EURASIP Journal on Applied Signal Processing 2003:11, 1157–1166

Equivalence between Frequency-Domain Blind Source Separation and Frequency-Domain Adaptive Beamforming for Convolutive Mixtures

Shoko Araki
NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
Email: shoko@cslab.kecl.ntt.co.jp

Shoji Makino
NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
Email: maki@cslab.kecl.ntt.co.jp

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

Ryo Mukai
NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
Email: ryo@cslab.kecl.ntt.co.jp

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

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

Received 2 December 2002 and in revised form 16 March 2003

Frequency-domain blind source separation (BSS) is shown to be equivalent to two sets of frequency-domain adaptive beamformers (ABFs) under certain conditions. The zero search of the off-diagonal components in the BSS update equation can be viewed as the minimization of the mean square error in the ABFs. The unmixing matrix of the BSS and the filter coefficients of the ABFs converge to the same solution if the two source signals are ideally independent. If they are dependent, this results in a bias for the correct unmixing filter coefficients. Therefore, the performance of the BSS is limited to that of the ABF if the ABF can use exact geometric information. This understanding gives an interpretation of BSS from a physical point of view.

Keywords and phrases: blind source separation, convolutive mixtures, adaptive beamformers.

1. INTRODUCTION

Blind source separation (BSS) is an approach for estimating source signals $s_i(t)$ using only the information on mixed signals $x_j(t)$ observed at each input channel. BSS can be applied to achieve noise-robust speech recognition and high-quality hands-free telecommunication. It might also become one of the cues for auditory scene analysis.

Several methods have been proposed for BSS of convolutive mixtures [1, 2]. Some approaches consider the impulse responses of a room h_{ji} as FIR filters, and estimate those filters in the time domain [3, 4, 5]; other approaches

transform the problem into the frequency domain to solve an instantaneous BSS problem for every frequency simultaneously [6, 7]. Here, we consider the BSS of convolutive mixtures of speech in the frequency domain.

In this paper, we provide an interpretation of BSS from a physical point of view showing the equivalence between frequency-domain BSS and two sets of frequency-domain adaptive beamformers (ABFs).

Signal separation by using a noise cancellation framework with signal leakage into the noise reference was discussed in [8, 9]. These studies showed that the least squares criterion is equivalent to the decorrelation criterion of a noise-free signal estimate and a signal-free noise estimate. The error minimization was shown to be completely equivalent to a zero search in the cross correlation.

Inspired by the discussions in [8, 9], but apart from the noise cancellation framework, we attempt to compare the frequency-domain BSS problem with the frequency-domain ABF framework. In earlier work, Dinc and Bar-Ness [10] and Cardoso and Souloumiac [11] indicated the connection between blind identification and beamforming in a narrowband context. Kurita et al. [12] and Parra and Alvino [13] utilized the relationship between BSS and ABFs to achieve better BSS performance; however, they did not discuss this relationship theoretically. We discuss this relationship more closely and more quantitatively, focusing on BSS with second-order statistics (SOS), and we show that BSS and ABFs have equivalent functions despite their completely different adaptation procedures. Moreover, we provide a physical understanding of frequency-domain BSS [14]. From the equivalence between BSS and ABFs, we can make it clear that the physical behavior of BSS is to reduce jammer signal by forming a spatial null in the jammer direction. Knaak and Filbert [15] have also provided a somewhat quantitative discussion of the relationship between frequency-domain ABF and frequency-domain BSS. Beyond their discussions, in this paper, we are also able to explain the effect of collapse of the independence assumption in BSS.

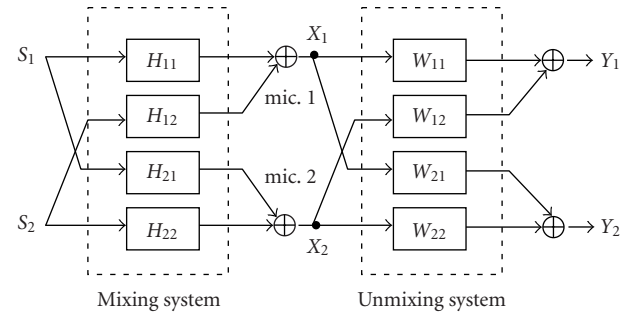
In Section 2, we summarize the framework of frequency-domain BSS for convolutive mixtures. In Section 3, the frequency-domain ABF is summarized. In Section 4, we show the equivalence between BSS and ABFs theoretically. In Section 5, we confirm this equivalence and the limitation with experiments using measured impulse responses in a real room and six combinations of male and female speech. Section 6 concludes this paper.

2. FREQUENCY-DOMAIN BSS OF CONVOLUTIVE MIXTURES OF SPEECH

2.1. Mixed signal model

In real environments, the signals are affected by reverberation and observed by the microphones. Therefore, N signals recorded by M microphones are modeled as

$$x_j(n) = \sum_{i=1}^{N} \sum_{p=1}^{P} h_{ji}(p) s_i(n-p+1) \quad (j = 1, \ldots, M), \qquad (1)$$



F 1: BSS system configuration.

where s_i is the source signal from a source i, x_j is the signal received by a microphone j, and h_{ji} is the P-taps impulse response from source i to microphone j.

2.2. Unmixed signal model

In order to obtain unmixed signals, we estimate unmixing filters $w_{ij}(k)$ of Q-taps, and the unmixed signals are obtained as

$$y_i(n) = \sum_{j=1}^{M} \sum_{q=1}^{Q} w_{ij}(q) x_j(n-q+1) \quad (i = 1, \ldots, N). \qquad (2)$$

The unmixing filters are estimated such that the unmixed signals become mutually independent.

In this paper, we consider a two-input, two-output convolutive BSS problem, that is, $N = M = 2$ (Figure 1).

2.3. Frequency-domain approach

The frequency-domain approach to convolutive mixtures is to transform the problem into an instantaneous BSS problem in the frequency domain [6, 7]. Using a T-point short-time Fourier transformation for (1), we obtain

$$\mathbf{X}(\omega, m) = \mathbf{H}(\omega)\mathbf{S}(\omega, m), \qquad (3)$$

where ω denotes the frequency, m represents the time-dependence of the short-time Fourier transformation, $\mathbf{S}(\omega, m) = [S_1(\omega, m), S_2(\omega, m)]^T$ is the source signal vector, and $\mathbf{X}(\omega, m) = [X_1(\omega, m), X_2(\omega, m)]^T$ is the observed signal vector. We assume that the (2×2) mixing matrix $\mathbf{H}(\omega)$ is invertible and that $H_{ji}(\omega) \neq 0$. Also, $\mathbf{H}(\omega)$ does not depend on time m.

The unmixing process can be formulated in a frequency bin ω:

$$\mathbf{Y}(\omega, m) = \mathbf{W}(\omega)\mathbf{X}(\omega, m), \qquad (4)$$

where $\mathbf{Y}(\omega, m) = [Y_1(\omega, m), Y_2(\omega, m)]^T$ is the estimated source signal vector and $\mathbf{W}(\omega)$ represents a (2×2) unmixing matrix at frequency bin ω. The unmixing matrix $\mathbf{W}(\omega)$ is determined so that $Y_1(\omega, m)$ and $Y_2(\omega, m)$ become mutually independent. The above calculation is carried out at each frequency independently. In this paper, we consider the DFT frame size T to be equal to the length Q of the unmixing filter.

2.4. Frequency-domain BSS of convolutive mixtures using SOS

In [9], it is pointed out that nonstationary signals provide enough additional information to enable us to estimate all $W_{ij}(\omega)$. Some authors have utilized SOS for mixed speech signals [16, 17].

The source signals $S_1(\omega, m)$ and $S_2(\omega, m)$ are assumed to be zero mean, nonstationary, and mutually uncorrelated.

In order to determine $\mathbf{W}(\omega)$ so that $Y_1(\omega, m)$ and $Y_2(\omega, m)$ become mutually uncorrelated, we seek a $\mathbf{W}(\omega)$ that diagonalizes the covariance matrices $\mathbf{R}_Y(\omega, k)$ simultaneously for all time blocks k:

$$\begin{aligned}\mathbf{R}_Y(\omega,k) &= \mathbf{W}(\omega)\mathbf{R}_X(\omega,k)\mathbf{W}^*(\omega)\\ &= \mathbf{W}(\omega)\mathbf{H}(\omega)\mathbf{\Lambda}_s(\omega,k)\mathbf{H}^*(\omega)\mathbf{W}^*(\omega)\\ &= \mathbf{\Lambda}_c(\omega,k),\end{aligned} \tag{5}$$

where $*$ denotes the conjugate transpose and, $\mathbf{R}_X$ is the covariance matrix of $\mathbf{X}(\omega)$, represented as follows:

$$\mathbf{R}_X(\omega,k) = \frac{1}{M}\sum_{m=0}^{M-1}\mathbf{X}(\omega, Mk+m)\mathbf{X}^*(\omega, Mk+m), \tag{6}$$

$\mathbf{\Lambda}_s(\omega, k)$ is the diagonal covariance matrix of the source signals that is different for each k, and $\mathbf{\Lambda}_c(\omega, k)$ is an arbitrary diagonal matrix.

The diagonalization of $\mathbf{R}_Y(\omega, k)$ can be written as an overdetermined least squares problem:

$$\arg\min_{W(\omega)}\sum_k \left|\left|\text{off-diag}\mathbf{W}(\omega)\mathbf{R}_X(\omega,k)\mathbf{W}^*(\omega)\right|\right|^2, \tag{7}$$

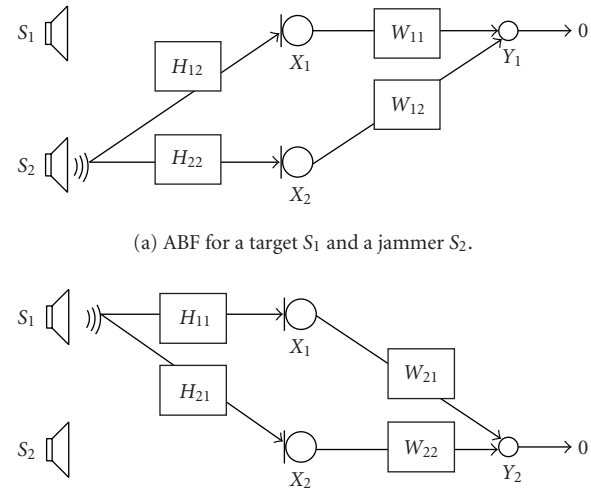
where $\|\cdot\|^2$ is the squared Frobenius norm. In order to avoid a trivial solution, $\mathbf{W}(\omega) = 0$, we use a constraint, for example, $\sum_k \|\text{diag}\mathbf{W}(\omega)\mathbf{R}_X(\omega,k)\mathbf{W}^*(\omega)\|^2 = c$ or $\|\mathbf{W}(\omega)\|^2 = c$, where c is a positive constant. While these constraints for determining a nontrivial $\mathbf{W}(\omega)$ give rise to a different solution, they still have the same function.

3. FREQUENCY-DOMAIN ABF

Here, we consider the frequency-domain ABF which can remove a jammer signal. Since our aim is to separate two signals S_1 and S_2 with two microphones, we use two sets of ABFs (see Figure 2). That is, an ABF that forms a null directivity pattern towards source S_2 by using filter coefficients W_{11} and W_{12}, and an ABF that forms a null directivity pattern towards source S_1 by using filter coefficients W_{21} and W_{22}. Note that the ABF can be adapted when only a jammer exists but a target does not exist, and that the direction of the target or the impulse responses from the target to the microphones should be known. In this section, we attach more importance to an intuitive explanation of the ABF mechanism than to a strict mathematical explanation.

3.1. ABF for target S_1 and jammer S_2

In order to estimate the coefficients W_{ij} of an ABF, we minimize the output signal power when a jammer is active but a target is not.



(a) ABF for a target S_1 and a jammer S_2.

(b) ABF for a target S_2 and a jammer S_1.

F 2: Two sets of ABF-system configurations.

First, we consider the case of a target S_1 and a jammer S_2 [see Figure 2a]. When target $S_1 = 0$, the output $Y_1(\omega, m)$ is expressed as

$$Y_1(\omega, m) = \mathbf{W}(\omega)\mathbf{X}(\omega, m), \tag{8}$$

where

$$\begin{aligned}\mathbf{W}(\omega) &= [W_{11}(\omega), W_{12}(\omega)]\\ \mathbf{X}(\omega, m) &= [X_1(\omega, m), X_2(\omega, m)]^T.\end{aligned} \tag{9}$$

To minimize jammer $S_2(\omega, m)$ in the output $Y_1(\omega, m)$ when target $S_1 = 0$, the mean square error $J(\omega)$ is introduced as

$$\begin{aligned}J(\omega) &= E[Y_1^2(\omega, m)]\\ &= \mathbf{W}(\omega)E[\mathbf{X}(\omega, m)\mathbf{X}^*(\omega, m)]\mathbf{W}^*(\omega)\\ &= \mathbf{W}(\omega)\mathbf{R}(\omega)\mathbf{W}^*(\omega),\end{aligned} \tag{10}$$

where $E[\cdot]$ is the expectation operator and

$$\mathbf{R}(\omega) = E\begin{bmatrix}X_1(\omega,m)X_1^*(\omega,m) & X_1(\omega,m)X_2^*(\omega,m)\\ X_2(\omega,m)X_1^*(\omega,m) & X_2(\omega,m)X_2^*(\omega,m)\end{bmatrix}. \tag{11}$$

By differentiating the cost function $J(\omega)$ with respect to $\mathbf{W}$ and setting the gradient to zero, we obtain (hereafter (ω, m) and (ω) are omitted for convenience)

$$\frac{\partial J(\omega)}{\partial \mathbf{W}} = 2\mathbf{R}\mathbf{W}^* = 0. \tag{12}$$

Using $X_1 = H_{12}S_2$, $X_2 = H_{22}S_2$, we get

$$W_{11}H_{12} + W_{12}H_{22} = 0. \tag{13}$$

With (13) only, we have a trivial solution $W_{11} = W_{12} = 0$. Therefore, an additional constraint should be added to

ensure that target signal S_1 is in the output Y_1, that is,

$$Y_1 = (W_{11}H_{11} + W_{12}H_{21})S_1 = c_1S_1, \tag{14}$$

which leads to

$$W_{11}H_{11} + W_{12}H_{21} = c_1, \tag{15}$$

where c_1 is an arbitrary complex constant. In the ABF framework, this constraint is usually approximately given by the steering vector under the condition that the direction of a target signal is known. This constraint can also be given by the measured impulse responses from a target source to microphones. In this paper, we assume that the target direction or impulse responses between a target and microphones are known correctly.

The ABF solution is derived from the simultaneous equations (13) and (15).

In practice, **R** is a positive definite matrix due to the effect of ambient noise and a finite length DFT. Here, however, we consider the ideal case. That is, we assume that **R** is not invertible. Moreover, for a practical ABF, **W** is calculated by solving the constrained minimization problem; the constraint is included in advance. Therefore, (13) usually includes an estimation error and does not become 0 in a strict sense. Although we should evaluate and compare this error for ABF and BSS quantitatively, in this paper, we stress the qualitative equivalence between ABFs and BSS.

3.2. *ABF for target* S_2 *and jammer* S_1

Similarly, for a target S_2, a jammer S_1, and an output Y_2 (see Figure 2b), we obtain

$$W_{21}H_{11} + W_{22}H_{21} = 0, \tag{16}$$

$$W_{21}H_{12} + W_{22}H_{22} = c_2. \tag{17}$$

3.3. *Two sets of ABFs*

By combining (13), (15), (16), and (17), we can summarize the simultaneous equations for two sets of ABFs as follows:

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}. \tag{18}$$

4. EQUIVALENCE BETWEEN BSS AND ABFs

As we showed in (7), the SOS-BSS algorithm works to minimize off-diagonal components in

$$E\begin{bmatrix} Y_1Y_1^* & Y_1Y_2^* \\ Y_2Y_1^* & Y_2Y_2^* \end{bmatrix}, \tag{19}$$

(see (5)) for all time blocks k. Using **H** and **W**, the outputs Y_1 and Y_2 are expressed in each frequency bin as

$$Y_1 = aS_1 + bS_2, \qquad Y_2 = cS_1 + dS_2, \tag{20}$$

where

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}. \tag{21}$$

These paths are shown in Figure 3. Here, a and d represent the paths for targets, and b and c are the paths for jammers.

4.1. *When* $S_1 \neq 0$ *and* $S_2 \neq 0$

We now analyze what is occurring in the BSS framework. After convergence, the expectation of the off-diagonal component $E[Y_1Y_2^*]$ is expressed as

$$\begin{aligned} E[Y_1Y_2^*]^2 = \{&ad^*E[S_1S_2^*] + bc^*E[S_2S_1^*] \\ &+ (ac^*E[S_1^2] + bd^*E[S_2^2])\}^2 = 0. \end{aligned} \tag{22}$$

Since S_1 and S_2 are assumed to be uncorrelated, the first and second terms become zero. Then, the BSS adaptation should drive the third term of (22) to zero for all time blocks k. That is, (22) is an identical equation with regard to $E[S_1^2]$ and $E[S_2^2]$ for all time blocks k. This leads to

$$ac^* = bd^* = 0. \tag{23}$$

Case 1. When $a = c_1$, $c = 0$, $b = 0$, and $d = c_2$,

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}. \tag{24}$$

This equation is identical to (18) in ABFs.

Case 2. When $a = 0$, $c = c_1$, $b = c_2$, and $d = 0$,

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} 0 & c_2 \\ c_1 & 0 \end{bmatrix}. \tag{25}$$

This equation leads to a permutation solution $Y_1 = c_2S_2$, $Y_2 = c_1S_1$; the estimated source signal components are recovered with a different order.

Case 3. When $a = 0$, $c = c_1$, $b = 0$, and $d = c_2$,

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ c_1 & c_2 \end{bmatrix}. \tag{26}$$

This equation leads to an undesirable solution $Y_1 = 0$, $Y_2 = c_1S_1 + c_2S_2$.

Case 4. When $a = c_1$, $c = 0$, $b = c_2$, and $d = 0$,

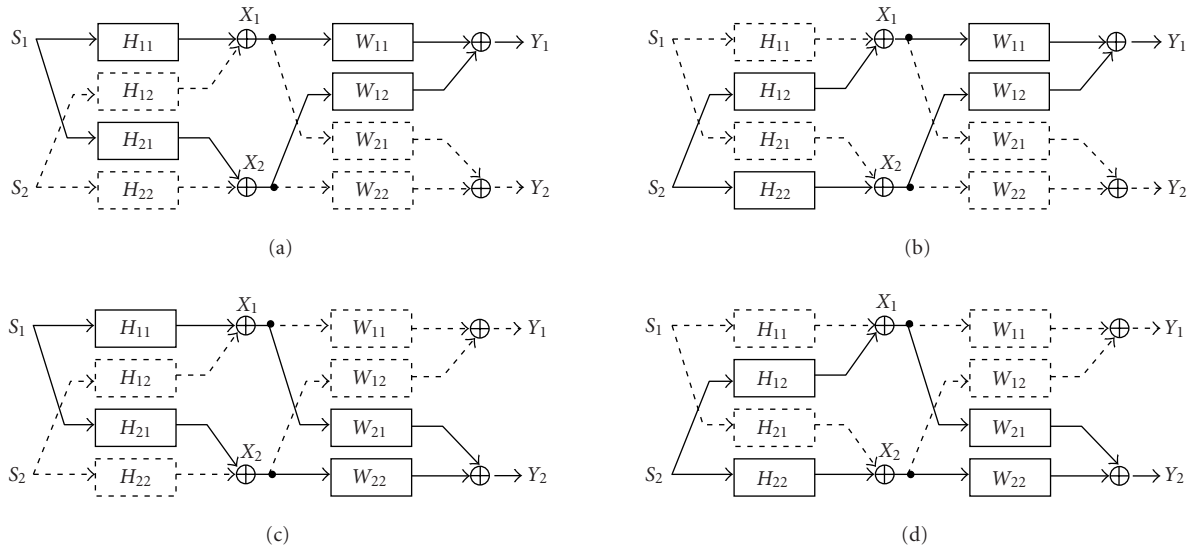
$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ 0 & 0 \end{bmatrix}. \tag{27}$$

This equation leads to an undesirable solution $Y_1 = c_1S_1 + c_2S_2$, $Y_2 = 0$.

Note that Cases 3 and 4 do not appear in general because we assume that $\mathbf{H}(\omega)$ is invertible and $H_{ji}(\omega) \neq 0$. That is, if $a = 0$, then $b \neq 0$ (Case 2), and if $c = 0$, then $d \neq 0$ (Case 1).

4.2. *When* $S_1 \neq 0$ *and* $S_2 = 0$

BSS can adapt even if there is only one active source. In this case, only one set of ABF is achieved.



F 3: Paths in (21).

When $S_2 = 0$, we have

$$Y_1 = aS_1, \qquad Y_2 = cS_1, \tag{28}$$

then

$$E[Y_1 Y_2^*] = E[aS_1 c^* S_1^*] = ac^* E[S_1^2] = 0, \tag{29}$$

and therefore, the BSS adaptation should drive

$$ac^* = 0. \tag{30}$$

Case 5. When $c = 0$ and $a = c_1$,

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} c_1 & - \\ 0 & - \end{bmatrix}, \tag{31}$$

where $-$ shows a don't care. Since $S_2 = 0$, the output can be derived correctly, $Y_1 = c_1 S_1$, $Y_2 = 0$, as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} c_1 & - \\ 0 & - \end{bmatrix} \begin{bmatrix} S_1 \\ 0 \end{bmatrix} = \begin{bmatrix} c_1 S_1 \\ 0 \end{bmatrix}. \tag{32}$$

Case 6. When $c = c_1$ and $a = 0$,

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} 0 & - \\ c_1 & - \end{bmatrix}. \tag{33}$$

This equation leads to the permutation solution which is $Y_1 = 0$, $Y_2 = c_1 S_1$:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 0 & - \\ c_1 & - \end{bmatrix} \begin{bmatrix} S_1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ c_1 S_1 \end{bmatrix}. \tag{34}$$

4.3. When $S_1 = 0$ and $S_2 \neq 0$

Similarly, only one set of ABF is achieved in this case.

Case 7. When $b = 0$ and $d = c_2$,

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} - & 0 \\ - & c_2 \end{bmatrix}. \tag{35}$$

We can obtain the result

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} - & 0 \\ - & c_2 \end{bmatrix} \begin{bmatrix} 0 \\ S_2 \end{bmatrix} = \begin{bmatrix} 0 \\ c_2 S_2 \end{bmatrix}. \tag{36}$$

Case 8. When $b = c_2$ and $d = 0$,

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} - & c_2 \\ - & 0 \end{bmatrix}. \tag{37}$$

This equation leads to the permutation solution

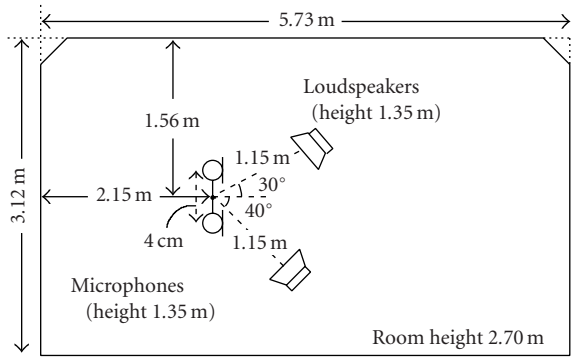
$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} - & c_2 \\ - & 0 \end{bmatrix} \begin{bmatrix} 0 \\ S_2 \end{bmatrix} = \begin{bmatrix} c_2 S_2 \\ 0 \end{bmatrix}. \tag{38}$$

The values c_1 and c_2 in Sections 3 and 4 are not the same due to the scaling problem in BSS: the estimated source signal components are recovered with a different gain in different frequency bins. Although the outputs obtained by BSS are filtered versions of the source signals, the behavior whereby they make a null towards the jammer signal is still the same as the two sets of ABFs. Moreover, we can scale the output signals in the same way as the constraint in an ABF (15) and (17) by using the directivity pattern obtained by the unmixing matrix (e.g., with the method described in Section 5.3).

5. EXPERIMENTS AND DISCUSSIONS

5.1. Limitation of frequency-domain BSS

Frequency-domain BSS and frequency-domain ABFs are equivalent (see (18) and (24)) in an ideal case if the inde-

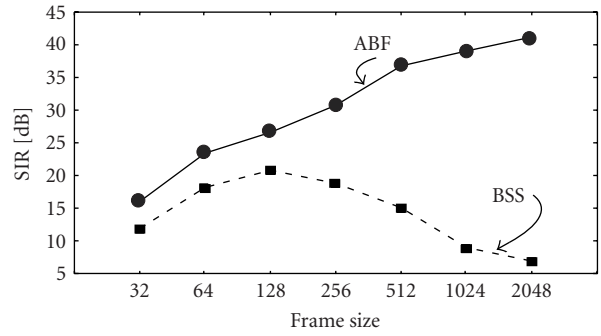


F 4: Layout of the room used in experiments.

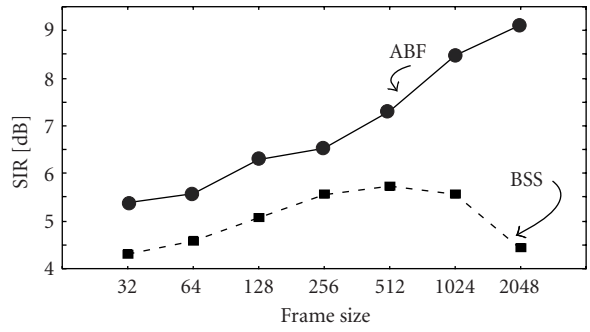
pendence assumption ideally holds (see (22)). If not, the first and second terms of (22) behave as a bias when calculating the correct coefficients a, b, c, and d in (22). We have shown in [18] that a long frame size works poorly in frequency-domain BSS for speech data of a few seconds. This is because when we use a long frame, the number of samples in each frequency bin becomes small. This makes the estimation of statistics, such as the zero mean and independent assumptions, difficult [19]. Therefore, the first and second terms of (22) are not equal to zero. Therefore, the upper bound of the BSS performance is given by that of the ABF. However, note that BSS does not need the absence of a target signal: BSS can adapt in the presence of target and jammer and also in the presence of only one active source, whereas an ABF can be adapted only when there is a jammer but no target. Note also that an ABF needs to know the array manifold and the target direction but BSS does not need these for the adaptation.

5.1.1. Simulation conditions and evaluation measurement

We compared the separation performance of BSS with that of an ABF. These experiments were conducted using speech data convolved with impulse responses recorded in two environments specified by different reverberation times: $T_R = 0$ millisecond and 300 milliseconds. Since the sampling rate was 8 kHz, 300 milliseconds correspond to 2400 taps. The size of the room used to measure the impulse responses was $5.73\,\text{m} \times 3.12\,\text{m} \times 2.70\,\text{m}$ and the distance between the loudspeakers and microphones was 1.15 m (Figure 4). We used a two-element array with an interelement spacing of 4 cm. The speech signals arrived from two directions, $-30°$ and $40°$. As the original speech, we used two sentences spoken by two male and two female speakers. The investigations were carried out for six combinations of speakers. The length of the speech data was about eight seconds. We used the first three seconds of the data for learning, and the entire eight seconds for separation. We changed the DFT frame size T from 32 to 2048 and investigated the performance for each condition. The frame shift was half the frame size T, and the analysis window was a Hamming window. To evaluate the performance, we used the *signal to interference ratio* (SIR), defined



(a) $T_R = 0$ ms.



(b) $T_R = 300$ ms.

F 5: Results of SIR for different frame sizes. The solid lines are for ABF and the broken lines are for BSS. (a) Nonreverberant test ($T_R = 0$ ms), (b) reverberant test ($T_R = 300$ ms).

as follows:

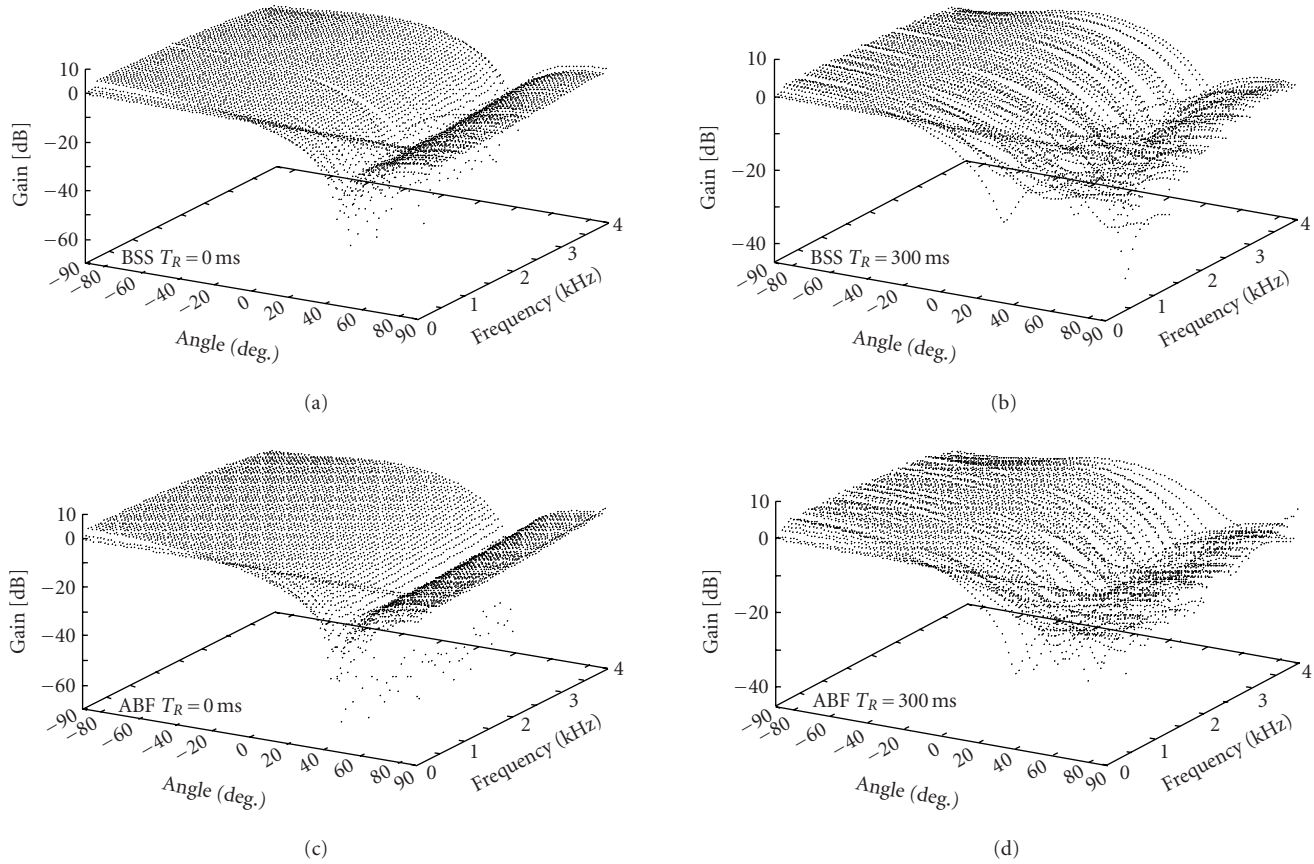
$$\begin{aligned}
\text{SIR}_i &= \text{SIR}_{O_i} - \text{SIR}_{Ii}, \\
\text{SIR}_{Oi} &= 10 \log \frac{\sum_\omega \left| A_{ii}(\omega) S_i(\omega) \right|^2}{\sum_\omega \left| A_{ij}(\omega) S_j(\omega) \right|^2}, \\
\text{SIR}_{Ii} &= 10 \log \frac{\sum_\omega \left| H_{ii}(\omega) S_i(\omega) \right|^2}{\sum_\omega \left| H_{ij}(\omega) S_j(\omega) \right|^2},
\end{aligned} \tag{39}$$

where $\mathbf{A}(\omega) = \mathbf{W}(\omega)\mathbf{H}(\omega)$ and $i \neq j$. SIR means the ratio of a target-originated signal to a jammer-originated signal. These values were averaged over all six combinations with respect to the speakers, and SIR_1 and SIR_2 were averaged.

The ABF we used was that proposed by Frost [20].

5.1.2. Simulation results

Figure 5 shows the separation performance of BSS and the ABF. With BSS, when the frame size was too long, the separation performance deteriorated. This is because the number of samples in each frequency bin is too small to estimate the statistics correctly when the frame size is long [19]. In this case, the first and second terms of (22) are not equal zero and behave as a bias noise as mentioned in Section 5.1. Therefore, the performance is degraded when we use a long frame in BSS.



F 6: Directivity patterns (a) obtained by BSS ($T_R = 0$ ms), (b) obtained by BSS ($T_R = 300$ ms), (c) obtained by ABF ($T_R = 0$ ms), and (d) obtained by ABF ($T_R = 300$ ms).

By contrast, an ABF does not employ the assumption of independence of the source signals. With the ABF, therefore, the separation performance increased as the frame size became longer. Figure 5 confirms that the performance of the BSS is limited by that of the ABF.

5.2. Physical interpretation of BSS

Now, we can understand the behavior of BSS as two sets of ABFs. Figure 6 shows the directivity patterns obtained by BSS and ABF. Figures 6a and 6b are the directivity patterns obtained by BSS after solving the permutation and scaling problem with the method described in Section 5.3, and Figures 6c and 6d show the directivity patterns by $\mathbf{W}$ obtained by ABF. When $T_R = 0$, a sharp spatial null is obtained with both BSS and ABF (see Figures 6a and 6c). When $T_R = 300$ milliseconds, the directivity pattern becomes duller (see Figures 6b and 6d).

BSS removes the sound from the jammer direction and reduces the reverberation of the jammer signal to some extent [21] in the same way as an ABF does. This understanding clearly explains the poor performance of the BSS in a real acoustic environment with a long reverberation.

The BSS was shown to outperform a null beamformer that forms a steep null directivity pattern towards a jammer [21, 22]. It is well known that an adaptive beamformer outperforms a null beamformer in long reverberation. Our understanding also clearly explains the result.

Although the ABF and BSS procedures are different, their essential behavior is the same: they make a null towards the jammer direction. The relationship between ABF and BSS is summarized in Table 1.

5.3. Improvement in separation performance with equivalence of BSS and ABFs

So far, we have described the equivalence of BSS and ABFs: an unmixing system obtained by BSS removes the sound from the jammer direction in the same way as ABFs do. In order to improve the separation performance of BSS, we should exploit this relationship between BSS and ABFs. In this section, we outline our successful examples of achieving this.

Permutation and scaling solution with directivity patterns

A scaling and permutation problem occurs in frequency-domain BSS, that is, the estimated source signal components are recovered with a different order and gain in different frequency bins. When we know the array manifold, we can solve

Tab 1: The relationship between ABF and BSS.

	ABF	BSS
Prior knowledge	Array manifold and look direction or acoustic transfer function are needed	Not needed in itself, but to solve the permutation/scaling problem, some is needed (e.g., array manifold)
Adaptation	When only jammer exist	Whenever
Sensitivity to independence	Insensitive (however sensitive to double-talk errors)	Highly sensitive
Behavior	Make a null towards the jammer direction and reduce the jammer signal	

the permutation and scaling problem in frequency-domain BSS with directivity patterns obtained by the unmixing system $\mathbf{W}(\omega)$ [12]. First, from the directivity pattern obtained by $\mathbf{W}(\omega)$, we estimate the source directions and reorder the row of $\mathbf{W}(\omega)$ so that the directivity pattern forms a null towards the same direction in all frequency bins, then we normalize the row of $\mathbf{W}(\omega)$ so that the target direction gains become 0 dB.

Source direction estimation with directivity pattern

After solving the permutation and scaling problem, we can roughly estimate the source directions by analyzing the null directions, for example, clustering and averaging the null directions for all frequency bins.

Initial value of unmixing system with null beamformers

Because the solution of BSS makes a spatial null towards a jammer, we can use this characteristics for designing the initial value of an unmixing system. As an initial value, we can use constraint null beamformers, which can make a sharp null towards a given jammer and maintain the gain and phase of a given target direction.

We can apply this method to frequency-domain BSS [23], time-domain BSS [24], and subband-domain BSS [23].

Design of appropriate microphone spacing for each frequency [25]

If the spacing is longer than half the wavelength, spatial aliasing occurs: nulls are formed in several directions. By contrast, when the sensors are very closely spaced, the phase difference at a low frequency becomes too small and it becomes difficult to obtain good separation. Generally speaking, a long spacing is suitable for low frequencies and a short spacing for high frequencies. If we arrange sensors according to frequency, we can obtain better BSS performance.

6. CONCLUSION

We provided an interpretation of BSS from a physical point of view showing the equivalence between frequency-domain BSS and two sets of frequency-domain ABFs. The unmixing matrix of the BSS and the filter coefficients of the ABFs converge to the same solution in the ideal case if the two source signals are ideally independent. If they are not independent, the dependency results in bias noise in estimating the correct unmixing filter coefficients. Therefore, the performance of the BSS is limited by that of the ABF. Moreover, BSS mainly removes sound from the jammer direction. Since we can understand the behavior of BSS as two sets of ABFs, BSS reduces the reverberation of the jammer signal to some extent in the same way as an ABF. This understanding clearly explains the poor performance of the BSS in a real acoustic environment with long reverberation.

ACKNOWLEDGMENT

We would like to thank Drs. Shigeru Katagiri and Kiyohiro Shikano for their continuous encouragement.

REFERENCES

[1] 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.
[2] S. Haykin, *Unsupervised Adaptive Filtering*, John Wiley & Sons, New York, NY, USA, 2000.
[3] T.-W. Lee, *Independent Component Analysis: Theory and Applications*, Kluwer Academic Publishers, Boston, Mass, USA, 1998.
[4] M. Kawamoto, A. K. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Real world blind separation of convolved non-stationary signals," in *Proc. International Workshop on Independence Component Analysis and Signal Separation (ICA '99)*, pp. 347–352, Aussois, France, January 1999.
[5] X. Sun and S. Douglas, "A natural gradient convolutive blind source separation algorithm for speech mixtures," in *Proc. 3rd International Conference on Independent Component Analysis and Blind Signal Separation (ICA '01)*, pp. 59–64, San Diego, Calif, USA, December 2001.
[6] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, 1998.
[7] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. International Workshop on Independence Component Analysis and Signal Separation (ICA '99)*, pp. 365–370, Aussois, France, January 1999.
[8] S. Van Gerven and D. Van Compernolle, "Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness," *IEEE Trans. Signal Processing*, vol. 43, no. 7, pp. 1602–1612, 1995.
[9] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation by decorrelation," *IEEE Trans. Speech, and Audio Processing*, vol. 1, no. 4, pp. 405–413, 1993.
[10] A. Dinc and Y. Bar-Ness, "Bootstrap: a fast blind adaptive signal separator," in *Proc. IEEE Int. Conf. Acoustics, Speech,*

Signal Processing, vol. 2, pp. 325–328, San Francisco, Calif, USA, March 1992.
[11] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings Part F: Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, 1993.
[12] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, pp. 3140–3143, Istanbul, Turkey, June 2000.
[13] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," in *Proc. IEEE International Workshop on Neural Networks for Signal Processing (NNSP '01)*, pp. 273–282, Falmouth, Mass, USA, September 2001.
[14] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers," in *Proc. Eurospeech 2001*, pp. 2595–2598, Aalborg, Denmark, September 2001.
[15] M. Knaak and D. Filbert, "Acoustical semi-blind source separation for machine monitoring," in *Proc. 3rd International Conference on Independent Component Analysis and Blind Signal Separation*, pp. 361–366, San Diego, Calif, USA, December 2001.
[16] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech, and Audio Processing*, vol. 8, no. 3, pp. 320–327, 2000.
[17] M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, pp. 1041–1044, Istanbul, Turkey, June 2000.
[18] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, pp. 2737–2740, Salt Lake City, Utah, USA, May 2001.
[19] S. Araki, S. Makino, R. Mukai, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolved mixture of speech," in *Proc. 3rd International Conference on Independent Component Analysis and Blind Signal Separation*, pp. 132–137, San Diego, Calif, USA, December 2001.
[20] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.
[21] R. Mukai, S. Araki, and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment," in *Proc. Eurospeech 2001*, pp. 2599–2602, Aalborg, Denmark, September 2001.
[22] H. Saruwatari, S. Kurita, and K. Takeda, "Blind source separation combining frequency-domain ICA and beamforming," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, pp. 2733–2736, Salt Lake City, Utah, USA, May 2001.
[23] S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Blind source separation for convolutive mixtures of speech using subband processing," in *Proc. 2nd International Workshop on Spectral Methods and Multirate Signal Processing (SMMSP '02)*, pp. 195–202, Barcelona, Spain, September 2002.
[24] R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming," in *Proc. IEEE International Workshop on Neural Networks for Signal Processing (NNSP '02)*, pp. 445–454, Martigny, Valais, Switzerland, September 2002.
[25] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind source separation with different sensor spacing and filter length for each frequency range," in *Proc. IEEE International Workshop on Neural Networks for Signal Processing (NNSP '02)*, pp. 465–474, Martigny, Valais, Switzerland, September 2002.

Shoko Araki received the B.E. and 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. She is a member of the IEEE and the Acoustical Society of Japan (ASJ).

Shoji Makino received the B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1979, 1981, and 1993, respectively. He joined NTT in 1981. He is now an Executive Manager of 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. He received 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 the author or coauthor of more than 170 articles in journals and conference proceedings and has been responsible for more than 140 patents. He is a member of the Conference Board of the IEEE SP Society and an Associate Editor of the IEEE Transactions on Speech and Audio Processing. He is a member of the Technical Committee on Audio and Electroacoustics as well as on Speech of the IEEE SP Society. Dr. Makino is a senior member of the IEEE, a member of the ASJ, and the IEICE.

Yoichi Hinamoto was born in Kobe, Japan in 1979. He received the B.E. degree in electrical and electronic engineering from the University of Tokushima in 2001 and M.E. degree in information science from Nara Institute of Science and Technology (NAIST) in 2003. Presently, he is a candidate for the Ph.D. degree in the Graduate School of Informatics, Kyoto University. His research interests include digital signal processing and adaptive filter algorithm. He is a member of the IEICE and the ASJ.

Ryo Mukai received the B.S. and M.S. degrees in information science from the University of Tokyo, Tokyo, Japan, in 1990 and 1992, respectively. His research interests include digital signal processing and blind source separation. He is a member of the IEEE, the ACM, the IEICE, the IPSJ, and the ASJ.

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

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