TWO-STAGE BLIND SOURCE SEPARATION COMBINING SIMO-MODEL-BASED ICA AND ADAPTIVE BEAMFORMING

Hiroshi Saruwatari, Satoshi Ukai, Tomoya Takatani, Tsuyoki Nishikawa, and Kiyohiro Shikano

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma, Nara, 630-0192, JAPAN
phone: +081-743-72-5281, fax: +81-743-72-5289, email: sawatari@is.naist.jp

ABSTRACT

A new two-stage blind source separation (BSS) for convolutive mixtures of speech is proposed, in which a Single-Input Multiple-Output (SIMO)-model-based ICA (SIMO-ICA) and an adaptive beamforming (ABF) are combined. SIMO-ICA can separate the mixed signals, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Thus, the separated signals of SIMO-ICA can maintain the spatial qualities of each sound source, and directions-of-arrival (DOAs) of the sources can be estimated after the separation by SIMO-ICA. Owing to the attractive property, the supervised ABF can be applied to efficiently remove the residual interference components after SIMO-ICA and the DOA estimation procedures. The experimental results reveal that the separation performance can be considerably improved by using the proposed method. In addition, the proposed method outperforms the combination of the conventional SIMO-output-type ICA and ABF, as well as both of the simple ICA and ABF.

1. INTRODUCTION

Blind source separation (BSS) is the approach taken to estimate original source signals using only the information of the mixed signals observed in each input channel. In recent works of BSS based on independent component analysis (ICA), various methods have been proposed for acoustic-sound separation [1, 2, 3, 4]. In this paper, we mainly address the BSS problem under highly reverberant conditions which often arise in many practical audio applications. The separation performance of the conventional ICA is far from being sufficient in such a case because too long separation filters is required but the unsupervised learning of the filter is not so easy. Therefore, one possible improvement is to partly combine ICA with another supervised signal enhancement technique, e.g., spectral subtraction. However, in the traditional ICA framework, each of the separated outputs is a *monaural* signal, and this leads to the drawback that many kinds of superior multichannel supervised techniques such as an adaptive beamforming (ABF) [5] cannot be applied.

To solve the problem, we propose a novel two-stage BSS algorithm. In this approach, the BSS problem is resolved into two stages: (a) previously proposed blind separation technique using a Single-Input Multiple-Output (SIMO)-model-based ICA (SIMO-ICA) [6, 7], and (b) the ABF in the supervised filtering framework. Here the term "SIMO" represents the specific transmission system in which the input is a single source signal and the outputs are its transmitted signals observed at multiple microphones. SIMO-ICA can separate the mixed signals, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Thus, the separated signals of SIMO-ICA can maintain the spatial qualities of each sound source, and directions-of-arrival (DOAs) information of the sources can be estimated after the separation by SIMO-ICA. Also, the most important and attractive property is that the residual components of the interference, which are often staying in the output of SIMO-ICA as well as the conventional ICA, maintain the *distinct spatial distribution* from that of the target signal. This is very different from the existing SIMO-output-type ICAs [1, 8, 9] which have no separability. Therefore, the supervised ABF can be applied to efficiently remove the residual interference components after SIMO-ICA and the DOA estimation procedures.

In this paper, we can explicitly reveal that (a) only the specific combination of SIMO-ICA and ABF is valid for the improvement of the separation performance, and (b) the proposed method can successfully achieve the BSS for both speech-speech and speech-noise mixtures even under a realistic reverberant condition.

2. MIXING PROCESS

In this study, a straight-line array is assumed. The number of microphones is K and the number of multiple sound sources is L. The coordinates of the elements are designated as d_k $(k = 1, \cdots, K)$, and the directions of arrival of multiple signals are designated as θ_l $(l = 1, \cdots, L)$. Hereafter, we deal with discrete time series, and symbols t, n and d are used as the discrete time indexes. Disregarding an additive background noise, we can express the observed signals in which multiple source signals are mixed linearly as

$$\boldsymbol{x}(t) = \sum_{n=0}^{N-1} \boldsymbol{a}(n)\boldsymbol{s}(t-n) = \boldsymbol{A}(z)\boldsymbol{s}(t), \tag{1}$$

where $\boldsymbol{s}(t) = [s_1(t), \cdots, s_L(t)]^{\mathrm{T}}$ is the source signal vector, and $\boldsymbol{x}(t) = [x_1(t), \cdots, x_K(t)]^{\mathrm{T}}$ is the observed signal vector. Also, $\boldsymbol{a}(n) = [a_{kl}(n)]_{kl}$ is the mixing filter matrix with the length of N, and $\boldsymbol{A}(z) = [A_{kl}(z)]_{kl} = [\sum_{n=0}^{N-1} a_{kl}(n)z^{-n}]_{kl}$ is the z-transform of $\boldsymbol{a}(n)$, where z^{-1} is used as the unit-delay operator, i.e., $z^{-n} \cdot x(t) = x(t-n)$, $a_{kl}(n)$ is the impulse response between the k-th microphone and the l-th sound source, and $[X]_{ij}$ denotes the matrix which includes the element X in the i-th row and the j-th column. In a common acoustical sound mixing, $a_{kl}(n)$ is not a single delayed impulse but a *decayed impulse train* which corresponds to the multi-path sound propagation caused by the room reflection and reverberation. Hereafter we only deal with the case of $K = L$ in this paper.

3. PROPOSED TWO-STAGE BSS ALGORITHM

3.1 Motivation and Strategy

In the previous research, SIMO-ICA has been proposed by Takatani et al. [6], and they showed that SIMO-ICA can separate the mixed signals into SIMO-model-based signals at the microphone points. This finding has motivated us to combine the unsupervised adaptive filtering (SIMO-ICA) and the multichannel supervised adaptive filtering (ABF). That is, the following post-processing can be applied after SIMO-ICA: (a) the DOA estimation of each sound source, and (b) speech-break detection for the target signal. The above-mentioned (a) and (b) can provide sufficient information for conducting the supervised adaptive filter learning in ABF. The ABF which follows SIMO-ICA can remove the residual component of the interference effectively. It is worth mentioning that the proposed algorithm is still *blind* although the supervised filtering is included in the second stage because the supervision for ABF is given by

SIMO-ICA automatically. The detailed process using the proposed algorithm is as follows.

3.2 First Stage: SIMO-ICA for Source Separation

In this stage, SIMO-ICA [6] is conducted for extracting the SIMO-model-based signals corresponding to each of sources. A brief explanation of the SIMO-ICA is given in the following. The SIMO-ICA consists of $(L-1)$ TDICA parts and a *fidelity controller*, and each ICA runs in parallel under the fidelity control of the entire separation system (see Fig. 1). The separated signals of the l-th ICA $(l = 1, \cdots L-1)$ in SIMO-ICA are defined by

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = [y_k^{(\mathrm{ICA}l)}(t)]_{k1} = \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(n)\boldsymbol{x}(t-n), \qquad (2)$$

where $\boldsymbol{w}_{(\mathrm{ICA}l)}(n) = [w_{ij}^{(\mathrm{ICA}l)}(n)]_{ij}$ is the separation filter matrix in the l-th ICA, and D is the length of the filter.

Regarding the fidelity controller, we calculate the following signal vector $\boldsymbol{y}_{(\mathrm{ICA}L)}(t)$, in which the all elements are to be mutually independent,

$$\boldsymbol{y}_{(\mathrm{ICA}L)}(t) \quad = \quad \boldsymbol{x}(t-D/2) - \sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}(t). \qquad (3)$$

Hereafter, we regard $\boldsymbol{y}_{(\mathrm{ICA}L)}(t)$ as an output of a *virtual* "L-th" ICA. The reason we use the word "*virtual* " here is that the L-th ICA does not have own separation filters unlike the other ICAs, and $\boldsymbol{y}_{(\mathrm{ICA}L)}(t)$ is subject to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ $(l = 1, \cdots, L-1)$. By transposing the second term $(-\sum_{l=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}l)}(t))$ in the right-hand side into the left-hand side, we can show that (3) means a constraint to force the sum of all ICAs' output vectors $\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t)$ to be the sum of all SIMO components $[\sum_{l=1}^{L} A_{kl}(z)s_l(t-D/2)]_{k1} (= \boldsymbol{x}(t-D/2))$.

If the independent sound sources are separated by (2), and simultaneously the signals obtained by (3) are also mutually independent, then the output signals converge on unique solutions, up to the permutation, as

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \mathrm{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l \boldsymbol{s}(t-D/2), \qquad (4)$$

where $\boldsymbol{P}_l$ $(l = 1, \cdots, L)$ are exclusively-selected permutation matrices which satisfy $\sum_{l=1}^{L} \boldsymbol{P}_l = [1]_{ij}$. Regarding a proof of this, see [6]. Obviously the solutions given by (4) provide necessary and sufficient SIMO components, $A_{kl}(z)s_l(t-D/2)$, for each l-th source. Thus, the separated signals of SIMO-ICA can maintain the spatial qualities of each sound source.

In order to obtain (4), the natural gradient of Kullback-Leibler divergence of (3) with respect to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ should be added to the existing TDICA-based iterative learning rule [3] of the separation filter in the l-th ICA $(l = 1, \cdots, L-1)$. The new iterative algorithm of the l-th ICA part $(l = 1, \cdots, L-1)$ in SIMO-ICA is given as

$$\begin{aligned}
&\boldsymbol{w}_{(\mathrm{ICA}l)}^{[j+1]}(n) \\
&= \boldsymbol{w}_{(\mathrm{ICA}l)}^{[j]}(n) - \alpha \sum_{d=0}^{D-1} \Bigg[\text{off-diag}\left\{ \left\langle \varphi\left(\boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t)\right) \right.\right. \\
&\quad \left.\left. \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}_{(\mathrm{ICA}l)}^{[j]}(d) \\
&\quad - \text{off-diag}\left\{ \left\langle \varphi\left(\boldsymbol{x}(t-\frac{D}{2}) - \sum_{k=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}k)}^{[j]}(t)\right) \right.\right. \\
&\quad \left.\left. \left(\boldsymbol{x}(t-n+d-\frac{D}{2}) - \sum_{k=1}^{L-1} \boldsymbol{y}_{(\mathrm{ICA}k)}^{[j]}(t-n+d)\right)^{\mathrm{T}} \right\rangle_t \right\} \\
&\quad \left(\boldsymbol{I}\delta(d-\frac{D}{2}) - \sum_{k=1}^{L-1} \boldsymbol{w}_{(\mathrm{ICA}k)}^{[j]}(d)\right) \Bigg], \qquad (5)
\end{aligned}$$

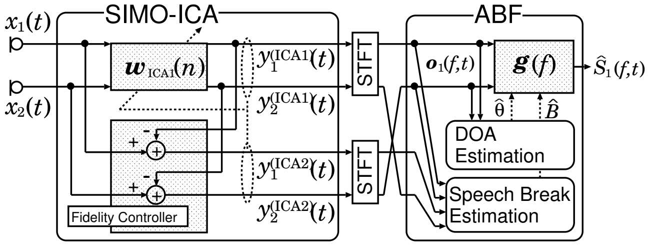

Figure 1: Proposed two-stage BSS, where $K = L = 2$.

where $\langle\cdot\rangle_t$ denotes the time-averaging operator, α is the step-size parameter, and $\delta(n)$ is a delta function, i.e., $\delta(0) = 1$ and $\delta(n) = 0$ $(n \neq 0)$. Also, the initial values of $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ for all l should be different.

3.3 Second Stage: Supervised Beamforming Using SIMO-Model-Based Signals

3.3.1 DOA Estimation and Speech-Break Detection:

In this paper we deal with the case of $K = L = 2$ and $\boldsymbol{P}_1 = \boldsymbol{I}$ in (4), but note that it can be easily extended to more general case of $K = L > 2$. Both the DOA estimation and speech-break detection for target signal are indispensable pre-processing to perform the supervised adaptive beamforming. We propose a specific DOA estimation method which utilizes the SIMO-model-based signals obtained by SIMO-ICA. The output signals of SIMO-ICA contain spatial information of each source, i.e., we can use the SIMO-model-based signals like dealing with multichannel signals observed at the microphone array. This can be still possible even if the SIMO-model-based separation is not completely but partly achieved to some extent; e.g., indeed the SIMO-ICA could provide the SNR improvement of more than 10 dB in our previous study. In this proposed system, the DOA of the l-th source is simply estimated on the basis of the phase difference among array signals in each frequency band, and is given by

$$\hat{\theta}_1(f) \quad = \quad \left\langle \sin^{-1}\left[\frac{\arg[Y_1^{(\mathrm{ICA1})}(f,t)/Y_2^{(\mathrm{ICA2})}(f,t)]}{2\pi f|d_1-d_2|c^{-1}}\right] \right\rangle_t, \qquad (6)$$

$$\hat{\theta}_2(f) \quad = \quad \left\langle \sin^{-1}\left[\frac{\arg[Y_1^{(\mathrm{ICA2})}(f,t)/Y_2^{(\mathrm{ICA1})}(f,t)]}{2\pi f|d_1-d_2|c^{-1}}\right] \right\rangle_t, \qquad (7)$$

where $Y_k^{(\mathrm{ICA}l)}(f,t)$ is the time-frequency representation of $y_k^{(\mathrm{ICA}l)}(t)$, and c is the velocity of sound. The resultant (fullband) DOA is obtained by averaging (6) or (7) within the specific frequency range, e.g., f=1000$\sim$4000 Hz, and we designate them as $\hat{\theta}_1$ and $\hat{\theta}_2$.

Regarding the speech-break detection, the separated signals from SIMO-ICA can be also used in which we check the absence or presence of the target speech signal with an appropriate threshold in the waveform domain.

3.3.2 ABF for Reduction of Residual Interference:

ABF proposed by Frost [5] is applied to the separated SIMO-model-based signals after SIMO-ICA and the DOA estimation. First, consider an ABF for enhancing the first sound source $s_1(t)$, where we obtain the array output by adding the weighted SIMO-model-based signals at each element. The resultant output signal of the ABF is described in the time-frequency domain as

$$\hat{S}_1(f,t) \quad = \quad \boldsymbol{g}(f)\boldsymbol{o}(f,t), \qquad (8)$$

$$\boldsymbol{g}(f) \quad \equiv \quad \left[g_1(f),\, g_2(f)\right], \qquad (9)$$

$$\boldsymbol{o}(f,t) \quad \equiv \quad \left[Y_1^{(\mathrm{ICA1})}(f,t),\, Y_2^{(\mathrm{ICA2})}(f,t)\right]^{\mathrm{T}}, \qquad (10)$$

where $\hat{S}_1(f,t)$ is the array output signal, $\boldsymbol{o}(f,t)$ is the SIMO-model-based signal vector in regard to $s_1(t)$, and $\boldsymbol{g}(f)$ is the weight vector of the array elements.

In the adaptive procedure, when the target signal is absent, the weight vector $\boldsymbol{g}(f)$ is optimized so as to minimize the array output power of interference arriving from outside of the look direction $\hat{\theta}_1$. This can be achieved by solving the following constrained minimization problem:

$$\min_{\boldsymbol{g}(f)} \boldsymbol{g}(f)\boldsymbol{R}(f)\boldsymbol{g}(f)^{\mathrm{H}}, \text{ subject to } \boldsymbol{g}(f)\boldsymbol{a}_{\hat{\theta}_1}(f) = 1, \tag{11}$$

$$\boldsymbol{R}(f) \equiv \left\langle \boldsymbol{o}(f,t)\boldsymbol{o}(f,t)^{\mathrm{H}} \right\rangle_{t\in\hat{B}}, \tag{12}$$

$$\boldsymbol{a}_{\hat{\theta}_1}(f) \equiv \begin{bmatrix} \exp(j2\pi f d_1 \sin(\hat{\theta}_1)/c) \\ \exp(j2\pi f d_2 \sin(\hat{\theta}_1)/c) \end{bmatrix}, \tag{13}$$

where $\boldsymbol{R}(f)$ is the array correlation matrix, $\boldsymbol{g}(f)\boldsymbol{R}(f)\boldsymbol{g}(f)^{\mathrm{H}}$ is equal to the array output power $\langle|\hat{S}_1(f,t)|^2\rangle_t$, and the superscript $^{\mathrm{H}}$ denotes the Hermitian transposition. $\hat{B}$ is the set of the frame numbers of speech-break, Also, $\boldsymbol{a}_{\hat{\theta}_1}(f)$ is generally called the steering vector.

The solution of the constrained minimization problem given by (11) yields the optimal weight vector

$$\boldsymbol{g}^{(\mathrm{opt})}(f) = \frac{\boldsymbol{a}_{\hat{\theta}_1}(f)^{\mathrm{H}}\boldsymbol{R}(f)^{-1}}{\boldsymbol{a}_{\hat{\theta}_1}(f)^{\mathrm{H}}\boldsymbol{R}(f)^{-1}\boldsymbol{a}_{\hat{\theta}_1}(f)}. \tag{14}$$

Using the ABF technique, we can realize the optimal directivity patterns for each interference, and the residual component of the interference can be reduced efficiently. For the second source, we can reduce the residual component of the interference which stays in $Y_1^{(\mathrm{ICA2})}(f,t)$ and $Y_2^{(\mathrm{ICA1})}(f,t)$ in the same manner.

3.4 Discussion on Separability in SIMO-ICA and Conventional SIMO-Output-Type Methods [1, 8, 9]

Note that there exists some alternative popular methods for obtaining the SIMO components in which the separated signals are projected back on to the microphones. Hereafter we simply abbreviate this class of methods to "PB".

The first example of PB is a method which utilizes the inverse of $\boldsymbol{W}(z)$ (see, e.g., [1]). In this PB method, the following operation is performed:

$$\begin{aligned} y_k^{(l)}(t) &= \{\boldsymbol{W}(z)^{-1}[\overbrace{0,\cdots,0}^{l-1},y_l(t),\overbrace{0,\cdots,0}^{L-l}]^{\mathrm{T}}\}_k \\ &= (\det \boldsymbol{W}(z))^{-1}\Delta_{lk}\cdot y_l(t), \end{aligned} \tag{15}$$

where $y_l(t)$ is a separated *monaural* signal obtained in the ICA, $y_k^{(l)}(t)$ represents the l-th resultant SIMO signal which is projected back onto the k-th microphone, $\{\cdot\}_k$ denotes the k-th element of the argument, and Δ_{kl} is a cofactor of the matrix $\boldsymbol{W}(z)$.

The second example of PB is a "deflation-type method" (see, e.g., [8, 9]). In this PB method, we extract a specific monaural source signal $y_l(t)$, and then $y_l(t)$ is projected back onto the k-th microphone as follows:

$$y_k^{(l)}(t) = \sum_n \langle x_k(t)y_l(t)z^{-n}\rangle_t / \langle|y_l(t)|^2\rangle_t \cdot y_l(t). \tag{16}$$

These PB methods are simpler than SIMO-ICA. However, *the separability among the target signal and the residual interference is lost* because the projection operator, $(\det \boldsymbol{W}(z))^{-1}\Delta_{lk}$ or $\sum_n \langle x_k(t)y_l(t)z^{-n}\rangle_t/\langle|y_l(t)|^2\rangle_t$, is applied to not only the target signal component but also the interference component in $y_l(t)$, as shown in (15) and (16). In other words, spatial information in the target signal is just similar to that in the interference, and this fact yields the negative result that the PB is *not* available for combination of SIMO-model-based signals and adaptive beamforming. In contrast to PB, SIMO-ICA holds the separability because the separation filter of SIMO-ICA cannot be always represented in the PB-form. This will be explicitly shown in Section 4.

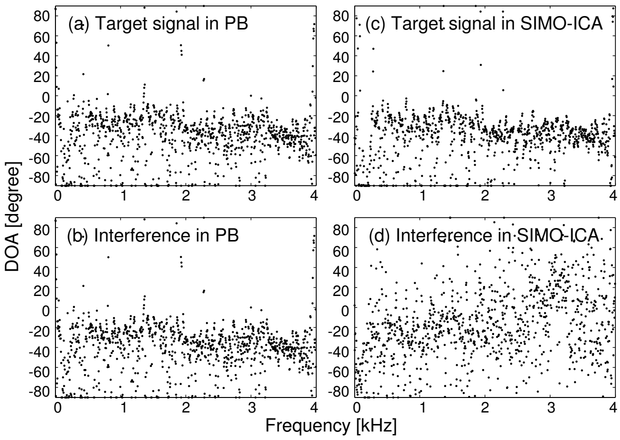

Figure 2: Examples of DOA estimation for target signal $s_1(t)$ (θ_1=$-30°$) and residual interference in PB or SIMO-ICA.

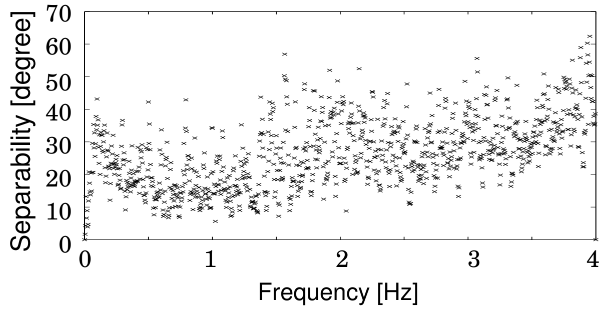

Figure 3: Separability of SIMO-ICA (average of different 12 speaker combinations).

4. EXPERIMENT UNDER REVERBERANT CONDITION

4.1 Conditions for Experiment

A two-element array with an interelement spacing of 4 cm is assumed. The speech signals are assumed to arrive from two directions, $-30°$ and $40°$. The distance between the microphone array and the loudspeakers is 1.15 m. Two kinds of sentences spoken by two male and two female speakers are used as the source speech samples. Using these sentences, we obtain 12 combinations. The sampling frequency is 8 kHz and the length of speech is limited to 3.6 seconds. To simulate the convolutive mixtures, the source signals are convolved with the impulse responses recorded in the real room which has a reverberation time of 300 ms. The length of the separation filter is set to 2048. The initial value in SIMO-ICA is generated by frequency-domain ICA [4, 7]. *Noise reduction rate* (NRR) [4], defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB, is used as the objective indication of separation performance. The SNRs are calculated under the assumption that the speech signal of the undesired speaker is regarded as noise.

4.2 Results of Separability

To explicitly visualize the separability, we depict the example of DOA-estimation results for the target signal component ($s_1(t)$; θ_1=$-30°$) and the residual interference in PB ($y_1^{(1)}(t)$ and $y_2^{(1)}(t)$) or SIMO-ICA ($y_1^{(\mathrm{ICA1})}(t)$ and $y_2^{(\mathrm{ICA2})}(t)$). Figures 2 (a) and (b) illustrate the DOAs in the conventional PB [1]. From the results, we can confirm that DOA information in (a) and (b) are the same, and consequently there is no separability in PB. On the other hand, Figs. 2 (c) and (d) show the DOA-estimation results in the SIMO-ICA of the proposed method. The DOAs of the interference, (d), are distinctly scattered from those of the target signal, (c), i.e., SIMO-

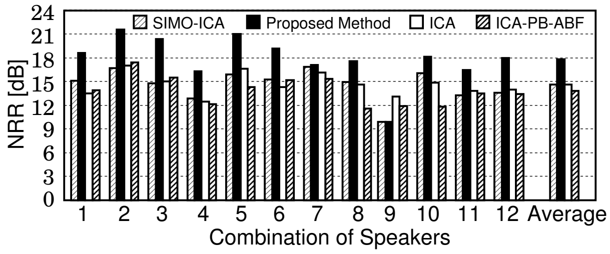

Figure 4: Result of NRR for different speaker combinations.

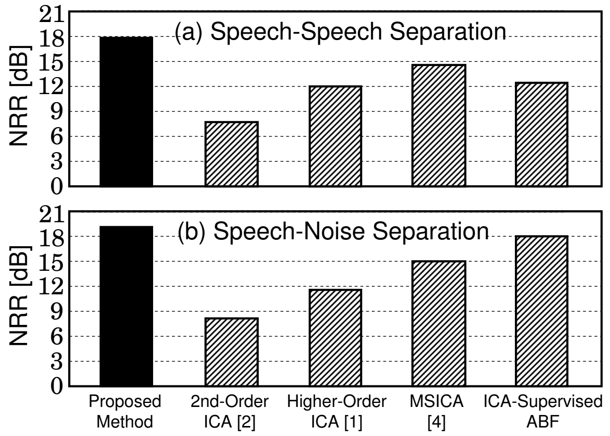

Figure 5: NRR comparison with conventional methods: (a) separation of two speech, and (b) separation of speech and stationary noise.

model-based signals outputted by SIMO-ICA have the possibility to be used as the inputs for ABF. Figure 3 shows the difference between DOA of the target signal and that of the interference signal in SIMO-ICA, which is the average of the 12 speaker combinations. The result indicates that the output signals of SIMO-ICA are almost separable in all of frequency bands. In contrast to SIMO-ICA, the difference of the DOA in PB was zero degree in all frequency bands, i.e, there is no separability.

4.3 Results of Separation Experiments for Two Speech

At first, the proposed two-stage BSS ("**Proposed Method**") is compared with the combination method of the conventional PB [1] and ABF ("**ICA-PB-ABF**"). As the conventional simple ICA, we select Multistage ICA ("**ICA**") proposed by Nishikawa [4]. Figure 4 shows the result of NRR for different speaker combinations. From the results, we can confirm that the proposed ABF which follows SIMO-ICA can remarkably and consistently improve the separation performance. This fact is a promising evidence on the feasibility of the proposed combination technique of SIMO-ICA and ABF. On the contrary, the ABF which follows conventional PB [1] could not contribute to the improvement of NRR. These results are well consistent with the discussion on the separability provided in Section 3.4. From the adobe-mentioned discussion about separability and the result, we can conclude as follows. **(a)** As far as we know, all of the existing ICA methods for obtaining the SIMO components are based on the PB operation, except SIMO-ICA. Thus, any combinations of the conventional ICA and ABF are *not* valid for improvement of the separation performance. **(b)** Only the specific combination of SIMO-ICA and ABF is valid owing to the separability between the target component and the residual component of the interference.

Secondly, we also compare the separation performance of the proposed method with those of many kinds of conventional BSS methods. As the conventional methods based on ICA, we prepared the second-order-based ICA proposed by Parra ("**2nd-Order ICA**") [2], Infomax-type higher-order-based frequency-domain ICA proposed by Murata ("**Higher-Order ICA**") [1], and Nishikawa's Multistage ICA ("**MSICA**") [4]. The other is an ABF ("**ICA-Supervised ABF**"). In General, the conventional ABF is directly applied to the observed signal, i.e., the input signal of ABF $\boldsymbol{o}(f,t)$ in Section 3.3.2 is replaced with $\boldsymbol{X}(f,t)$ which is the time-frequency representation of $\boldsymbol{x}(t)$. The conventional ABF requires two kind of supervisions; DOA and the speech-break-segments of the target signal. We carried out the experiments for ABF with the supervisions which are estimated by SIMO-ICA. Figure 5(a) shows the average of NRRs for 12 speaker combinations. From the result, we can confirm that the proposed method overtakes all of methods in separation performance.

4.4 Results of Separation Experiments for Speech and Noise

In Figure 5(b), we add the NRR results obtained in the separation of speech and a stationary human-like-colored noise. As can be seen, the proposed method still outperforms all of ICA-based methods although the performances of the proposed method and the ICA-Supervised ABF are comparable (the proposed method is slightly better). These facts from Figs. 5(a) and (b) imply that the proposed method is beneficial to both applications of speech-speech and speech-noise separations.

5. CONCLUSION

We proposed a new BSS framework in which the SIMO-model-based ICA and the multichannel supervised adaptive filtering (ABF) are efficiently combined. In order to evaluate its effectiveness, a separation experiment was carried out under a reverberant condition. The experimental results revealed that the NRR can be considerably improved by using the proposed two-stage BSS algorithm. In addition, we could find the fact that the proposed method outperforms the combination of the conventional SIMO-output-type ICA and ABF as well as the simple SIMO-ICA.

6. ACKNOWLEDGEMENT

This work was partly supported by CREST "Advanced Media Technology for Everyday Living" of JST in Japan.

REFERENCES

[1] N. Murata, and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. NOLTA98* Vol.3, pp.923–926, 1998.

[2] L. Parra, and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech & Audio Processing* Vol.8, pp.320–327, 2000.

[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications,* John Wiley & sons, Ltd, West Sussex, 2002.

[4] T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA," *IEICE Trans. Fundamentals,* Vol.E86-A, No.4, pp.846–858, 2003.

[5] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE* Vol.60, pp.926–935, 1972.

[6] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fidelity blind separation of acoustic signals using SIMO-model-based ICA with information-geometric learning," *Proc. IWAENC2003*, pp.251–254, 2003.

[7] S. Ukai, H. Saruwatari, T. Takatani, R. Mukai, and H. Sawada, "Multistage SIMO-model-based blind source separation combining frequency-domain ICA and time-domain ICA," *Proc. ICASSP2004*, 2004.

[8] J. Tugnait, "Identification and deconvolution of multichannel linear non-gaussian processes using higher order statistics and inverse filter criteria," *IEEE Trans. on Signal Processing* Vol.45, pp.658–672, 1997.

[9] C. Simon, P. Loubaton, and C. Jutten, "Separation of a class of convolutive mixtures: a contrast function approach," *Signal Processing* Vol.81, pp.883–888, 2001.