity, there is a diagonal matrix $\mathbf{D}(f)$, and $\mathbf{D}(f)\mathbf{W}(f)\mathbf{H}(f) = \mathbf{I}$ holds. Because of the scaling ambiguity, we cannot obtain $\mathbf{H}(f)$ simply from the ICA solution $\mathbf{W}(f)$. However, the ratio of elements in the same column $H_{ji}(f)/H_{j'i}(f)$ is invariable in relation to $\mathbf{D}(f)$, and is given by

$$\frac{H_{ji}(f)}{H_{j'i}(f)} = \frac{[\mathbf{W}^{-1}(f)\mathbf{D}^{-1}(f)]_{ji}}{[\mathbf{W}^{-1}(f)\mathbf{D}^{-1}(f)]_{j'i}} = \frac{[\mathbf{W}^{-1}(f)]_{ji}}{[\mathbf{W}^{-1}(f)]_{j'i}}, \quad (6)$$

where $[\cdot]_{ji}$ denotes the j th element of the matrix. By using this invariant, we can estimate several types of geometric information (e.g., DOA, range) related to separated signals. The estimated information can be used to solve the permutation problem.

If we have more sensors than sources ($N < M$), principal component analysis (PCA) is performed before ICA so that the N -dimensional subspace spanned by the row vectors of $\mathbf{W}(f)$ is almost identical to the signal subspace, and the Moore-Penrose pseudoinverse $\mathbf{W}^+ \triangleq \mathbf{W}^T(\mathbf{W}\mathbf{W}^T)^{-1}$ is used instead of \mathbf{W}^{-1} .

3.2. DOA estimation with far-field model

We can estimate the DOA of source signals by using the above invariant $H_{ji}(f)/H_{j'i}(f)$. With a far-field model, a frequency response is formulated as

$$H_{ji}(f) = e^{j2\pi f c^{-1} \mathbf{a}_i^T \mathbf{p}_j}, \quad (7)$$

where c is the wave propagation speed, \mathbf{a}_i is a unit vector that points to the direction of source i , and \mathbf{p}_j represents the location of sensor j . According to this model, we have

$$\frac{H_{ji}(f)}{H_{j'i}(f)} = e^{j2\pi f c^{-1} \mathbf{a}_i^T (\mathbf{p}_j - \mathbf{p}_{j'})} \quad (8)$$

$$= e^{j2\pi f c^{-1} \|\mathbf{p}_j - \mathbf{p}_{j'}\| \cos \theta_{i,jj'}}, \quad (9)$$

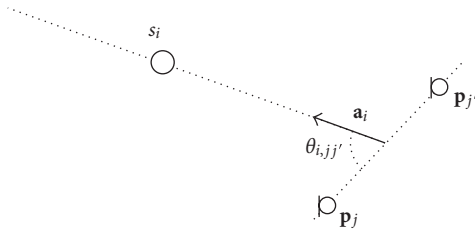


FIGURE 2: Direction of source i relative to the sensor pair j and j' .

where $\theta_{i,jj'}$ is the direction of source i relative to the sensor pair j and j' (Figure 2). By using the argument of (9) and (6), we can estimate

$$\begin{aligned}\hat{\theta}_{i,jj'}(f) &= \arccos \frac{\arg(H_{ji}/H_{j'i})}{2\pi f c^{-1} \|(\mathbf{p}_j - \mathbf{p}_{j'})\|} \\ &= \arccos \frac{\arg([\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i})}{2\pi f c^{-1} \|(\mathbf{p}_j - \mathbf{p}_{j'})\|}.\end{aligned}\quad (10)$$

This procedure is valid for sensor pairs with a small spacing that does not cause spatial aliasing. $\hat{\theta}_{i,jj'}(f)$ is estimated for each frequency bin f , but we omit the argument f for simplicity of notation in the following sections.

3.3. Sensitivity of DOA estimation and a solution

DOA estimation is sensitive to source locations. Figure 3 shows examples of DOA estimation using (10) with two different source locations. When the source signals are almost in front of a sensor pair, their directions can be estimated robustly. However, when the signals are nearly horizontal to the axis of the pair, the estimated directions tend to have large errors. This can be explained as follows.

When we denote an error in calculated $\arg(H_{ji}/H_{j'i})$ as $\Delta \arg(\hat{H})$, and an error in $\hat{\theta}_{i,jj'}$ as $\Delta \hat{\theta}$, the ratio $|\Delta \hat{\theta}/\Delta \arg(\hat{H})|$ can be approximated by the partial derivative of (10):

$$\left| \frac{\Delta \hat{\theta}}{\Delta \arg(\hat{H})} \right| \approx \left| \frac{1}{2\pi f c^{-1} \|(\mathbf{p}_j - \mathbf{p}_{j'})\| \sin(\hat{\theta}_{i,jj'})} \right|. \quad (11)$$

Figure 4 shows examples of this value for several frequency bins. We can see that $\Delta \arg(\hat{H})$ causes a large error in the estimated DOA when the direction is near the axis of the sensor pair. Therefore, we should consider the estimated DOA to be unreliable in such cases. If we use multiple sensor pairs with various axis directions, we can reject unreliable estimation [24]. More sophisticated estimation, such as a density estimation of θ instead of a point estimation, might be possible by using the error distribution as prior knowledge.

3.4. Ambiguity of DOA estimation and a new solution

DOA estimation involves some ambiguities. When we use only one pair of sensors or a linear array, the estimated $\hat{\theta}_{i,jj'}$ determines a cone rather than a direction. If we assume a horizontal plane on which sources exist, the cone is reduced to

two half-lines. However, the ambiguity of two directions that are symmetrical with respect to the axis of the sensor pair still remains. This is a fatal problem when the source locations are omnidirectional. When the spacing between sensors is larger than half a wavelength, spatial aliasing causes another ambiguity, but we do not consider this here.

The ambiguity can be solved by using multiple sensor pairs (Figure 5). If we use sensor pairs that have different axis directions, we can estimate cones with various vertex angles for one source direction. If the *relative* DOA $\hat{\theta}_{i,jj'}$ is estimated without any error, the *absolute* DOA \mathbf{a}_i satisfies

$$\frac{(\mathbf{p}_j - \mathbf{p}_{j'})^T \mathbf{a}_i}{\|\mathbf{p}_j - \mathbf{p}_{j'}\|} = \cos \hat{\theta}_{i,jj'}. \quad (12)$$

When we use L sensor pairs whose indexes are $j(l)j'(l)$ ($1 \leq l \leq L$), \mathbf{a}_i is given by the solution of the following equation:

$$\mathbf{V} \mathbf{a}_i = \mathbf{c}_i, \quad (13)$$

where $\mathbf{V} \triangleq (\mathbf{v}_1, \dots, \mathbf{v}_L)^T$, $\mathbf{v}_l \triangleq (\mathbf{p}_{j(l)} - \mathbf{p}_{j'(l)})/\|\mathbf{p}_{j(l)} - \mathbf{p}_{j'(l)}\|$ is a normalized axis, and $\mathbf{c}_i \triangleq [\cos(\hat{\theta}_{i,j(1)j'(1)}), \dots, \cos(\hat{\theta}_{i,j(L)j'(L)})]^T$. Sensor pairs should be selected so that $\text{rank}(\mathbf{V}) \geq 3$ if the potential source locations are three-dimensional, or $\text{rank}(\mathbf{V}) \geq 2$ if we assume a plane on which sources exist.

In a practical situation, $\hat{\theta}_{i,j(l)j'(l)}$ has an estimation error, and (13) has no exact solution. Thus we adopt an optimal solution by employing certain criteria such as

$$\hat{\mathbf{a}}_i = \arg \min_{\mathbf{a}} \|\mathbf{V} \mathbf{a} - \mathbf{c}_i\| \quad (\text{subject to } \|\mathbf{a}\| = 1). \quad (14)$$

This can be solved approximately by using the Moore-Penrose pseudoinverse $\mathbf{V}^+ \triangleq (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T$, and we have

$$\hat{\mathbf{a}}_i \approx \frac{\mathbf{V}^+ \mathbf{c}_i}{\|\mathbf{V}^+ \mathbf{c}_i\|}. \quad (15)$$

Accordingly, we can determine a unit vector $\hat{\mathbf{a}}_i$ pointing to the direction of source s_i .

3.5. Estimation of sphere with near-field model

The interpretation of the ICA solution with a near-field model yields other geometric information. When we adopt the near-field model, including the attenuation of the wave, $H_{ji}(f)$ is formulated as

$$H_{ji}(f) = \frac{1}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{-j2\pi f c^{-1} (\|\mathbf{q}_i - \mathbf{p}_j\|)}, \quad (16)$$

where \mathbf{q}_i represents the location of source i . By taking the ratio of (16) for a pair of sensors j and j' , we obtain

$$\frac{H_{ji}(f)}{H_{j'i}(f)} = \frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{-j2\pi f c^{-1} (\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i - \mathbf{p}_{j'}\|)}. \quad (17)$$

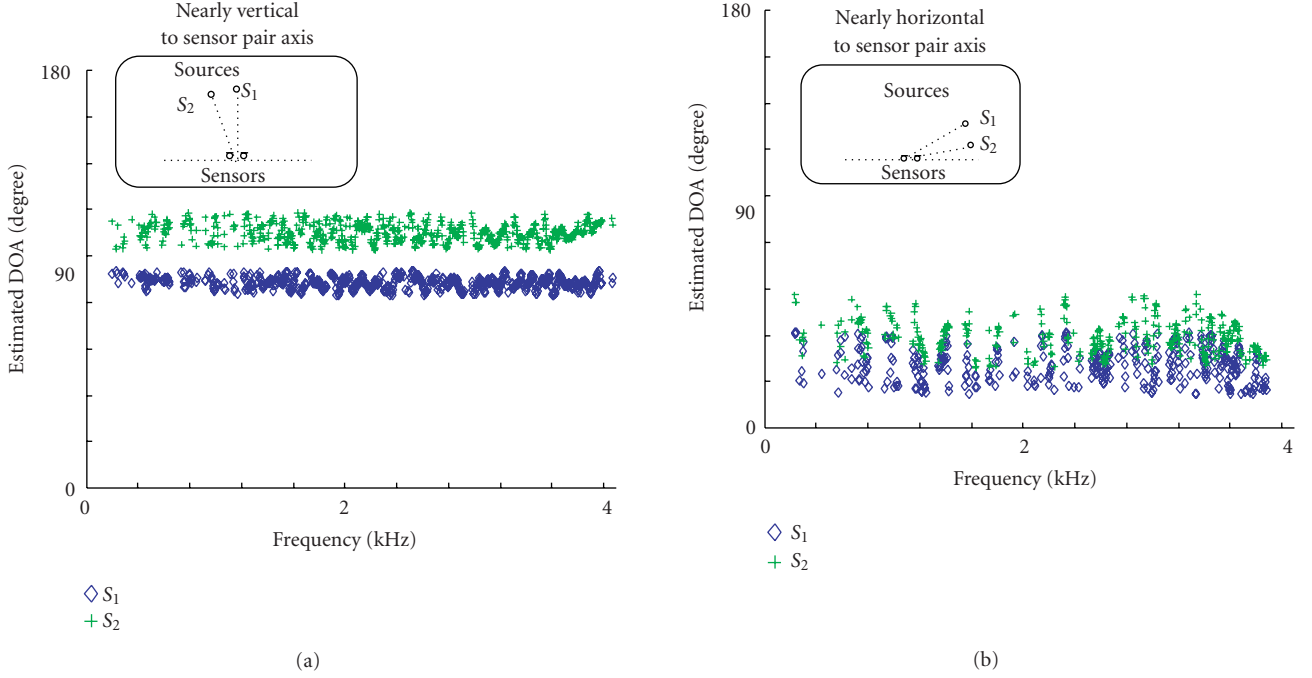


FIGURE 3: Source locations and estimated DOAs.

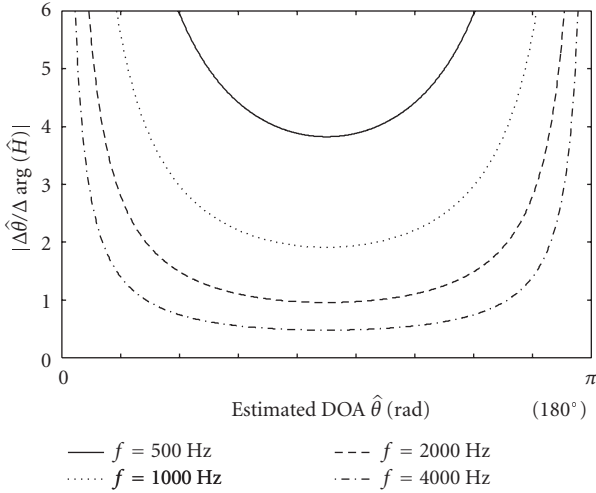


FIGURE 4: Sensitivity of DOA estimation.

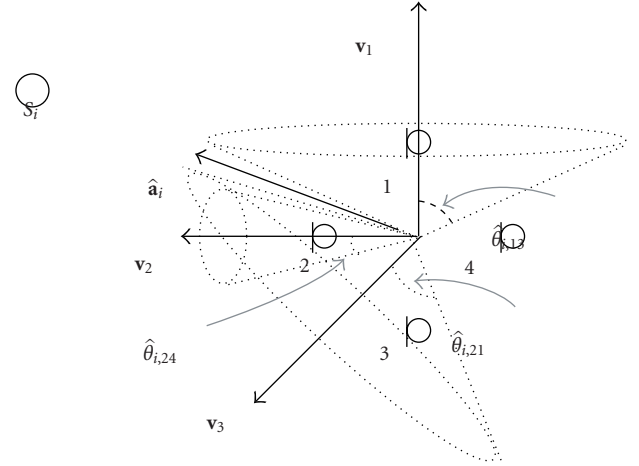
By using the modulus of (17) and (6) we have

$$\frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} = \left| \frac{[\mathbf{W}^{-1}]_{ji}}{[\mathbf{W}^{-1}]_{j'i}} \right|. \quad (18)$$

By solving (17) for \mathbf{q}_i , we have a sphere whose center $O_{i,jj'}$ and radius $R_{i,jj'}$ are given by

$$O_{i,jj'} = \mathbf{p}_j - \frac{1}{r_{i,jj'}^2 - 1} (\mathbf{p}_{j'} - \mathbf{p}_j), \quad (19)$$

$$R_{i,jj'} = \left\| \frac{r_{i,jj'}}{r_{i,jj'}^2 - 1} (\mathbf{p}_{j'} - \mathbf{p}_j) \right\|, \quad (20)$$


 FIGURE 5: Solving ambiguity of estimated DOAs. Index of sensor pairs $j(1)j'(1) = 13, j(2)j'(2) = 24, j(3)j'(3) = 21$.

where $r_{i,jj'} \triangleq |[\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i}|$. Thus, we can estimate a sphere $(\hat{O}_{i,jj'}, \hat{R}_{i,jj'})$ on which \mathbf{q}_i exists by using the result of ICA \mathbf{W} and the locations of the sensors \mathbf{p}_j and $\mathbf{p}_{j'}$. Figure 6 shows an example of the spheres determined by (18) for various ratios $r_{i,jj'}$. This procedure is valid for sensor pairs with a spacing large enough to cause a level difference.

3.6. Permutation alignment

This subsection outlines the procedure for permutation alignment by integrating a localization approach and a correlation approach. The procedure, which uses DOA as geometric information, has been detailed in [22].

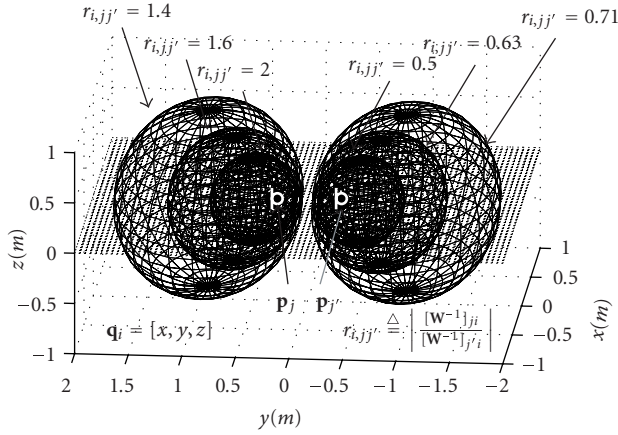


FIGURE 6: Example of spheres determined by (18) ($\mathbf{p}_j = [0, 0.3, 0]$, $\mathbf{p}_{j'} = [0, -0.3, 0]$).

The procedure consists of the following steps.

- (1) Cluster separated frequency components $y_k(f, m)$ for all k and all f by using geometric information such as (10), (15), (19), and (20), and decide the permutations at certain frequencies where the confidence of source localization is sufficiently high.
- (2) Decide the permutations to maximize the sum of the interfrequency correlation of separated signals. The correlation should be calculated for the amplitude $|y_k(f, m)|$ or (log-scaled) power $|y_k(f, m)|^2$ instead of the raw complex-valued signals $y_k(f, m)$, since the correlation of raw signals would be very low because of the short-time DFT property. The sum of the correlations between $|y_k(f, m)|$ and $|y_k(g, m)|$ within distance δ (i.e., $|f - g| < \delta$) is used as a criterion. The permutations are decided for frequencies where the criterion gives a clear-cut decision.
- (3) Calculate the correlations between $|y_k(f, m)|$ and its harmonics $|y_k(g, m)|$ ($g = 2f, 3f, 4f, \dots$), and decide the permutations to maximize the sum of the correlations. The permutations are decided for frequencies where the correlation among harmonics is sufficiently high.
- (4) Decide the permutations for the remaining frequencies based on neighboring correlations.

Let us discuss the advantages of the integrated method. The main advantage is that it does not cause a large misalignment as long as the permutations fixed by the localization approach are correct. Moreover, the correlation part (steps (2), (3), and (4)) compensates for the lack of preciseness of the localization approach. The correlation part consists of three steps for two reasons. First, the harmonics part (step (3)) works well if most of the other permutations are fixed. Second, the method becomes more robust by quitting step (2) if there is no clear-cut decision. With this structure, we can avoid fixing the permutations for consecutive frequencies without high confidence. As shown in the experimental results (Section 5.2), this integrated method is effective at separating many sources.

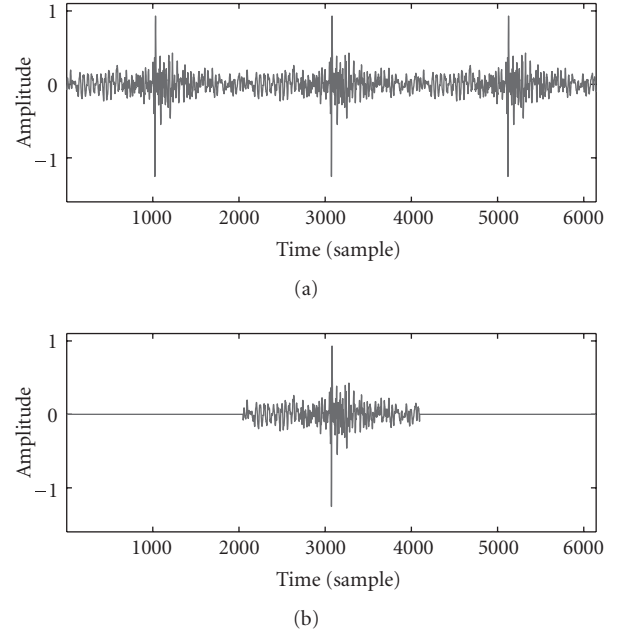


FIGURE 7: Periodic time-domain filter represented by frequency responses sampled at $L = 2048$ points (a) and its one-period realization (b).

4. SPECTRAL SMOOTHING WITH ERROR MINIMIZATION

Frequency-domain BSS is influenced by the circularity of discrete-frequency representation. Circularity refers to the fact that frequency responses sampled at L points with an interval f_s/L (f_s : sampling frequency) represent a periodic time-domain signal whose period is L/f_s . Figure 7 shows two time-domain filters. The upper part of the figure shows a periodic infinite-length filter represented by frequency responses $w_{kj}(f) = [\mathbf{W}(f)]_{kj}$ calculated by ICA at L points. Since this filter is unrealistic, we usually use its one-period realization shown in the lower part of the figure.

However, such one-period filters may cause a problem. Figure 8 shows impulse responses from a source $s_i(t)$ to an output $y_k(t)$ defined by

$$u_{ki}(l) = \sum_{j=1}^m \sum_{\tau=0}^{L-1} w_{kj}(\tau) h_{ji}(l - \tau). \quad (21)$$

The responses on the left $u_{11}(l)$ correspond to the extraction of a target signal, and those on the right $u_{14}(l)$ correspond to the suppression of an interference signal. The upper responses are obtained with infinite-length filters, and the lower ones with one-period filters. We see that the one-period filters create spikes, which distort the target signal and degrade the separation performance.

4.1. Windowing

To solve this problem, we need to control the frequency responses $w_{kj}(f)$ so that the corresponding time-domain filter

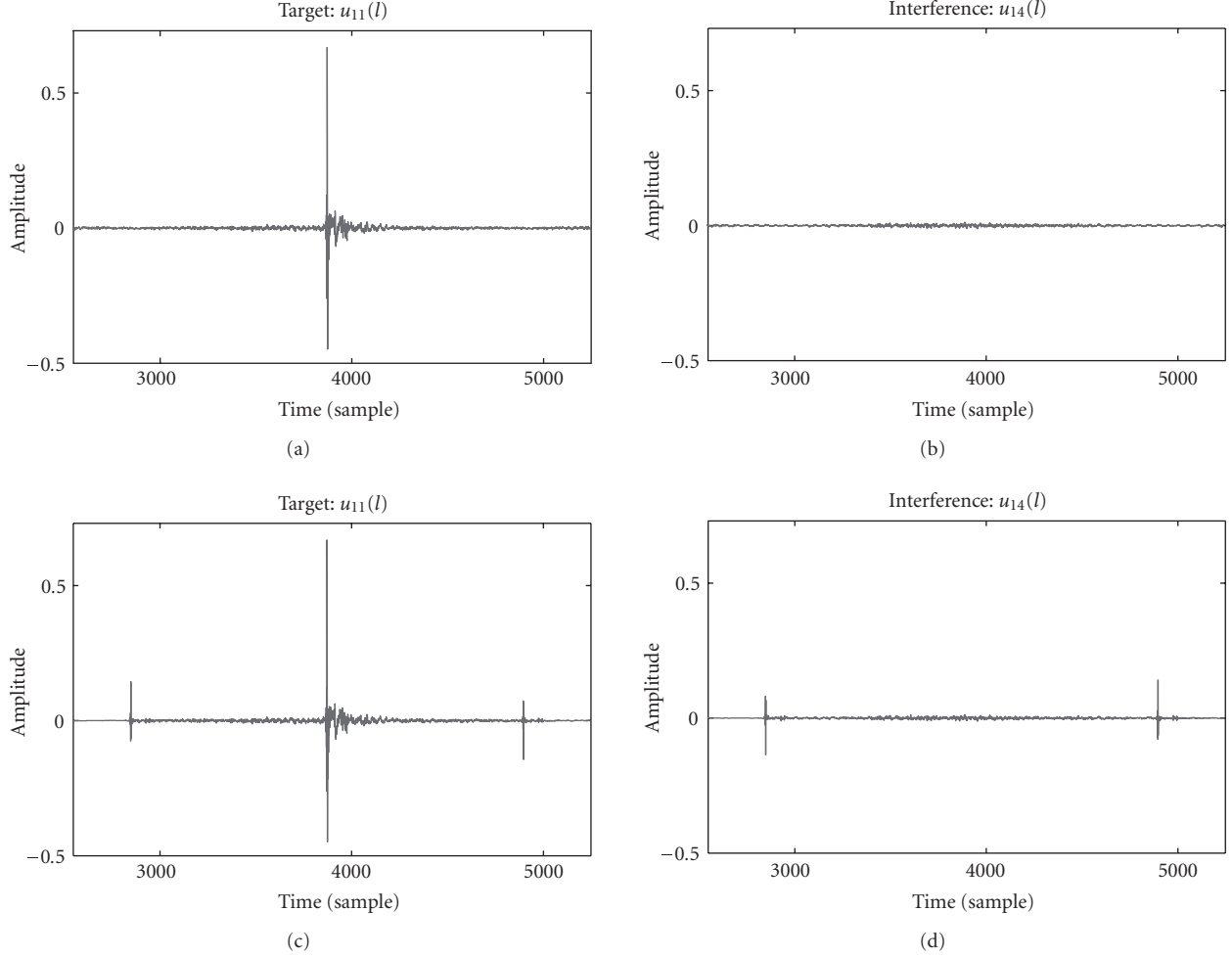


FIGURE 8: Impulse responses $u_{ki}(l)$ obtained with the periodic filters (above) and with their one-period realization (below).

$w_{kj}(l)$ does not rely on the circularity effect whereby adjacent periods work together to perform some filtering. The most widely used approach is spectral smoothing, which is realized by multiplying a window $g(l)$ that tapers smoothly to zero at each end, such as a Hanning window $g(l) = (1/2)(1 + \cos(2\pi l/L))$. This makes the resulting time-domain filter $w_{kj}(l) \cdot g(l)$ fit length L and have a small amplitude around the ends [33]. As a result, the frequency responses $w_{kj}(f)$ are smoothed as

$$\tilde{w}_{kj}(f) = \sum_{\phi=0}^{f_s-\Delta f} g(\phi) w_{kj}(f - \phi), \quad (22)$$

where $g(f)$ is the frequency response of $g(l)$ and $\Delta f = f_s/L$. If a Hanning window is used, the frequency responses are smoothed as

$$\tilde{w}_{kj}(f) = \frac{1}{4} [w_{kj}(f - \Delta f) + 2w_{kj}(f) + w_{kj}(f + \Delta f)] \quad (23)$$

since the frequency responses $g(f)$ of the Hanning window are $g(0) = 1/2$, $g(\Delta f) = g(f_s - \Delta f) = 1/4$, and zero for the other frequency bins.

The windowing successfully eliminates the spikes. However, it changes the frequency response from $w_{kj}(f)$ to $\tilde{w}_{kj}(f)$ and causes an error. Let us evaluate the error for each row $\mathbf{w}_k(f) = [w_{k1}(f), \dots, w_{kM}(f)]^T$ of the ICA solution $\mathbf{W}(f)$. The error is

$$\begin{aligned} \mathbf{e}_k(f) &= \min_{\alpha_k} [\tilde{\mathbf{w}}_k(f) - \alpha_k \mathbf{w}_k(f)] \\ &= \tilde{\mathbf{w}}_k(f) - \frac{\tilde{\mathbf{w}}_k(f)^H \mathbf{w}_k(f)}{\|\mathbf{w}_k(f)\|^2} \mathbf{w}_k(f), \end{aligned} \quad (24)$$

where $\tilde{\mathbf{w}}_k(f) = [\tilde{w}_{k1}(f), \dots, \tilde{w}_{kM}(f)]^T$ and α_k is a complex-valued scalar representing the scaling ambiguity of the ICA solution. The minimization \min_{α_k} is based on the least-squares, and can be represented by the projection of $\tilde{\mathbf{w}}_k$ to \mathbf{w}_k . We can evaluate the error for the Hanning window case by substituting (23) for $\tilde{\mathbf{w}}_k$ of (24):

$$\mathbf{e}_k(f) = \frac{1}{4} [\mathbf{e}_k^-(f) + \mathbf{e}_k^+(f)], \quad (25)$$

where

$$\mathbf{e}_k^-(f) = \mathbf{w}_k(f - \Delta f) - \frac{\mathbf{w}_k(f - \Delta f)^H \mathbf{w}_k(f)}{\|\mathbf{w}_k(f)\|^2} \mathbf{w}_k(f), \quad (26)$$

$$\mathbf{e}_k^+(f) = \mathbf{w}_k(f + \Delta f) - \frac{\mathbf{w}_k(f + \Delta f)^H \mathbf{w}_k(f)}{\|\mathbf{w}_k(f)\|^2} \mathbf{w}_k(f). \quad (27)$$

Here \mathbf{e}_k^- (or \mathbf{e}_k^+) represents the difference between two vectors $\mathbf{w}_k(f)$ and $\mathbf{w}_k(f - \Delta f)$ (or $\mathbf{w}_k(f + \Delta f)$). Since these differences are usually not very large, the error \mathbf{e}_k does not seriously affect the separation if we use a Hanning window for spectral smoothing.

4.2. Minimizing error by adjusting scaling ambiguity

Even if the error caused by the windowing is not very large, the separation performance is improved by its minimization [25]. This is performed by adjusting the scaling ambiguity of the ICA solution before the windowing. Let $d_k(f)$ be a complex-valued scalar for the scaling adjustment:

$$\mathbf{w}_k(f) \leftarrow d_k(f) \mathbf{w}_k(f). \quad (28)$$

We want to find $d_k(f)$ such that the error (24) is minimized. The scalar $d_k(f)$ should be close to 1 to avoid any great change in the predetermined scaling. Thus, an appropriate total cost to be minimized is

$$\mathcal{J} = \sum_f J_k(f), \quad (29)$$

where

$$J_k(f) = \frac{\|\mathbf{e}_k(f)\|^2}{\|\mathbf{w}_k(f)\|^2} + \beta |d_k(f) - 1|^2, \quad (30)$$

and β is a parameter indicating the importance of maintaining the predetermined scaling. With the Hanning window, the error after the scaling adjustment is easily calculated by substituting (28) for (25):

$$\mathbf{e}_k(f) = \frac{1}{4} [d_k(f - \Delta f) \mathbf{e}_k^-(f) + d_k(f + \Delta f) \mathbf{e}_k^+(f)], \quad (31)$$

where \mathbf{e}_k^- and \mathbf{e}_k^+ are defined in (26) and (27), respectively.

The minimization of the total cost can be performed iteratively by

$$d_k(f) = d_k(f) - \mu \frac{\partial \mathcal{J}}{\partial d_k(f)} \quad (32)$$

with a small step size μ . With the Hanning window, the gradient is

$$\begin{aligned} \frac{\partial \mathcal{J}}{\partial d_k(f)} &= \frac{\partial J_k(f - \Delta f)}{\partial d_k(f)} + \frac{\partial J_k(f + \Delta f)}{\partial d_k(f)} + \frac{\partial J_k(f)}{\partial d_k(f)} \\ &= \frac{\mathbf{e}_k(f - \Delta f)^H \mathbf{e}_k^+(f - \Delta f) + \mathbf{e}_k(f + \Delta f)^H \mathbf{e}_k^-(f + \Delta f)}{8 \cdot \|\mathbf{w}_k(f)\|^2} \\ &\quad + 2\beta(d_k(f) - 1). \end{aligned} \quad (33)$$

With (31) to (33), we can optimize the scalar $d_k(f)$ for the scaling adjustment, and minimize the error caused by spectral smoothing (23) with the Hanning window.

5. EXPERIMENTS AND DISCUSSIONS

We carried out two kinds of experiments. The first involves the separation of two source signals arriving from the same direction. The purpose of this experiment is to show that spheres estimated by near-field model can substitute for DOAs when solving permutation problem in such a case. Iwaki and Ando [34] have proposed a BSS system for a case where signals and microphones are located on the same line. In our experiment, the signals and microphones are not necessarily on the same line, and thus represent a more realistic situation.

The second experiment consists of the separation of six source signals that come from various directions with two of them coming from the same direction. In this experiment, we used a combination of small and large spacing microphone pairs. The small spacing microphone pairs with various axis directions enable us to estimate DOA robustly and without ambiguity. Large spacing microphone pairs give us the geometric information we need to distinguish signals arriving from the same direction. We utilize this information to solve the permutation problem. We also show the effectiveness of the spectral smoothing with error minimization in this experiment.

The performance is measured by the signal-to-inference ratio (SIR). When we solve the permutation problem so that $s_k(t)$ is output to $y_k(t)$, the output SIR for $y_k(t)$ is defined as

$$\text{SIR}_k \triangleq 10 \log \left[\frac{\sum_t y_{kk}(t)^2}{\sum_t \{ \sum_{i \neq k} y_{ki}(t) \}^2} \right] \text{ (dB)}, \quad (34)$$

where $y_{ki}(t)$ is the portion of $y_k(t)$ that comes from $s_i(t)$ that is calculated by

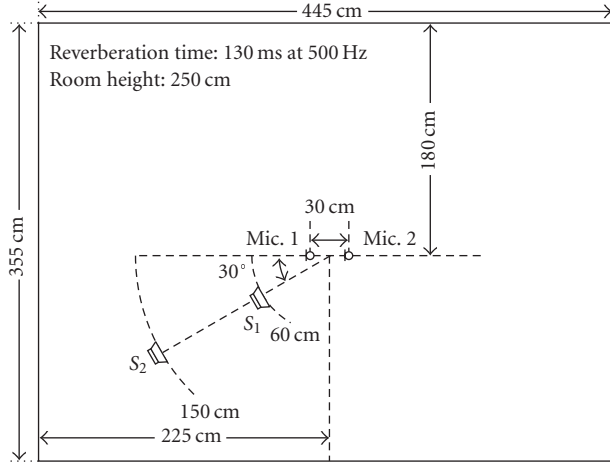
$$y_{ki}(t) = \sum_{j=1}^M \sum_{l=0}^{L-1} u_{ki}(l) s_i(t-l), \quad (35)$$

where $u_{ki}(l)$ is a system impulse response defined by (21).

5.1. Two sources arriving from the same direction

We began by carrying out experiments with two sources and two microphones using speech signals convolved with impulse responses measured in a room. The room layout is shown in Figure 9. The sources are located in the same direction from the microphone pair. The reverberation time of the room was 130 milliseconds at 500 Hz. Other conditions are summarized in Table 1. The experimental procedure is as follows.

First, we apply ICA to observed signals $x_j(t)$ ($j = 1, 2$), and calculate separation matrix $\mathbf{W}(f)$ for each frequency bin. Then we estimate radiuses $\hat{R}_{1,12}$ and $\hat{R}_{2,12}$ of two spheres on which each source signal exists by using $\mathbf{W}^{-1}(f)$ and (20), and the permutation is aligned so that $\hat{R}_{2,12} \geq \hat{R}_{1,12}$. In order to evaluate the reliability of the solution provided by the estimated spheres, we introduce a threshold parameter $\text{thr} \geq 1$, and we accept solutions only for frequency bins that satisfy the condition $\hat{R}_{2,12}/\hat{R}_{1,12} \geq \text{thr}$. We then apply the



○ Microphones (omnidirectional, height: 135 cm)
 □ Loudspeakers (height: 135 cm)

FIGURE 9: Room layout.

TABLE 1: Experimental conditions.

Sampling rate	8 kHz
Data length	2 seconds
Window	Hanning
Frame length	1024 points (128 ms)
Frame shift	256 points (32 ms)
ICA algorithm	InfoMax (complex-valued)

correlation-based method to the remaining frequency bins. The permutation problem is solved simply by using the geometric information when $th_R = 1$, and simply by using the correlation when $th_R = \infty$.

We define SIR as the average of the SIR_1 and SIR_2 in order to cancel out the effect of the input SIR. We measured SIRs for 12 combinations of source signals using two male and two female speakers and varying the threshold parameter th_R .

Figure 10 shows the experimental results. When we solve the permutation problem using only the estimated spheres ($th_R = 1$), the performance is insufficient. In contrast, the performance we obtain using only the correlation ($th_R = \infty$) is unstable. The combination of both methods yields good and stable performance. These tendencies are similar to the results we obtain when we use DOAs as geometric information [22].

We obtained good performance when the threshold parameter th_R was relatively large. When th_R was 8 to 16, the permutation of about 1/5 to 1/10 of the frequency bins was determined by the geometric information. This result suggests that we should use this geometric information for frequency bins where the estimation is highly reliable.

Figure 11 shows the spatial gain patterns of the separation filters in one frequency bin ($f = 1000$ Hz) drawn with the near-field model. The gain of the observed signal

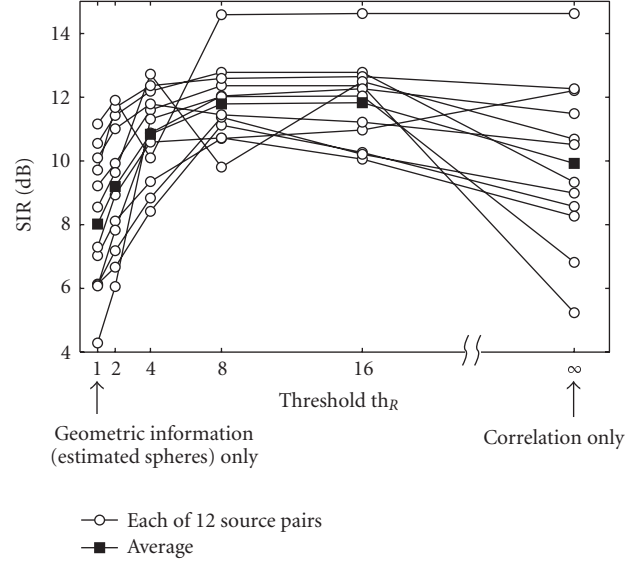


FIGURE 10: Experimental results. SIRs are evaluated for 12 combinations of source signals with various values for threshold parameter th_R

at microphone 1 is defined as 0 dB. We can see that the separation filter forms a spot null beam focusing on the interference signal. When source signals are located in different directions, a separation filter utilizes the phase difference of the input signals and makes a directive null towards the interference signal [35], whereas both the phase and level differences are utilized to make a regional null when signals come from the same direction.

5.2. Separation of six sources

Next, we carried out experiments with six sources and eight microphones using speech signals convolved with impulse responses measured in a room with a reverberation time of 130 milliseconds. In general, we can separate up to N sources with N microphones unless the mixing system is singular. However, $N \times N$ mixing systems tend to be singular or nearly singular depending on the locations of the source signals. One or two degrees of freedom relax such a critical situation.

The program was coded in Matlab and run on an AMD Athlon 64 FX-53 Processor (2.4 GHz CPU clock). The computation time was about 30 seconds for 6 second data. This is much faster than a time-domain approach. The room layout is shown in Figure 12. Other conditions are summarized in Table 2. We assume that the number of source signals $N = 6$ is known. The experimental procedure is as follows.

First, we apply ICA to $x_j(t)$ ($j = 1, \dots, 8$), and calculate separation matrix $\mathbf{W}(f)$ for each frequency bin. The initial value of $\mathbf{W}(f)$ is calculated by PCA. Then we estimate the DOAs by using the rows of $\mathbf{W}^+(f)$ (pseudoinverse) corresponding to the small spacing microphone pairs (1-3, 2-4, 1-2, and 2-3). Figure 13 shows a histogram of the estimated DOAs of all the frequency components. The DOAs can be

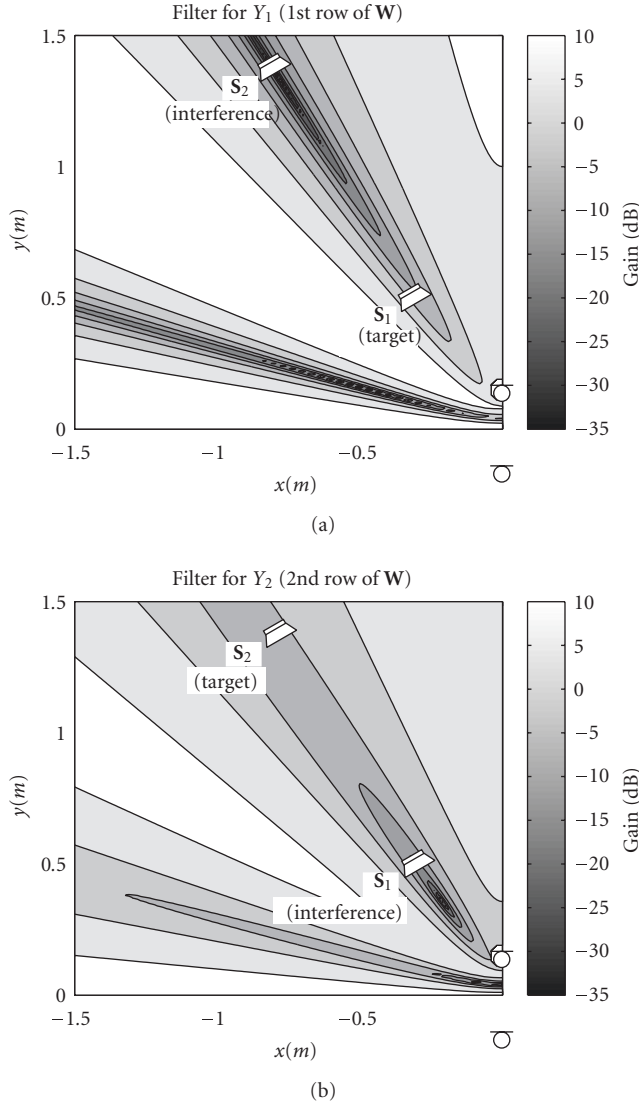
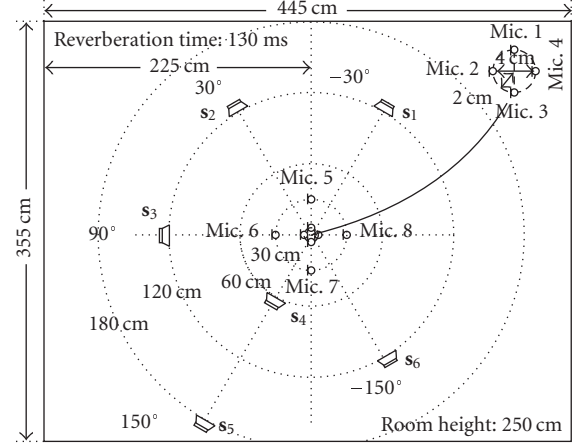


FIGURE 11: Example spatial gain patterns of separation filters ($f = 1000$ Hz).

clustered by using an ordinary clustering method such as the k -means algorithm [36]. There are five clusters in this histogram, and one cluster is twice the size of the others. This implies that two signals come from the same direction (about 150°). We can solve the permutation problem for the other four sources by using this DOA information (Figure 14).

Then, we apply the estimation of spheres to the signals that belong to the large cluster by using the rows of $\mathbf{W}^+(f)$ corresponding to the large spacing microphone pairs (7-5, 7-8, 6-5, and 6-8). Figure 15 shows estimated radiuses for s_4 and s_5 for the microphone pair 7-5. Although the radius estimation includes a large error, it provides sufficient information to distinguish two signals. Accordingly, we can classify the signals into six clusters. We determine the permutation only for frequency bins with a consistent classification, and we employ a correlation-based method for the rest. Finally, we construct separation filters in the time domain from the



○ Microphones (omnidirectional, height: 135 cm)

□ Loudspeakers (height: 135 cm)

FIGURE 12: Room layout for experiments.

TABLE 2: Experimental conditions.

Sampling rate	8 kHz
Data length	6 seconds
Frame length	2048 points (256 ms)
Frame shift	512 points (64 ms)
ICA algorithm	InfoMax (complex-valued)

ICA result. We solve the scaling problem by (5), and then perform a scaling adjustment to minimize the windowing error described in Section 4.2 before multiplying a Hanning window for the spectral smoothing.

We measured SIRs for three permutation solving strategies: the correlation-based method (C), estimated DOAs and correlation (D + C), and a combination of estimated DOAs, spheres, and correlation (D + S + C, proposed method). We also measured input SIRs by using the mixture observed by microphone 1 for the reference (Input SIR).

The experimental results are summarized in Table 3. Method C scored a good SIR only for s_4 and failed for all other signals. This shows the lack of robustness of the correlation-based method. Method D + C improved the separation performance as we had expected. However, it failed to separate s_4 , which came from the same direction as s_5 . Our proposed method (D + S + C) succeeded in separating all the signals with good score. We can see again that the discrimination obtained by using estimated spheres is effective in improving SIRs for signals coming from the same direction. The introduced sphere information contributes only to SIR_4 and SIR_5 , therefore the improvement in the average SIR appears superficially small. However this is a significant improvement overall. We have carried out some experiments with various combinations of source signals and obtained similar results.

In this experiment, since the input SIR was very bad (-7.1 dB), the average of the output SIRs was at most 11 dB.

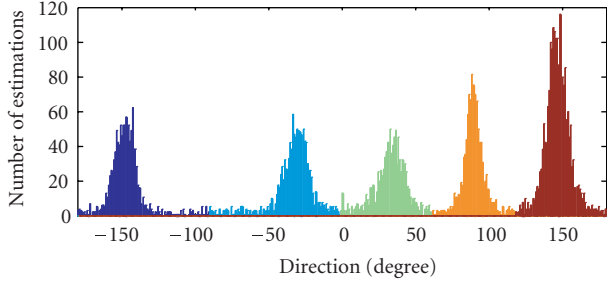


FIGURE 13: Histogram of estimated DOAs obtained by using small spacing microphone pairs.

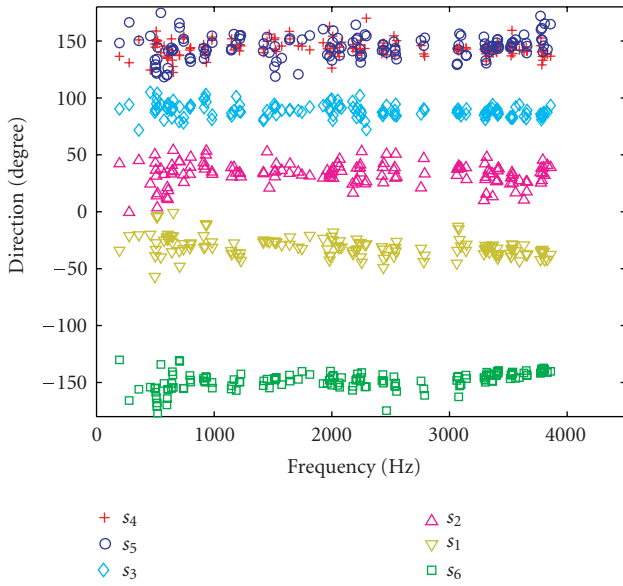


FIGURE 14: Permutation solved by using DOAs.

However, the SIR improvement (difference between the input and output SIRs) was about 18 dB. This score is comparable to that obtained in an ordinary two-source case.

Table 4 shows the results of the experiments we undertook to examine the effectiveness of the spectral smoothing and the scaling adjustment proposed in Section 5. We compared cases where the spectral smoothing was applied differently: no smoothing, simply multiplying a Hanning window (win), and with the scaling adjustment before multiplying a Hanning window (adj + win). The permutation problem was solved by D + S + C in all cases, and the frequency components are correctly aligned in most frequency bins. We can see that the spectral smoothing is essential for frequency-domain BSS in addition to solving the permutation problem, and that the scaling adjustment used for minimizing error improves SIR.

Finally we complement the room layout for the experiments. One reason for the regular speaker layouts is that we wanted to demonstrate the ability to separate symmetrically located source signals, which cannot be separated with a conventional linear array. Another reason is that we need a large

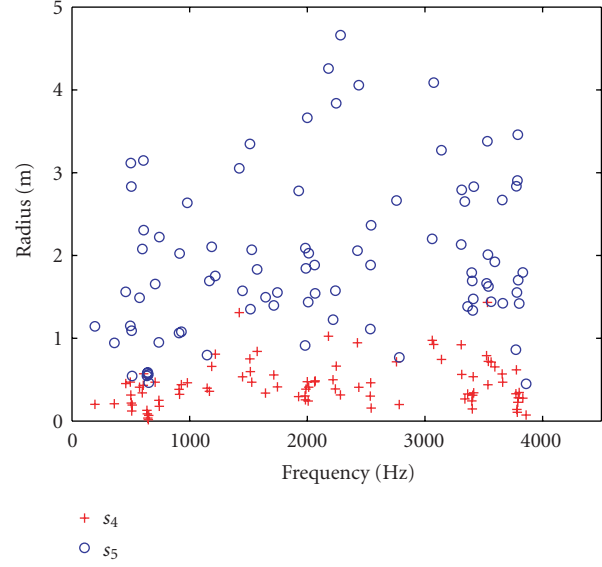


FIGURE 15: Estimated radiuses for s_4 and s_5 .

TABLE 3: Experimental results.

							(dB)
	SIR ₁	SIR ₂	SIR ₃	SIR ₄	SIR ₅	SIR ₆	Ave.
Input SIR	-8.3	-6.8	-7.8	-7.7	-6.7	-5.2	-7.1
C	4.4	2.6	4.0	9.2	3.6	-2.0	3.7
D + C	9.6	9.3	14.7	2.7	6.5	14.0	9.4
D + S + C (proposed method)	10.8	10.4	14.5	7.0	11.0	12.2	11.0

TABLE 4: Experimental results (permutation was solved by D + S + C).

							(dB)
	SIR ₁	SIR ₂	SIR ₃	SIR ₄	SIR ₅	SIR ₆	Ave.
No smoothing	5.4	7.1	8.9	2.3	6.3	7.4	6.2
win	8.9	9.8	14.3	5.2	7.4	11.2	9.5
Adj + win (proposed method)	10.8	10.4	14.5	7.0	11.0	12.2	11.0

enough angle between two sources to obtain good separation performance. This is not just the limitation of our permutation solving method, but also the limitation of the separation filter obtained by ICA that forms spatial directivity. Improving the robustness against the source locations is one of the most important issues for the future.

6. CONCLUSION

In this paper, we discussed the practical problems arising with frequency-domain BSS when the number of source signals is large and the source locations are omnidirectional. We proposed a method for obtaining proper geometric information with which to solve the permutation problem.

The interpretation of the ICA solution by a near-field model yields information about spheres on which source signals exist. This information can be used as an alternative to the DOA when signals come from the same or similar directions. Experimental results showed that the proposed method can robustly separate a mixture of signals arriving from the same direction. We also proposed the combination of small and large spacing sensor pairs with various axis directions. We can solve the problems of the sensitivity and ambiguity of the DOA estimation by using multiple sensor pairs. In experiments, our method succeeded in separating six speech signals with eight microphones, even when two came from the same direction. In addition, we confirmed the importance of spectral smoothing and the effectiveness of scaling adjustment in the frequency-domain BSS of many signals. Our techniques have been applied to a prototype system that performs an on-the-spot BSS of live recorded signals [37]. We believe that the proposed techniques enhance the usefulness of frequency-domain BSS for real audio applications.

REFERENCES

- [1] S. Haykin, Ed., *Unsupervised Adaptive Filtering*, John Wiley & Sons, New York, NY, USA, 2000.
- [2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, New York, NY, USA, 2002.
- [3] J. Benesty, S. Makino, and J. Chen, Eds., *Speech Enhancement*, Springer, New York, NY, USA, 2005.
- [4] P. Comon, "Independent component analysis. A new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [5] 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.
- [6] T. W. Lee, *Independent Component Analysis*, Kluwer Academic, Boston, Mass, USA, 1998.
- [7] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.
- [8] C. G. Puntonet and A. Prieto, Eds., *Independent Component Analysis and Blind Signal Separation*, vol. 3195 of *Lecture Notes in Computer Science*, Springer, New York, NY, USA, 2004.
- [9] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proceedings of 3rd International Conference on Independent Component Analysis and Blind Source Separation (ICA '01)*, pp. 722–727, San Diego, Calif, USA, December 2001.
- [10] S. C. Douglas and X. Sun, "Convolutional blind separation of speech mixtures using the natural gradient," *Speech Communication*, vol. 39, no. 1-2, pp. 65–78, 2003.
- [11] K. Matsuoka, Y. Ohba, Y. Toyota, and S. Nakashima, "Blind separation for convolutional mixture of many voices," in *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC '03)*, pp. 279–282, Kyoto, Japan, September 2003.
- [12] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fidelity blind separation of acoustic signals using SIMO-model-based independent component analysis," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E87-A, no. 8, pp. 2063–2072, 2004.
- [13] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutional mixtures based on second-order statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 120–134, 2005.
- [14] V. C. Soon, L. Tong, Y. F. Huang, and R. Liu, "A robust method for wideband signal separation," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '93)*, vol. 1, pp. 703–706, Chicago, Ill, USA, May 1993.
- [15] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1–3, pp. 21–34, 1998.
- [16] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutional blind source separation," in *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation (ICA '00)*, pp. 215–220, Helsinki, Finland, June 2000.
- [17] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '00)*, vol. 5, pp. 3140–3143, Istanbul, Turkey, June 2000.
- [18] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1–4, pp. 1–24, 2001.
- [19] M. Z. Ikram and D. R. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, vol. 1, pp. 881–884, Orlando, Fla, USA, May 2002.
- [20] L. C. Parra and C. V. Alvino, "Geometric source separation: merging convolutional source separation with geometric beamforming," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 6, pp. 352–362, 2002.
- [21] D. W. E. Schobben and P. C. W. Sommen, "A frequency domain blind signal separation method based on decorrelation," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1855–1865, 2002.
- [22] 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, 2004.
- [23] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Near-field frequency domain blind source separation for convolutional mixtures," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '04)*, vol. 4, pp. 49–52, Montreal, Que, Canada, May 2004.
- [24] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Frequency domain blind source separation using small and large spacing sensor pairs," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '04)*, vol. 5, pp. 1–4, Vancouver, BC, Canada, May 2004.
- [25] H. Sawada, R. Mukai, S. de la Kethulle, S. Araki, and S. Makino, "Spectral smoothing for frequency-domain blind source separation," in *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC '03)*, pp. 311–314, Kyoto, Japan, September 2003.
- [26] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Convolutional blind source separation for more than two sources in the frequency domain," *Acoustical Science and Technology*, vol. 25, no. 4, pp. 296–298, 2004.
- [27] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Frequency-domain blind source separation," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds., chapter 13, pp. 299–327, Springer, New York, NY, USA, 2005.
- [28] S. Makino, H. Sawada, R. Mukai, and S. Araki, "Blind source separation of convolutional mixtures of speech in frequency

domain,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 7, pp. 1640–1654, 2005, (Invited).

- [29] H. Sawada, S. Winter, R. Mukai, S. Araki, and S. Makino, “Estimating the number of sources for frequency-domain blind source separation,” in *Proceedings of 5th International Conference on Independent Component Analysis (ICA '04)*, vol. 3195 of *Lecture Notes in Computer Science*, pp. 610–617, Springer, Granada, Spain, September 2004.
- [30] E. Bingham and A. Hyvärinen, “A fast fixed-point algorithm for independent component analysis of complex valued signals,” *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, 2000.
- [31] S.-I. Amari, “Natural gradient works efficiently in learning,” *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.
- [32] H. Sawada, R. Mukai, S. Araki, and S. Makino, “Polar coordinate based nonlinear function for frequency-domain blind source separation,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E86-A, no. 3, pp. 590–596, 2003.
- [33] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, “Combined approach of array processing and independent component analysis for blind separation of acoustic signals,” *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 3, pp. 204–215, 2003.
- [34] M. Iwaki and A. Ando, “Selective microphone system using blind separation by block decorrelation of output signals,” in *Proceedings of the 4th International Conference on Independent Component Analysis and Blind Signal Separation (ICA '03)*, pp. 1023–1028, Nara, Japan, April 2003.
- [35] 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.
- [36] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley Interscience, New York, NY, USA, 2nd edition, 2000.
- [37] R. Mukai, H. Sawada, S. Araki, and S. Makino, “Blind source separation and DOA estimation using small 3-D microphone array,” in *Proceedings of the Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA '05)*, pp. d.9–10, Piscataway, NJ, USA, March 2005.

Ryo Mukai received the B.S. and the M.S. degrees in information science from the University of Tokyo, Japan, in 1990 and 1992, respectively. He joined NTT Corporation in 1992. From 1992 to 2000, he was engaged in research and development of processor architecture for network service systems and distributed network systems. Since 2000, he has been with NTT Communication Science Laboratories, where he is engaged in research of blind source separation. His current research interests include digital signal processing and its applications. He is a Senior Member of the IEEE, and a Member of the ACM, the Acoustical Society of Japan (ASJ), Institute of Electronics, Information and Communication Engineers (IEICE), and Information Processing Society of Japan (IPSJ). He is also a Member of the Technical Committee on Blind Signal Processing of the IEEE Circuits and Systems Society, and the Organizing Committee of the ICA 2003 in Nara. He is the Publications Chair of the IWAENC 2003 in Kyoto and the WASPAA 2007 in Mohonk.



He received the Sato Paper Award of the ASJ in 2005 and the Paper Award of the IEICE in 2005.

Hiroshi Sawada received the B.E., M.E., and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993, and 2001, respectively. In 1993, he joined NTT Communication Science Laboratories, where he is now a Senior Research Scientist. From 1993 to 2000, he was engaged in research on the computer-aided design of digital systems, logic synthesis, and computer architecture. Since 2000, he has been engaged in research on signal processing, microphone array, and blind source separation (BSS). More specifically, he is working on the frequency-domain BSS for acoustic convolutive mixtures using independent component analysis (ICA). He serves as an Associate Editor of the IEEE Transactions on Audio, Speech, and Language Processing. He is a Senior Member of the IEEE, and a Member of the Institute of Electronics, Information and Communication Engineers (IEICE), and the Acoustical Society of Japan (ASJ). He received the 9th TELECOM System Technology Award for Student from the Telecommunications Advancement Foundation in 1994, and the Best Paper Award of the IEEE Circuit and System Society in 2000.



Shoko Araki received the B.E. and the M.E. degrees in mathematical engineering and information physics from the University of Tokyo, Japan, in 1998 and 2000, respectively. In 2000, she joined NTT Communication Science Laboratories, Kyoto. Her research interests include array signal processing, blind source separation applied to speech signals, and auditory scene analysis. She received the TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2004, the Best Paper Award of the IWAENC in 2003, and the 19th Awaya Prize from Acoustical Society of Japan (ASJ) in 2001. She is a Member of the IEEE, IEICE, and the ASJ.



Shoji Makino received the B.E., M.E., and Ph.D. degrees from Tohoku University, Japan, in 1979, 1981, and 1993, respectively. He is an Executive Manager at the NTT Communication Science Laboratories. He is also a Guest Professor at the Hokkaido University. His research interests include blind source separation of convolutive mixtures of speech, adaptive filtering technologies, and realization of acoustic echo cancellation. He is the author or coauthor of more than 200 articles in journals and conference proceedings and has been responsible for more than 150 patents. He is a Member of both the Awards Board and the Conference Board of the IEEE SP Society. He is an Associate Editor of the IEEE Transactions on Speech and Audio Processing and an Associate Editor of the EURASIP Journal on Applied Signal Processing. He is a Member of the Technical Committee on Audio and Electroacoustics of the IEEE SP Society as well as the Technical Committee on Blind Signal Processing of the IEEE CAS Society. He is also the General Chair of the WASPAA 2007 in Mohonk, the Organizing Chair of the ICA2003 in Nara, the General Chair of the IWAENC2003 in Kyoto. He is an IEEE Fellow, a Council Member of the ASJ, and the Chair of the Technical Committee on Engineering Acoustics of the IEICE.

