EVALUATION OF SIMO SEPARATION METHODS FOR BLIND DECOMPOSITION OF BINAURAL MIXED SIGNALS

[1]*Tomoya Takatani,* [1]*Satoshi Ukai,* [1]*Tsuyoki Nishikawa,* [1]*Hiroshi Saruwatari, and* [1]*Kiyohiro Shikano*

[1]{tomoya-t, sawatari, shikano}@is.naist.jp

[1] Nara Institute of Science and Technology, Graduate School of Information Science,
8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192, JAPAN

ABSTRACT

High-fidelity blind source separation (BSS) using Single-Input Multiple-Output (SIMO)-model-based Independent Component Analysis (SIMO-ICA) is now being studied by the authors. This paper describes a comparison of two types of SIMO-ICAs with different constrains and the conventional methods, and gives explicit discussion on the sensitivity of the parameters settings in the methods. In order to discuss the difference, the source-separation experiments using the mixed binaural sounds are carried out under the same real acoustic conditions. The experiment results reveal that SIMO-ICA-IG outperforms SIMO-ICA-LS and the conventional methods, and the parameter setting in SIMO-ICA-IG does not depend on the source signals' properties compared with that of SIMO-ICA-LS.

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 to deal with a means of separation of acoustic sounds [1], [3]–[5]. 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 method from being applied to binaural signal processing [2]

Generally speaking, human beings listen to the sounds by their two ears. These sounds detected at both ears called “binaural sounds.” This binaural sound involves the information about the localization, directivity, and spatial qualities of each sound source. Also, if the several *undesired* sources exist around the target sound, we listen to the mixed binaural sound from the sources, not the binaural sound from the single source. Our research goal is to realize the audio augmented reality system which can extract the target binaural sound component of the mixed binaural sound without the loss of the information about the spatial qualities. In order to realize this system, we use the special apparatus, earphone-microphone system, shown in Figure 1, for picking up the sounds at the entrance of ear canal (cf. Figure 2). In this system, it is essential to blindly decompose the mixed sounds, not into the monaural signals but into the binaural sounds at ear points.

In order to solve the above-mentioned fundamental problems, we have recently proposed high-fidelity BSS methods using two kinds of the Single-Input Multiple-Output (SIMO)-model-based ICAs, SIMO-ICA with least squares (SIMO-ICA-LS) [6] and SIMO-ICA with information-geometric learning (SIMO-ICA-IG) [7]. 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. The SIMO-ICA consists of multiple ICA parts and a fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system. The 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 the SIMO-ICA can maintain the spatial qualities of each sound source.

Our previous works [6, 7] only provided the experimental results in the simple microphone array's framework. In this paper, we newly discuss the feasibility and usability of the two SIMO-ICAs from the point of view of *binaural signal separation*. The source-separation experiments using two SIMO-ICAs and conventional monaural output methods are carried out under the same real acoustic conditions. The experiment results reveal that SIMO-ICA-IG outperforms SIMO-ICA-LS and conventional methods, and the parameter setting in SIMO-ICA-IG does not depend on the source signals' properties compared with that in SIMO-ICA-LS.

2. MIXING PROCESS

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

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

where $\boldsymbol{s}(t) = [s_1(t), s_2(t)]^{\mathrm{T}}$ 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

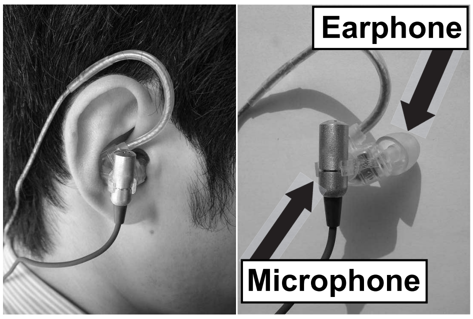

Figure 1: Detail of the earphone-microphone. This apparatus picks up the binaural sounds detected at both ears.

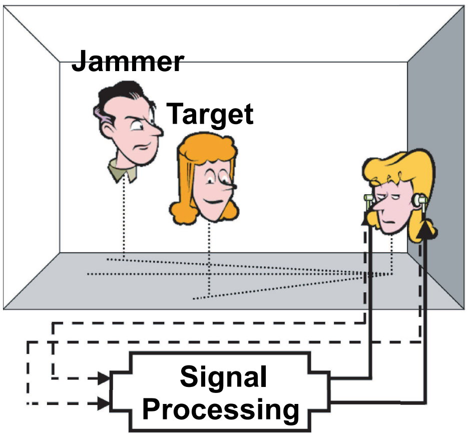

Figure 2: The concept of audio augmented reality which can reproduce only the target sound. This system aims to extract the target component of the mixed sounds detected at both ears without the loss of information about directivity, localization , and the spatial qualities of target source.

and the j-th column. The observed signal $\boldsymbol{x}(t)$ is generally represented as a superposition of the SIMO-model-based signals as follows:

$$\begin{aligned}\boldsymbol{x}(t) \;=\;& [A_{11}(z)s_1(t), \cdots, A_{K1}(z)s_1(t)]^{\mathrm{T}} \\ &+ [A_{12}(z)s_2(t), \cdots, A_{K2}(z)s_2(t)]^{\mathrm{T}} \\ &\vdots \\ &+ [A_{1L}(z)s_L(t), \cdots, A_{KL}(z)s_L(t)]^{\mathrm{T}}, \qquad (2)\end{aligned}$$

where $[A_{1l}(z)s_l(t), \cdots, A_{Kl}(z)s_l(t)]^{\mathrm{T}}$ is a specific vector which includes SIMO-model-based signals with respect to the l-th sound source; the k-th element corresponds to the k-th microphone's signal.

3. CONVENTIONAL SEPARATION METHODS

The conventional ICA is basically a means of extracting each of the independent sound sources as a monaural signal [3, 4]. In addition, the quality of the separated sound cannot be guaranteed, i.e., the separated signals can possibly include spectral distortions because the modified separated signals which convolved with arbitrary linear filters are still mutually independent. Therefore, the conventional ICA has a serious drawback in that the separated sounds cannot maintain information about the directivity, localization, or spatial qualities of each sound source. In order to resolve the problems, particularly for the sound quality, Matsuoka et al. have proposed a modified ICA based on the Minimal Distortion Principle [5]. However, this method is valid for only monaural outputs, and the fidelity of the output signals as SIMO-model-based signals cannot be guaranteed.

4. TWO KINDS OF SIMO-ICA ALGORITHMS

In order to solve the above-mentioned fundamental problems, we have recently proposed SIMO-model-based blind separation methods using two kinds of SIMO-ICAs, SIMO-ICA-LS and SIMO-ICA-IG.

4.1. Proposed algorithm1: SIMO-ICA-LS [6]

The SIMO-ICA-LS consists of L ICA parts and a *fidelity controller*, and each ICA runs in parallel under the fidelity control of the entire separation system. The separated signals of the l-th ICA ($l = 1, \cdots, L$) in SIMO-ICA-LS are defined by

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA}l)}(n)\boldsymbol{x}(t-n), \qquad (3)$$

where $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ is the separation filter matrix in the l-th ICA. Regarding the fidelity controller, we introduce the following cost function to be minimized,

$$\begin{aligned}&C(\boldsymbol{w}_{(\mathrm{ICA}1)}(n), ..., \boldsymbol{w}_{(\mathrm{ICA}L)}(n)) \\ &\equiv \left\langle \| \sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t) - \boldsymbol{x}(t - D/2) \|^2 \right\rangle_t, \qquad (4)\end{aligned}$$

where $\| \boldsymbol{x} \|$ is the Euclidean norm of vector $\boldsymbol{x}$. The cost function Eq. (4) means a degree of similarity between the sum of all ICA's output $\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}(t)$ and the sum of all SIMO components $[\sum_{l=1}^{L} A_{kl}(t-D/2)]_{k1} (= \boldsymbol{x}(t-D/2)$. Here the delay of $D/2$ is used to deal with nonminimum phase systems. Using Eq. (3) and Eq. (4), we can obtain the unique SIMO solutions, up to the permutation, as

$$\boldsymbol{y}_{(\mathrm{ICA}l)}(t) = \mathrm{diag}\left[\boldsymbol{A}(z)\boldsymbol{P}_l^{\mathrm{T}}\right] \boldsymbol{P}_l \boldsymbol{s}(t - D/2), \qquad (5)$$

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

$$\sum_{l=1}^{L} \boldsymbol{P}_l = [1]_{ij}. \qquad (6)$$

In order to obtain SIMO-model-based signals, the natural gradient [1] of Eq. (4) with respect to $\boldsymbol{w}_{\mathrm{ICA}l}(n)$ should be added to the iterative learning rule of the separation filter. The iterative algorithm of SIMO-ICA-LS is expressed as

$$\begin{aligned}&\boldsymbol{w}_{(\mathrm{ICA}l)}^{[j+1]}(n) \\ &= \boldsymbol{w}_{(\mathrm{ICA}l)}^{[j]}(n) - \alpha \sum_{d=0}^{D-1} \Bigg\{ \text{off-diag} \left\langle \boldsymbol{\varphi}(\boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t)) \right. \\ &\quad \left. \cdot \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t-n+d)^{\mathrm{T}} \right\rangle_t \\ &\quad + \beta \Big\langle (\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t) - \boldsymbol{x}(t - \frac{D}{2})) \\ &\quad \cdot \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t-n+d)^{\mathrm{T}} \Big\rangle_t \Bigg\} \cdot \boldsymbol{w}_{(\mathrm{ICA}l)}^{[j]}(d), \qquad (7)\end{aligned}$$

where α and β are the step-size parameters; α is for the control of the total update quantity and β is for the fidelity control. In Eq. (7) the updating $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ should be simultaneously performed in parallel because each iterative equation is associated with the others via $\sum_{l=1}^{L} \boldsymbol{y}_{(\mathrm{ICA}l)}^{[j]}(t)$. Also, the initial values of $\boldsymbol{w}_{(\mathrm{ICA}l)}(n)$ for all l should be different.

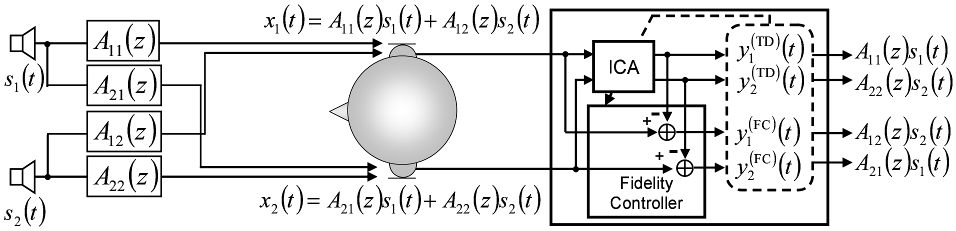

Figure 3: Example of input and output relations in the proposed SIMO-ICA-IG.

4.2. Proposed algorithm2: SIMO-ICA-IG [7]

The SIMO-ICA-IG consists of an ICA part and a *fidelity controller*. The separated signals of the l-th ICA ($l = 1, \cdots, L-1$) in the SIMO-ICA are defined by

$$\boldsymbol{y}_{(\mathrm{ICA})}(t) = [y_k^{(\mathrm{ICA})}(t)]_{k1} = \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{ICA})}(n)\boldsymbol{x}(t-n), \quad (8)$$

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

$$\boldsymbol{y}_{(\mathrm{FC})}(t) = [y_k^{(\mathrm{FC})}(t)]_{k1} = \boldsymbol{x}(t - \frac{D}{2}) - \boldsymbol{y}_{(\mathrm{ICA})}(t). \quad (9)$$

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

$$\boldsymbol{w}_{(\mathrm{FC})}(n) = \boldsymbol{I}\delta(n - \frac{D}{2}) - \boldsymbol{w}_{(\mathrm{ICA})}(n), \quad (10)$$

where $\delta(n)$ is a delta function, where $\delta(0) = 1$ and $\delta(n) = 0$ $(n \neq 0)$. From (10), we can rewrite (9) as

$$\boldsymbol{y}_{(\mathrm{FC})}(t) = \sum_{n=0}^{D-1} \boldsymbol{w}_{(\mathrm{FC})}(n)\boldsymbol{x}(t-n). \quad (11)$$

The reason why we use the word “virtual” here is that fidelity controller does not have own separation filters unlike the ICA, and $\boldsymbol{w}_{(\mathrm{FC})}(n)$ is subject to $\boldsymbol{w}_{(\mathrm{ICA})}(n)$. To explicitly show the meaning of the fidelity controller, (9) is rewritten as

$$\boldsymbol{y}_{(\mathrm{ICA})}(t) + \boldsymbol{y}_{(\mathrm{FC})}(t) - \boldsymbol{x}(t - D/2) = [0]_{k1}. \quad (12)$$

Equation (12) means a constraint to force the sum of the all of output vectors $\boldsymbol{y}_{(\mathrm{ICA})}(t) + \boldsymbol{y}_{(\mathrm{FC})}(t)$ 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 (8), and simultaneously the signals obtained by (9) are also mutually independent, then the output signals converge on unique solutions,

$$\boldsymbol{y}_{(\mathrm{ICA})}(t) = [A_{11}(z)s_1(t - D/2), A_{22}(z)s_2(t - D/2)]^{\mathrm{T}}, \quad (13)$$

$$\boldsymbol{y}_{(\mathrm{FC})}(t) = [A_{12}(z)s_2(t - D/2), A_{21}(z)s_1(t - D/2)]^{\mathrm{T}}, \quad (14)$$

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 [7]. Equations (13) and (14) represent necessary and sufficient SIMO components of all source signals.

In order to obtain the above-mentioned solutions, the natural gradient [1] of Kullback-Leibler divergence of (9) with respect to $\boldsymbol{w}_{(\mathrm{ICA})}(n)$ should be added to the iterative learning rule of the separation filter in the ICA. The iterative algorithm of the ICA part in SIMO-ICA is given as

$$\begin{aligned}\boldsymbol{w}_{(\mathrm{TD})}^{[j+1]}(n) = \boldsymbol{w}_{(\mathrm{TD})}^{[j]}(n) - \alpha \sum_{d=0}^{D-1} \Big[&\text{off-diag}\Big\{\Big\langle \boldsymbol{\varphi}(\boldsymbol{y}_{(\mathrm{ICA})}^{[j]}(t)) \\ &\boldsymbol{y}_{(\mathrm{ICA})}^{[j]}(t-n+d)^{\mathrm{T}}\Big\rangle_t\Big\}\boldsymbol{w}_{(\mathrm{ICA})}^{[j]}(d) \\ &-\text{off-diag}\Big\{\Big\langle \boldsymbol{\varphi}(\boldsymbol{y}_{(\mathrm{FC})}^{[j]}(t))\boldsymbol{y}_{(\mathrm{FC})}^{[j]}(t-n+d)^{\mathrm{T}}\Big\rangle_t\Big\} \\ &\Big(\boldsymbol{I}\delta(d - \frac{D}{2}) - \boldsymbol{w}_{(\mathrm{ICA})}^{[j]}(d)\Big)\Big], \end{aligned} \quad (15)$$

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 (15), the initial values of $\boldsymbol{w}_{(\mathrm{ICA})}(n)$ and $\boldsymbol{w}_{(\mathrm{FC})}(n)$ are arbitrary, but should be different each other.

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 two directions, -30° and 45°. 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 6 combinations. The sampling frequency is 8 kHz and the length of speech is limited to 3 seconds. The length of $\boldsymbol{w}(n)$ in each method is 1024, and the initial values are inverse filters of HRTFs whose directions of sources are -60° and 60°. The step-size parameters η and α are 5.0×10^{-2} and 1.0×10^{-6}. SIMO-model accuracy (SA) [8] 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

In each method, the step-size parameter α is changed from 5.0×10^{-8} to 5.0×10^{-6}. Also, the balancing parameter β is changed from 1.0×10^{-5} to 1.0×10^{-2} in MDP-ICA and SIMO-ICA-LS, and that is changed from 0.1 to 100 in SIMO-ICA-IG.

Figure 4 provides the results of SIMO-model accuracy for each speaker combination in 2nd-order ICA by Parra, NH-ICA by Choi, MDP-ICA by Matsuoka, SIMO-ICA-LS, SIMO-ICA-IG,and SIMO-ICA-IG ($\beta = 1$). As shown in this figure, we can recognize the proposed SIMO-ICA-IG’s superiority to the other methods.

Figure 5 shows the combinations of optimum step-size parameter and the balancing parameter, which give the best separation performances, for different speaker combinations in MDP-ICA,

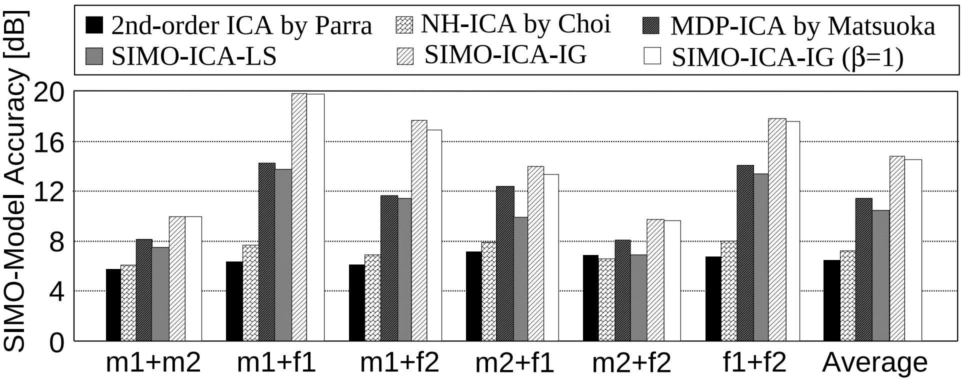

Figure 4: Results of SIMO-model accuracy for each speaker combination in 2nd-order ICA by Parra, NH-ICA by Choi, MDP-ICA by Matsuoka, SIMO-ICA-LS, SIMO-ICA-IG,and SIMO-ICA-IG ($\beta = 1$). Here "m1" and "m2" correspond to male speakers, and "f1" and "f2" correspond to female speakers.

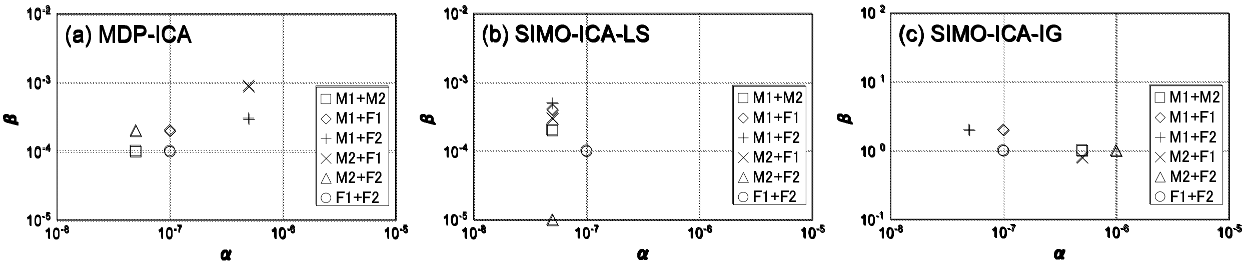

Figure 5: Optimal step-size parameter and balancing parameter for different speaker combinations in (a) MDP-ICA, (b) SIMO-ICA-LS, and (c) SIMO-ICA-IG.

SIMO-ICA-LS, and SIMO-ICA-IG. From this result, the optimum step-size parameter α among these methods is within the range between 5.0×10^{-8} and 1.0×10^{-6}.

In Figs. 5 (a) and (b), the optimum balancing parameter β in MDP-ICA and SIMO-ICA-LS is dispersed in the huge range between 1.0×10^{-4} and 1.0×10^{-3}, and 1.0×10^{-5} and 5.0×10^{-4}. In Fig. 5 (c), however, the optimum balancing parameter β in SIMO-ICA-IG is within the range between 0.8 to 2.0, i.e., almost around 1. From this result, the range of balancing parameter β in SIMO-ICA-IG is narrower than that in MDP-ICA and SIMO-ICA-LS, and consequently the parameter setting in SIMO-ICA-IG does not depend on the source signals' properties. In addition, we can mention the attractive feature that β in SIMO-ICA-IG is negligible because the separation performance of SIMO-ICA-IG with $\beta = 1$ is almost the same as that of optimal SIMO-ICA-IG (see white bars in Fig. 4).

Overall, the results indicate that the proposed SIMO-ICA-IG outperforms other methods, and there is no deterioration in performance of SIMO-ICA-IG even if the balancing parameter β is set to 1.

6. CONCLUSION

We discuss and compare SIMO-model-based BSS methods for audio augmented reality. The experiment results reveal that SIMO-ICA-IG outperforms SIMO-ICA-LS and conventional methods, and the parameter setting in SIMO-ICA-IG does not depend on the source signals' properties compared with that in SIMO-ICA-LS. Therefore, we can conclude that the SIMO-ICA-IG is a robuster and easy-to-use algorithm than SIMO-ICA-LS.

7. ACKNOWLEDGMENT

This research is partly supported by CREST program "Advanced Media Technology for Everyday Living" of JST.

8. REFERENCES

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

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

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

[4] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," *Proc. International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99),* pp.371–376, 1999.

[5] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. International Conference on Independent Component Analysis and Blind Signal Separation,* pp.722–727, Dec. 2001.

[6] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fidelity blind separation of acoustic signals using SIMO-model-based independent component analysis," *IEICE Trans. Fundam.* Vol.E87-A, No.8, pp.2063–2072, 2004.

[7] 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.

[8] 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.