MULTISTAGE SIMO-MODEL-BASED BLIND SOURCE SEPARATION COMBINING FREQUENCY-DOMAIN ICA AND TIME-DOMAIN ICA

Satoshi Ukai, Hiroshi Saruwatari, Tomoya Takatani, Ryo Mukai†, and Hiroshi Sawada†

Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara, 630-0192, JAPAN (E-mail: sato-uk@is.aist-nara.ac.jp)
†NTT Communication Science Laboratories, 2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237, JAPAN

ABSTRACT

In this paper, single-input multiple-output (SIMO)-model-based blind source separation (BSS) is addressed, where unknown mixed source signals are detected at the microphones, and these signals can be separated, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. This technique is highly applicable to high-fidelity signal processing such as binaural signal processing. First, we provide an experimental comparison between two kinds of the SIMO-model-based BSS methods, namely, traditional frequency-domain ICA with projection-back processing (FDICA-PB), and SIMO-ICA recently proposed by the authors. Secondly, we propose a new combination technique of the FDICA-PB and SIMO-ICA, which can achieve a more higher separation performance in comparison to two methods. The experimental results reveal that the accuracy of the separated SIMO signals in the simple SIMO-ICA is inferior to that of FDICA-PB, but the proposed combination technique can outperform both simple FDICA-PB and SIMO-ICA.

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. Owing to the attractive features of BSS, much attention has been paid to the BSS technique in various fields of signal processing. In this paper, we mainly address the BSS problem lying on acoustic signal processing.

In recent works based on independent component analysis (ICA) [1], various methods have been proposed to deal with the BSS for acoustical sounds. However, the existing ICA-based BSS approaches are basically means of extracting each of the independent sound sources as a *monaural* signal. Accordingly they have a serious drawback 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 [2], or any high-fidelity acoustic signal processing.

In order to solve the problem, we should adopt a new blind separation framework in which Single-Input Multiple-Output (SIMO)-model-based BSS is considered. 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 sensors. In the SIMO-model-based separation scenario, unknown multiple source signals which are mixed through unknown acoustical transmission channels are detected at the microphones, and these signals can be separated, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Thus, the SIMO-model-based separated signals can maintain the spatial qualities of each sound source. Obviously the attractive feature is highly applicable to high-fidelity acoustic signal processing.

The first objective of this paper is to provide an experimental comparison between two kinds of the SIMO-model-based BSS methods as follows. (a) Traditional frequency-domain ICA (FDICA) with projection-back processing (hereafter we call it *FDICA-PB*) proposed by Murata and Ikeda [3]. (b) *SIMO-ICA* which consists of multiple time-domain ICAs (TDICAs) recently proposed by the authors [4]. The second objective of this paper is to propose a new combination technique of the FDICA-PB and SIMO-ICA, which can achieve a more higher separation performance with the low computational complexity in comparison to two methods. The experiments are carried out under a reverberant condition, and the results explicitly reveal that the advantages and disadvantages of the each method, and the superiority of the proposed combination technique over the FDICA-PB or SIMO-ICA.

2. MIXING PROCESS

In this study, the number of microphones is K and the number of multiple sound sources is L. The observed signals in which multiple source signals are mixed linearly are expressed as

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

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. Hereafter, we only deal with the case of $K = L$ in this paper.

3. SIMO-MODEL-BASED BSS 1: CONVENTIONAL FDICA-PB

In the conventional FDICA-PB, 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 $\boldsymbol{X}(f,t)$ $=[X_1(f,t), \cdots, X_K(f,t)]^{\mathrm{T}}$.

Next, we perform signal separation using the complex-valued unmixing matrix, $\boldsymbol{W}(f) = [W_{lk}(f)]_{lk}$, so that the L time-series

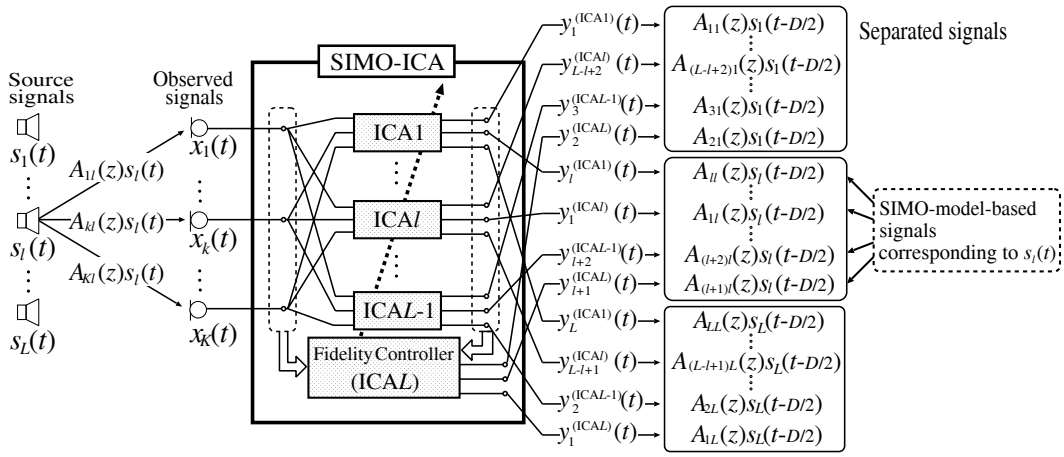

Figure 1: Example of input and output relations in proposed SIMO-ICA performed in first stage, where permutation $\boldsymbol{P}_l$ is given by (12).

output $\boldsymbol{Y}(f,t)$=$[Y_1(f,t),\cdots,Y_L(f,t)]^{\mathrm{T}}$ becomes mutually independent; this procedure can be given as

$$\boldsymbol{Y}(f,t) \quad = \quad \boldsymbol{W}(f)\boldsymbol{X}(f,t). \tag{2}$$

We perform this procedure with respect to all frequency bins. The optimal $\boldsymbol{W}(f)$ is obtained by, e.g., the following iterative updating:

$$\begin{aligned}\boldsymbol{W}^{[i+1]}(f) \quad &= \quad \eta\big[\boldsymbol{I} - \big\langle \boldsymbol{\Phi}(\boldsymbol{Y}(f,t))\boldsymbol{Y}^{\mathrm{H}}(f,t)\big\rangle_t\big]\boldsymbol{W}^{[i]}(f) \\ &\quad + \boldsymbol{W}^{[i]}(f),\end{aligned} \tag{3}$$

where $\langle\cdot\rangle_t$ denotes the time-averaging operator, $[i]$ is used to express the value of the i th step in the iterations, and η is the step-size parameter. In our research, we define the nonlinear vector function $\boldsymbol{\Phi}(\cdot)$ as [5]

$$\boldsymbol{\Phi}(\boldsymbol{Y}(f,t)) \quad \equiv \quad \big[e^{\jmath\cdot\arg(Y_1(f,t))},\cdots,e^{\jmath\cdot\arg(Y_L(f,t))}\big]^{\mathrm{T}}, \tag{4}$$

where $\arg[\cdot]$ represents an operation to take the argument of the complex value. After the iterations, the permutation problem, i.e., indeterminacy in ordering sources, can be solved by [6].

Finally, in order to obtain the SIMO components, the separated signals are projected back onto the microphones by using the inverse of $\boldsymbol{W}(f)$ [3]. In this method, the following operation is performed.

$$Y_k^{(l)}(f,t) = \Big\{\boldsymbol{W}(f)^{-1}\big[\overbrace{0,\cdots,0}^{l-1},Y_l(f,t),\overbrace{0,\cdots,0}^{L-l}\big]^{\mathrm{T}}\Big\}_k, \tag{5}$$

where $Y_k^{(l)}(f,t)$ represents the l-th resultant separated source signal which is projected back onto the k-th microphone, and $\{\cdot\}_k$ denotes the k-th element of the argument.

The FDICA-PB has the advantage that **(F1)** this method is very fast and nonsensitive to the initial value in the iterative updating because the calculation of FDICA given by (3) and the projection-back processing given by (5) are simple. There exists, however, the disadvantages that **(F2)** the inversion of $\boldsymbol{W}(f)$ often fails and yields harmful results because the invertibility of every $\boldsymbol{W}(f)$ cannot be guaranteed, and **(F3)** the circular convolution effect inherent in FDICA is likely to cause the deterioration of the separation performance.

4. SIMO-MODEL-BASED BSS 2: PROPOSED SIMO-ICA

The SIMO-ICA [4] 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

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = [y_k^{(\mathrm{ICA}l)}(t)]_{k1} = \sum_{n=0}^{D-1}\boldsymbol{w}_{(\mathrm{ICA}l)}(n)\boldsymbol{x}(t-n), \tag{6}$$

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 filter length.

Regarding the fidelity controller, we calculate the following signal vector, in which the all elements are to be mutually independent,

$$\boldsymbol{y}_{(\mathrm{ICA}L)}(t) \quad = \quad \boldsymbol{x}(t-D/2) - \sum_{l=1}^{L-1}\boldsymbol{y}_{(\mathrm{ICA}l)}(t). \tag{7}$$

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

$$\boldsymbol{w}_{(\mathrm{ICA}L)}(n) \;=\; \boldsymbol{I}\delta(n-\frac{D}{2}) - \sum_{l=1}^{L-1}\boldsymbol{w}_{(\mathrm{ICA}l)}(n), \tag{8}$$

where $\delta(n)$ is a delta function, i.e., $\delta(0)=1$ and $\delta(n)=0$ ($n\neq 0$). 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 $\boldsymbol{w}_{(\mathrm{ICA}L)}(n)$ is subject to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ $(l=1,\cdots,L-1)$.

To explicitly show the meaning of the fidelity controller, we rewrite (7) as

$$\sum_{l=1}^{L}\boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t-D/2) = 0. \tag{9}$$

Equation (9) 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 $[\sum_{l=1}^{L}A_{kl}(z)s_l(t-D/2)]_{k1}(=\boldsymbol{x}(t-D/2))$. Here the delay of $D/2$ is used as to deal with nonminimum phase systems. Using (6) and (7), we can obtain the appropriate separated signals and maintain their spatial qualities as follows.

Theorem: If the independent sound sources are separated by (6), and simultaneously the signals obtained by (7) are also mutually independent, then the output signals converge on unique solutions, up to the permutation, as

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \mathrm{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{s}(t-D/2), \tag{10}$$

where $\boldsymbol{P}_l$ $(l = 1, \cdots, L)$ are exclusively-selected permutation matrices which satisfy

$$\sum_{l=1}^{L}\boldsymbol{P}_l = [1]_{ij}\,. \tag{11}$$

Regarding a proof of the theorem, see [4].

Obviously the solutions given by (10) provide necessary and sufficient SIMO components, $A_{kl}(z)s_l(t-D/2)$, for each l-th source. However, the condition (11) allows multiple possibilities for the combination of $\boldsymbol{P}_l$. For example, one possibility is shown in Fig. 1 and this corresponds to

$$\boldsymbol{P}_l = [\delta_{im(k,l)}]_{ki}, \tag{12}$$

where δ_{ij} is Kronecker's delta function, and

$$m(k,l) = \begin{cases} k+l-1 & (k+l-1 \le L) \\ k+l-1-L & (k+l-1 > L) \end{cases}. \tag{13}$$

In this case, (10) yields

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = [A_{km(k,l)}s_{m(k,l)}(t-D/2)]_{k1} \quad (l=1,\cdots,L). \tag{14}$$

In order to obtain (10), the natural gradient of Kullback-Leibler divergence of (7) with respect to $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ should be added to the existing TDICA-based iterative learning rule [7] 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

$$\begin{aligned}
&\boldsymbol{w}_{(\mathrm{ICA}l)}^{[i+1]}(n) \\
&= \boldsymbol{w}_{(\mathrm{ICA}l)}^{[i]}(n) - \alpha\sum_{d=0}^{D-1}\Bigg[\Bigg\{\text{off-diag}\left\langle\boldsymbol{\varphi}\big(\boldsymbol{y}_{(\mathrm{ICA}l)}^{[i]}(t)\big)\right. \\
&\quad \left.\boldsymbol{y}_{(\mathrm{ICA}l)}^{[i]}(t-n+d)^{\mathrm{T}}\right\rangle_t\Bigg\}\cdot\boldsymbol{w}_{(\mathrm{ICA}l)}^{[i]}(d) \\
&\quad -\Bigg\{\text{off-diag}\left\langle\boldsymbol{\varphi}\big(\boldsymbol{x}(t-\frac{D}{2})-\sum_{l=1}^{L-1}\boldsymbol{y}_{(\mathrm{ICA}l)}^{[i]}(t)\big)\right. \\
&\quad \left.\cdot\Big(\boldsymbol{x}(t-n+d-\frac{D}{2})-\sum_{l=1}^{L-1}\boldsymbol{y}_{(\mathrm{ICA}l)}^{[i]}(t-n+d)^{\mathrm{T}}\Big)\right\rangle_t\Bigg\} \\
&\quad \cdot\Big(\boldsymbol{I}\delta(d-\frac{D}{2})-\sum_{l=1}^{L-1}\boldsymbol{w}_{(\mathrm{ICA}l)}^{[i]}(d)\Big)\Bigg],
\end{aligned} \tag{15}$$

where α is the step-size parameter, and $\boldsymbol{\varphi}(\cdot)$ is the nonlinear vector function, e.g., the l-th element is set to be $\tanh(y_l(t))$. In (15), the updating $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ for all l should be simultaneously performed in parallel in terms of l because each iterative equation is associated with the others via $\sum_{l=1}^{L-1}\boldsymbol{y}_{(\mathrm{ICA}l)}^{[i]}(t)$. Also, the initial values of $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ for all l should be different. After the iterations, the separated signals should be classified into SIMO components of each source because the permutation arises. This can be easily achieved by using a cross correlation among separated signals [4].

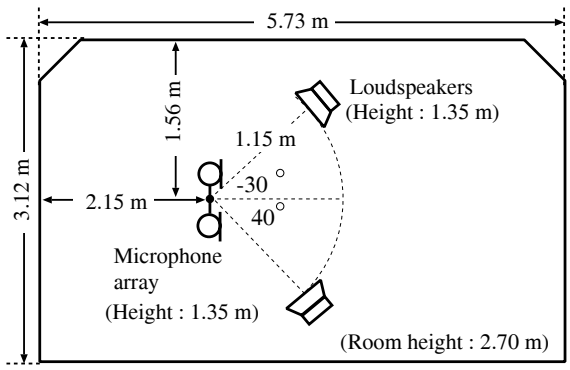

Figure 2: Layout of reverberant room used in experiments.

The SIMO-ICA has the following advantage and disadvantage. **(T1)** This method is free from both the circular convolution effect and the invertibility of the separation filter matrix. **(T2)** Since the SIMO-ICA is based on TDICA which involves more complex calculations than FDICA, the convergence of the SIMO-ICA is very slow, and the sensitivity to the initial settings of separation filter matrices is very high.

5. PROPOSED COMBINATION TECHNIQUE OF FDICA-PB AND SIMO-ICA

As described above, two kinds of SIMO-model-based BSS methods have some disadvantages. However, note that the advantages and disadvantages of FDICA-PB and SIMO-ICA are mutually complementary, i.e., (F2) and (F3) can be resolved by (T1), and (T2) can be resolved by (F1). Therefore, we propose a new multistage technique combining FDICA-PB and SIMO-ICA.

The proposed multistage technique is conducted with the following steps. In the first step, we perform FDICA to separate the source signals to some extent with the fast- and robust-convergence advantage (F1). After the FDICA, we generate a specific initial value $\boldsymbol{w}_{(\mathrm{ICA}l)}^{[0]}(n)$ for SIMO-ICA performed in the next step by using $\boldsymbol{W}(f)$ obtained from FDICA. This procedure is given by

$$\boldsymbol{w}_{(\mathrm{ICA}l)}^{[0]}(n) \;=\; \mathrm{IFFT}\Big[\mathrm{diag}\left[\boldsymbol{W}(f)^{-1}\boldsymbol{P}_l^{\mathrm{T}}\right]\boldsymbol{P}_l\boldsymbol{W}(f)\Big], \tag{16}$$

where $\boldsymbol{P}_l$ are set to be, e.g., (12), and IFFT$[\cdot]$ represents an inverse DFT with the time shift of $D/2$ samples. In the final step, we perform SIMO-ICA (15) to obtain resultant SIMO components with the advantage (T1).

Compared with the simple SIMO-ICA, this combination algorithm is not so sensitive to the initial value of the separation filter because FDICA is used for estimating the good initial value. Also, this technique has the possibility to provide a more accurate separation result over the simple FDICA because the resultant quality of the output signal is determined by the separation ability of the SIMO-ICA starting from the good initial state.

6. EXPERIMENTS AND RESULTS

6.1. 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 direc-

tions, -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 6 combinations. The sampling frequency is 8 kHz and the length of speech is limited to 7.5 seconds. To simulate the convolutive mixtures, the source signals are convolved with impulse responses recorded in the experimental room (see Fig. 2) which has a reverberation time (RT) of 150 ms. The length of the separation filter is set to be 2048 in both FDICA-PB and SIMO-ICA. The initial value in the simple SIMO-ICA case is null-beamformer whose directional null is steered to $\pm 45^\circ$. The initial value in FDICA is generated by PCA and FastICA [8].

As an objective evaluation score, *SIMO-model accuracy* (SA) [9] is used to indicate a degree of similarity (mean-squared-error) between the SIMO-model-based BSSs' outputs and the original SIMO-model-based signals ($A_{kl}(z)s_l(t - D/2)$).

6.2. Results and Discussion

In order to evaluate the SIMO-ICA's sensitivity to the initial value of the separation filter matrices $\boldsymbol{w}^{[0]}_{(\mathrm{ICA}l)}(n)$, we carry out separation experiments with the different initial matrices. According to (16), we generate multiple distinct $\boldsymbol{w}^{[0]}_{(\mathrm{ICA}l)}(n)$ from $\boldsymbol{W}(f)$ of various qualities by changing the number of iterations in FDICA.

Figure 3 shows the typical results of SAs for SIMO-ICA with different initial values in the case of a specific male-male combination. The SAs in these initial states are set to be from 0.6 to 18.2 dB, where 18.2 dB was maximum and an upper limit in FDICA-PB (see the horizontal solid line). From this figure, it is evident that the performances of SIMO-ICA is inferior to those of FDICA-PB under low-quality initial value conditions (0.6–15.3 dB), but SIMO-ICA can outperform FDICA-PB especially when the initial value is improved over 16.3 dB. This is a promising evidence on the feasibility of the proposed combination technique of FDICA-PB and SIMO-ICA, i.e., we can obtain accurate SIMO signals by using SIMO-ICA which follows a saturated FDICA-PB with the sufficient iterative updating.

Figure 4 shows the results of SAs for FDICA-PB, SIMO-ICA, and the proposed combination technique in all speaker combinations. In the results of the proposed combination technique, there exists a consistent improvement of SA compared with FDICA-PB as well as the simple SIMO-ICA. The average score of the improvement is 10.1 dB over SIMO-ICA, and is 4.5 dB over FDICA-PB. From these results, we can conclude that the proposed combination technique can assist the SIMO-ICA in improving the separation performance, and successfully achieve the SIMO-model-based BSS.

7. CONCLUSION

In this paper, first, the conventional FDICA-PB and the proposed SIMO-ICA were compared under a reverberant condition to evaluate the feasibility of SIMO-model-based BSS. Secondly, we proposed a new combination technique of FDICA-PB and SIMO-ICA to achieve the more higher separation performance compared with each of two methods. The experimental results revealed that the accuracy of the separated SIMO signals in the simple SIMO-ICA is inferior to that of FDICA-PB under low-quality initial value conditions, but the proposed combination technique of FDICA-PB and SIMO-ICA can outperform both simple FDICA-PB and SIMO-ICA. The average of the improvement was 10.1 dB over SIMO-ICA, and was 4.5 dB over FDICA-PB.

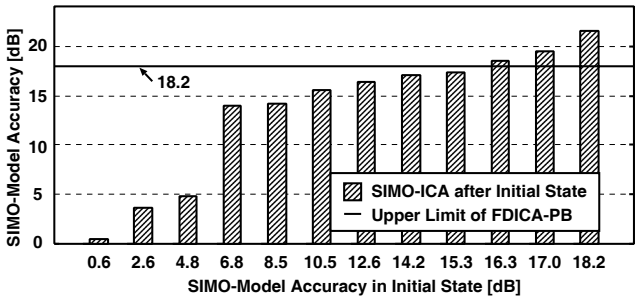

Figure 3: SIMO-model accuracy of SIMO-ICA under different initial value conditions.

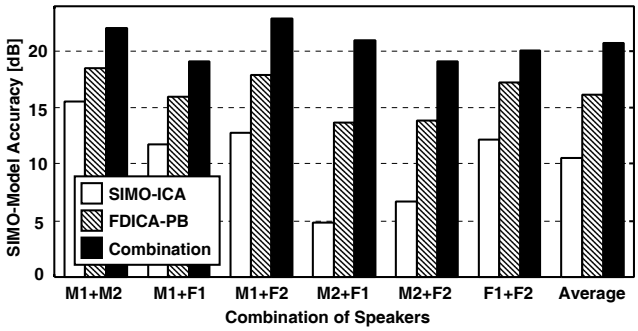

Figure 4: Comparison of SIMO-model accuracy among conventional FDICA-PB, proposed SIMO-ICA, proposed combination technique. "M1" and "M2" mean two male speakers, and "F1" and "F2" mean two female speakers.

8. ACKNOWLEDGEMENT

This work was partly supported by Core Research for Evolutional Science and Technology Program "Advanced Media Technology for Everyday Living" of Japan Science and Technology Agency.

9. REFERENCES

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

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

[3] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. Int. Sympo. on Nonlinear Theory and its Application (NOLTA '98)*, vol.3, pp.923–926, 1998.

[4] 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. Int. Workshop on Acoustic Echo and Noise Control*, pp.251–254, 2003.

[5] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency domain blind source separation," *IEICE Trans. Fundam.*, vol.E86-A, no.3, pp.590–596, 2003.

[6] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *Proc. Int. Sympo. on ICA and BSS*, pp.505–510, 2003.

[7] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," *Proc. Int. Workshop on ICA and BSS (ICA'99),* pp.371–376, 1999.

[8] A. Hyvarinen, "Fast and robust fixed-point algorithm for independent component analysis," *IEEE Trans. Neural Networks*, vol.10, no.3, pp.626–634, 1999.

[9] 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. Int. Workshop on Acoustic Echo and Noise Control*, pp.299–302, 2003.