

# A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation

Hiroshi Sawada, *Member, IEEE*, Ryo Mukai, *Member, IEEE*, Shoko Araki, *Member, IEEE*, and Shoji Makino, *Fellow, IEEE*

**Abstract**—Blind source separation (BSS) for convolutive mixtures can be solved efficiently in the frequency domain, where independent component analysis (ICA) is performed separately in each frequency bin. However, frequency-domain BSS involves a permutation problem: the permutation ambiguity of ICA in each frequency bin should be aligned so that a separated signal in the time-domain contains frequency components of the same source signal. This paper presents a robust and precise method for solving the permutation problem. It is based on two approaches: direction of arrival (DOA) estimation for sources and the interfrequency correlation of signal envelopes. We discuss the advantages and disadvantages of the two approaches, and integrate them to exploit their respective advantages. Furthermore, by utilizing the harmonics of signals, we make the new method robust even for low frequencies where DOA estimation is inaccurate. We also present a new closed-form formula for estimating DOAs from a separation matrix obtained by ICA. Experimental results show that our method provided an almost perfect solution to the permutation problem for a case where two sources were mixed in a room whose reverberation time was 300 ms.

**Index Terms**—Blind source separation (BSS), convolutive mixture, direction of arrival (DOA) estimation, frequency domain, independent component analysis (ICA), permutation problem, signal envelope.

## I. INTRODUCTION

**B**LIND SOURCE separation (BSS) [1] is a technique for estimating original source signals from their mixtures at sensors. Independent component analysis (ICA) [2], [3] is one of the major statistical tools used for solving this problem. If signals are mixed instantaneously, we can directly employ an instantaneous ICA algorithm to separate the mixed signals. In a real room environment, however, signals are mixed in a convolutive manner with reverberations. This makes the BSS problem difficult since we need a matrix of FIR filters, not just a matrix of scalars, to separate convolutively mixed signals. We need thousands of filter taps to separate acoustic signals mixed in a room. Many methods have been proposed to solve the convolutive BSS problem, and they can be classified into two approaches based on how we apply ICA.

The first approach is time-domain BSS, where ICA is applied directly to the convolutive mixture model [4]–[7]. The approach achieves good separation once the algorithm converges, since the ICA algorithm correctly evaluates the independence of separated signals. However, ICA for convolutive mixtures is not as simple as ICA for instantaneous mixtures, and computationally expensive for long FIR filters because it includes convolution operations.

The other approach is frequency-domain BSS, where complex-valued ICA for instantaneous mixtures is applied in each frequency bin [8]–[17]. The merit of this approach is that the ICA algorithm becomes simple and can be performed separately at each frequency. Also, any complex-valued instantaneous ICA algorithm can be employed with this approach. However, the permutation ambiguity of the ICA solution becomes a serious problem. We need to align the permutation in each frequency bin so that a separated signal in the time domain contains frequency components from the same source signal. This problem is well known as the permutation problem of frequency-domain BSS.

Some methods have been proposed where filter coefficients are updated in the frequency domain but nonlinear functions for evaluating independence are applied in the time domain [18]–[20]. There is also a frequency-domain implementation of time-domain BSS where time-domain convolution is speeded up by the overlap-save method [21], [22]. In either case, the permutation problem does not occur since the independence of separated signals is evaluated in the time domain. However, the algorithm moves back and forth between the two domains in every iteration, spending nonnegligible time for discrete Fourier transform (DFT) and inverse DFT. Therefore, we consider that the permutation problem is essential if we want to benefit from the merit of frequency-domain BSS mentioned above.

Various methods have been proposed for solving the permutation problem. Making separation matrices smooth in the frequency domain is one solution. This has been realized by averaging separation matrices with adjacent frequencies [8], limiting the filter length in the time domain [8], [16], [17], [22], or considering the coherency of separation matrices at adjacent frequencies [14]. Another approach is based on direction of arrival (DOA) estimation in array signal processing. By analyzing the directivity patterns formed by a separation matrix, source directions can be estimated and therefore, permutations can be aligned [9], [10]. If the sources are audio signals such as speech, we can employ the interfrequency correlations of output signal

Manuscript received August 29, 2003; revised April 3, 2004. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Walter Kellermann.

The authors are with NTT Communication Science Laboratories, NTT Corporation, Kyoto 619-0237, Japan (e-mail: sawada@cslab.kecl.ntt.co.jp; ryo@cslab.kecl.ntt.co.jp; shoko@cslab.kecl.ntt.co.jp; maki@cslab.kecl.ntt.co.jp).

Digital Object Identifier 10.1109/TSA.2004.832994

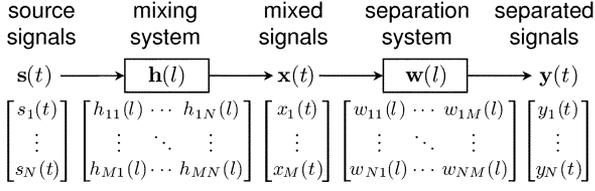


Fig. 1. BSS for convolutive mixtures.

envelopes to align the permutations [12], [13]. Each of these approaches has different characteristics, and may perform well under certain specific conditions but not others.

We consider that integrating some of these approaches is one way of obtaining better performance. In this paper, we propose a new method for solving the permutation problem robustly and precisely by integrating two of the approaches outlined above. The first is the DOA approach, which is discussed in Section III-A, as this will provide the new method with robustness. The second is based on interfrequency correlations of output signal envelopes, which is discussed in Section III-B, and will make the new method precise. The proposed method is described in Section IV. Experimental results are reported in Section V and are very promising.

As another contribution, we propose a method of estimating the direction of sources analytically in Section IV-C. Unlike conventional methods [9], [10], this method does not require the calculation of directivity patterns. Instead, it calculates the directions of target signals directly from an estimated mixing matrix, which is basically the inverse of a frequency-domain separation matrix obtained by ICA. This method can estimate the directions of more than two sources, thus enabling us to separate more than two sources practically by frequency-domain BSS.

## II. BSS FOR CONVOLUTIVE MIXTURES

### A. Problem Formulation

Fig. 1 shows the block diagram of BSS. Suppose that  $N$  source signals  $s_k(t)$  are mixed and observed at  $M$  sensors

$$x_j(t) = \sum_{k=1}^N \sum_l h_{jk}(l) s_k(t-l) \quad (1)$$

where  $h_{jk}(l)$  represents the impulse response from source  $k$  to sensor  $j$ . We assume that the number of sources  $N$  is known or can be estimated in some way, and the number of sensors  $M$  is more than or equal to  $N$  ( $N \leq M$ ). The goal is to separate the mixtures  $x_j(t)$  and to obtain a filtered version of a source  $s_{\Pi(i)}(t)$  at each output

$$y_i(t) = \sum_l \alpha_i(l) s_{\Pi(i)}(t-l) \quad (2)$$

where  $\alpha_i(l)$  is a filter and  $\Pi: \{1, \dots, N\} \rightarrow \{1, \dots, N\}$  is a permutation. The separation system typically consists of a set of FIR filters  $w_{ij}(l)$  of length  $L$  to produce separated signals

$$y_i(t) = \sum_{j=1}^M \sum_{l=0}^{L-1} w_{ij}(l) x_j(t-l). \quad (3)$$

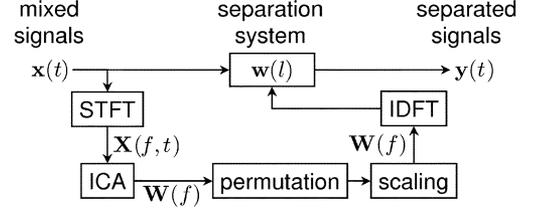


Fig. 2. Flow of frequency-domain BSS.

The filter  $\alpha_i(l)$  and the permutation  $\Pi$  in (2) represents the scaling and the permutation ambiguity of BSS, respectively. We assume that the permutation ambiguity is decided based on the directions of sources estimated by the method discussed in this paper. Thus, let  $\Pi$  be identity mapping to have a source  $s_i(t)$  at output  $i$  for simplicity. As for the scaling ambiguity, it is desirable to obtain just a delayed version, not a filtered version (2), of  $s_i(t)$  at the output  $y_i(t)$ . However, it is very difficult to achieve this with the ICA scheme unless  $s_i(t)$  is white, which is not the case for separating natural sounds such as speech [7]. Hence, we allow a filter

$$\alpha_i(l) = h_{ii}(l) \quad (4)$$

following the minimal distortion principle (MDP) [6].

The separation system can be analyzed by using the impulse responses from a source  $s_k(t)$  to a separated signal  $y_i(t)$

$$u_{ik}(l) = \sum_{j=1}^M \sum_{\tau=0}^{L-1} w_{ij}(\tau) h_{jk}(l-\tau). \quad (5)$$

The separation performance is evaluated by using a signal-to-interference ratio (SIR). It is calculated as the ratio of the power of a target component  $\sum_l u_{ii}(l) s_i(t-l)$  and interference components  $\sum_{k \neq i} \sum_l u_{ik}(l) s_k(t-l)$ .

### B. Frequency-Domain BSS

We employ frequency-domain BSS where ICA is applied separately in each frequency bin to obtain the frequency responses  $W_{ij}(f)$  of separation filters  $w_{ij}(l)$  [8]–[15]. Fig. 2 shows the flow. First, time-domain signals  $x_j(t)$  are converted into frequency-domain time-series signals  $X_j(f, t)$  with an  $L$ -point short-time Fourier transform (STFT)

$$X_j(f, t) = \sum_{\tau=-\frac{L}{2}}^{\frac{L}{2}-1} x_j(\tau+t) \text{win}(\tau) e^{-j2\pi f\tau} \quad (6)$$

where  $f$  is one of  $L$  frequencies  $f = 0, (1/L)f_s, \dots, ((L-1)/L)f_s$  ( $f_s$ : the sampling frequency),  $\text{win}(\tau)$  is a window that tapers smoothly to zero at each end, such as a Hanning window  $(1/2)(1 + \cos(2\pi\tau/L))$ , and  $t$  is now down-sampled with the distance of the window shift.

Then, to obtain the frequency responses  $W_{ij}(f)$  of filters  $w_{ij}(l)$ , complex-valued ICA

$$\mathbf{Y}(f, t) = \mathbf{W}(f) \mathbf{X}(f, t) \quad (7)$$

is solved, where  $\mathbf{X}(f, t) = [X_1(f, t), \dots, X_M(f, t)]^T$ ,  $\mathbf{Y}(f, t) = [Y_1(f, t), \dots, Y_N(f, t)]^T$ , and  $\mathbf{W}(f)$  is an  $N \times M$

separation matrix whose elements are  $W_{ij}(f)$ . If we have more sensors than sources ( $N < M$ ), principal component analysis (PCA) is typically performed as a preprocessing of ICA [23] so that the  $N$  dimensional subspace spanned by the row vectors of  $\mathbf{W}(f)$  is almost identical to the signal subspace.

One of the advantages of frequency-domain BSS is that we can employ any ICA algorithm for instantaneous mixtures, such as the information maximization approach [24] combined with the natural gradient [25], FastICA [26], JADE [27], or an algorithm based on the nonstationarity of signals [28]. We use the information maximization approach combined with the natural gradient in this paper. The separation matrix  $\mathbf{W}$  is improved by the learning rule

$$\Delta \mathbf{W} = \mu [\mathbf{I} - \langle \Phi(\mathbf{Y})\mathbf{Y}^H \rangle_t] \mathbf{W} \quad (8)$$

where  $\mu$  is a step-size parameter,  $\langle \cdot \rangle_t$  denotes the averaging operator over time, and  $\Phi(\cdot)$  is an element-wise nonlinear function for a complex signal  $Y_i = |Y_i|e^{j\arg(Y_i)}$ . We use

$$\Phi(Y_i) = -\frac{\partial}{\partial |Y_i|} \log p(|Y_i|) e^{j\arg(Y_i)} \quad (9)$$

as a nonlinear function assuming that the density  $p(Y_i)$  is independent of the argument of  $Y_i$  [15]. Note that the subspace identified by the PCA for an  $N < M$  case is not changed by the update (8).

The ICA solution in each frequency bin has permutation and scaling ambiguity: even if we permute the rows of  $\mathbf{W}(f)$  or multiply a row by a constant, it is still an ICA solution. In matrix notation,  $\mathbf{\Lambda}(f)\mathbf{P}(f)\mathbf{W}(f)$  is also an ICA solution for any permutation  $\mathbf{P}(f)$  and diagonal  $\mathbf{\Lambda}(f)$  matrix. The permutation matrix  $\mathbf{P}(f)$  should be decided so that  $Y_i(f, t)$  at all frequencies correspond to the same source  $s_i(t)$  by the update of the mixing matrix

$$\mathbf{W}(f) \leftarrow \mathbf{P}(f)\mathbf{W}(f). \quad (10)$$

This is the permutation problem, which is the main topic of this paper. The scaling ambiguity can be decided so that the MDP [6] is realized in the frequency-domain [13]. Let  $\mathbf{H}(f)$  be the unknown mixing matrix. Considering (4), the diagonal matrix  $\mathbf{\Lambda}(f)$  should satisfy

$$\mathbf{\Lambda}(f)\mathbf{W}(f)\mathbf{H}(f) = \text{diag}[\mathbf{H}(f)]. \quad (11)$$

Although  $\mathbf{H}(f)$  is unknown, there is another diagonal matrix  $\mathbf{D}(f)$  that satisfies  $\mathbf{W}(f)\mathbf{H}(f) = \mathbf{D}(f)$  if ICA is successfully solved. Thus,  $\mathbf{H}(f)$  can be estimated up to scaling by  $\mathbf{H}(f) = \mathbf{W}^{-1}(f)\mathbf{D}(f)$ . By substituting this estimation with (11), we have  $\mathbf{\Lambda}(f) = \text{diag}[\mathbf{W}^{-1}(f)]$ . The scaling ambiguity is therefore, decided by

$$\mathbf{W}(f) \leftarrow \text{diag}[\mathbf{W}^{-1}(f)] \mathbf{W}(f). \quad (12)$$

If  $N < M$ , the Moore-Penrose pseudoinverse  $\mathbf{W}^+(f)$  is used instead of  $\mathbf{W}^{-1}(f)$  (see Appendix A). Finally, separation filters  $w_{ij}(l)$  are obtained by applying inverse DFT to  $W_{ij}(f)$ .

### III. TWO EXISTING APPROACHES

This section discusses the two approaches that are integrated into our new method for solving the permutation problem.

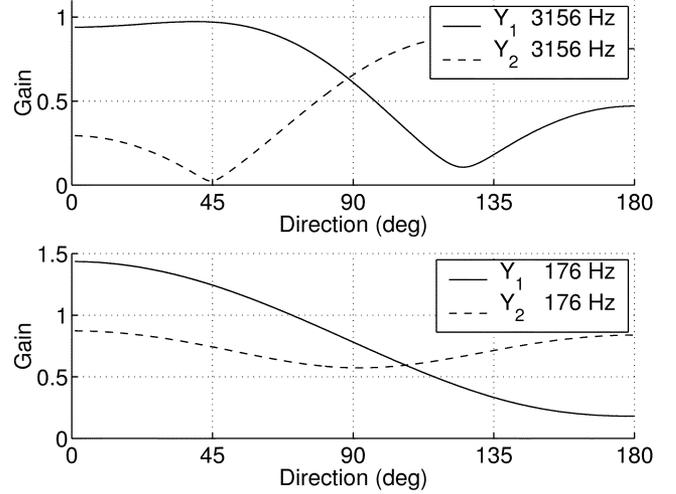


Fig. 3. Directivity patterns for two sources.

#### A. Direction of Arrival (DOA) Approach

We first discuss the DOA approach where the directions of source signals are estimated and permutations are aligned based on them. If half the wavelength of a frequency is longer than the sensor spacing, there is no spatial aliasing. In most such cases, each row of  $\mathbf{W}(f)$  forms spatial nulls in the directions of jammer signals and extracts a target signal in another direction [11]. Once we have estimated the directions  $\Theta(f) = [\theta_1(f), \dots, \theta_N(f)]^T$  of target signals extracted by every row of  $\mathbf{W}(f)$ , we can obtain a permutation matrix  $\mathbf{P}(f)$  by sorting or clustering  $\Theta(f)$ .

Now, we review the method [9], [10] that estimates the directions of sources and aligns permutations by plotting the directivity pattern of each output  $Y_i(f, t)$ . Let  $d_j$  be the position of sensor  $x_j$  (we assume linearly arranged array sensors), and  $\theta_k$  be the direction of source  $s_k$  (the direction orthogonal to the array is  $90^\circ$ ). In beamforming theory [29], the frequency response of an impulse response  $h_{jk}(t)$  is approximated as

$$H_{jk}(f) = e^{j2\pi f c^{-1} d_j \cos \theta_k} \quad (13)$$

where  $c$  is the propagation velocity. In this approximation, we assume a plane wavefront and no reverberation. The frequency response  $U_{ik}(f)$  of (5) can be expressed as  $U_{ik}(f) = \sum_{j=1}^M W_{ij}(f) H_{jk}(f) = \sum_{j=1}^M W_{ij}(f) \exp[j2\pi f c^{-1} d_j \cos \theta_k]$ . If we regard  $\theta_k$  as a variable  $\theta$ , the formula is expressed as the following:

$$U_i(f, \theta) = \sum_{j=1}^M W_{ij}(f) e^{j2\pi f c^{-1} d_j \cos \theta}. \quad (14)$$

This formula changes according to the direction  $\theta$ , and is thus called a directivity pattern.

Fig. 3 shows the gain  $|U_i(f, \theta)|$  of directivity patterns for two sources mixed under the conditions shown in Table I. The upper part (3156 Hz) shows that output  $Y_1$  extracts a source signal originating from around  $45^\circ$  and suppresses the other signal coming from around  $125^\circ$ , which is called a null direction. A null direction is obtained by searching a directivity pattern for the minimum. With a similar consideration regarding  $Y_2$ , we estimate the directions  $\Theta(3156) = [45^\circ, 125^\circ]^T$  of the target

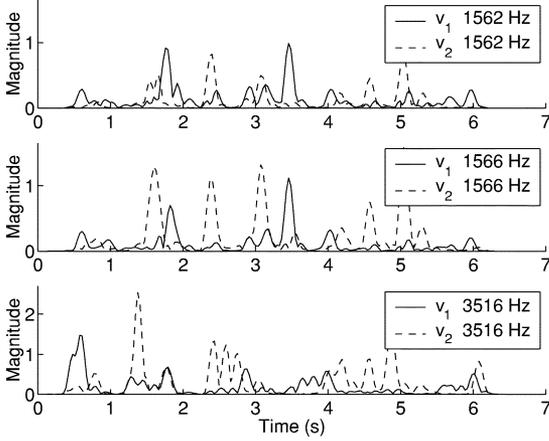


Fig. 4. Envelopes of two output signals at different frequencies.

signals. Even if the approximation (13) was used for the reverberant condition, the estimation was good enough to decide the permutation. However, not every frequency bin gives us such an ideal directivity pattern. The lower part of Fig. 3 is the pattern at a low frequency (176 Hz). We see that the null is not well formed for  $Y_1$  and the null of  $Y_2$  is in an obscure direction. In fact, we cannot estimate  $\Theta(176)$  or decide a permutation for this frequency with confidence.

There are three problems with this method:

- 1) directions of arrival cannot be well estimated at some frequencies, especially at low frequencies where the phase difference caused by the sensor spacing is very small, and also at high frequencies where spatial aliasing might occur;
- 2) calculation of null directions by plotting directivity patterns is time consuming;
- 3) estimating DOAs from null directions is difficult when there are more than two sources.

The first problem reveals the limitation of the DOA approach, and will be solved in Sections IV-A and IV-B. The other two problems are caused by using directivity patterns, and will be solved in Section IV-C.

### B. Correlation Approach

We discuss an approach to permutation alignment based on interfrequency correlations of signals [12], [13]. We use the envelope

$$v_i^f(t) = |Y_i(f, t)| \quad (15)$$

of a separated signal  $Y_i(f, t)$  to measure the correlations. We define the correlation of two signals  $x(t)$  and  $y(t)$  as

$$\text{cor}(x, y) = \frac{(\mu_{x \cdot y} - \mu_x \cdot \mu_y)}{(\sigma_x \cdot \sigma_y)} \quad (16)$$

where  $\mu_x$  is the mean and  $\sigma_x$  is the standard deviation of  $x$ . Based on this definition,  $\text{cor}(x, x) = 1$ , and  $\text{cor}(x, y) = 0$  if  $x$  and  $y$  are uncorrelated. Envelopes have high correlations at neighboring frequencies if separated signals correspond to the same source signal. Fig. 4 shows an example. Two envelopes

$v_1^{1562}$  and  $v_1^{1566}$ , as well as  $v_2^{1562}$  and  $v_2^{1566}$ , are highly correlated. Thus, calculating such correlations helps us to align permutations.

Let  $\Pi_f$  be a permutation corresponding to the inverse  $\mathbf{P}^{-1}(f)$  of the permutation matrix of (10). A simple criterion for deciding  $\Pi_f$  is to maximize the sum of the correlations between neighboring frequencies within distance  $\delta$

$$\Pi_f = \underset{\Pi}{\text{argmax}} \sum_{|g-f| \leq \delta} \sum_{i=1}^N \text{cor} \left( v_{\Pi(i)}^f, v_{\Pi_g(i)}^g \right) \quad (17)$$

where  $\Pi_g$  is the permutation at frequency  $g$ . This criterion is based on local information and has a drawback in that mistakes in a narrow range of frequencies may lead to the complete misalignment of the frequencies beyond the range. To avoid this problem, the method in [13] does not limit the frequency range in which correlations are calculated. It decides permutations one by one based on the criterion

$$\Pi_f = \underset{\Pi}{\text{argmax}} \sum_{i=1}^N \text{cor} \left( v_{\Pi(i)}^f, \sum_{g \in \mathcal{F}} v_{\Pi_g(i)}^g \right) \quad (18)$$

where  $\mathcal{F}$  is a set of frequencies in which the permutation is decided. This method assumes high correlations of envelopes even between frequencies that are not close neighbors. This assumption is not satisfied for all pairs of frequencies, although a high correlation can be assumed for a fundamental frequency and its harmonics. As shown in Fig. 4,  $v_i^{1566}$  and  $v_i^{3516}$  do not have a high correlation. Therefore, this method still has a drawback in that permutations may be misaligned at many frequencies.

## IV. NEW ROBUST AND PRECISE METHOD

This section presents our new method that integrates the two approaches discussed above to solve the permutation problem robustly and precisely.

### A. Basic Idea of the Method

We begin by reviewing the characteristics of the two existing approaches.

- **Robustness:** The DOA approach is robust since a misalignment at a frequency does not affect other frequencies. The correlation approach is not robust since a misalignment at a frequency may cause consecutive misalignments.
- **Preciseness:** The DOA approach is not precise since the evaluation is based on an approximation of a mixing system. The correlation approach is precise as long as signals are well separated by ICA since the measurement is based on separated signals.

To benefit from both advantages, namely the robustness of the DOA approach and the preciseness of the correlation approach, our method basically solves the permutation problem in the two following steps:

- 1) fix the permutations at some frequencies where the confidence of the DOA approach is sufficiently high;
- 2) decide the permutations for the remaining frequencies based on neighboring correlations (17) without changing the permutations fixed by the DOA approach.

```

/* Fix permutations by the DOA approach */
for (∀ f) {
  Θ(f) = DOA(f, W(f))
  Θs(f) = sort(Θ(f))
}
Θ̄s = ⟨Θs(f)⟩f /* averaged directions */
F = ∅ /* the set of fixed frequencies */
for (∀ f) {
  if (confident(Θ(f), Θ̄s, W(f))) {
    Πf = getPermutation(Θ(f))
    F = F ∪ {f}
  }
}
/* Fix permutations by neighboring correlations */
while (∃ f ∉ F) {
  for (∀ f ∉ F) {
    Rf = maxΠ ∑|g-f|≤δ, g∈F ∑i=1N cor(vΠ(i)f, vΠg(i)g)
    Πf = argmaxΠ ∑|g-f|≤δ, g∈F ∑i=1N cor(vΠ(i)f, vΠg(i)g)
  }
  ψ = argmaxf Rf
  F = F ∪ {ψ}
  Rψ = 0
}

```

Fig. 5. Pseudo-code for the first version of the proposed method.

Fig. 5 shows the pseudo-code. A set  $\mathcal{F}$  contains frequencies where the permutation is decided. The key point in the first DOA approach is that we fix a permutation only if the confidence of the permutation is sufficiently high. The procedure *confident* decides whether the confidence is high enough. Our criteria for the decision are the following:

- 1) number of estimated directions is the same as the number of sources;
- 2) directions  $\Theta_s(f)$  do not differ greatly from the averaged directions  $\bar{\Theta}_s$ , i.e.  $|\Theta_s(f) - \bar{\Theta}_s|$  is smaller than a threshold  $th_{\Theta}$ ;
- 3) SIR calculated by using (14) is sufficiently large, i.e.  $10 \log_{10} |U_i(f, \theta_i(f))|^2 - 10 \log_{10} \sum_{k \neq i} |U_i(f, \theta_k(f))|^2$  is larger than a threshold  $th_U$ .

In the second step, permutations are decided one by one for the frequencies where the permutation is not fixed. The measurement for deciding permutations is given by the sum of correlations with fixed frequencies  $g \in \mathcal{F}$  within distance  $|g - f| \leq \delta$ .

This method does not cause a large misalignment as long as the permutations fixed by the DOA approach are correct. Moreover, the correlation part compensates for the lack of preciseness of the DOA approach.

### B. Exploiting the Harmonic Structure of Signals

The method proposed above works very well in many cases. However, there is a case where the DOA approach does not provide any fixed permutation with confidence in a certain range of frequencies. This occurs particularly at low frequencies where it is hard to estimate DOAs as discussed in Section III-A. In such a case, the proposed method has to align permutations for the range solely through the use of neighboring correlations, and may yield consecutive misalignments.

To cope with this problem, we exploit the harmonic structure of a signal. As alluded to in Section III-B, there are strong correlations between the envelopes of a fundamental frequency  $f$

```

/* Fix permutations by harmonic structure */
for (∀ f ∉ F) {
  Ha = setOfHarmonics(f);
  Rf = maxΠ ∑g∈Ha∩F ∑i=1N cor(vΠ(i)f, vΠg(i)g)
  if (Rf ≥ thha) {
    Πf = argmaxΠ ∑g∈Ha∩F ∑i=1N cor(vΠ(i)f, vΠg(i)g)
    F = F ∪ {f}
  }
}

```

Fig. 6. Pseudo-code for the harmonic part of the proposed method.

and its harmonics  $2f$ ,  $3f$ , and so forth. Suppose that the permutation is not fixed at frequency  $f$  but fixed at its harmonics. If the correlation

$$R_f = \max_{\Pi} \sum_{g=\text{setOfHarmonics}(f)} \sum_{i=1}^N \text{cor}(v_{\Pi(i)}^f, v_{\Pi_g(i)}^g) \quad (19)$$

is larger than a threshold  $th_{ha}$ , we fix the permutation at frequency  $f$  with confidence. Fig. 6 shows the pseudo-code for the harmonic part. The procedure *setOfHarmonics*( $f$ ) provides a set of harmonic frequencies of  $f$ .

To incorporate the above idea, the final version of our method fixes all permutations with four steps:

- Step 1) by the DOA approach (the upper part of Fig. 5);
- Step 2) by neighboring correlations (the lower part of Fig. 5) with the exception that the while loop terminates if the maximum  $R_f$  is smaller than a threshold  $th_{cor}$ ;
- Step 3) by the harmonic method (Fig. 6);
- Step 4) by neighboring correlations (the lower part of Fig. 5) again without the exception.

There are two important points as regards the final version. The first is that the method becomes more robust because of the exception in Step 2. We do not fix the permutations for consecutive frequencies without high confidence. The second point is that Step 3 works well only if most of the other permutations are fixed. This means that the harmonic method alone does not work well and we need Steps 1 and 2 to fix most of the permutations.

### C. Closed-Form Formula for Estimating DOAs

The DOA estimation method reviewed in Section III-A has two problems, a high computational cost and the difficulty of using it for mixtures of more than two sources. Instead of plotting directivity patterns and searching for the minimum as a null direction, we propose a new method of estimating the directions  $\Theta(f)$  of source signals. In principle, this method can be applied to any number of source signals.

It starts by estimating the frequency response  $\mathbf{H}(f)$  of the mixing system from a separation matrix  $\mathbf{W}(f)$  obtained by ICA. If the ICA is successfully solved, there are a permutation matrix  $\mathbf{P}(f)$  and a diagonal matrix  $\mathbf{\Lambda}(f)$  that satisfy  $\mathbf{\Lambda}(f)\mathbf{P}(f)\mathbf{W}(f)\mathbf{H}(f) = \mathbf{I}$ . Thus,  $\mathbf{H}(f)$  can be estimated by  $\mathbf{H} = \mathbf{W}^{-1}\mathbf{P}^{-1}\mathbf{\Lambda}^{-1}$  up to permutation and scaling ambiguities: the  $\mathbf{H}(f)$  columns can be permuted arbitrarily and have arbitrary scales compared with the real frequency response. Again, if  $N < M$ , the Moore–Penrose pseudoinverse  $\mathbf{W}^+$  is used instead of  $\mathbf{W}^{-1}$  (see Appendix A). It should be noted

TABLE I  
EXPERIMENTAL CONDITIONS

Reverberation time	$T_R = 300$ ms
Source signal	speech of 6 s
Direction of sources	$120^\circ$ and $50^\circ$ (2 sources)
Distance between 2 sensors	$d = 4$ cm
Sampling rate	$f_s = 8$ kHz
Frame size of DFT	$L = 2048$
Frequency resolution	$\Delta f = f_s/L = 3.90625$ Hz
Distance to take correlation	$\delta = 3 \cdot \Delta f$
Harmonics to take correlation	$Ha = \{2f - \Delta f, 2f, 2f + \Delta f, 3f - \Delta f, 3f, 3f + \Delta f\}$
Thresholds for the DOA part	$th_\Theta = 1.5 \cdot \sigma_\Theta$ (standard deviation) $th_U = 10$ dB
Threshold for the harmonic part	$th_{ha} = 0.1 \cdot N \cdot  Ha $
Step size	$\mu = 0.1$
Number of iterations for (8)	100
Nonlinear function (9)	$\Phi(Y_i) = e^{j \cdot \arg(Y_i)}$

that the scaling ambiguity is canceled out by calculating the ratio between two elements  $H_{jk}(f)$  and  $H_{j'k}(f)$  of the same column of  $\mathbf{H}(f)$

$$\frac{H_{jk}}{H_{j'k}} = \frac{[\mathbf{W}^{-1}\mathbf{P}^{-1}\mathbf{\Lambda}^{-1}]_{jk}}{[\mathbf{W}^{-1}\mathbf{P}^{-1}\mathbf{\Lambda}^{-1}]_{j'k}} = \frac{[\mathbf{W}^{-1}]_{j\Pi(k)}}{[\mathbf{W}^{-1}]_{j'\Pi(k)}} \quad (20)$$

where  $\Pi$  is the permutation corresponding to postmultiplying  $\mathbf{P}^{-1}$ .

An element  $H_{jk}(f)$  of the matrix  $\mathbf{H}(f)$  obtained in the above manner may have an arbitrary amplitude. Since the approximation (13) of the mixing system does not suit this situation, we remodel the mixing system with attenuation  $A_{jk}$  (real-valued) and phase modulation  $\exp[j\varphi_k]$  at the origin:

$$H_{jk}(f) = A_{jk} e^{j\varphi_k} e^{j2\pi f c^{-1} d_j \cos \theta_k}. \quad (21)$$

From (20) and (21), we have

$$\frac{[\mathbf{W}^{-1}]_{j\Pi(k)}}{[\mathbf{W}^{-1}]_{j'\Pi(k)}} = \frac{A_{jk}}{A_{j'k} e^{j2\pi f c^{-1} (d_j - d_{j'}) \cos \theta_k}}. \quad (22)$$

Then, taking the argument yields a formula for estimating  $\theta_k$

$$\theta_k = \arccos \frac{\arg \left( \frac{[\mathbf{W}^{-1}]_{j\Pi(k)}}{[\mathbf{W}^{-1}]_{j'\Pi(k)}} \right)}{2\pi f c^{-1} (d_j - d_{j'})}. \quad (23)$$

If the absolute value of the input variable of arccos is larger than 1,  $\theta_k$  becomes complex and no direction is obtained. In this case, formula (23) can be tested with another pair  $j$  and  $j'$ . By calculating  $\theta_k$  for all  $k = 1, \dots, N$ , we can obtain the directions of all source signals whatever permutation  $\Pi$  may be.

The new method offers an advantage in terms of computational cost. Estimated directions are provided by the closed-form formula (23), whereas the minima of  $|U_i(f, \theta)|$  should be searched for with the previous method using directivity patterns (14). For a two-source case, we prove that  $\theta_k$  calculated by the above formula is the same as a null direction that is the minimum of directivity patterns (see Appendix B).

## V. EXPERIMENTAL RESULTS

We performed experiments to separate speech signals in a reverberant environment whose conditions are summarized in Table I. The sensor spacing was selected so that there was no

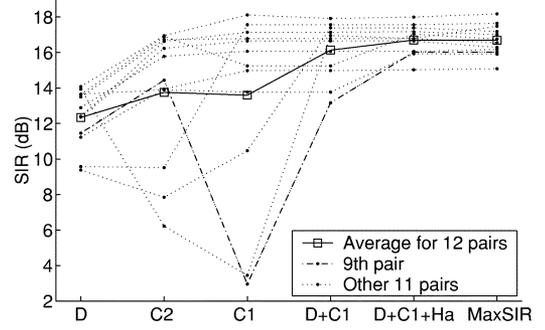


Fig. 7. Separation results for 12 pairs of speech signals with six different methods for the permutation problem: the DOA approach “D,” the correlation approach “C2” based on (18), the correlation approach “C1” based on (17), the first version “D+C1” of the proposed method, the final version “D+C1+Ha” of the proposed method, and the permutations maximizing the SIR at each frequency “MaxSIR” (see Appendix C). Although “MaxSIR” is not a realistic solution, it gives a rough estimate of the upper bound of performance.

TABLE II  
COMPUTATIONAL TIME (S)

STFT	ICA	Permutation alignment					
		D	C2	C1	D+C1	D+C1+Ha	MaxSIR
0.18	11.72	0.07	0.39	0.48	0.60	0.63	0.43

spatial aliasing for any frequency. We generated mixed signals by convolving speech signals  $s_k(t)$  and impulse responses  $h_{jk}(t)$  so that we could calculate SIRs defined in Section II-A. Fig. 7 shows the overall separation results in terms of SIR. We separated 12 pairs of speech signals with six different methods for the permutation problem, as explained in the caption. Table II shows the computational time for STFT, ICA and each of the six different methods used for permutation alignment. They are for source signals of 6 s, and averaged over the 12 pairs. The BSS program was coded in Matlab and run on Athlon XP 3200+.

The performance with “D” is stable, but not sufficient. The results with “C1” and “C2” are not stable and sometimes very poor, although most of the time they are very good. Both proposed methods “D + C1” and “D + C1 + Ha” offer good stable results. In particular, the method exploiting the harmonic structure “D + C1 + Ha” offers almost the same results as “MaxSIR.” As regards computational time, method “D,” where DOA is calculated by (23) instead of plotting directivity patterns, is very fast. The other methods, including both proposed methods, can be performed in a feasible computational time.

Now we examine the effectiveness of the proposed methods by looking at the ninth pair of speech signals in detail. Fig. 8 shows the SIRs at each frequency for “C1,” “D + C1” and “D + C1 + Ha.” We see a large region (from 450 to 1400 Hz) of permutation misalignments for the “C1” case, where permutations were decided only with neighboring correlations. Fig. 9 shows the difference between the correlation sums  $\sum_{i=1}^N \text{cor}(v_{\Pi_f(i)}^f, v_{\Pi_g(i)}^g)$  for different permutations. We see that the difference is very small around 1400 Hz, and the criterion based on (17) does not provide a clear-cut decision. Therefore, the risk of the permutations being misaligned is very high around 1400 Hz only with (17) in this case.

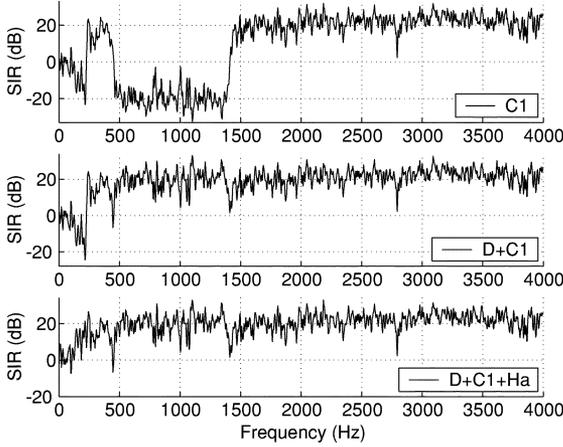


Fig. 8. SIRs measured at frequencies. The permutation problems were solved by three different methods, the correlation approach “C1,” the first version “D+C1” of the proposed method, and the final version “D+C1+Ha.”

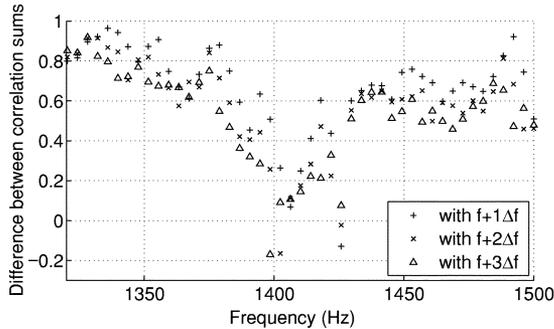


Fig. 9. Differences between the correlation sums of two possible permutations,  $[\text{cor}(v_1^f, v_1^g) + \text{cor}(v_2^f, v_2^g)] - [\text{cor}(v_1^f, v_2^g) + \text{cor}(v_2^f, v_1^g)]$ , with neighboring frequencies  $g = f + 1\Delta f, f + 2\Delta f, f + 3\Delta f$ .

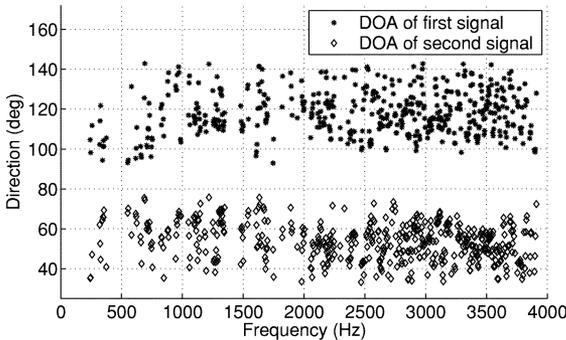


Fig. 10. DOA estimations with confidence.

With the “D + C1” method, the misalignments of the region (from 450 to 1400 Hz) were corrected. This is because the DOA approach provided correct permutations for some frequencies in the region. Fig. 10 shows the DOA estimations for each frequency with confidence. We see many estimations from 450 to 1400 Hz. However, there was no DOA estimation with confidence at frequencies lower than 250 Hz. This is why consecutive misalignments occurred even for “D + C1.” As shown at the bottom of Fig. 8, the misalignments were corrected with the “D + C1 + Ha” method. This shows the effectiveness of exploiting the harmonic structure for low frequencies.

Although the results shown here are only for two sources, the method can also be applied for more than two sources. In [30], the separation performance for four sources was compared. The results corresponding to “D + C1 + Ha” were satisfactory and superior to the others. In [31], six sources were separated with a planar array of eight sensors. Again, the separation performance is superior when both the DOA approach and the correlation approach were used.

## VI. CONCLUSION

We have proposed a robust and precise method for solving the permutation problem. Our method effectively integrates two approaches: the DOA approach and the correlation approach. The criterion of the DOA approach is based on directions that are absolute. This makes the approach robust. By contrast, the criterion of the correlation approach is calculated from the separated signals themselves. This makes the approach precise. Our proposed method benefits from both advantages. In the experiments, the proposed method solved permutation problems almost perfectly under the conditions shown in Table I. The method even performs well for more than two sources [30], [31]. We consider that the proposed method has expanded the applicability of frequency-domain BSS.

## APPENDIX A

### ESTIMATING THE MIXING MATRIX FOR AN $N < M$ CASE

As discussed in Section II-B and IV-C, estimating the mixing matrix  $\mathbf{H}$  up to scaling and permutation by the inverse  $\mathbf{W}^{-1}$  is very useful in frequency-domain BSS. When the number of sensors  $M$  is larger than the number of sources  $N$  ( $N < M$ ), the Moore–Penrose pseudoinverse  $\mathbf{W}^+$  is used instead of  $\mathbf{W}^{-1}$ . This appendix discusses the condition where this operation gives a proper estimation of the mixing matrix.

If ICA is solved, there is a permutation matrix  $\mathbf{P}$  and a diagonal matrix  $\mathbf{\Lambda}$  that satisfy  $\mathbf{\Lambda P W H} = \mathbf{I}$ . If  $N = M$ ,  $\mathbf{H}$  is uniquely given by  $\mathbf{H} = \mathbf{W}^{-1} \mathbf{P}^{-1} \mathbf{\Lambda}^{-1}$ . However, if  $N < M$ , there are an infinite number of solutions  $\mathbf{H}$  for  $\mathbf{W H} = \mathbf{P}^{-1} \mathbf{\Lambda}^{-1}$ . Among them, the solution  $\mathbf{H} = \mathbf{W}^+ \mathbf{P}^{-1} \mathbf{\Lambda}^{-1}$  realized by the Moore–Penrose pseudoinverse has a special property: the subspace  $\mathcal{C}(\mathbf{H})$  spanned by the  $N$  column vectors of  $\mathbf{H}$  is identical to the subspace  $\mathcal{R}(\mathbf{W})$  spanned by the  $N$  row vectors of  $\mathbf{W}$  [32]. Therefore, if the subspace  $\mathcal{R}(\mathbf{W})$  is properly selected,  $\mathbf{H}$  can be used as an estimation of the mixing matrix up to scaling and permutation. Otherwise,  $\mathbf{H}$  does not give a good estimation, and the frequency-domain version of MDP (12) and the DOA estimation (23) may fail. It is safe to employ PCA to decide the subspace  $\mathcal{R}(\mathbf{W})$  as described in Section II-B, since it is almost identical to the subspace spanned by the  $N$  column vectors of the mixing matrix.

## APPENDIX B

### EQUIVALENCE BETWEEN $\theta_k$ AND A NULL DIRECTION

For a two-source case, we prove that  $\theta_k$  calculated by (23) is the same as a null direction that is the minimum of a directivity pattern (14). When  $|U_i(f, \theta)|$  is minimized,  $\theta$  corresponds to a

null direction. Let  $\varphi_j = 2\pi f c^{-1} d_j$  and  $f$  be omitted in (14). The value to be minimized is

$$\begin{aligned} J(\theta) &= U_i(\theta) \cdot U_i(\theta)^* \\ &= (W_{i1} e^{j\varphi_1 \cos \theta} + W_{i2} e^{j\varphi_2 \cos \theta}) \\ &\quad \cdot (W_{i1}^* e^{-j\varphi_1 \cos \theta} + W_{i2}^* e^{-j\varphi_2 \cos \theta}). \end{aligned} \quad (24)$$

Let  $\varphi = \varphi_2 - \varphi_1$ . The first and second derivatives are

$$\frac{dJ}{d\theta} = -\varphi \sin \theta \cdot 2\text{Im}(W_{i1} W_{i2}^* e^{-j\varphi \cos \theta}) \quad (25)$$

$$\begin{aligned} \frac{d^2 J}{d\theta^2} &= -\varphi \cos \theta \cdot 2\text{Im}(W_{i1} W_{i2}^* e^{-j\varphi \cos \theta}) \\ &\quad - \varphi^2 \sin^2 \theta \cdot 2\text{Re}(W_{i1} W_{i2}^* e^{-j\varphi \cos \theta}) \end{aligned} \quad (26)$$

where Re and Im extract the real and imaginary part of a complex, respectively. If  $\arg(W_{i1} W_{i2}^* e^{-j\varphi \cos \theta}) = \pi$ ,  $dJ/d\theta$  is zero and  $d^2 J/d\theta^2$  is positive, and  $J(\theta)$  is minimized. Thus, the null direction formed by the  $i$ -th row of  $\mathbf{W}$  is given by

$$\begin{aligned} \arg(-W_{i1} W_{i2}^*) &= \varphi \cos \theta_i^{null} \Leftrightarrow \\ \theta_i^{null} &= \arccos \frac{\arg\left(-\frac{W_{i1}}{W_{i2}}\right)}{2\pi f c^{-1} (d_2 - d_1)}. \end{aligned} \quad (27)$$

Considering  $[\mathbf{W}^{-1}]_{21} = -W_{21}/\det(\mathbf{W})$  and  $[\mathbf{W}^{-1}]_{11} = W_{22}/\det(\mathbf{W})$ , we see that  $\theta_1$  and  $\theta_2^{null}$  are the same:

$$\begin{aligned} \theta_1 &= \arccos \frac{\arg\left(\frac{[\mathbf{W}^{-1}]_{21}}{[\mathbf{W}^{-1}]_{11}}\right)}{2\pi f c^{-1} (d_2 - d_1)} \\ &= \arccos \frac{\arg\left(-\frac{W_{21}}{W_{22}}\right)}{2\pi f c^{-1} (d_2 - d_1)} \\ &= \theta_2^{null}. \end{aligned} \quad (28)$$

The derivation of  $\theta_i^{null}$  is based on derivatives. We have another derivation of  $\theta_i^{null}$  based on the graphical interpretation of a directivity pattern [33].

#### APPENDIX C CALCULATION OF MaxSIR

If we know the individual observations

$$q_{jk}(t) = \sum_l h_{jk}(l) s_k(t-l) \quad (29)$$

of source signals  $s_k(t)$  at sensors  $x_j$ , a good permutation can be calculated by maximizing the SIR in each frequency bin. We first apply STFT to  $q_{jk}(t)$  in the same manner as (6):

$$Q_{jk}(f, t) = \sum_{\tau=-\frac{L}{2}}^{\frac{L}{2}-1} q_{jk}(\tau+t) \text{win}(\tau) e^{-j2\pi f \tau}. \quad (30)$$

Let  $\mathbf{Q}(f, t)$  be a matrix whose elements are  $Q_{jk}(f, t)$ , and  $\mathbf{P}_\Pi$  be a permutation matrix that permutes the rows of the right hand matrix according to a permutation  $\Pi$ . Then, the permutation at frequency  $f$  is obtained by

$$\Pi_f = \underset{\Pi}{\text{argmax}} \text{trace} \left( \left\langle |\mathbf{P}_\Pi \mathbf{W}(f) \mathbf{Q}(f, t)|^2 \right\rangle_t \right) \quad (31)$$

where  $|\cdot|^2$  calculates the power of each element, and  $\text{trace}(\cdot)$  returns the sum of the diagonal elements of a matrix.

#### ACKNOWLEDGMENT

The authors wish to thank H. Saruwatari, for his valuable discussions and for providing the impulse responses used in the experiments, T. Nakatani, for valuable discussions on the harmonic structure of speech, S. Katagiri, for continuous encouragement, and the anonymous reviewers who helped to improve the quality of this paper.

#### REFERENCES

- [1] S. Haykin, Ed., *Unsupervised Adaptive Filtering (Volume I: Blind Source Separation)*. New York: Wiley, 2000.
- [2] T. W. Lee, *Independent Component Analysis—Theory and Applications*. Norwell, MA: Kluwer, 1998.
- [3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
- [4] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. IEEE Workshop Signal Processing Advances Wireless Communications*, Apr. 1997, pp. 101–104.
- [5] M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved nonstationary signals," *Neurocomput.*, vol. 22, pp. 157–171, 1998.
- [6] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. ICA*, Dec. 2001, pp. 722–727.
- [7] S. C. Douglas and X. Sun, "Convolutional blind separation of speech mixtures using the natural gradient," *Speech Commun.*, vol. 39, pp. 65–78, 2003.
- [8] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomput.*, vol. 22, pp. 21–34, 1998.
- [9] 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. ICASSP*, June 2000, pp. 3140–3143.
- [10] M. Z. Ikram and D. R. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation," in *Proc. ICASSP*, May 2002, pp. 881–884.
- [11] 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 J. Appl. Signal Process.*, no. 11, pp. 1157–1166, 2003.
- [12] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. ICA*, June 2000, pp. 215–220.
- [13] 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, Oct. 2001.
- [14] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "A combined approach of array processing and independent component analysis for blind separation of acoustic signals," in *Proc. ICASSP*, May 2001, pp. 2729–2732.
- [15] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency domain blind source separation," *IEICE Trans. Fund.*, vol. E86-A, no. 3, pp. 590–596, Mar. 2003.
- [16] L. Parra and C. Spence, "Convolutional blind separation of nonstationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 320–327, May 2000.
- [17] L. Schobben and W. Sommen, "A frequency domain blind signal separation method based on decorrelation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1855–1865, Aug. 2002.
- [18] A. D. Back and A. C. Tsoi, "Blind deconvolution of signals using a complex recurrent network," *Proc. Neural Networks Signal Processing*, pp. 565–574, 1994.
- [19] R. H. Lambert and A. J. Bell, "Blind separation of multiple speakers in a multipath environment," in *Proc. ICASSP*, Apr. 1997, pp. 423–426.
- [20] T. W. Lee, A. J. Bell, and R. Orglmeister, "Blind source separation of real world signals," in *Proc. ICNN*, June 1997, pp. 2129–2135.

- [21] M. Joho and P. Schniter, "Frequency domain realization of a multi-channel blind deconvolution algorithm based on the natural gradient," in *Proc. ICA*, Apr. 2003, pp. 543–548.
- [22] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutive mixtures," in *Proc. ICA*, Apr. 2003, pp. 945–950.
- [23] S. Winter, H. Sawada, and S. Makino, "Geometrical understanding of the PCA subspace method for overdetermined blind source separation," in *Proc. ICASSP*, Apr. 2003, pp. 769–772.
- [24] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [25] S. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.
- [26] A. Hyvärinen, "Fast and robust fixed-point algorithm for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, pp. 626–634, May 1999.
- [27] J. F. Cardoso and A. Souloumiac, "Blind beamforming for nongaussian signals," *Proc. Instit. Elec. Eng.-F*, pp. 362–370, Dec. 1993.
- [28] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411–419, 1995.
- [29] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Mag.*, pp. 2–24, Apr. 1988.
- [30] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Convolutive blind source separation for more than two sources in the frequency domain," in *Proc. ICASSP*, May 2004, pp. III-885–III-888.
- [31] R. Mukai, H. Sawada, S. de la Kethulle, S. Araki, and S. Makino, "Array geometry arrangement for frequency domain blind source separation," in *Proc. IWAENC*, Sept. 2003, pp. 219–222.
- [32] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York: Springer-Verlag, 1997.
- [33] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust approach to the permutation problem of frequency-domain blind source separation," in *Proc. ICASSP*, Apr. 2003, pp. 381–384.



**Hiroshi Sawada** (M'02) 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, Kyoto. 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 and blind source separation for convolutive mixtures using independent component

analysis.

Dr. Sawada received the 9th TELECOM System Technology Award for Students from the Telecommunications Advancement Foundation in 1994, and the Best Paper Award of the IEEE Circuit and System Society in 2001. He is a member of the IEICE and the ASJ.



**Ryo Mukai** (A'95–M'01) 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, Kyoto, Japan, 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 member of ACM, the Acoustical Society of Japan (ASJ), IEICE, and IPSJ.



**Shoko Araki** (M'01) received the B.E. and the M.E. degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1998 and 2000, respectively.

Her research interests include array signal processing, blind source separation applied to speech signals, and auditory scene analysis.

Ms. Araki 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 2002. She is a member of the ASJ.



**Shoji Makino** (A'89–M'90–SM'99–F'04) was born in Nikko, Japan, on June 4, 1956. He received the B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1979, 1981, and 1993, respectively.

He joined NTT, Kyoto, Japan, in 1981. He is now an Executive Manager at the NTT Communication Science Laboratories. His research interests include blind source separation of convolutive mixtures of speech, acoustic signal processing, and adaptive filtering and its applications such as acoustic echo cancellation. He is the author or coauthor of more than 170 articles in journals and conference proceedings and has been responsible for more than 140 patents.

Dr. Makino received the TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2004, the Best Paper Award of the IWAENC in 2003, the Paper Award of the IEICE in 2002, the Paper Award of the ASJ in 2002, the Achievement Award of the IEICE in 1997, and the Outstanding Technological Development Award of the ASJ in 1995. He is a member of the Conference Board of the IEEE Signal Processing Society and an Associate Editor of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING. Dr. Makino is a member of the ASJ and the IEICE. He is a member of the Technical Committee on Audio and Electroacoustics as well as Speech of the IEEE Signal Processing Society. He is a member of the International ICA Steering Committee and the Organizing Chair of the 2003 International Conference on Independent Component Analysis and Blind Signal Separation. He is the General Chair of the 2003 International Workshop on Acoustic Echo and Noise Control. He was a Vice Chair of the Technical Committee on Engineering Acoustics of the IEICE.