HIGH-PRESENCE HEARING-AID SYSTEM USING DSP-BASED REAL-TIME BLIND SOURCE SEPARATION MODULE

Y. Mori, T. Takatani, H. Saruwatari, K. Shikano

Nara Institute of Science and Technology
Nara, 630-0192, Japan
(e-mail: yoshim-m@is.naist.jp)

T. Hiekata, T. Morita

Kobe Steel,Ltd.
Kobe, 651-2271, Japan
(e-mail: t-hiekata@kobelco.jp)

ABSTRACT

Real-time two-stage blind source separation (BSS) method for convolutive mixtures of speech is now being studied by the authors, in which a single-input multiple-output (SIMO)-model-based independent component analysis (ICA) and a SIMO-model-based binary masking are combined. In addition, we have developed a pocket-size real-time DSP module implementing the two-stage BSS method. In this paper, we introduce a high-presence hearing-aid system which can reduce the interference sound and reproduce the target sound while keeping the directivity, and realize the system with the real-time BSS module. To evaluate it, we carried out the objective and subjective experiments using 9 users. From these results, it is revealed that the decomposition performance and the directivity maintenance of the proposed system are superior to those of conventional methods.

Index Terms— Hearing aids, Real time systems, User interfaces, Digital signal processors

1. INTRODUCTION

Generally speaking, human beings listen to the sounds by their two ears. These sounds detected at both ears called "binaural sounds." There binaural sounds involve information about the localization, directivity, and spatial qualities of each sound source. Also, if several *undesired* sources, for example noise, interference speech and so on, exist around us, we listen to the mixed binaural sounds from the sources, not the binaural sounds from the single source. Our research goal is to realize the high-presence hearing-aid system (see Fig. 1) which can extract the target components of the mixed binaural sounds without the loss of information about the spatial qualities in real-time. In order to realize this system, we use the special apparatus, earphone-microphone system, shown in Fig. 1, for picking up the sounds at the entrance of ear canal. Also, we should construct the real-time blind system which can separate the target binaural sounds, not into the monaural target sound.

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 researches of BSS based on independent component analysis (ICA), various methods have been presented for acoustic-sound separation [1–3]. On the other hand, binary masking based BSS [4] technique has also been proposed. However, those conventional 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.

In order to attack this problem, one of the authors has proposed the single-input multiple-output (SIMO)-model-based ICA (SIMO-ICA) [5]. 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 enable us to separate the mixed signals, not into monaural source signals but into SIMO-model-based signals from independent sources as if these sources were at the microphones. Thus, the separated signals of SIMO-ICA can maintain the rich spatial qualities of each sound source. However, the most important issue for a development of hearing-aid system is to make the system working in real-time. From this point of view, the above-mentioned SIMO-ICA is an insufficient solution because of needing huge amount of calculations. Alternatively, we have proposed real-time two-stage BSS algorithm [6]. This approach resolves the BSS problem into two stages: (a) SIMO-ICA and (b) SIMO-model-based binary masking (SIMO-BM) for the SIMO signals obtained from the preceding SIMO-ICA. Whereas this method is aimed to output the monaural signal and cannot provide SIMO-model-based signals, the part of the method contains SIMO-ICA. Thus, our proposed method dismissed SIMO information by accidentally.

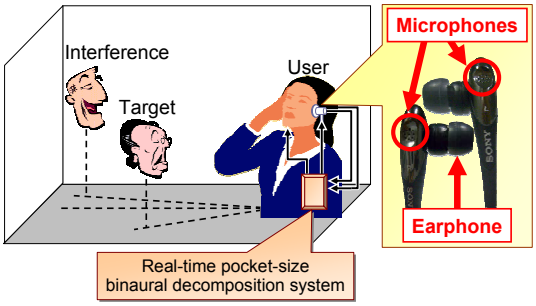

Fig. 1. The concept of high-presence hearing-aid system which can reproduce only the target sound with the earphone-microphone.

In this paper, we extend the real-time two-stage BSS method to be able to output SIMO-model based binaural signals, and apply the method to the hearing-aid system, which can reproduce the target binaural sounds extracted from the mixed binaural signals. Also we verify its effectiveness through objective and subjective evaluation experiments for multiple users. From these experimental results, we can confirm that the decomposition performance and the directivity maintenance of the proposed system are superior to those of the conventional method.

2. MIXING PROCESS

In this study, the number of microphones is $K = 2$ and the number of multiple sound sources is $L = 2$. By applying the short-time discrete-time Fourier transform, we can express the observed signals

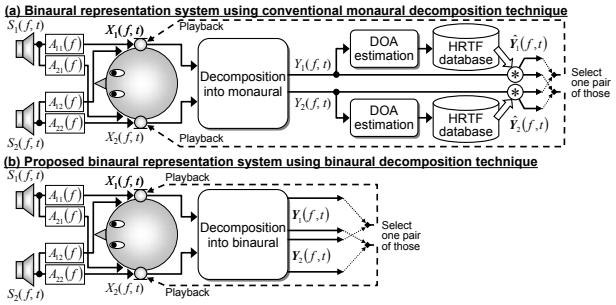

Fig. 2. Overview of (a) conventional decomposition and representation system and (b) proposed binaural decomposition and representation system.

as follows in the time-frequency domain:

$$\boldsymbol{X}(f,t) = [X_1(f,t), \cdots, X_K(f,t)]^{\mathrm{T}} = \boldsymbol{A}(f)\boldsymbol{S}(f,t), \tag{1}$$

$$\boldsymbol{S}(f,t) = [S_1(f,t), \cdots, S_L(f,t)]^{\mathrm{T}}, \tag{2}$$

where $\boldsymbol{X}(f,t)$ is the observed signal vector, and $\boldsymbol{S}(f,t)$ is the source signal vector. Also, $\boldsymbol{A}(f) = [A_{kl}(f)]_{kl}$ is the mixing matrix, where $[X]_{ij}$ denotes the matrix which includes the element X in the i-th row and the j-th column. The mixing matrix $\boldsymbol{A}(f)$ is complex-valued because we introduce a model to deal with the room reverberations.

3. HEARING-AID SYSTEM

3.1. Relevant study and problem

In this study, we construct the hearing-aid system, which can decompose the mixed binaural signals into some pairs of the binaural signals w.r.t. each sound source in real-time. The blind decomposition techniques [1–4], those were basically proposed for the preprocessing of the speech recognition system or the speech communication system, can only output monaural signals. Therefore the configuration of the hearing-aid system using conventional decomposition techniques is often designed as in Fig. 2(a). This system consists of three parts; in the first part, the decomposition method makes the input mixed binaural signals into monaural signals w.r.t. the source signals. After decomposition, the direction-of-arrival (DOA) estimator detects DOAs of the source signals. In the final part, the decomposed signals are convoluted with head-related transfer function (HRTF) data which are selected by DOAs, and the convoluted signals are represented to the user. This HRTF database is recorded and constructed in advance, e.g., using a dummy head or someone's personal data. However, this system has some problems. First of all, the DOA estimator cannot work well when the residual error is remained in the decomposed signals. In addition, the HRTF database is not matched to the user at all. Since the construction of the database is very complicated work and the measurement costs very long time, a user-specific HRTF database is difficult to be built in short time.

3.2. Our approach

To maintain directivity of target sources, we estimate the SIMO-model-based signals. Here SIMO-model-based signal represents the specific transmitted signal in which only one source signal exists in the acoustic field and the observed signals are binaural sounds. Those signals include the information about directivity and localization of the source signal, and also involve the user-specific HRTF. SIMO-ICA [5] 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. However, this ICA is unfitted for real-time processing, and thus we cannot utilize this method to realize the hearing-aid system directory. On the other hand, we have proposed real-time BSS method [6], which extracts SIMO-model-based signals and use them internally and temporally in the algorithm. Therefore, in this paper, we extend this method to be able to output binaural signals, that is SIMO-model-based signals. In addition, we use the special apparatus, earphone-microphone (modification of noise canceling headphone, SONY MDR-NC11A), shown in Fig. 1, for picking up the binaural mixed sounds at the entrance of ear canal. The outline of this approach is shown in Fig. 2(b). Hereby, this system can observe the mixed signals with not only spatial information of the room but also HRTF of the user, then decompose the observed signals, and finally represents the output signals to the user. The prospective advantages are as follows; (a) no need to measure each user's HRTF in advance, and (b) if the complete SIMO decomposition is achieved, there exists no mismatch regarding HRTF, unlike the conventional methods.

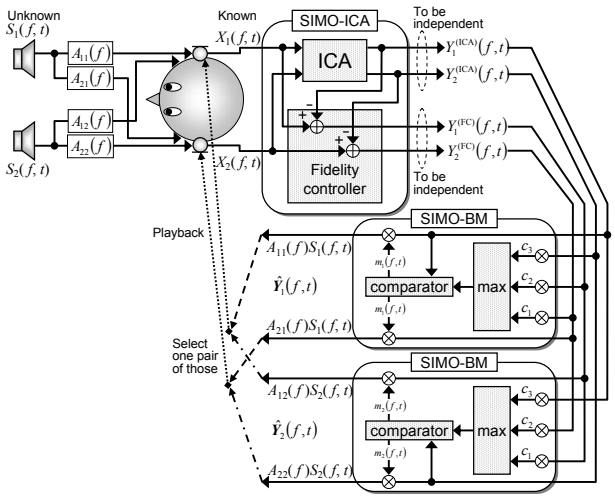

Fig. 3. Input and output relations in proposed two-stage BSS and sound reproduction system.

4. ALGORITHM

The configuration of the proposed method is shown in Fig. 3. In the following subsections, we explicate each signal processing part of the proposed method.

4.1. 1st stage: SIMO-ICA part

Time-domain SIMO-ICA [5] has recently been proposed by some of the authors as a means of obtaining SIMO-model-based signals directly in ICA updating. In this study, we use frequency-domain SIMO-ICA (FD-SIMO-ICA). FD-SIMO-ICA is conducted for extracting the SIMO-model-based signals corresponding to each of the sources. FD-SIMO-ICA consists of one FDICA parts and a *fidelity controller* (FC). The separated signals of ICA in FD-SIMO-ICA are defined by

$$\boldsymbol{Y}_{(\mathrm{ICA})}(f,t) = \boldsymbol{W}_{(\mathrm{ICA})}(f)\,\boldsymbol{X}(f,t), \tag{3}$$

where $\boldsymbol{W}_{(\mathrm{ICA})}(f) = [W_{ij}^{(\mathrm{ICA})}(f)]_{ij}$ is the separation filter matrix in ICA.

Regarding the FC, we obtain the following signal vector $\boldsymbol{Y}_{(\mathrm{ICA2})}(f,t)$:

$$\boldsymbol{Y}_{(\mathrm{FC})}(f,t) = \boldsymbol{Y}_{(\mathrm{ICA2})}(f,t) = \boldsymbol{X}(f,t) - \boldsymbol{Y}_{(\mathrm{ICA})}(f,t). \tag{4}$$

Hereafter, we regard $\boldsymbol{Y}_{(\mathrm{FC})}(f,t)$ as an output of a *virtual* "*2nd*" ICA. The reason we use the word "*virtual*" here is that the *2nd* ICA does not have its own separation filters unlike another ICA, and $\boldsymbol{Y}_{(\mathrm{ICA2})}(f,t)$ is subject to $\boldsymbol{W}_{(\mathrm{ICA})}(f)$. By transposing the second term $(-\boldsymbol{Y}_{(\mathrm{ICA})}(f,t))$ on the right-hand side to the left-hand side, we can show that (4) suggests a constraint to force the ICAs' output vectors $\boldsymbol{Y}_{(\mathrm{ICA})}(f,t)$ to be the sum of all SIMO components $[\sum_{l=1}^{L} A_{kl}(f)\, S_l(f,t)]_{k1}$ $(= \boldsymbol{X}(f,t))$.

If the independent sound sources are separated by (3), and simultaneously the signals obtained by (4) are also mutually independent, then the output signals converge on unique solutions [5]:

$$\begin{cases} \boldsymbol{Y}_{(\mathrm{ICA})}(f,t) = [A_{11}(f)S_1(f,t), A_{22}(f)S_2(f,t)]^{\mathrm{T}} \\ \boldsymbol{Y}_{(\mathrm{FC})}(f,t) = [A_{12}(f)S_2(f,t), A_{21}(f)S_1(f,t)]^{\mathrm{T}} \end{cases}. \quad (5)$$

Obviously, the solutions provide necessary and sufficient SIMO components, $A_{kl}(f)S_l(f,t)$, 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 (5), the natural gradient of Kullback-Leibler divergence of (4) with respect to $\boldsymbol{W}_{(\mathrm{ICA})}(f)$ should be added to the existing nonholonomic iterative learning rule [1] of the separation filter in the ICA. The new iterative algorithm of the ICA part in FD-SIMO-ICA is given as

$$\begin{aligned} \boldsymbol{W}^{[j+1]}_{(\mathrm{ICA})}(f) = \boldsymbol{W}^{[j]}_{(\mathrm{ICA})}(f) \\ -\alpha \left\{ \text{off-diag} \left\langle \boldsymbol{\Phi}\left(\boldsymbol{Y}^{[j]}_{(\mathrm{ICA})}(f,t)\right) \boldsymbol{Y}^{[j]}_{(\mathrm{ICA})}(f,t)^{\mathrm{H}} \right\rangle_t \right\} \cdot \boldsymbol{W}^{[j]}_{(\mathrm{ICA})}(f) \\ +\alpha \left\{ \text{off-diag} \left\langle \boldsymbol{\Phi}\left(\boldsymbol{Y}^{[j]}_{(\mathrm{FC})}(f,t)\right) \cdot \boldsymbol{Y}^{[j]}_{(\mathrm{FC})}(f,t)^{\mathrm{H}} \right\rangle_t \right\} \\ \cdot \left(\boldsymbol{I} - \boldsymbol{W}^{[j]}_{(\mathrm{ICA})}(f)\right), \end{aligned} \quad (6)$$

where α is the step-size parameter, $\boldsymbol{I}$ is the identity matrix, $\langle\cdot\rangle_t$ denotes the time-averaging operator, $[i]$ is used to express the value of the i th step in the iterations, and $\boldsymbol{\Phi}(\cdot)$ is the appropriate nonlinear vector function [7].

4.2. 2nd stage: SIMO-BM part [6] extended to binaural output

After FD-SIMO-ICA, SIMO-model-based binary masking processing is applied to (5). In [6], this part can only outputs monaural signals, thus in this paper, we extend this part to be able to output binaural signals. The resultant output signal corresponding to the source 1 is determined in the proposed SIMO-BM as follows:

$$\hat{\boldsymbol{Y}}_1(f,t) = \left[m_1(f,t)\, Y_1^{(\mathrm{ICA})}(f,t),\ m_1(f,t)\, Y_2^{(\mathrm{FC})}(f,t) \right]^{\mathrm{T}}, \quad (7)$$

where $m_1(f,t)$ is the *SIMO-model-based* binary mask operation which is defined as $m_1(f,t) = 1$ if

$$\begin{aligned} & Y_1^{(\mathrm{ICA})}(f,t) \\ & > \max\left[\left| c_1 Y_2^{(\mathrm{FC})}(f,t) \right|,\ \left| c_2 Y_1^{(\mathrm{FC})}(f,t) \right|,\ \left| c_3 Y_2^{(\mathrm{ICA})}(f,t) \right|, \right] \end{aligned} \quad (8)$$

otherwise $m_1(f,t) = 0$. Here $\max[\cdot]$ represents the function to pick up the maximum value among the arguments, and $c_1, \cdots, c_3$ are the weight for enhancing the contribution of each SIMO component to the masking decision.

The resultant output corresponding to the source 2 is given by

$$\hat{\boldsymbol{Y}}_2(f,t) = \left[m_2(f,t)\, Y_1^{(\mathrm{FC})}(f,t),\ m_2(f,t)\, Y_2^{(\mathrm{ICA})}(f,t) \right]^{\mathrm{T}}, \quad (9)$$

where $m_2(f,t)$ is defined as $m_2(f,t) = 1$ if

$$\begin{aligned} & Y_2^{(\mathrm{ICA})}(f,t) \\ & > \max\left[\left| c_1 Y_1^{(\mathrm{FC})}(f,t) \right|,\ \left| c_2 Y_2^{(\mathrm{FC})}(f,t) \right|,\ \left| c_3 Y_1^{(\mathrm{ICA})}(f,t) \right|, \right] \end{aligned} \quad (10)$$

otherwise $m_2(f,t) = 0$.

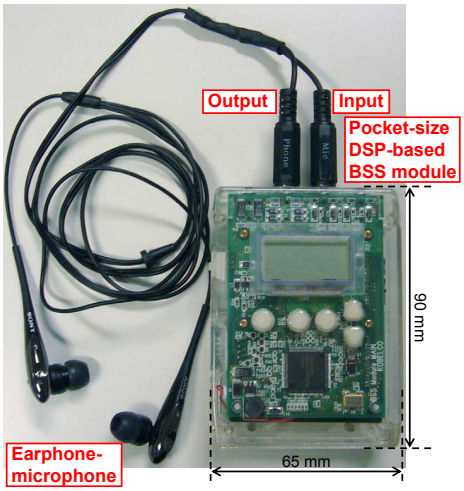

Fig. 4. Overview of proposed hearing-aid system, which consists of the earphone-microphone and the pocket-size real-time BSS module.

Table 1. Specifications of pocket-size real-time BSS module

Processor	TI TMS320VC6713 (clock frequency: 200 MHz)
Input/output interfaces	Binaural mics. in
	Binaural out
Sampling frequency	8 kHz (expandable to 16 / 32 kHz)
Power supply	AA cell battery × 2
Amount of memory	Flash ROM: 100 K Byte used
	SDRAM: 1 M Byte used
Weight	150 g (including buttery)

5. EXPERIMENTS AND RESULTS

5.1. Conditions of experiments

We carried out binaural-sound-separation experiments using source signals which are convolved with impulse responses recorded at each user's ears 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; $\theta_1 = \{-75°, -60°, -45°, -30°, -15°, 0°\}$ and $\theta_2 = \{0°, 15°, 30°, 45°, 60°, 75°\}$. The distance between a user and the sound source is 1.5 m. The number of users is 9 (8 males and 1 female). The sampling frequency is 8 kHz and the length of each speech sample is limited to 9 seconds. The length of $\boldsymbol{W}(f)$ in each method is 1024. SIMO-BM parameters are set to $[c_1, c_2, c_3] = [1, 0, 0.5]$. This choice is the well-balanced between the separation performance and the sound distortion. As the conventional method for comparison, we use the binary masking [4] and conventional ICA [1]. In these evaluations, since the output signals of binary masking and ICA are monaural w.r.t. each sound source, we attach the information of the specific HRTF (dummy head) w.r.t. the estimated target direction from separated signals after separation processing, and generate the SIMO-model-based signals. Noise reduction rate (NRR) [3] is used as an objective evaluation scores. The NRR is defined as the output SNR minus input SNR in dB and indicates the separation performance.

5.2. Real-time hearing-aid system

We have already built a pocket-size real-time BSS module [6], where the proposed two-stage BSS algorithm can work on a general-purpose DSP as shown in Fig. 4 and Table 1. In this module, SIMO-ICA is conducted using current 3-s-duration data for estimating the separation matrix, that is applied to the next (*not current*) 3 s samples.

This is because the hardware resource is lack of calculating the filter learning in real-time. Unlike the learning, SIMO-ICA filtering and SIMO-BM can be conducted just in the current input signals. Thus, total system of this module can work in real-time.

We connect the earphone-microphone to this BSS module, and as a result, the whole system realizes real-time hearing-aid system. In the following experiments, the examinee actually wears this earphone-microphone and BSS module.

5.3. Results and discussion

Figure 5 shows the results of NRR for different θ_1 and θ_2. These are the averaged scores for all speaker combinations. From Fig. 5, it is revealed that interference reduction performance of binary masking is overall low. In case of ICA, for $(\theta_1, \theta_2) = (-45°, 45°)$, the performance is comparatively high. This is because the initial value of ICA is steered to $(-45°, 45°)$, so that ICA's filter learning is easily than the other directions. On the other hand, NRR of proposed method is higher than those of the conventional methods. Thus, these objective evaluation experiments, we can confirm that the performance of output signals of the propose system is superior to that of the conventional methods.

In order to confirm that the output signals maintain the spatial qualities of each sound source, we carried out the subjective evaluation experiments by 9 users. Figure 6 shows the relations (confusion matrix) between perceptual directions and target directions using (a) real SIMO-model-based signals (no interference signals), (b) output signals of binary masking which are convolved with HRTF w.r.t. the estimated target direction, (c) output signals of conventional ICA which are convolved with HRTF w.r.t. the estimated target direction, (d) output signals of proposed method, respectively. On this experiment, in case that the examinee can answer all correct directions, the confusion matrix shows a diagonal line with the same radius circles. From Fig. 6, we can confirm the following facts: (I) Conventional methods (b) and (c) show the perceptual confusions, especially in the right lateral angle, due to the mismatch among HRTFs of database (dummy head) and each users. (II) The sound image in the proposed system is more identical with that in real SIMO-model-based signals, compared with the conventional methods.

Overall it is asserted that the proposed method can reduce the interference component of the mixed binaural signals and reproduce the target component while keeping information about the directivity and spatial qualities of target source. In addition, it should be emphasized that the proposed system can work without measuring any user-specific HRTFs; this contributes toward the great reduction of costs for fitting hearing-aid systems to the user.

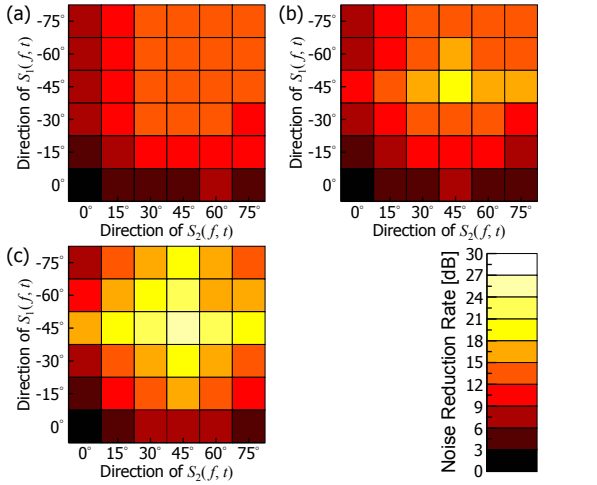

Fig. 5. Experimental results of NRR in (a) binary masking, (b) ICA, (c) proposed method.

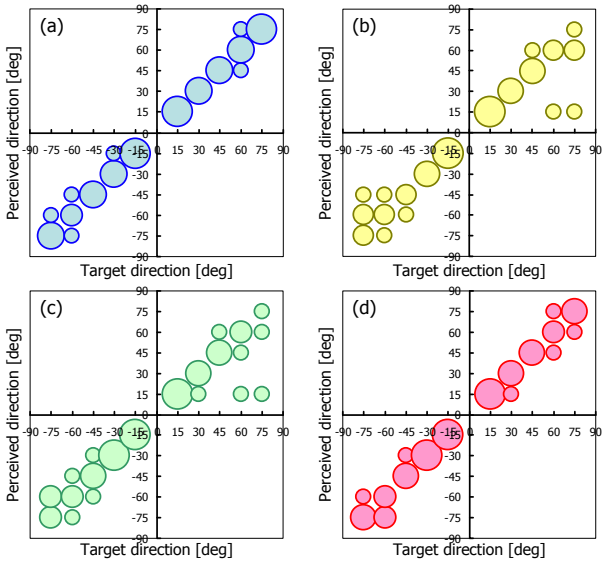

Fig. 6. Confusion matrices of subjective evaluation: (a) real binaural signals (non mixed signals), (b) output signals of binary masking, (c) output signals of ICA, (d) output signals of proposed method.

6. CONCLUSION

In this paper, we propose high-presence hearing-aid system which can extract and reproduce the target component from mixed binaural sounds using pocket-size real-time BSS module. The proposed system contains SIMO-ICA pert and SIMO-BM part, and decompose the mixed binaural sounds into binaural sounds w.r.t. each sound source. To evaluate its effectiveness, separation experiments are carried out under a reverberant condition for different users. The objective and subjective experimental results reveal the performance of the proposed method is superior to that of the conventional methods.

7. REFERENCES

[1] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. ICA and BSS (ICA99)*, pp. 365–371, 1999.

[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] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, and T. Nishikawa, "Blind source separation combining independent component analysis and beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2003, pp. 1135–1146, 2003.

[4] R. Lyon, "A computational model of binaural localization and separation," in *Proc. ICASSP83*, pp. 1148–1151, 1983.

[5] 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," in *Proc. IWAENC2003*, pp. 251–254, 2003.

[6] Y. Mori, T. Takatani, H. Saruwatari, K. Shikano, T. Hiekata, and T. Morita, "Two-stage blind separation of moving sound sources with pocket-size real-time dsp module," in *Proc. EUSIPCO2006*, 2006.

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