BLIND SOURCE SEPARATION COMBINING SIMO-ICA AND SIMO-MODEL-BASED BINARY MASKING

Yoshimitsu Mori†, Tomoya Takatani†, Hiroshi Saruwatari†, Takashi Hiekata‡, Takashi Morita‡

† Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192, Japan (e-mail: yoshim-m@is.naist.jp)
‡ Kobe Steel,Ltd.
Kobe, 651-2271, Japan (e-mail: t-hiekata@kobelco.jp)

ABSTRACT

A new two-stage blind source separation (BSS) for convolutive mixtures of speech is proposed, in which a Single-Input Multiple-Output (SIMO)-model-based ICA and a new SIMO-model-based binary mask processing are combined. SIMO-model-based ICA can separate the mixed signals, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Thus, the separated signals of SIMO-model-based ICA can maintain the spatial qualities of each sound source. Owing to the attractive property, novel SIMO-model-based binary mask processing can be applied to efficiently remove the residual interference components after SIMO-model-based ICA. The experimental results reveal that the separation performance can be considerably improved by using the proposed method compared with the conventional BSS methods.

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. Basically BSS is classified into *unsupervised* filtering technique, and much attention has been paid to BSS in many fields of signal processing.

In recent researches of BSS based on independent component analysis (ICA), various methods have been presented for acoustic-sound separation [1, 2, 3]. This paper also addresses the BSS problem under highly reverberant conditions which often arise in many practical audio applications. The separation performance of the conventional ICA is far from being sufficient in the reverberant case because too long separation filters is required but the unsupervised learning of the filter is not so easy. Therefore, one possible improvement is to partly combine ICA with another signal enhancement technique, but in the conventional ICA, each of the separated outputs is a *monaural* signal, and this leads to the drawback that many kinds of superior *multichannel* techniques cannot be applied.

In order to attack the tough problem, we propose a novel two-stage BSS algorithm which is applicable to an array of directional microphones. This approach resolves the BSS problem into two stages: (a) a Single-Input Multiple-Output (SIMO)-model-based ICA proposed by the authors [4] and (b) new SIMO-model-based binary mask processing for the SIMO signals obtained from the preceding SIMO-model-based ICA. 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-model-based ICA can separate the mixed signals, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Thus, the separated signals of SIMO-model-based ICA can maintain rich spatial qualities of each sound source. After the SIMO-model-based ICA, the residual components of the interference, which are often staying in the output of SIMO-model-based ICA as well as the conventional ICA, can be efficiently removed by the following binary mask processing. The experimental results reveal the proposed method's efficacy in a realistic reverberant condition.

2. MIXING PROCESS AND CONVENTIONAL BSS

2.1. Mixing Process

In this study, the number of microphones is K and the number of multiple sound sources is L, where we deal with the case of $K = L$. In the frequency domain, the observed signals in which multiple sources are mixed are given by $\boldsymbol{X}(f) = \boldsymbol{A}(f)\boldsymbol{S}(f)$, where $\boldsymbol{X}(f)$ = $[X_1(f), \cdots, X_K(f)]^{\mathrm{T}}$ is the observed signal vector, and $\boldsymbol{S}(f)$ = $[S_1(f), \cdots, S_L(f)]^{\mathrm{T}}$ 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.

2.2. Conventional ICA-Based BSS

In the frequency-domain ICA (FDICA), 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 for 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 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 $\boldsymbol{Y}(f,t) = \boldsymbol{W}(f)\boldsymbol{X}(f,t)$. We perform this procedure with respect to all frequency bins. The optimal $\boldsymbol{W}(f)$ is obtained by, e.g., the following iterative updating equation:

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

where $\boldsymbol{I}$ 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, η is the step-size parameter, and $\boldsymbol{\Phi}(\cdot)$ is the appropriate nonlinear vector function [5]. After the iterations, the source permutation and the scaling indeterminacy problem can be solved by, e.g., [1, 3].

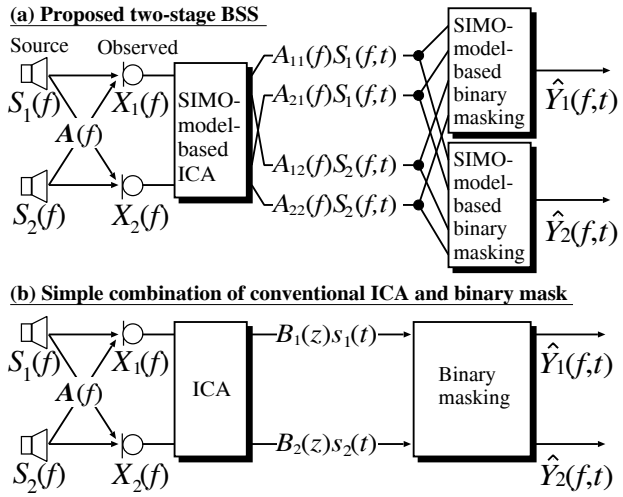

Fig. 1. Input and output relations in (a) proposed two-stage BSS and (b) simple combination of conventional ICA and binary mask processing. This corresponds to the case of $K = L = 2$.

2.3. Conventional Binary-Mask-Based BSS

Binary mask processing [6, 7] is one of the alternative approach which is aimed to solve the BSS problem, but is not based on ICA. We estimate a binary mask by comparing the amplitudes of the observed signals, and pick up the target sound component which arrives at the *better microphone* closer to the target speech. This procedure is performed in time-frequency regions, and is to pass the specific regions where target speech is dominant and mask the other regions. Under the assumption that the l-th sound source is close to the l-th microphone and $L = 2$, the l-th separated signal is given by

$$\hat{Y}_l(f,t) \quad = \quad m_l(f,t)X_l(f,t), \tag{2}$$

where $m_l(f,t)$ is the binary mask operation which is defined as $m_l(f,t) = 1$ if $X_l(f,t) > X_k(f,t)$ $(k \neq l)$; otherwise $m_l(f,t) = 0$.

This method requires very few computational complexities, and this property is well applicable to real-time processing. The method, however, needs a sparseness assumption in the sources' spectral components, i.e., there are no overlaps in time-frequency components of the sources. Indeed the assumption does not hold in an usual audio application, e.g., a mixture of speech and a common broadband stationary noise.

3. PROPOSED TWO-STAGE BSS ALGORITHM

3.1. Motivation and Strategy

In the previous research, SIMO-model-based ICA (SIMO-ICA) was proposed by, e.g., Takatani et al. [4], and they showed that the SIMO-ICA can separate the mixed signals into SIMO-model-based signals at the microphone points. This finding has motivated us to combine the SIMO-ICA and a *new* binary masking strategy, so called *SIMO-model-based binary masking* (SIMO-BM). That is, the masking function is determined by whole information of the SIMO components of all sources obtained from SIMO-ICA. The configuration of the proposed method is shown in Fig. 1(a). SIMO-BM which follows SIMO-ICA can remove the residual component of the interference effectively without adding huge computational complexities.

It is worth mentioning that the novelty of this strategy mainly lies in the two-stage idea of the unique combination of SIMO-ICA and the SIMO-model-based binary mask. To illustrate the novelty of the proposed method, we hereinafter compare the proposed combination with a simple two-stage combination of a conventional monaural-output ICA and the conventional binary masking (see Fig. 1(b)).

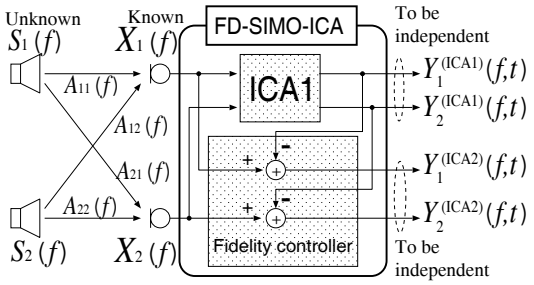

Fig. 2. Input and output relations in the proposed FD-SIMO-ICA, where $K = L = 2$.

In general, the conventional ICAs can only supply the source signals $Y_l(f,t) = B_l(f)S_l(f,t) + E_l(f,t)$ $(l = 1, \cdots, L)$, where $B_l(f)$ is an unknown arbitrary distortion filter and $E_l(f,t)$ is a residual separation error which is mainly caused by an insufficient convergence in ICA. The residual error $E_l(f,t)$ should be removed by binary mask processing in the next post-processing stage. However, the combination is very problematic and cannot function well because of the existence of the spectral overlaps in the time-frequency domain. For instance, if all sources have nonzero spectral components (i.e., sparseness assumption does not hold) in the specific frequency subband and these are comparable, the decision in binary mask processing for $Y_1(f,t)$ and $Y_2(f,t)$ is vague and the output results in a ravaged (very distorted) signal. Thus the simple combination of the conventional ICA and binary mask processing is not valid for solving the BSS problem.

On the other hand, our proposed combination contains the special SIMO-ICA in the first stage, where the SIMO-ICA can supply the specific SIMO signals with respect to each of sources, $A_{kl}(f)S_l(f,t)$, up to the possible delay of the filters and the residual error. The obtained SIMO components is very beneficial to the decision of the masking function. For example, the binary masking between $A_{11}(f)S_1(f,t)$ and $A_{21}(f)S_1(f,t)$ is more acoustically reasonable rather than the conventional combination because the spatial properties that the separated SIMO component at the specific microphone closer to the target sound still maintains the large gain, are kept. Thus, after having the SIMO components, we can introduce the SIMO-BM for the efficient reduction of the remaining error in ICA, even when the sparseness assumption does not hold.

3.2. Algorithm: SIMO-ICA in the 1st Stage

Time-domain *SIMO-ICA* [4] has recently been proposed by one of the authors as a means of obtaining SIMO-model-based signals directly in the ICA updating. In this paper, we extend the time-domain SIMO-ICA to frequency-domain SIMO-ICA (FD-SIMO-ICA). FD-SIMO-ICA is conducted for extracting the SIMO-model-based signals corresponding to each of sources. The FD-SIMO-ICA consists of $(L-1)$ FDICA parts and a *fidelity controller*, and each ICA runs in parallel under the fidelity control of the entire separation system (see Fig. 2). The separated signals of the l-th ICA $(l = 1, \cdots L-1)$ in FD-SIMO-ICA are defined by

$$\boldsymbol{Y}_{(\mathrm{ICA}l)}(f,t) = [\,Y_k^{(\mathrm{ICA}l)}(f,t)]_{k1} = \boldsymbol{W}_{(\mathrm{ICA}l)}(f)\boldsymbol{X}(f,t), \tag{3}$$

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

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

$$\boldsymbol{Y}_{(\mathrm{ICA}L)}(f,t) \quad = \quad \boldsymbol{X}(f,t) - \sum_{l=1}^{L-1} \boldsymbol{Y}_{(\mathrm{ICA}l)}(f,t). \qquad (4)$$

Hereafter, we regard $\boldsymbol{Y}_{(\mathrm{ICA}L)}(f,t)$ as an output of a *virtual* "L-th" ICA. The reason we use the word "*virtual*" here is that the L-th ICA does not have own separation filters unlike the other ICAs, and $\boldsymbol{Y}_{(\mathrm{ICA}L)}(f,t)$ is subject to $\boldsymbol{W}_{(\mathrm{ICA}l)}(f)$ $(l = 1, \cdots, L-1)$. By transposing the second term $(-\sum_{l=1}^{L-1} \boldsymbol{Y}_{(\mathrm{ICA}l)}(f,t))$ in the right-hand side into the left-hand side, we can show that (4) means a constraint to force the sum of all ICAs' output vectors $\sum_{l=1}^{L} \boldsymbol{Y}_{(\mathrm{ICA}l)}(f,t)$ 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, up to the permutation, as $\boldsymbol{Y}_{(\mathrm{ICA}l)}(f,t) = \mathrm{diag}\big[\boldsymbol{A}(f)\boldsymbol{P}_l^{\mathrm{T}}\big]\boldsymbol{P}_l\boldsymbol{S}(f,t)$, where $\boldsymbol{P}_l$ $(l = 1, \cdots, L)$ are exclusively-selected permutation matrices which satisfy $\sum_{l=1}^{L} \boldsymbol{P}_l = [1]_{ij}$. Regarding a proof of this, see [4] with an appropriate modification into the frequency-domain representation. 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. For example in the case of $L = K = 2$, one possibility is given by

$$\begin{aligned} &\big[Y_1^{(\mathrm{ICA}1)}(f,t),\ Y_2^{(\mathrm{ICA}1)}(f,t)\big]^{\mathrm{T}} \\ &\quad = \quad \big[A_{11}(f)S_1(f,t),\ A_{22}(f)S_2(f,t)\big]^{\mathrm{T}}, \qquad (5) \\ &\big[Y_1^{(\mathrm{ICA}2)}(f,t),\ Y_2^{(\mathrm{ICA}2)}(f,t)\big]^{\mathrm{T}} \\ &\quad = \quad \big[A_{12}(f)S_2(f,t),\ A_{21}(f)S_1(f,t)\big]^{\mathrm{T}}, \qquad (6) \end{aligned}$$

where $\boldsymbol{P}_1 = \boldsymbol{I}$ and $\boldsymbol{P}_2 = [1]_{ij} - \boldsymbol{I}$.

In order to obtain (5) and (6), the natural gradient of Kullback-Leibler divergence of (4) with respect to $\boldsymbol{W}_{(\mathrm{ICA}l)}(f)$ should be added to the existing nonholonomic iterative learning rule [1] 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 FD-SIMO-ICA is given as

$$\begin{aligned} &\boldsymbol{W}_{(\mathrm{ICA}l)}^{[j+1]}(f) \\ &\quad = \quad \boldsymbol{W}_{(\mathrm{ICA}l)}^{[j]}(f) - \alpha \Bigg[\Bigg\{ \text{off-diag} \Big\langle \boldsymbol{\Phi}\big(\boldsymbol{Y}_{(\mathrm{ICA}l)}^{[j]}(f,t)\big) \\ &\qquad \boldsymbol{Y}_{(\mathrm{ICA}l)}^{[j]}(f,t)^{\mathrm{H}} \Big\rangle_t \Bigg\} \cdot \boldsymbol{W}_{(\mathrm{ICA}l)}^{[j]}(f) \\ &\qquad - \Bigg\{ \text{off-diag} \Big\langle \boldsymbol{\Phi}\big(\boldsymbol{X}(f,t) - \sum_{l=1}^{L-1} \boldsymbol{Y}_{(\mathrm{ICA}l)}^{[j]}(f,t)\big) \\ &\qquad \cdot \Big(\boldsymbol{X}(f,t) - \sum_{l=1}^{L-1} \boldsymbol{Y}_{(\mathrm{ICA}l)}^{[j]}(f,t)\Big)^{\mathrm{H}} \Big\rangle_t \Bigg\} \\ &\qquad \cdot \Big(\boldsymbol{I} - \sum_{l=1}^{L-1} \boldsymbol{W}_{(\mathrm{ICA}l)}^{[j]}(f)\Big) \Bigg], \qquad (7) \end{aligned}$$

where α is the step-size parameter, and we define the nonlinear vector function $\boldsymbol{\Phi}(\cdot)$ as $[\tanh(|Y_l(f,t)|)e^{j\cdot\arg(Y_l(f,t))}]_{l1}$ [5]. Also, the initial values of $\boldsymbol{W}_{(\mathrm{ICA}l)}(f)$ for all l should be different.

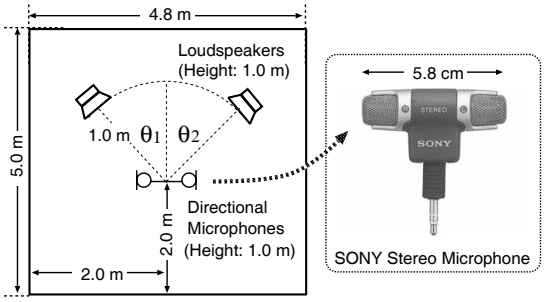

Fig. 3. Layout of reverberant room used in experiments.

3.3. Algorithm: SIMO-BM in the 2nd Stage

After FD-SIMO-ICA, SIMO-model-based binary masking processing is applied. Here we consider the case of (5) and (6). The resultant output signal corresponding to the source 1 is determined in the proposed SIMO-BM as follows:

$$\hat{Y}_1(f,t) \quad = \quad m_1(f,t)Y_1^{(\mathrm{ICA}1)}(f,t), \qquad (8)$$

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

$$\begin{aligned} &Y_1^{(\mathrm{ICA}1)}(f,t) \\ &> \max\big[|c_1 Y_2^{(\mathrm{ICA}2)}(f,t)|,\ |c_2 Y_1^{(\mathrm{ICA}2)}(f,t)|,\ |c_3 Y_2^{(\mathrm{ICA}1)}(f,t)|\big]; \end{aligned} \qquad (9)$$

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. For example, $[c_1, c_2, c_3]$ = $[0, 0, 1]$ yields the simple combination of the conventional ICA and the conventional binary mask. Otherwise if we set $[c_1, c_2, c_3]$ = $[1, 0, 0]$, we can utilize better (acoustically reasonable) SIMO information of each source as described in Sect. 3.1. If we change another pattern of c_i, we can generate various SIMO-model-based maskings with different separation and distortion properties.

The resultant output corresponding to the source 2 is given by

$$\hat{Y}_2(f,t) \quad = \quad m_2(f,t)Y_2^{(\mathrm{ICA}1)}(f,t), \qquad (10)$$

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

$$\begin{aligned} &Y_2^{(\mathrm{ICA}1)}(f,t) \\ &> \max\big[|c_1 Y_1^{(\mathrm{ICA}2)}(f,t)|,\ |c_2 Y_2^{(\mathrm{ICA}2)}(f,t)|,\ |c_3 Y_1^{(\mathrm{ICA}1)}(f,t)|\big]; \end{aligned} \qquad (11)$$

otherwise $m_2(f,t) = 0$. Also the extension to the general case of $L = K > 2$ can be easily implemented in the same manner.

4. SOUND SEPARATION EXPERIMENT

4.1. Experimental Conditions

First, to evaluate the feasibility to general hands-free applications, we carried out sound-separation experiments in a real reverberant room illustrated in Fig. 3, where two sources and two directional microphones (stereo-microphone) are set. The reverberation time in this room is 200 ms. Two speech signals are assumed to arrive from different directions, θ_1 and θ_2, where we prepare three kinds of source direction patterns as follows; $(\theta_1, \theta_2) = (-40^\circ, 30^\circ)$,

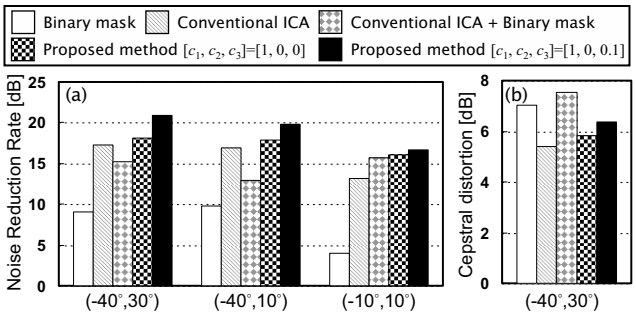

Fig. 4. (a) Results of NRR under different speaker allocations, and (b) results of CD for $(\theta_1, \theta_2) = (-40^\circ, 30^\circ)$.

$(-40^\circ, 10^\circ)$, or $(-10^\circ, 10^\circ)$. We used the speech signals spoken by two male and two female speakers as the source samples, and we generated 12 combinations of speakers. The sampling frequency is 8 kHz and the length of each sound sample is limited to 3 seconds. The DFT size of $\boldsymbol{W}(f)$ is 1024. We use a null-beamformer-based initial value [3] which is steered to $(-60^\circ, 60^\circ)$.

We basically compare the following methods: (A) the conventional binary-mask-based BSS given in Sect. 2.3, (B) the conventional ICA-based BSS given in Sect. 2.2, where the scaling ambiguity can be properly solved by [1], (C) simple combination of the conventional ICA and binary mask processing, and (D) the proposed two-stage BSS method.

4.2. Experimental Evaluation on Separation Performance

Noise reduction rate (NRR) [3], defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB, is used as the objective indication of separation performance. Figure 4(a) shows the results of NRR under different speaker allocations. These scores are the averages of 12 speaker combinations. From the results, we can confirm that the proposed two-stage BSS can improve the separation performance regardless the speaker directions, and the proposed BSS outperforms all of the conventional methods.

To assess the distortion of the separated signals, we measure a *Cepstral Distortion* (CD) which indicates the distance between the spectral envelope of the original source signal and the target component in the separated output. The CD cannot take into account the degree of interference reduction unlike NRR, and thus the CD and NRR are complementary scores. The FFT-based 10th-order cepstrum is used, and the CD is decreased as the distortion is reduced. Figure 4(b) depicts the examples of CD (average of 12 speaker combinations) under $(\theta_1, \theta_2) = (-40^\circ, 30^\circ)$. As can be confirmed, the CDs of both the conventional ICA and the proposed method are smaller than those of the binary masking and its simple combination with ICA. This means that (a) the conventional binary-mask-based methods involve a heavy distortion due to the improper time variant masking arising in the non-sparse frequency subband, (b) but the proposed method cannot be affected by such an improperness.

These results are promising evidences that the proposed combination of SIMO-ICA and SIMO-BM is well applicable to the low-distortion sound segregation, e.g., hands-free telecommunication via mobile phones.

5. SPEECH RECOGNITION EXPERIMENT

Next, to evaluate the applicability to the speech enhancement, we performed large vocabulary speech recognition experiments utilizing the proposed BSS as a preprocessing for noise reduction. Regarding the decoder, Julius[8] is used. We use a Phonetic Tied Mixture (PTM) model trained via 260 speakers selected from JNAS database. The test sets include 200 sentences.

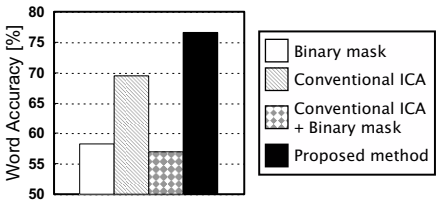

Fig. 5. Result of word accuracy for $(\theta_1, \theta_2) = (-40^\circ, 30^\circ)$.

Figure 5 indicates the results of word recognition performance (word accuracy) for each methods, where we can see the proposed method's superiority. The score of the proposed method with $[c_1, c_2, c_3]$ = $[1, 0, 0.1]$ is obviously better that those of the binary masking and its simple combination with ICA, and significantly outperforms the conventional ICA. Thus the proposed method is potentially beneficial to noise-robust speech recognition as well as the hands-free telephony.

6. CONCLUSION

We proposed a new BSS framework in which the SIMO-ICA and a new SIMO-BM are efficiently combined. In order to evaluate its effectiveness, a separation experiment was carried out under a reverberant condition. The experimental results revealed that the SNR can be considerably improved by using the proposed two-stage BSS algorithm. In addition, we could find that the proposed method outperforms the combination of the conventional ICA and binary mask processing as well as the simple ICA and binary mask processing.

7. ACKNOWLEDGMENT

The authors thank Dr. H. Sawada, Mr. R. Mukai, and Mrs. S. Araki of NTT CS-lab. for their kind discussion on this work. This work was partly supported by CREST "Advanced Media Technology for Everyday Living" of JST in Japan.

8. REFERENCES

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

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

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

[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. IWAENC2003*, 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. Fundamentals*, vol.E86-A, no.3, pp.590–596, 2003.

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

[7] N. Roman, D. Wang and G. Brown, "Speech segregation based on sound localization," *Proc. IJCNN01*, pp.2861–2866, 2001.

[8] A. Lee, T. Kawahara and K. Shikano, "Julius – An open source real-time large vocabulary recognition engine," *Proc. EUROSPEECH*, pp.1691–1694, 2001.