Blind Source Separation of Acoustic Signals Based on Multistage ICA Combining Frequency-Domain ICA and Time-Domain ICA

Tsuyoki NISHIKAWA\(^{a}\), Student Member, Hiroshi SARUWATARI\(^{b}\), and Kiyohiro SHIKANO\(^{c}\), Regular Members

SUMMARY We propose a new algorithm for blind source separation (BSS), in which frequency-domain independent component analysis (FDICA) and time-domain ICA (TDICA) are combined to achieve a superior source-separation performance under reverberant conditions. Generally speaking, conventional TDICA fails to separate source signals under heavily reverberant conditions because of the low convergence in the iterative learning of the inverse of the mixing system. On the other hand, the separation performance of conventional FDICA also degrades significantly because the independence assumption of narrow-band signals collapses when the number of subbands increases. In the proposed method, the separated signals of FDICA are regarded as the input signals for TDICA, and we can remove the residual crosstalk components of FDICA by using TDICA. The experimental results obtained under the reverberant condition reveal that the separation performance of the proposed method is superior to those of TDICA and FDICA-based BSS methods.

key words: blind source separation, time-domain independent component analysis, frequency-domain independent component analysis, reverberation, microphone array

1. Introduction

Source separation of acoustic signals is to estimate the original sound source signals from among the mixed signals observed in each input channel. This technique is applicable to the realization of noise-robust speech recognition and high-quality hands-free telecommunication systems. The methods of achieving the source separation can be classified into two groups: methods based on a single-channel input, and those based on multichannel inputs. As single-channel types of source separation, a method of tracking a formant structure [1], the organization technique for hierarchical perceptual sounds [2], and a method based on auditory scene analysis [3] have been proposed. As multichannel-type source separation, the method based on array signal processing, e.g., a microphone array system, is one of the most effective techniques [4]. In this system, the directions of arrival (DOAs) of the sound sources are estimated and then each of the source signals is separately obtained using the directivity of the array. The delay-and-sum (DS) array and the adaptive beamformer (ABF) are conventional and popular microphone arrays currently used for source separation and noise reduction.

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

In recent years, alternative source-separation approaches have been proposed by researchers using not array signal processing but a specialized branch of information theory, i.e., information-geometry theory [5], [6]. Blind source separation (BSS) is the approach for estimating original source signals using only the information of the mixed signals observed in each input channel, where the independence among the source signals is mainly used for the separation. This technique is based on unsupervised adaptive filtering [6], and provides us with extended flexibility in that the source-separation procedure requires no training sequences and no a priori information on the DOAs of the sound sources. The early contributory works on BSS were performed by Cardoso and Jutten [7], [8], where high-order statistics of the signals are used for measuring the independence. Common has clearly defined the term independent component analysis (ICA), and presented an algorithm that measures independence among the source signals [9]. This report on ICA was later followed by Bell and Se-
Following a discussion on the results of the experiments, we give conclusions in Sect. 6.

2. Sound Mixing Model of Microphone Array

In this study, a straight-line array is assumed. The number of array elements (microphones) is $K$ and the number of multiple sound sources is $L$ (see Fig. 1), and we deal with the case of $K = L = 2$.

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

$$ X(f) = A(f)S(f), $$

where $X(f)$ is the observed signal vector, $S(f)$ is the source signal vector, and $A(f)$ is the mixing matrix (see Fig. 2); these are given as

$$ X(f) = [X_1(f), \ldots, X_K(f)]^T, $$

$$ S(f) = [S_1(f), \ldots, S_L(f)]^T, $$

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

In this case, $A(f)$ is the mixing matrix which is assumed to be complex-valued because we introduce a model to deal with the arrival lags among the elements of the microphone array and the room reverberations.

![Fig. 1 Configuration of microphone array and signals.](image)

**Fig. 2** Blind source separation procedure performed in multi-stage ICA.
3. Conventional ICA and Its Problems

3.1 Frequency-Domain ICA

The conventional BSS based on FDICA is conducted with the following steps: (1) transform the observed fullband signals into the narrow-band signals, (2) optimize the inverse of the mixing matrix $A(f)$ in each subband, and (3) reconstruct the fullband separated signal from the narrow-band separated signals. FDICA has the following advantages and disadvantages.

**Advantages:**

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

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

**Disadvantages:**

(F3) The separation performance is saturated before reaching a sufficient performance because the independence assumption collapses in each narrowband [15] (see, e.g., Sect. 5.2).

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

As for disadvantage (F4), various solutions have already been proposed [11], [16]-[18]. However, the collapse of the independence assumption, (F3), is a serious and inherent problem, and this prevents us from applying FDICA in a real acoustic environment with a long reverberation.

3.2 Time-Domain ICA

In the conventional BSS based on TDICA, each element of the mixing matrix is represented as a FIR filter. We can optimize its inverse, i.e., form an inverse filter system, by using the fullband observed signals themselves. TDICA has the following advantages and disadvantages.

**Advantages:**

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

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

**Disadvantages:**

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

(T4) The convergence degrades under reverberant conditions.

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

4. Proposed Method: Multistage ICA

As described above, the conventional ICA methods have some disadvantages. However, note that the advantages and disadvantages of FDICA and TDICA are mutually complementary, i.e., (F3) can be resolved by (T1) and (T2), and (T3) and (T4) can be resolved by (F1) and (F2). Hence, in order to resolve the disadvantages, we propose a new algorithm, MSICA, in which FDICA and TDICA are combined (see Fig. 2).

MSICA is conducted with the following steps. In the first stage, we perform FDICA to separate the source signals to some extent with the high-stability advantages of FDICA, (F1) and (F2). In the second stage, we regard the separated signals of FDICA as the input signals for TDICA, and we remove the residual crosstalk components of FDICA by using TDICA. Finally, we regard the output signals of TDICA as the resultant separated signals. MSICA can achieve a high stability and a separation performance superior to that of conventional FDICA and TDICA. In the following sections, we describe details of the ICA-learning rules for each stage.

4.1 First-Stage ICA: Frequency-Domain ICA

In the first-stage ICA, we introduce the fast-convergence FDICA proposed by one of the authors [13]. We perform the signal separation procedure as described below (see FDICA in Fig. 2). In FDICA, first, the short-time analysis of observed signals is conducted by frame-by-frame discrete Fourier transform (DFT). By plotting the spectral values in a frequency bin of each microphone input frame by frame, we consider them as a time series. Hereafter, we designate the time series as $X(f,t) = [X_1(f,t), \ldots, X_K(f,t)]^T$. Next, we perform signal separation using the complex-valued inverse of the mixing matrix, $W(f)$, so that the time series output $Y(f,t) = [Y_1(f,t), \ldots, Y_L(f,t)]^T$ becomes mutually independent; this procedure can be given as

$$Y(f,t) = W(f)X(f,t).$$  \hspace{1cm} (5)

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

In conventional FDICA, the optimal $W(f)$ is obtained by the following iterative equation [11]:

$$W_{i+1}(f) = \eta \left[ \text{diag} \left( \langle \Phi(Y(f,t))Y(f,t)^H \rangle_t \right) \right. \hspace{1cm} \right.$$  \hspace{1cm} \hspace{1cm} (6)

$$- \langle \Phi(Y(f,t))Y(f,t)^H \rangle_t W_i(f), \hspace{1cm} + W_i(f),$$

$$W_i(f) \right.$$
where $(\cdot)_i$ denotes the time-averaging operator, $i$ is used to express the value of the $i$-th step in the iterations, and $\eta$ is the step-size parameter. Also, we define the nonlinear vector function $\Phi(\cdot)$ as

$$\Phi(Y(t), f, t) = \begin{bmatrix} \Phi(Y_1(t), f, t) \\ \Phi(Y_2(t), f, t) \\ \vdots \\ \Phi(Y_L(t), f, t) \end{bmatrix}^T,$$  

$$\Phi(Y_i(t), f, t) = \left[ 1 + \exp(-\text{Re}[Y_i(t), f, t]) \right]^{-1} + j \cdot \left[ 1 + \exp(-\text{Im}[Y_i(t), f, t]) \right]^{-1},$$

where $\text{Re}[Y_i(t), f, t]$ and $\text{Im}[Y_i(t), f, t]$ are the real and imaginary parts of $Y_i(t, f, t)$, respectively.

### 4.2 Second-Stage ICA: Time-Domain ICA

In the second-stage ICA, we introduce the TDICA which uses nonstationarity of the source signals (see TDICA in Fig. 2). We separate the sources by minimizing the nonnegative cost function which takes the minimum value only when the second-order cross-correlation becomes zero if the source signals are nonstationary. The cost function can be given as [14]

$$Q(W(t), z) = \frac{1}{2B} \sum_{b=1}^{B} \left\{ \log \left( \det \text{diag} R_y^{(b)}(0) \right) - \log \left( \det R_y^{(b)}(0) \right) \right\},$$

where $B$ is the number of local analysis blocks. $R_y^{(b)}(n)$ is the correlation matrix of the separated signals, i.e., $R_y^{(b)}(n) = \langle y(t) y(t - n) \rangle_r^{(b)}$, where $(\cdot)_r^{(b)}$ denotes the time-averaging operator for the $b$-th local analysis block, $y(t) = [y_1(t), \ldots, y_L(t)]^T$ is the resultant separated signal vector, and $W(t)(z)$ is the z-transform of the separation filter coefficient $w(t)(n)$ ($n = 0, \ldots, N - 1$); these are given as

$$y(t) = \sum_{n=0}^{N-1} w(t)(n) y(t - n) = W(t)(z) \cdot y(t),$$

$$W(t)(z) = \sum_{n=0}^{N-1} w(t)(n) z^{-n},$$

$$w(t)(n) = \begin{bmatrix} w_{11}(n) & \cdots & w_{1L}(n) \\ \vdots & \ddots & \vdots \\ w_{L1}(n) & \cdots & w_{LL}(n) \end{bmatrix}^T,$$

$$\Delta u_i^{(n)}(t) = -\frac{\partial Q(W(t), z)}{\partial u_i^{(n)}(t)} W(t)(z)^{-1} W(t)(z),$$

The standard gradient, $\partial Q(W(t), z)/\partial u_i^{(n)}(n)$, on the right-hand side in Eq. (14) is rewritten as

$$\frac{\partial Q(W(t), z)}{\partial u_i^{(n)}(n)} = \frac{1}{B} \sum_{b=1}^{B} \left\{ \left( \text{diag} R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) - \left( R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) \right\} W(t)(z)^{-1},$$

where $-T$ represents transpose of inverse matrix. The derivation of Eq. (15) is given by Appendix. Substituting Eq. (15) into Eq. (14), we obtain the following iterative equation of the separation filter (hereafter we designate the iterative equation as “TDICA 1”):

$$w_i^{(t)}(n) = \alpha \Delta u_i^{(n)}(n)$$

$$w_i^{(t)}(n) = \alpha \left( \sum_{b=1}^{B} \left( \text{diag} R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) \right) W_i^{(t)}(z),$$

where $\alpha$ is the step-size parameter. In this study, we define the convolution operation $R_y^{(n)}(H)$ as

$$R_y^{(n)}(H)(z) = \sum_{m=0}^{M-1} R_y^{(n-m)}(h(m)) z^{-m},$$

$$= \sum_{m=0}^{M-1} \langle y(t) y(t - n + m) \rangle_r^{(b)} h(m),$$

where $H(z)$ is the z-transform of an arbitrary FIR-filter matrix $h(m)$. In this equation, when $m$ is larger than $n$, we calculate the future correlation between the separated signals. On the other hand, when $m$ is larger than $n$, we calculate the past correlation. Therefore we can achieve the iterative leaning for a both-side filter, and we can treat the mixing condition even including the non-minimum phase systems [20].
Since Eq. (16) evaluates only off-diagonal of $R_y^{(b)}(0)$, we confirmed that the iterative equation of Eq. (16) could not achieve a superior separation performance under the reverberant condition (see Sect. 5.3). Namely, the source separation is not achieved by only using nonstationarity of signals. Therefore we use not only nonstationarity of signals but also time-delayed decorrelation approach. We expand Eq. (16) to the following equation to evaluate the off-diagonal of $R_y^{(b)}(n)$ for all time delays $n$ (hereafter we designate the iterative equation as "TDICA 2"):

$$\begin{align*}
\mathbf{w}_{i+1}(n) &= \mathbf{w}_i(n) \\
&+ \frac{\alpha}{B} \sum_{b=1}^{B} \left( \left( \text{diag} R_y^{(b)}(0) \right)^{-1} \text{diag} R_y^{(b)}(n) - \left( \text{diag} R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) \right) \mathbf{W}_i^{(s)}(z). \quad (18)
\end{align*}$$

We confirmed that the separation performance was improved by using both nonstationarity and time-delayed decorrelation approach (see Sect. 5.3).

5. Experiments and Results

5.1 Experimental Setup

A two-element array with the interelement spacing of 4 cm is assumed. The speech signals are assumed to arrive from two directions, $-30^\circ$ and $40^\circ$ (direction normal to the array is set to be $0^\circ$). The distance between the microphone array and the loudspeakers is 1.15 m (see Fig. 3). Two kinds of sentences, spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research [21], are used as the original speech samples. The sampling frequency is 8 kHz and the length of speech is limited to within 3 seconds. Using these sentences, we obtain 12 combinations with respect to speakers and source directions. As for the mixing system, we use the impulse responses recorded in a real room with the reverberation time of 300 ms. In order to evaluate the performance, we used the noise reduction rate (NRR), defined as the output SNR in dB minus input SNR in dB.

5.2 Relation between Separation Performance and Number of Subbands in FDICA

In order to confirm the independence property of narrow-band signals in FDICA ((F3) described in Sect. 3.1), we carried out the preliminary experiment under the following analysis conditions. The number of subbands (frame length in DFT) is set to be from 32 to 4096, the frame shift is 16 taps, the window function is a Hamming window, the number of iterations in ICA is 30, and the step-size parameter $\eta$ for iterations is set to be $1.0 \times 10^{-5}$.

Figure 4 shows the NRR results for different numbers of subbands in FDICA. As shown in Fig. 4, the NRR of FDICA obviously degrades when the number of subbands becomes too large, and the separation performance is saturated before reaching a sufficient performance. This is because we transform the fullband signals into the narrow-band signals and the independence assumption collapses in each frequency band, particularly when the number of subbands is large.

In order to confirm the fact, we newly define the following objective measure to quantify an independence, and investigate the relation between the number of subbands and the independence among subband signals.

$$J = \left( \|	ext{diag} \left( \Phi(\mathbf{Y}(f,t)) \mathbf{Y}(f,t) \right) \|_F \right)_{f}, \quad (19)$$

where $\| \cdot \|$ is frobenius norm of matrix. This measure $J$ is a part of the iterative Eq. (6) and has no dimension. Therefore the absolute value of $J$ is meaningless itself, however, the relative value between the different numbers of subbands is important. If narrow-band signals become mutually independent, the measure $J$ becomes zero. Also we can consider that the independence of subband signals is high when $J$ is small. In order to evaluate the independence of real narrow-band...
speech signals, we carried out the experiment in which the input signal, \( Y(f,t) \), in Eq. (19) is regarded as the perfectly separated sources, i.e., original speech samples. Figure 5 shows the relation between the number of subbands and the value of \( J \) which corresponds to the independence of subband signals. Figure 5 shows that the independence decreases as the number of subbands increases, especially when the number of subbands is large.

Above-mentioned experimental results clarify the disadvantage that the separation performance is saturated in FDICA because we transform the fullband signals into the narrow-band signals. On the basis of these results, we should cascade another signal processing analysis, e.g., TDICA, with FDICA to obtain the further separation performances.

5.3 Relation between Separation Performance and Filter Length in TDICA

We carried out the experiments using TDICA and MSICA to evaluate the contribution of increments of separation-filter length for improving the separation performances under reverberant conditions. As for TDICA, we used the iterative equations by substituting \( R^{(b)}_{(x)}(n) = (x(t)x(t-n))^T \) for \( R^{(b)}_{(y)}(n) \), where \( x(t) \) is the time-domain signal of \( X(f) \). The analysis conditions of these experiments are as follows: the filter length \( N \) is set to be from 10 to 2000 taps, the maximum number of iterations is 500, and the step-size parameter \( \alpha \) for iterations is set to be \( 1/N \). As for the initial value of \( \mathbf{w}^{(0)}(n) \), the diagonal components of \( \mathbf{w}^{(0)}(n) \) are set to be \( \delta(n-N/2) \), where \( \delta(0) = 1 \) and \( \delta(m) = 0, (m \neq 0) \), and the off-diagonal components are set to be 0. As for the local analysis block, we divided the signals equally into \( B \) parts (\( B = 1-10 \)). We chose the optimal \( B \) and number of iterations for each filter length because the convergence is different for every filter length. As for the FDICA part in MSICA, the analysis conditions are the same as those given in Sect. 5.2, except for the number of subbands (which is fixed at 1024 bands).

Figures 6(a), (b) and (c) show the NRR results in the TDICA 1, TDICA 2 and MSICA for different filter lengths, respectively. Figure 6(a) shows that TDICA 1 can not achieve a signal separation under the reverberant condition. Comparing Fig. 6(a) with Fig. 6(b), we confirm that TDICA 2 can achieve a superior separation performance to TDICA 1. These results show that it is necessary to evaluate correlations of different times to achieve a superior performance. As shown in Fig. 6(b), when the separation filter is lengthened, the separation performance of the TDICA degrades. This also implies that the simple TDICA separates only the direct and several reflected components of arriving signals. On the other hand, in Fig. 6(c), the separation performance of MSICA is improved when the filter length is longer. This reveals that the TDICA part in MSICA can separate the source signals even with the reverberation components, and the TDICA is still useful near the optimal point.

Figures 7 and 8 show the filter coefficients in which the best separation performance is performed after the iterative learning in the simple TDICA 2 and TDICA part in MSICA. In these figures, (a), (b), (c), and (d)
correspond to the filter coefficients of \( w_{11}^{(l)}(n) \), \( w_{12}^{(l)}(n) \), \( w_{21}^{(l)}(n) \), and \( w_{22}^{(l)}(n) \), respectively. As shown in these figures, we can confirm that the separation FIR-filter \( w^{(l)}(n) \) is learned as a both-side filter. As shown in Fig. 7, the filter coefficients are obviously changed from the initial value by the iterative leaning of the simple
TDICA 2. On the other hand, as shown in Fig. 8, the filter coefficients are not so changed from the initial value in TDICA part in MSICA. The reason for this phenomenon is that FDICA reduced the component through the direct path from the interference source in MSICA and consequently TDICA part in MSICA only has to reduce the residual crosstalk components.

5.4 Comparison between Conventional ICA and MSICA

We compared the performance of the proposed MSICA with that of the conventional ICA under the reverberant condition. As for FDICA, the analysis conditions are the same as those given in Sect. 5.2, except for the number of subbands (which is fixed at 1024 bands). As for TDICA, the number of local analysis blocks, B, is fixed at 3 blocks, the number of iterations is 400, and the filter length is 10 taps. As for the TDICA part in MSICA, the number of local analysis blocks, B, is fixed at 9 blocks, the number of iterations is 400, and the filter length is 1000 taps.

Figure 9 shows the NRRs of the conventional FDICA, TDICA, and MSICA. In this figure, we separately plot the NRRs for different combination of speakers, and the averages of their NRRs. The results reveal that the separation performances of the proposed MSICA are superior to those of the conventional FDICA and TDICA with every combination. Specifically, compared with the conventional ICA, the proposed method can improve the NRR by about 2.7 dB over that of FDICA and by about 6.2 dB over that of TDICA, for an average of 12 combinations.

As described in Sect. 5.2, the FDICA in this study showed the saturation of NRR when we used the 1024-subband analysis. As described in Sect. 5.3, the simple TDICA could not separate the source signals accurately under the reverberant condition. These findings indicate the practical limitations of the separation performances of conventional ICA-based BSS methods. From the results of Fig. 9, however, we can confirm that the proposed MSICA can inherently remove these limitations, and is effective for improving the separation performance and convergence under reverberant conditions.

5.5 Discussion on Combination Order in MSICA

As described in the previous section, the combination of FDICA and TDICA can contribute to the improvement of separation. In this combination, the advantage (F2) of FDICA is useful in the initial step of separation procedure and the advantage (T2) of TDICA is also useful in the later step. Therefore we use FDICA as the first-stage ICA and TDICA as the second-stage ICA. In order to confirm the availability of this combination order, we compare the proposed combination (hereafter we designate this combination as "MSICA1") with the combination in which TDICA is used in the first stage and FDICA is used in the second stage (hereafter we designate this swapped combination as "MSICA2").

The experiment of MSICA2 was carried out in the following manner. As for TDICA part in MSICA2, the number of local analysis blocks, B, is fixed at 3 blocks, the number of iterations is 400, and the filter length is 10 taps. As for FDICA part in MSICA2, the analysis conditions are the same as those given in Sect. 5.2, except for the number of subbands (which is fixed at 1024 bands). As the result, the NRR of 7.5 dB is obtained in MSICA2, and this performance is better than that of simple TDICA but is poorer than that of MSICA1 and simple FDICA. In MSICA2, the separation performance is still improved by using FDICA in the second stage, however, the separation performance is saturated because of the disadvantage (F3) of FDICA. MSICA2 can not achieve the separation performance of 9.4 dB which corresponds to NRR of simple FDICA. This reason is that FDICA in this paper uses the beamforming technique and the directivity pattern of the array which provide a good initial value of the separation matrix to improve the convergence [13], however, such kind of information is no longer valid in the combination order of MSICA2 because we can not know the effective positions of the array elements after the first-stage TDICA and can not depict the directivity pattern. Thus the separation performance of MSICA2 is almost equal to that of a raw FDICA without the beamforming technique (from [13] we can see the NRR of about 7.5 dB at the 30-iteration point). This fact indicates that the swapped combination order of MSICA2 has no contribution to the improvement of the separation performance, and the proposed combination order of MSICA1 (FDICA in the first stage and TDICA in the second stage) is essential.

6. Conclusion

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

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

Secondly, the results of the signal separation experiment with TDICA reveals that the separation performance of TDICA is not sufficient and is not improved even when the separation filter is lengthened.

Finally, the results of the signal separation experiment with the proposed method reveals that the separation performance of the proposed algorithm is superior to that of conventional ICA-based BSS methods, and the combination of FDICA and TDICA is inherently effective for improving the separation performance. Specifically, the proposed method can improve the SNR by about 2.7 dB over that of FDICA and by about 6.2 dB over that of TDICA, for an average of 12 speaker-combinations.

Acknowledgement

The authors are grateful to Dr. Shoji Makino, Miss Shoko Araki and Mr. Ryo Mukai of NTT. CO., LTD, and Dr. Mitsuru Kawamoto of Shimane University for their useful discussions. This work was partly supported by CREST (Core Research for Evolutional Science and Technology) of JST (Japan Science and Technology Corporation).

References


\[ I: \text{an adaptive algorithm based on neuroimimetic architecture,} \]
\[ \text{Signal Processing, vol.24, pp.1–10, 1991.} \]

Appendix: Derivation of Eq. (15)

The cost function Eq. (9) is

\[
\begin{align*}
\frac{\partial Q(W^{(i)}(z))}{\partial w^{(i)}(n)} &= \frac{1}{2B} \sum_{b=1}^{B} \left\{ \frac{\partial}{\partial \widehat{w}^{(i)}(n)} \log \left( \det \text{diag} \widehat{R}_y^{(b)}(0) \right) \right. \\
& \quad - \left. \frac{\partial}{\partial w^{(i)}(n)} \log \left( \det \widehat{R}_y^{(b)}(0) \right) \right\}. \\
& \quad \text{(A-1)}
\end{align*}
\]

In this study, we deal with the case in which the number of the source signals and that of the observed signals are 2, i.e., the separation filter matrix \( w^{(i)}(n) \) is a 2 \times 2 matrix.
The partial differentiation $\partial \log(\det R^{(b)}_y(0)) / \partial w^{(t)}(n)$ is given by the following equation:

$$
\frac{\partial \log(\det R^{(b)}_y(0))}{\partial w^{(t)}(n)} = \frac{\partial \log(\det R^{(b)}_y(0))}{\partial R^{(b)}_y(0)} \cdot \frac{\partial R^{(b)}_y(0)}{\partial w^{(t)}(n)}
= \left( \det R^{(b)}_y(0) \right)^{-1} \frac{\partial \det R^{(b)}_y(0)}{\partial w^{(t)}(n)}.
$$  \hspace{2cm} (A.2)

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

$$
R^{(b)}_y(0) = (y(t)y(t)^T)^{(b)}_x,
$$  \hspace{2cm} (A.4)

$$
W^{(t)}(z^{-1}) = \sum_{n=0}^{N-1} w^{(t)}(n)z^n.
$$  \hspace{2cm} (A.5)

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

$$
[R^{(b)}_y(0)]_{11} = [W^{(t)}(z^{-1})]_{11}^2 [R^{(b)}_{y^{(t)}}(0)]_{11}
+ [W^{(t)}(z^{-1})]_{11} [W^{(t)}(z^{-1})]_{12} \cdot (R^{(b)}_{y^{(t)}}(0)]_{21} + (R^{(b)}_{y^{(t)}}(0)]_{12}
+ [W^{(t)}(z^{-1})]_{12}^2 [R^{(b)}_{y^{(t)}}(0)]_{22},
$$  \hspace{2cm} (A.6)

$$
[R^{(b)}_y(0)]_{12} = [W^{(t)}(z^{-1})]_{11} [W^{(t)}(z^{-1})]_{12} [R^{(b)}_{y^{(t)}}(0)]_{11}
+ [W^{(t)}(z^{-1})]_{12} [W^{(t)}(z^{-1})]_{21} [R^{(b)}_{y^{(t)}}(0)]_{22}
+ [W^{(t)}(z^{-1})]_{11} [W^{(t)}(z^{-1})]_{21} (R^{(b)}_{y^{(t)}}(0)]_{22}
+ [W^{(t)}(z^{-1})]_{12} [W^{(t)}(z^{-1})]_{21} (R^{(b)}_{y^{(t)}}(0)]_{22},
$$  \hspace{2cm} (A.7)

$$
[R^{(b)}_y(0)]_{21} = [W^{(t)}(z^{-1})]_{11} [W^{(t)}(z^{-1})]_{12} [R^{(b)}_{y^{(t)}}(0)]_{11}
+ [W^{(t)}(z^{-1})]_{12} [W^{(t)}(z^{-1})]_{21} [R^{(b)}_{y^{(t)}}(0)]_{22}
+ [W^{(t)}(z^{-1})]_{11} [W^{(t)}(z^{-1})]_{21} (R^{(b)}_{y^{(t)}}(0)]_{22}
+ [W^{(t)}(z^{-1})]_{12} [W^{(t)}(z^{-1})]_{21} (R^{(b)}_{y^{(t)}}(0)]_{22},
$$  \hspace{2cm} (A.8)

$$
[R^{(b)}_y(0)]_{22} = [W^{(t)}(z^{-1})]_{21}^2 [R^{(b)}_{y^{(t)}}(0)]_{11}
+ [W^{(t)}(z^{-1})]_{21} [W^{(t)}(z^{-1})]_{22} \cdot (R^{(b)}_{y^{(t)}}(0)]_{21} + (R^{(b)}_{y^{(t)}}(0)]_{12}
+ [W^{(t)}(z^{-1})]_{22} (R^{(b)}_{y^{(t)}}(0)]_{22}.
$$  \hspace{2cm} (A.9)

where $[\cdot]_{ij}$ denotes the $ij$-th element of the argument.

The partial differentiation $\partial \det R^{(b)}_y(0)/\partial w^{(t)}(n)$ in Eq. (A.2) is expanded as the following equation by substituting Eq. (A.3):

$$
\frac{\partial \det R^{(b)}_y(0)}{\partial w^{(t)}(n)} = \frac{\partial \det R^{(b)}_y(0)}{\partial w^{(t)}(n)}
= \left( \det R^{(b)}_y(0) \right)^{-1} \frac{\partial \det R^{(b)}_y(0)}{\partial w^{(t)}(n)}.
$$  \hspace{2cm} (A.3)

Hereafter, we resolve Eq. (A.10) into the partial differentiation for each element of $w^{(t)}(n)$ as

$$
\frac{\partial \det R^{(b)}_y(0)}{\partial w^{(t)}(n)} = 2^n \left( \frac{\partial \det R^{(b)}_y(0)}{\partial w^{(t)}(n)} \right)_{11}
= 2^n \left( \frac{\partial \det R^{(b)}_y(0)}{\partial w^{(t)}(n)} \right)_{12} \cdot \left( R^{(b)}_y(0) ]_{21} + ( R^{(b)}_y(0) ]_{12} \right)
+ 2^n \left( \frac{\partial \det R^{(b)}_y(0)}{\partial w^{(t)}(n)} \right)_{22} \cdot \left( R^{(b)}_y(0) ]_{22} + ( R^{(b)}_y(0) ]_{21} \right).
$$  \hspace{2cm} (A.10)
where we used the partial differentiation as
\[
\frac{\partial \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{ij}}{\partial \left[ \mathbf{w}^{(t)}(n) \right]_{ij}} = \frac{\partial}{\partial \left[ \mathbf{w}^{(t)}(n) \right]_{ij}} \left( \sum_{n'=0}^{N-1} \mathbf{w}^{(t)}(n') z^{n'} \right) = z^n, \quad (A.12)
\]
and the following relation defined from the symmetry of correlation matrix,
\[
\mathbf{R}^{(b)}(0)_{12} = \mathbf{R}^{(b)}(0)_{21}, \quad (A.13)
\]
\[
\mathbf{R}^{(b)}(0)_{12} = \mathbf{R}^{(b)}(0)_{21}. \quad (A.14)
\]
By calculating the other elements, we obtain
\[
\frac{\partial \det \mathbf{R}^{(b)}(0)}{\partial \left[ \mathbf{w}^{(t)}(n) \right]_{12}} = z^n \left( \left( \mathbf{R}^{(b)}(0) \right)_{12} \left\{ 2 \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{12} \left[ \mathbf{R}^{(b)}(0) \right]_{22} + \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{11} \left[ \mathbf{R}^{(b)}(0) \right]_{21} + \left[ \mathbf{R}^{(b)}(0) \right]_{12} \right\} \\
- \left[ \mathbf{R}^{(b)}(0) \right]_{21} \left\{ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{12} \left[ \mathbf{R}^{(b)}(0) \right]_{21} + \left[ \mathbf{R}^{(b)}(0) \right]_{12} \right\} \\
+ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{22} \left[ \mathbf{R}^{(b)}(0) \right]_{22} \\
- \left[ \mathbf{R}^{(b)}(0) \right]_{12} \left\{ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{12} + \left[ \mathbf{R}^{(b)}(0) \right]_{12} \right\} \\
+ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{22} \left[ \mathbf{R}^{(b)}(0) \right]_{22} \right) \right), \quad (A.15)
\]
\[
\frac{\partial \det \mathbf{R}^{(b)}(0)}{\partial \left[ \mathbf{w}^{(t)}(n) \right]_{12}} = z^n \left( \left( \mathbf{R}^{(b)}(0) \right)_{21} \left\{ 2 \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{22} + \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{11} \left[ \mathbf{R}^{(b)}(0) \right]_{21} + \left[ \mathbf{R}^{(b)}(0) \right]_{12} \right\} \\
\right)
\]
\[
= z^n \left( \left[ \mathbf{R}^{(b)}(0) \right]_{21} \left\{ 2 \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{22} + \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{11} \left[ \mathbf{R}^{(b)}(0) \right]_{21} + \left[ \mathbf{R}^{(b)}(0) \right]_{12} \right\} \\
+ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{22} \\
- \left[ \mathbf{R}^{(b)}(0) \right]_{12} \left\{ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{21} + \left[ \mathbf{R}^{(b)}(0) \right]_{12} \right\} \\
+ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{22} \right) \right), \quad (A.16)
\]
Here, the elements of \( \mathbf{W}^{(t)}(z^{-1}) \) \( \mathbf{R}^{(b)}(0) \) are given as the following equation by using the elements of \( \mathbf{R}^{(b)}(0) \) and \( \mathbf{W}^{(t)}(z^{-1}) \):
\[
\left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{11} \left[ \mathbf{R}^{(b)}(0) \right]_{11} = \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{11} \left[ \mathbf{R}^{(b)}(0) \right]_{11}, \\
+ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{12} \left[ \mathbf{R}^{(b)}(0) \right]_{11}, \quad (A.18)
\]
\[
\left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{12} \left[ \mathbf{R}^{(b)}(0) \right]_{12} = \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{12} \left[ \mathbf{R}^{(b)}(0) \right]_{12}, \quad (A.19)
\]
\[
\left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{11} = \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{11} + \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{22} \left[ \mathbf{R}^{(b)}(0) \right]_{11}, \quad (A.20)
\]
Substituting Eqs. (A.18)–(A.21) into Eqs. (A.11)–(A.17), the following equations are obtained:
\[
\frac{\partial \det \mathbf{R}^{(b)}(0)}{\partial \left[ \mathbf{w}^{(t)}(n) \right]_{11}} = 2z^n \left( \left[ \mathbf{R}^{(b)}(0) \right]_{21} \left\{ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{21} + \left[ \mathbf{R}^{(b)}(0) \right]_{22} \right\} \\
- \left[ \mathbf{R}^{(b)}(0) \right]_{12} \left\{ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{12} + \left[ \mathbf{R}^{(b)}(0) \right]_{22} \right\} \\
+ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{22} \left[ \mathbf{R}^{(b)}(0) \right]_{22} \right), \quad (A.22)
\]
\[
\frac{\partial \det \mathbf{R}^{(b)}(0)}{\partial \left[ \mathbf{w}^{(t)}(n) \right]_{12}} = 2z^n \left( \left[ \mathbf{R}^{(b)}(0) \right]_{21} \left\{ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{22} + \left[ \mathbf{R}^{(b)}(0) \right]_{22} \right\} \\
- \left[ \mathbf{R}^{(b)}(0) \right]_{12} \left\{ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{21} \left[ \mathbf{R}^{(b)}(0) \right]_{12} + \left[ \mathbf{R}^{(b)}(0) \right]_{22} \right\} \\
+ \left[ \mathbf{W}^{(t)}(z^{-1}) \right]_{22} \left[ \mathbf{R}^{(b)}(0) \right]_{22} \right), \quad (A.23)
\]
\[
\begin{align*}
\partial \det R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left( [R_y^{(0)}(0)]_{22} [W(t)(z^{-1})R_y^{(0)}(0)]_{12} \\
- [R_y^{(0)}(0)]_{12} [W(t)(z^{-1})R_y^{(0)}(0)]_{22} \right) \\
\end{align*}
\]

(A-23)

\[
\begin{align*}
\partial \det R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left( -[R_y^{(0)}(0)]_{21} [W(t)(z^{-1})R_y^{(0)}(0)]_{11} \\
+ [R_y^{(0)}(0)]_{11} [W(t)(z^{-1})R_y^{(0)}(0)]_{21} \right). \\
\end{align*}
\]

(A-24)

\[
\begin{align*}
\partial \det R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left( -[R_y^{(0)}(0)]_{21} [W(t)(z^{-1})R_y^{(0)}(0)]_{12} \\
+ [R_y^{(0)}(0)]_{12} [W(t)(z^{-1})R_y^{(0)}(0)]_{22} \right). \\
\end{align*}
\]

(A-25)

Thus, Eq. (A-10) is rewritten as the following matrix form:

\[
\begin{align*}
\partial \det R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left[ \\
\begin{array}{ll}
[R_y^{(0)}(0)]_{22} & -[R_y^{(0)}(0)]_{12} \\
-[R_y^{(0)}(0)]_{21} & [R_y^{(0)}(0)]_{11}
\end{array} \right] \\
\cdot \\
\begin{array}{l}
[W(t)(z^{-1})R_y^{(0)}(0)]_{11} [W(t)(z^{-1})R_y^{(0)}(0)]_{12} \\
[W(t)(z^{-1})R_y^{(0)}(0)]_{21} [W(t)(z^{-1})R_y^{(0)}(0)]_{22}
\end{array} \\
= 2z^n \text{adj} \left( R_y^{(0)}(0) \right) W(t)(z^{-1})R_y^{(0)}(0), \\
\end{align*}
\]

where adj[·] is adjoint matrix. By substituting Eq. (A-26) into Eq. (A-2), \( \partial \log(\det R_y^{(0)}(0))/\partial w(t)(n) \) is rewritten as

\[
\begin{align*}
\partial \log(\det R_y^{(0)}(0)) & \quad \partial w(t)(n) \\
= 2z^n (\det R_y^{(0)}(0))^{-1} \text{adj} \left( R_y^{(0)}(0) \right) \\
\cdot \\
W(t)(z^{-1})R_y^{(0)}(0) \\
= 2z^n \left( R_y^{(0)}(0) \right)^{-1} W(t)(z^{-1})R_y^{(0)}(0) \\
\cdot \\
W(t)(z^{-1})^{-T} W(t)(z^{-1})^{-T}. \\
\end{align*}
\]

(A-27)

By using relation Eq. (A-3), Eq. (A-27) is expanded as

\[
\begin{align*}
\partial \log(\det R_y^{(0)}(0)) & \quad \partial w(t)(n) \\
= 2 \left( R_y^{(0)}(0) \right)^{-1} z^n \langle y(t)y(t)^T \rangle_t^{(0)} W(t)(z^{-1})^{-T} \\
= 2 \left( R_y^{(0)}(0) \right)^{-1} \langle z^{-n} y(t)y(t)^T \rangle_t^{(0)} W(t)(z^{-1})^{-T} \\
= 2 \left( R_y^{(0)}(0) \right)^{-1} \langle y(t)y(t-n)^T \rangle_t^{(0)} W(t)(z^{-1})^{-T} \\
= 2 \left( R_y^{(0)}(0) \right)^{-1} R_y^{(0)}(n) W(t)(z^{-1})^{-T}. \\
\end{align*}
\]

(A-28)

On the other hand, \( \partial \log(\det \text{diag} R_y^{(0)}(0))/\partial w(t)(n) \) is also derived as the same manner in the above derivation:

\[
\begin{align*}
\frac{\partial \log(\det \text{diag} R_y^{(0)}(0))}{\partial w(t)(n)} \\
= \frac{\partial \log(\det \text{diag} R_y^{(0)}(0))}{\partial \text{diag} R_y^{(0)}(0)} \cdot \frac{\partial \text{diag} R_y^{(0)}(0)}{\partial w(t)(n)} \\
= \left( \text{diag} R_y^{(0)}(0) \right)^{-1} \partial \text{diag} R_y^{(0)}(0) \cdot \frac{\partial w(t)(n)}{\partial w(t)(n)}. \\
\end{align*}
\]

(A-29)

The elements of \( \partial \text{diag} R_y^{(0)}(0)/\partial w(t)(n) \) are obtained by

\[
\begin{align*}
\partial \text{diag} R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left( [R_y^{(0)}(0)]_{22} \left\{ [W(t)(z^{-1})]_{11} [R_y^{(0)}(0)]_{11} \\
+ [W(t)(z^{-1})]_{12} [R_y^{(0)}(0)]_{21} \right\} \right), \\
\end{align*}
\]

(A-30)

\[
\begin{align*}
\partial \text{diag} R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left( [R_y^{(0)}(0)]_{22} \left\{ [W(t)(z^{-1})]_{11} [R_y^{(0)}(0)]_{12} \\
+ [W(t)(z^{-1})]_{12} [R_y^{(0)}(0)]_{22} \right\} \right), \\
\end{align*}
\]

(A-31)

\[
\begin{align*}
\partial \text{diag} R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left( [R_y^{(0)}(0)]_{11} \left\{ [W(t)(z^{-1})]_{21} [R_y^{(0)}(0)]_{11} \\
+ [W(t)(z^{-1})]_{22} [R_y^{(0)}(0)]_{21} \right\} \right), \\
\end{align*}
\]

(A-32)

\[
\begin{align*}
\partial \text{diag} R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left( [R_y^{(0)}(0)]_{11} \left\{ [W(t)(z^{-1})]_{21} [R_y^{(0)}(0)]_{12} \\
+ [W(t)(z^{-1})]_{22} [R_y^{(0)}(0)]_{22} \right\} \right). \\
\end{align*}
\]

(A-33)

Thus, \( \partial \text{diag} R_y^{(0)}(0)/\partial w(t)(n) \) is given by the following matrix form:

\[
\begin{align*}
\partial \text{diag} R_y^{(0)}(0) & \quad \partial w(t)(n) \\
= 2z^n \left[ \\
\begin{array}{c}
[R_y^{(0)}(0)]_{22} \\
0
\end{array} \\
\begin{array}{c}
[R_y^{(0)}(0)]_{11} \\
[R_y^{(0)}(0)]_{12} \\
[R_y^{(0)}(0)]_{21} \\
[R_y^{(0)}(0)]_{22}
\end{array} \\
\end{array} \right] \\
\cdot \\
\begin{array}{l}
\left\{ [W(t)(z^{-1})]_{11} [R_y^{(0)}(0)]_{11} \\
+W(t)(z^{-1})]_{12} [R_y^{(0)}(0)]_{21} \right\} \\
\left\{ [W(t)(z^{-1})]_{11} [R_y^{(0)}(0)]_{12} \\
+W(t)(z^{-1})]_{12} [R_y^{(0)}(0)]_{22} \right\}
\end{array} \\
= 2z^n \text{adj} \left[ \text{diag} R_y^{(0)}(0) \right] W(t)(z^{-1})R_y^{(0)}(0). \\
\end{align*}
\]

(A-34)

\[
\begin{align*}
\partial \log(\det \text{diag} R_y^{(0)}(0))/\partial w(t)(n) \quad \text{is rewritten as following equation by substituting Eq. (A-34) into}
\end{align*}
\]
Eq. (A.29):

\[
\frac{\partial \log(\det \text{diag} R_y^{(b)}(0))}{\partial w^{(t)}(n)} = 2z^n \left( \text{det diag} R_y^{(b)}(0) \right)^{-1} \text{adj} \left[ \text{diag} R_y^{(b)}(0) \right] \\
\cdot W^{(t)}(z^{-1}) R_y^{(b)}(0) \\
= 2z^n \left( \text{diag} R_y^{(b)}(0) \right)^{-1} W^{(t)}(z^{-1}) R_y^{(b)}(0) \\
\cdot W^{(t)}(z^{-1})^T W^{(t)}(z^{-1})^{-T} \\
= 2 \left( \text{diag} R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) W^{(t)}(z^{-1})^{-T}. \quad (A.35)
\]

As a result, \( \partial Q(W^{(t)}(z)) / \partial w^{(t)}(n) \) is obtained by following equation by substituting Eqs. (A.28), (A.35) into Eq. (A.1):

\[
\frac{\partial Q(W^{(t)}(z))}{\partial w^{(t)}(n)} = \frac{1}{2B} \sum_{b=1}^{B} \left\{ 2 \left( \text{diag} R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) W^{(t)}(z^{-1})^{-T} \\
- \left( R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) W^{(t)}(z^{-1})^{-T} \right\} \\
= \frac{1}{B} \sum_{b=1}^{B} \left\{ \left( \text{diag} R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) \\
- \left( R_y^{(b)}(0) \right)^{-1} R_y^{(b)}(n) \right\} W^{(t)}(z^{-1})^{-T}. \quad (A.36)
\]

Tsuyoshi Nishikawa was born in Mie, Japan on 1978. He received the B.E. degree in electronic system and information engineering from Kinki University in 2000 and received 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 the Acoustical Society of Japan.

Hiroshi Saruwatari was born in Nagoya, Japan, on July 27, 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. 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, and the Acoustical Society of Japan.

Kiyohiro Shikano received the B.S., M.S., and Ph.D. degrees in electrical engineering from Nagoya University in 1970, 1972, and 1980, respectively. He is currently a professor of Nara Institute of Science and Technology (NAIST), where he is directing speech and acoustics laboratory. His major research areas are speech recognition, multi-modal dialog system, speech enhancement, adaptive microphone array, and acoustic field reproduction. From 1972 he had been working at NTT Laboratories, where he had been engaged in speech recognition research. During 1990–1993, he was the executive research scientist at NTT Human Interface Laboratories, where he supervised the research of speech recognition and speech coding. During 1986–1990, he was the Head of Speech Processing Department at ATR Interpreting Telephony Research Laboratories, where he was directing speech recognition and speech synthesis research. During 1984–1986, he was a visiting scientist in Carnegie Mellon University, where he was working on distance measures, speaker adaptation, and statistical language modeling. He received the Yonezawa Prize from IEICE in 1975, the Signal Processing Society 1990 Senior Award from IEEE in 1991, the Technical Development Award from ASJ in 1994, IPSJ Yamashita SIG Research Award in 2000, and Paper Award from the Virtual Reality Society of Japan in 2001. He is a member of the Information Processing Society of Japan, the Acoustical Society of Japan (ASJ), Japan VR Society, the Institute of Electrical and Electronics Engineers (IEEE), and International Speech Communication Society.