BLIND SEPARATION OF BINAURAL SOUND MIXTURES USING SIMO-ICA WITH SELF-GENERATOR FOR INITIAL FILTER

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

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192, JAPAN
phone: +81-743-72-5287, fax: +81-743-73-5289, email: tomoya-t@is.naist.jp
http://www.aist-nara.ac.jp/~tomoya-t

ABSTRACT

In this paper, we address the blind separation problem of binaural mixed signals, and propose a novel blind separation method using Single-Input-Multiple-Output-model-based independent component analysis (SIMO-ICA) with a self-generator (SG) for the initial filter. SIMO-ICA which has been proposed by the authors can separate mixed signals, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Although this attractive feature of SIMO-ICA is beneficial to the binaural sound separation, SIMO-ICA has a serious drawback in its high sensitivity to the initial settings of the separation filter. In the proposed method, the SG functions as the preprocessor of SIMO-ICA, and it can provide a valid initial filter for SIMO-ICA. To evaluate its effectiveness, binaural sound separation experiments are carried out under a reverberant condition. The experimental results reveal that the separation performance of the proposed method is superior to those of conventional methods.

1. INTRODUCTION

Blind source separation (BSS) is the approach taken to estimate original source signals using only information of the mixed signals observed in each input channel. This technique is applicable to high-quality hands-free telecommunication systems. In recent works of BSS based on ICA [1], various methods have been proposed to deal with a means of separation of acoustic sounds which corresponds to the convolutive mixture case [2, 3]. However, the conventional ICA-based BSS approaches are basically means of extracting each of the independent sound sources as a *monaural* signal, and consequently they have a serious drawback in that the separated sounds cannot maintain information about the directivity, localization, or spatial qualities of each sound source. This prevents any BSS methods from being applied to binaural signal processing [4] or high-fidelity sound reproduction systems [5].

In order to solve the above-mentioned fundamental problems, several high-fidelity BSSs using the ICA-based algorithm have been proposed, in which the convolutive mixtures of acoustic signals are decomposed into the Single-Input Multiple-Output (SIMO) components. 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. Murata et al. have proposed FDICA-PB [6]. In this algorithm, first, the source signals are estimated as a monaural signal by FDICA, and then projection-back processing projects the source signals estimated by FDICA onto the observed signal's space using the inverse filter of the separation filter in FDICA. However, this algorithm has some disadvantages [7] as follows. First, the invertibility of the separation filter cannot be guaranteed. Thus, the inversion of the separation filter often fails and yields harmful results. Secondly, the circular convolution effect in FDICA is likely to cause the deterioration of separation performance. To solve these problems, the authors have proposed SIMO-model-based ICA (SIMO-ICA) [8]. The SIMO-ICA consists of multiple time-domain ICA (TDICA) parts and a fidelity controller. Since SIMO-ICA estimates SIMO components of the observed signals directly, inversion problem does not arise. Also, since SIMO-ICA is constructed of TDICA, it is free from the circular convolution problem. However, the convergence of SIMO-ICA is very slow, and the sensitivity to the initial settings of the separation filter is very high.

In this paper, in order to improve the decomposition performance, we propose a novel multi-step learning algorithm combining FDICA-PB, DOA estimation, and SIMO-ICA. First, we perform FDICA-PB to decompose the observed signals to some extent. After the FDICA-PB, we estimate the DOAs of sources using outputs of FDICA-PB. Then the proposed method resets the separation filter to the valid initial value and re-optimizes the filter using both FDICA-PB and SIMO-ICA. In this procedure, a filter bank of previously measured head related transfer functions (HRTFs) for multiple DOAs is supplied to generate the valid initial filter. To evaluate its effectiveness, decomposition experiments are carried out under a reverberant condition. The experimental results reveal that the decomposition performance of the proposed method is superior to those of the conventional methods.

2. MIXING PROCESS

In this study, the number of microphones is $K = 2$ and the number of multiple sound sources is $L = 2$ (see Fig. 1). In general, the observed signals in which multiple source signals are mixed linearly are expressed as

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

$\boldsymbol{s}(t) = [s_1(t), s_2(t)]^{\mathrm{T}}$ is the source signal vector and $\boldsymbol{x}(t) = [x_1(t), x_2(t)]^{\mathrm{T}}$ 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} 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.

3. CONVENTIONAL DECOMPOSING ALGORITHMS: SIMO-MODEL-BASED ICA (SIMO-ICA) [8]

For extracting the SIMO components in the mixed signals, we have proposed blind decomposing framework for SIMO-model-based acoustic signals using the extended TDICA algorithm, SIMO-ICA. SIMO-ICA consists of the TDICA part and a *fidelity controller*, and the TDICA runs in parallel under the fidelity control of the entire separation system (see Fig. 1). The output signals of the TDICA part in SIMO-ICA are defined by

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

where $\boldsymbol{w}_{(\mathrm{TD})}(n)$ is the separation filter matrix of the TDICA. Regarding the fidelity controller, the following signal vector is calculated, in which all of the elements are to be mutually independent,

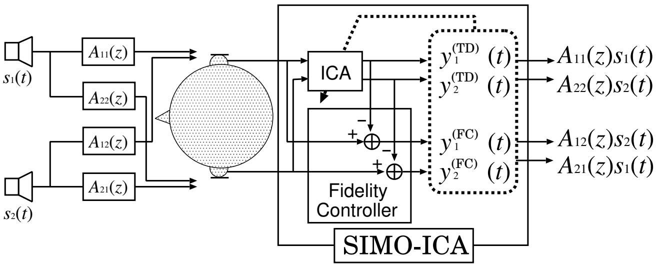

Figure 1: Example of input and output relations in the conventional SIMO-ICA under binaural recording.

$$\boldsymbol{y}_{(\mathrm{FC})}(t) = [y_k^{(\mathrm{FC})}(t)]_{k1} = \boldsymbol{x}(t-\frac{D}{2}) - \boldsymbol{y}_{(\mathrm{TD})}(t). \tag{3}$$

Hereafter, we regard $\boldsymbol{y}_{(\mathrm{FC})}(t)$ as an output of a *virtual* ICA, and define its virtual separation filter matrix as

$$\boldsymbol{w}_{(\mathrm{FC})}(n) = \boldsymbol{I}\delta(n-\frac{D}{2}) - \boldsymbol{w}_{(\mathrm{TD})}(n), \tag{4}$$

where $\delta(n)$ is a delta function, where $\delta(0)=1$ and $\delta(n)=0$ $(n \neq 0)$. From (4), we can rewrite (3) as

$$\boldsymbol{y}_{(\mathrm{FC})}(t) = \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{FC})}(n)\boldsymbol{x}(t-n). \tag{5}$$

The reason why we use the word "virtual" here is that fidelity controller does not have own separation filters unlike the TDICA, and $\boldsymbol{w}_{(\mathrm{FC})}(n)$ is subject to $\boldsymbol{w}_{(\mathrm{TD})}(n)$. To explicitly show the meaning of the fidelity controller, (3) is rewritten as

$$\boldsymbol{y}_{(\mathrm{TD})}(t) + \boldsymbol{y}_{(\mathrm{FC})}(t) - \boldsymbol{x}(t-D/2) = [0]_{k1}. \tag{6}$$

Equation (6) means a constraint to force the sum of the all of output vectors $\boldsymbol{y}_{(\mathrm{TD})}(t) + \boldsymbol{y}_{(\mathrm{FC})}(t)$ to be the sum of all of the SIMO components $[\sum_{l=1}^{L} A_{kl}(z)s_l(t-D/2)]_{k1} (= \boldsymbol{x}(t-D/2))$. Here the delay of $D/2$ is used as to deal with nonminimum phase systems.

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,

$$\boldsymbol{y}_{(\mathrm{TD})}(t) = [\ A_{11}(z)s_1(t-D/2), \quad A_{22}(z)s_2(t-D/2)\ ]^{\mathrm{T}}, \tag{7}$$

$$\boldsymbol{y}_{(\mathrm{FC})}(t) = [\ A_{12}(z)s_2(t-D/2), \quad A_{21}(z)s_1(t-D/2)\ ]^{\mathrm{T}}, \tag{8}$$

where diag$\{\boldsymbol{X}\}$ and off-diag$\{\boldsymbol{X}\}$ are the operation for setting every nondiagonal and diagonal elements of the matrix $\boldsymbol{X}$ to be zero. The proof of theorem and more details are given in [8]. Equations (7) and (8) represent necessary and sufficient SIMO components of all source signals.

In order to obtain the above-mentioned solutions, the natural gradient [3] of Kullback-Leibler divergence of (3) with respect to $\boldsymbol{w}_{(\mathrm{TD})}(n)$ should be added to the iterative learning rule of the separation filter in the TDICA. The iterative algorithm of the TDICA part in SIMO-ICA is given as

$$\begin{aligned}\boldsymbol{w}_{(\mathrm{TD})}^{[j+1]}(n) = \ & \boldsymbol{w}_{(\mathrm{TD})}^{[j]}(n) - \alpha \sum_{d=0}^{D-1} \Bigg[\text{off-diag} \left\{ \left\langle \varphi\left(\boldsymbol{y}_{(\mathrm{TD})}^{[j]}(t)\right) \boldsymbol{y}_{(\mathrm{TD})}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \boldsymbol{w}_{(\mathrm{TD})}^{[j]}(d) \\ & - \text{off-diag} \left\{ \left\langle \varphi\left(\boldsymbol{y}_{(\mathrm{FC})}^{[j]}(t)\right) \boldsymbol{y}_{(\mathrm{FC})}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \right\} \left(\boldsymbol{I}\delta(d-\frac{D}{2}) - \boldsymbol{w}_{(\mathrm{TD})}^{[j]}(d) \right) \Bigg],\end{aligned} \tag{9}$$

where α is the step-size parameter, the superscript $[j]$ is used to express the value of the j-th step in the iterations, and $\langle\cdot\rangle_t$ denotes the time-averaging operator. In (9), the initial values of $\boldsymbol{w}_{(\mathrm{TD})}(n)$ and $\boldsymbol{w}_{(\mathrm{FC})}(n)$ are arbitrary, but should be different each other.

4. PROPOSED ALGORITHM

4.1 Motivation and Strategy

The SIMO-ICA algorithm has the drawbacks of the arbitrariness with respect to the initial value of the separation filter, and the decomposing performances of this conventional method is deteriorated by the irrelevant initial value. Also, there is no theoretical strategy for the selection of the *valid* initial value in the conventional ICA framework. Thus, the development of the self-generation of good initial filters is a problem demanding attention.

Meanwhile, the binaural transfer function is roughly divided into the room transfer function and HRTF. Since the former depends on the components of the direct sound, the reflection sound, and reverberation, it is generally unknown in the blind setup. However, the latter depends on the only DOAs of sources and can be previously known because HRTF is an inherent feature in the recording apparatus itself and can be approximately measured by using, e.g., a dummy head. Thus, HRTF is an important factor which solves the blind decomposition problem of binaural mixed signals, and we can use HRTF as the valid initial value of the separation filter matrix if we can previously know the DOAs of sources.

These facts motivated us to combine FDICA-PB, DOA estimation, and SIMO-ICA. First, we perform the FDICA-PB to decompose the observed signals to some extent. Secondly, we estimate the DOA of sources using the output of FDICA-PB *blindly*. Then, the algorithm resets the separation filter to the valid initial value and re-optimize the separation filter using both FDICA-PB and SIMO-ICA. In this procedure, a filter bank of previously measured HRTFs for multiple DOAs is supplied to generate the valid initial filter. The important technology required here is an accurate DOA estimation of sources. The SIMO-output algorithm has one advantage that the output signals maintain the spatial qualities of each source. Thus, the output signals of FDICA-PB, which is one of the SIMO-output BSS, are synchronized with the observed signals. i.e., the time alignment has been taken. We can detect the single talk segments in the observed signals by using SIMO-output signals of FDICA-PB, and estimate the DOAs of sources using the observed signals corresponds to these segments.

4.2 Proposed Algorithm

The proposed algorithm is conducted by the following steps.

[Step 0: Early Initialization] Set DOAs of sources $\hat{\theta}_i$ to early initial (arbitrary) values, $\hat{\theta}_{\mathrm{init}i}$.

[Step 1: HRTF Matrix Bank] The HRTF matrix bank consists of multiple HRTF matrices. The single HRTF matrix for θ_1 and θ_2 is given as

$$\boldsymbol{H}(\theta_1,\theta_2,f) = \begin{bmatrix} H_L(\theta_1,f) & H_L(\theta_2,f) \\ H_R(\theta_1,f) & H_R(\theta_2,f) \end{bmatrix}, \tag{10}$$

where $H_L(\theta,f)$ (or $H_R(\theta,f)$) is the HRTF between the left (right) ear and the source whose direction is θ. To construct the HRTF matrix bank, we prepare the multiple HRTF matrices in advance by changing θ_1 and θ_2. Using the HRTF matrix bank and the DOAs of sources, we can automatically generate the valid initial value for FDICA as follows:

$$\boldsymbol{W}_{(\mathrm{FD})}^{[0]}(f) = \boldsymbol{H}^{-1}(\hat{\theta}_1,\hat{\theta}_2,f). \tag{11}$$

Note that the initial value is not an optimal separation filter matrix under a reverberant condition, but the separation filter matrix can be finally optimized through ICA iterations.

[Step 2: FDICA-PB [6]] Murata et al. have proposed an FDICA-PB method which can estimate the SIMO components of the observed signals on the basis of the monaural outputs of FDICA. In

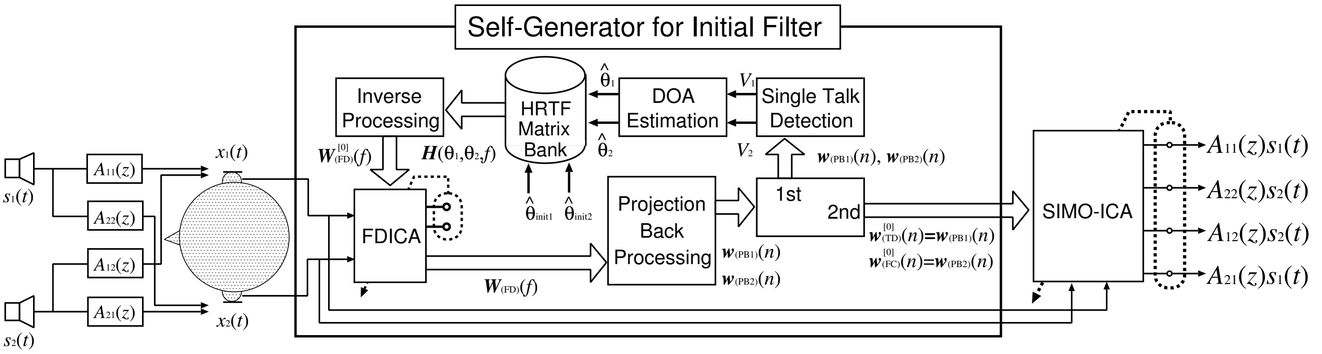

Figure 2: Example of input and output relations in the proposed method.

this method, first, the separation filter matrix $\boldsymbol{W}_{(\mathrm{FD})}(f)$ in the frequency domain is optimized to the separate source signals to obtain the monaural signals. The separated signals $\boldsymbol{Y}_{(\mathrm{FD})}(f,t)$ in the time-frequency domain are expressed as

$$\boldsymbol{Y}_{(\mathrm{FD})}(f,t) = \boldsymbol{W}_{(\mathrm{FD})}(f)\boldsymbol{X}(f,t), \tag{12}$$

where $\boldsymbol{X}(f,t)$ is the observed signal vector which is calculated by means of a frame-by-frame discrete Fourier transform (DFT). The iterative learning algorithm is expressed as

$$\boldsymbol{W}^{[i+1]}_{(\mathrm{FD})}(f) = \eta\left\{\boldsymbol{I} - \left\langle \Phi(\boldsymbol{Y}^{[i]}_{(\mathrm{FD})}(f,t))\boldsymbol{Y}^{[i]}_{(\mathrm{FD})}(f,t)^{\mathrm{H}}\right\rangle_t\right\} \boldsymbol{W}^{[i]}_{(\mathrm{FD})}(f) + \boldsymbol{W}^{[i]}_{(\mathrm{FD})}(f), \tag{13}$$

where the initial value of $\boldsymbol{W}_{(\mathrm{FD})}(f)$ is given by Eq. (11). However, the output signals of FDICA given by Eq. (12) are monaural signals with respect to the sound sources, not SIMO-model-based signals. Thus, using the following equations, we project the monaural separated signals onto the microphone signal's space.

$$\boldsymbol{w}_{(\mathrm{PB1})}(n) = \mathrm{IDFT}\left[\mathrm{diag}\left\{\boldsymbol{W}^{-1}_{(\mathrm{FD})}(f)\right\}\boldsymbol{W}_{(\mathrm{FD})}(f)\right], \tag{14}$$

$$\boldsymbol{w}_{(\mathrm{PB2})}(n) = \mathrm{IDFT}\left[\text{off-diag}\left\{\boldsymbol{W}^{-1}_{(\mathrm{FD})}(f)\right\}\boldsymbol{W}_{(\mathrm{FD})}(f)\right], \tag{15}$$

where $\mathrm{IDFT}[\cdot]$ represents an inverse DFT with the time shift of the $D/2$ samples. The separated signals of FDICA-PB in the time domain are expressed as

$$\boldsymbol{y}_{(\mathrm{PB}i)}(t) = \sum_{n=0}^{D-1}\boldsymbol{w}_{(\mathrm{PB}i)}(n)\boldsymbol{x}(t-n). \tag{16}$$

$\boldsymbol{y}_{(\mathrm{PB1})}(t)$ is a good approximation of the SIMO solution in Eq. (7) without permutation (also, $\boldsymbol{y}_{(\mathrm{PB2})}(t)$ corresponds to Eq. (8)).

[**Step 3: Single Talk Detection**] In order to detect the single talk segments of the observed signals, we divide the observed signals and output signals of FDICA-PB into multiple frames. Each frame of these signals is expressed as

$$\boldsymbol{x}(u,v) = \boldsymbol{x}(u+(v-1)\times U), \tag{17}$$

$$\boldsymbol{y}_{(\mathrm{PB}i)}(u,v) = \boldsymbol{y}_{(\mathrm{PB}i)}(u+(v-1)\times U), \tag{18}$$

where u is the time index in a frame, U is the number of samples in a frame, v is the frame index. Each single talk segment $\boldsymbol{V}_i$ of the observed signals is detected on the basis of the following criteria:

$$\boldsymbol{V}_1 = \Big\{v|Q_1^{(\mathrm{PB1})}(v) > T; Q_2^{(\mathrm{PB2})}(v) > T; Q_2^{(\mathrm{PB1})}(v) < T; Q_1^{(\mathrm{PB2})}(v) < T\Big\}, \tag{19}$$

$$\boldsymbol{V}_2 = \Big\{v|Q_1^{(\mathrm{PB1})}(v) < T; Q_2^{(\mathrm{PB2})}(v) < T; Q_2^{(\mathrm{PB1})}(v) > T; Q_1^{(\mathrm{PB2})}(v) > T\Big\}, \tag{20}$$

$$Q_k^{(\mathrm{PB}i)}(v) = 10\log_{10}\frac{\sum_{u=1}^{U}|y_k^{(\mathrm{PB}i)}(u,v)|^2}{\max_v\left\{\sum_{u=1}^{U}|y_k^{(\mathrm{PB}i)}(u,v)|^2\right\}}, \tag{21}$$

where T is a threshold which is experimentally determined.

[**Step 4: DOA Estimation Using Single Talk Segments**] We can obtain the DOAs $\hat{\theta}_i$ of sources by using the single talk segments. The estimated angle $\hat{\theta}_i$ is given as

$$\hat{\theta}_i = \arg\max_{\theta}\left\{\left\langle\sum_f X_1(f,v)X_2(f,v)^{\mathrm{H}}e^{-\frac{j2\pi f d\sin\theta}{c}}\right\rangle_{v\in V_i}\right\}, \tag{22}$$

where $\langle\cdot\rangle_{v\in V}$ is the frame-averaging operator which is composed of elements v in single talk segments V. Thus, we can obtain the valid initial value of the separation filter matrix using these estimated values, $\hat{\theta}_1$ and $\hat{\theta}_2$.

[**Step 5: Re-Optimization**] Using the DOAs of the sources estimated with Eq. (22), we reset the separation filter to the valid initial value, and re-optimize that in FDICA-PB (execute [**Steps 1 and 2**] again).

[**Step 6: SIMO-ICA**] Optimize the separation filter matrices $\boldsymbol{w}_{(\mathrm{TD})}(n)$ and $\boldsymbol{w}_{(\mathrm{FC})}(n)$ in the time domain, by using Eq. (9) to enhance the target components further. The separation filter matrices (14) and (15) are used as the initial values of the separation filter matrix $\boldsymbol{w}_{(\mathrm{TD})}(n)$ and $\boldsymbol{w}_{(\mathrm{FC})}(n)$ in SIMO-ICA.

If the early initialization, HRTF matrix bank, FDICA-PB, and SIMO-ICA (**[Step 0–2, 6]**) are executed without single talk detection, DOA estimation, nor re-optimization (**[Step 3–5]**), this algorithm corresponds to the multistage SIMO-ICA (MS-SIMO-ICA) algorithm [9] which has previously been proposed by one of the authors.

5. EXPERIMENTS AND RESULTS

5.1 Conditions for Experiments

We carried out binaural-sound-separation experiments using source signals which are convolved with impulse responses recorded with a head and torso simulator (HATS) (Brüel & Kjær) in the experimental room. The reverberation time in this room is 200 ms. Two speech signals are assumed to arrive from different directions, θ_1 and θ_2; $\theta_1 = \{-90^\circ, -75^\circ, -60^\circ, -45^\circ, -30^\circ, -15^\circ, 0^\circ\}$ and $\theta_2 = \{0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ, 90^\circ\}$. The distance between HATS and the sound source is 1.5 m. Two kinds of sentences, spoken by two male and two female speakers, are used as the original speech samples. Using these sentences, we obtain 12 combinations. The sampling frequency is 8 kHz and the length of speech is limited to 3 seconds. Regarding the conventional ICA for comparison, we use SIMO-ICA, FDICA-PB, and MS-SIMO-ICA. The length of $\boldsymbol{w}(n)$

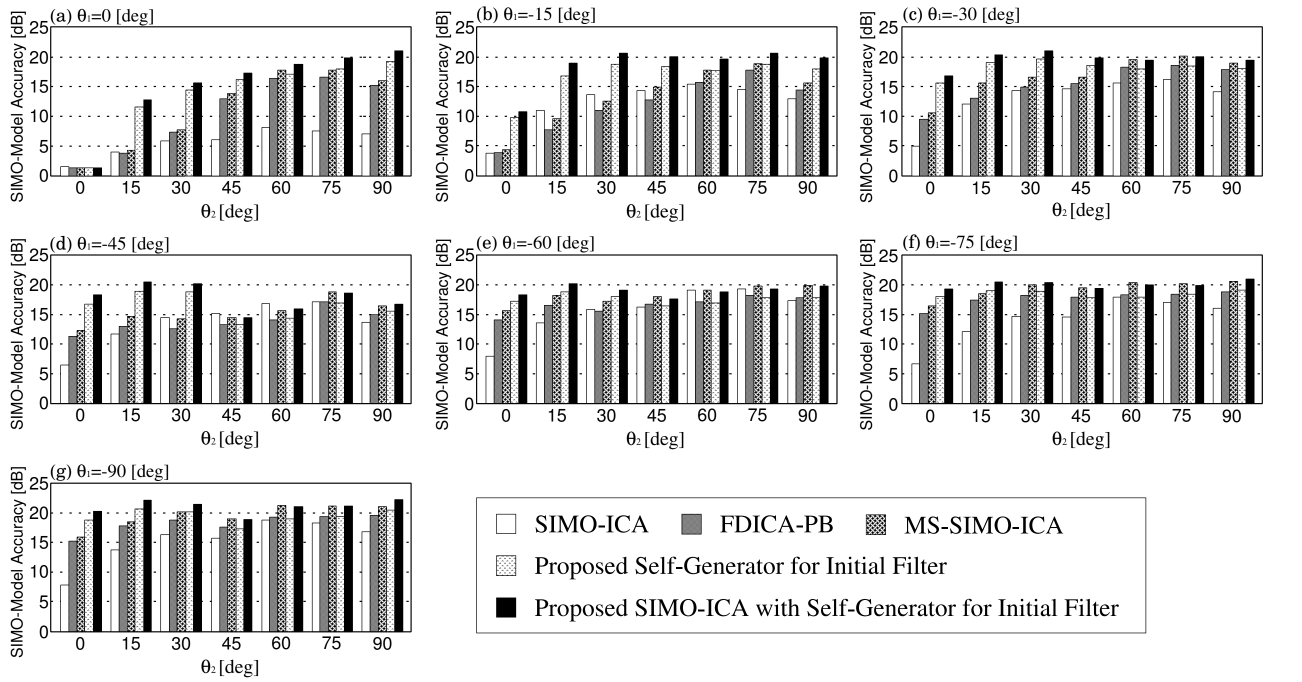

Figure 3: Experimental results of SIMO-ICA, FDICA-PB, MS-SIMO-ICA, proposed self-generator for the initial filter, and proposed SIMO-ICA with the self generator in reverberant condition.

in each method is 1024, and the initial values are inverse filters of HRTFs whose directions of sources, $\hat{\theta}_{\mathrm{init1}}$ and $\hat{\theta}_{\mathrm{init2}}$, are $-60°$ and $60°$. The step-size parameters α and η are 5.0×10^{-2} and 1.0×10^{-6}. SIMO-model accuracy (SA) [10] is used as an evaluation score. The SA indicates the degree of similarity between the outputs of SIMO-ICA and the real SIMO-model-based signals.

5.2 Results and Discussion

Figures 3 show the results of SA in the conventional SIMO-ICA, FDICA-PB, MS-SIMO-ICA and the proposed method. These are averaged scores of all speaker combinations. The following points are revealed.

- When θ_2 is smaller than 45°, decomposing performances in the proposed method are superior to those in the conventional methods regardless of θ_1, except for the trivial case of $\theta_1 = \theta_2 = 0$.
- The performance of the proposed method can be remarkably improved, especially when the angle between the speakers is narrow, e.g., $\theta_1 = 0° \sim -45°$ and $\theta_2 = 0° \sim 30°$. Also, the performances in the SG itself are superior to those of FDICA-PB. This means that the SG can contribute to the improvement of performances in the whole separation system.
- When θ_2 is larger than 45°, the decomposing performances of the proposed method are almost the same as those of the conventional methods.

Therefore, it can be asserted that the proposed algorithm for the self-generation of initial values works effectively and increases the SIMO-ICA's separation performance.

6. CONCLUSION

In order to improve the decomposition performance, we newly propose a method based on an alternation learning algorithm combining FDICA, Single-Input-Multiple-Output-mode-based independent component analysis (SIMO-ICA), and the direction of arrival (DOA) estimation. To evaluate its effectiveness, decomposition experiments are carried out under a reverberant condition. The experimental results reveal that the decomposition performance of the proposed method is superior to those of the conventional methods, especially when the angle between the sound sources is narrow.

7. ACKNOWLEDGMENT

This research is partly supported by CREST program "Advanced Media Technology for Everyday Living" of Japan Science and Technology Agency.

REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol.36, pp.287–314, 1994.

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol.22, pp.21–34, 1998.

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

[4] J. Blauert, *Spatial Hearing (revised edition)*, Cambridge, MA: The MIT Press, 1997.

[5] Y. Tatekura, H. Saruwatari, K. Shikano, "Sound reproduction system including adaptive compensation of temperature fluctuation effect for broad-band sound control," *IEICE Trans. Fundamentals*, vol.E85–A, no.8, pp.1851–1860, 2002.

[6] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. NOLTA '98*, vol. 3, pp.923–926, 1998.

[7] T. Nishikawa, H. Saruwatari, K. Shikano, "Stable leaning algorithm for blind separation of temporally correlated acoustic signals combining multistage ICA and linear prediction," *IEICE Trans. Fundamentals*, vol.E86-A, no.4, pp.846–858, 2003.

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

[9] S. Ukai, H. Saruwatari, T. Takatani, R. Mukai, H. Sawada, "Multistage SIMO-Model-Based Blind Source Separation Combining Frequency-Domain ICA and Time-Domain ICA," *Proc. ICASSP 2004*, pp.IV-109–IV-112, 2004.

[10] H. Yamajo, H. Saruwatari, T. Takatani, T. Nishikawa, and K. Shikano, "Evaluation of blind separation and deconvolution for convolutive speech mixture using SIMO-model-based ICA," *Proc. IWAENC 2003*, pp.299–302, 2003.