

BLIND SOURCE SEPARATION COMBINING FREQUENCY-DOMAIN ICA AND BEAMFORMING

Hiroshi SARUWATARI[†], Satoshi KURITA[‡], and Kazuya TAKEDA[‡]

[†]Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0101, JAPAN
E-mail: sawatari@is.aist-nara.ac.jp

[‡]Nagoya University / Center for Integrated Acoustic Information Research (CIAIR)
Furo-cho, Chikusa-ku, Nagoya, 464-8603, JAPAN

ABSTRACT

In this paper, we describe a new method of blind source separation (BSS) on a microphone array combining subband independent component analysis (ICA) and beamforming. The proposed array system consists of the following three sections: (1) subband-ICA-based BSS section with direction-of-arrival (DOA) estimation, (2) null beamforming section based on the estimated DOA information, and (3) integration of (1) and (2) based on the algorithm diversity. Using this technique, we can resolve the low-convergence problem through optimization in ICA. The results of the signal separation experiments reveal that the noise reduction rate (NRR) of about 18 dB is obtained under the nonreverberant condition, and NRRs of 8 dB and 6 dB are obtained in the case that the reverberation times are 150 msec and 300 msec. These performances are superior to those of both simple ICA-based BSS and simple beamforming method.

1. INTRODUCTION

Blind source separation (BSS) is the approach to estimate original source signals using only the information of the mixed signals observed in each input channel. This technique is applicable to the realization of noise-robust speech recognition and high-quality hands-free telecommunication systems. In the recent works, as for the BSS based on the independent component analysis (ICA) [1], the several methods, in which the inverse of the complex mixing matrices are calculated in the frequency domain, have been proposed to deal with the arriving lags among each element of the microphone array system [2, 3]. Since the calculations are carried out at each frequency independently, the following problems arise in these methods: (1) permutation of each sound source, (2) arbitrariness of each source gain. To resolve these problems, a priori assumption of similarity among the envelopes of source signal waveforms is necessary [2].

In this paper, a new method of BSS on a microphone array combining subband ICA and beamforming is proposed. The proposed array system consists of the following three sections (see Fig. 1 for the system configuration): (1) subband ICA section, (2) null beamforming section, and (3) integration of (1) and (2). First, a new subband ICA is introduced to achieve the frequency-domain BSS on the microphone array system, where directivity patterns of the array are explicitly used to estimate each direction of arrival (DOA) of the sound sources [4]. Using this method, we can resolve both permutation and arbitrariness problems simultaneously without the assumption for the source signal waveforms.

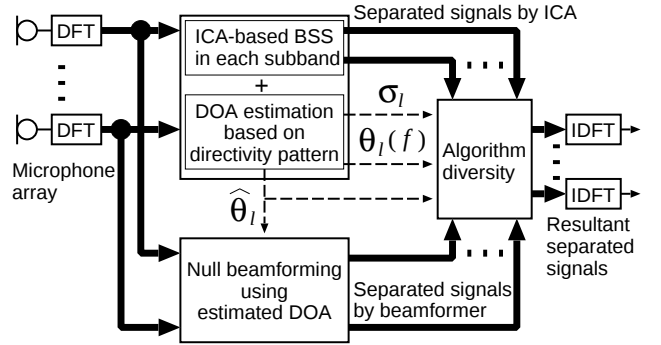


Figure 1: Configuration of proposed microphone array system based on subband ICA and beamforming.

Next, based on the DOA estimated in the above-mentioned ICA section, we construct a null beamformer, in which the directional null is steered to the direction of the undesired sound source, in parallel with the ICA-based BSS. This approach to signal separation has the advantage that there is no difficulty with respect to a low-convergence on optimization because the null beamformer is determined by only DOA information without independence between sound sources. Finally, both signal separation procedures are appropriately integrated by the algorithm diversity [5] in the frequency domain. The following sections describe the proposed method in detail, and can show that the signal separation performance of the proposed method is superior to those of both conventional beamforming and ICA-based BSS methods.

2. ALGORITHM

2.1. Subband ICA Section

In this study, a straight-line array is assumed. The coordinates of the elements are designated as d_k ($k = 1, \dots, K$), and the directions of arrival of multiple sound sources are designated as θ_l ($l = 1, \dots, L$) (see Fig. 2).

In general, the observed signals in which multiple source signals are mixed linearly are given by the following equation in the frequency domain:

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \quad (1)$$

where $\mathbf{X} = [X_1(f), \dots, X_K(f)]^T$ is the observed signal vector, and $\mathbf{S} = [S_1(f), \dots, S_L(f)]^T$ is the source signal vector. $\mathbf{A}$ is the

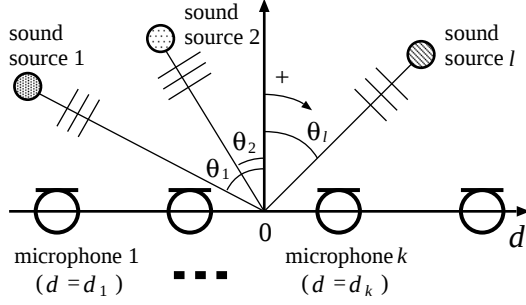


Figure 2: Configuration of microphone array and signals.

mixing matrix which is assumed to be complex-valued because we introduce the model to deal with the arriving lags among each of the element of the microphone array and room reverberations.

We perform the signal separation by using the complex-valued unmixing matrix, $\mathbf{W}$, so that the each element in the output $\mathbf{Y} = \mathbf{W}\mathbf{X}$ becomes mutually independent in the case of $K = L$. The optimal $\mathbf{W}$ can be obtained by using the following iterative equation [4, 6]:

$$\mathbf{W}_{i+1} = \eta \left(\text{diag} \left(\langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle \right) - \langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle \right) (\mathbf{W}_i^H)^{-1} + \mathbf{W}_i \quad (2)$$

where $\langle \cdot \rangle$ denotes the averaging operator, i is used to express the value of the i th step in the iterations, and η is the step size parameter. Also, we define the nonlinear vector function $\Phi(\cdot)$ as

$$\Phi(\mathbf{Y}) = 1 / \{1 + \exp(-\mathbf{Y}^{(R)})\} + j \cdot 1 / \{1 + \exp(-\mathbf{Y}^{(I)})\}, \quad (3)$$

where $\mathbf{Y}^{(R)}$ and $\mathbf{Y}^{(I)}$ are the real and imaginary parts of $\mathbf{Y}$, respectively.

Since the above-mentioned calculations are carried out at each frequency independently, problems about the source permutation and scaling indeterminacy arise at every frequency bin. In order to resolve the problems, we have already provided the solution [4] to utilize the directivity pattern of the array system, $F_l(f, \theta)$, which is given by

$$F_l(f, \theta) = \sum_{k=1}^K W_{lk}(f) \exp[j2\pi f d_k \sin \theta / c], \quad (4)$$

where c is the velocity of sound. Hereafter we assume the two-channel case without loss of generality, i.e., $K = L = 2$. In the directivity patterns, directional nulls exist in only two particular directions. Accordingly, by obtaining statistics with respect to the directions of nulls at all frequency bins, we can estimate the DOAs of the sound sources. The DOA of the l th sound source, $\hat{\theta}_l$, can be estimated as

$$\hat{\theta}_l = \frac{2}{N} \sum_{m=1}^{N/2} \theta_l(f_m), \quad (5)$$

where N is a total point of DFT, and $\theta_l(f_m)$ represents the DOA of the l th sound source at the m th frequency bin. These are given by

$$\theta_1(f_m) = \min[\arg\min_{\theta} |F_1(f_m, \theta)|, \arg\min_{\theta} |F_2(f_m, \theta)|], \quad (6)$$

$$\theta_2(f_m) = \max[\arg\min_{\theta} |F_1(f_m, \theta)|, \arg\min_{\theta} |F_2(f_m, \theta)|], \quad (7)$$

where $\min[x, y]$ ($\max[x, y]$) is defined as a function in order to obtain the smaller (larger) value among x and y . Based on these DOA informations, we can detect and correct the source permutation and the gain inconsistency.

2.2. Beamforming Section

In the beamforming section, we can construct an alternative unmixing matrix in parallel based on the null beamforming technique where the DOA information obtained in the ICA section is used. In the case that the look direction is $\hat{\theta}_1$ and the directional null is steered to $\hat{\theta}_2$, the elements of the unmixing matrix are given as

$$W_{11}^{(\text{BF})}(f_m) = \exp[-j2\pi f_m d_1 \sin \hat{\theta}_1 / c] \times \left\{ \exp[j2\pi f_m d_1 (\sin \hat{\theta}_2 - \sin \hat{\theta}_1) / c] - \exp[j2\pi f_m d_2 (\sin \hat{\theta}_2 - \sin \hat{\theta}_1) / c] \right\}^{-1}, \quad (8)$$

$$W_{12}^{(\text{BF})}(f_m) = -\exp[-j2\pi f_m d_2 \sin \hat{\theta}_1 / c] \times \left\{ \exp[j2\pi f_m d_1 (\sin \hat{\theta}_2 - \sin \hat{\theta}_1) / c] - \exp[j2\pi f_m d_2 (\sin \hat{\theta}_2 - \sin \hat{\theta}_1) / c] \right\}^{-1}. \quad (9)$$

Also in the case that the look direction is $\hat{\theta}_2$ and the directional null is steered to $\hat{\theta}_1$, the elements of the unmixing matrix are given as

$$W_{21}^{(\text{BF})}(f_m) = -\exp[-j2\pi f_m d_1 \sin \hat{\theta}_2 / c] \times \left\{ -\exp[j2\pi f_m d_1 (\sin \hat{\theta}_1 - \sin \hat{\theta}_2) / c] + \exp[j2\pi f_m d_2 (\sin \hat{\theta}_1 - \sin \hat{\theta}_2) / c] \right\}^{-1}, \quad (10)$$

$$W_{22}^{(\text{BF})}(f_m) = \exp[-j2\pi f_m d_2 \sin \hat{\theta}_2 / c] \times \left\{ -\exp[j2\pi f_m d_1 (\sin \hat{\theta}_1 - \sin \hat{\theta}_2) / c] + \exp[j2\pi f_m d_2 (\sin \hat{\theta}_1 - \sin \hat{\theta}_2) / c] \right\}^{-1}. \quad (11)$$

These elements given by Eqs. (8)–(11) are normalized so that the each gain for look direction is set to be 1.

2.3. Integration of Subband ICA with Null Beamforming

In order to integrate the subband ICA with null beamforming, we newly introduce the following strategy for selecting the most suitable unmixing matrix in each frequency bin, i.e., algorithm diversity in the frequency domain. (1) If the directional null is steered to the proper estimated DOA of the undesired sound source, we use the unmixing matrix obtained by the subband ICA, $W_{lk}^{(\text{ICA})}(f)$. (2) If the directional null deviates from the estimated DOA, we use the unmixing matrix obtained by the null beamforming, $W_{lk}^{(\text{BF})}(f)$, in preference to that of the subband ICA. The above strategy yields the following algorithm:

$$W_{lk}(f) = \begin{cases} W_{lk}^{(\text{ICA})}(f), & (|\theta_l(f) - \hat{\theta}_l| < h \cdot \sigma_l) \\ W_{lk}^{(\text{BF})}(f), & (|\theta_l(f) - \hat{\theta}_l| \geq h \cdot \sigma_l), \end{cases} \quad (12)$$

where h is a magnification parameter of the threshold, and σ_l represents the deviation with respect to the estimated DOA of the l th sound source; it can be given as

$$\sigma_l = \sqrt{\frac{2}{N} \sum_{m=1}^{N/2} (\theta_l(f_m) - \hat{\theta}_l)^2}. \quad (13)$$

Table 1: Analysis Conditions of Signal Separation

Sampling Frequency	8 kHz
Frame Length	32 msec
Frame Shift	16 msec
Window	Hamming window
Number of Iterations	500
Step Size Parameter	$\eta = 1.0 \times 10^{-4}$

Using the algorithm with an adequate value of h , we can recover the unmixing matrix trapped on a local minimizer of the optimization procedure in ICA. Also, by changing the parameter h , we can construct various types of array signal processing for BSS, e.g., a simple null beamforming with $h = 0$, and a simple ICA-based BSS procedure with $h = \infty$.

3. EXPERIMENTS AND RESULTS

3.1. Conditions for Experiments

A two-element array with the interelement spacing of 4 cm is assumed. The speech signals are assumed to arrive from two directions, -30° and 40° . Six sentences spoken by six male and six female speakers selected from the ASJ continuous speech corpus for research are used as the original speech. Using these sentences, we obtain 36 combinations with respect to speakers and source directions. In these experiments, we used the following signals as the source signals: (1) the original speech not convolved with the impulse responses, and (2) the original speech convolved with the impulse responses recorded in two environments specified by different reverberation times (RTs), 150 msec and 300 msec. The impulse responses are recorded in a variable reverberation time room as shown in Fig. 3. The analysis conditions of these experiments are summarized in Table 1.

3.2. Results 1: Effectiveness of Algorithm Diversity

In order to illustrate the behavior of the proposed array for different values of h , the *noise reduction rate* (NRR), defined as the output signal-to-noise ratio (SNR) in dB minus input SNR in dB, is shown in Figs. 4–6. These values are taken as the average of all of the combinations with respect to speakers and source directions. The SNRs correspond to the objective evaluation score in the case that the suppressed signal is regarded as noise.

From Fig. 4 for the nonreverberant tests, it can be seen that the NRRs monotonically increase as the parameter h decreases, i.e., the performance of the null beamformer is superior to that of ICA-based BSS. This indicates that the directions of the sound sources are estimated correctly by the proposed method, and thus the null beamforming technique is more suitable for the separation of directional sound sources under nonreverberant condition.

In contrast, from Figs. 5 and 6 for the reverberant tests, it is shown that (1) the NRR monotonically increases as the parameter h decreases in the case that the observed signals of 1 sec duration are used to learn the unmixing matrix, and (2) we can obtain the optimum performances by setting the appropriate value of h , e.g., $h = 2$, in the case that the learning durations are 3 and 5 sec. We can summarize from these results that the proposed combination algorithm of ICA and null beamforming is effective for the signal separation, particularly under the reverberant conditions.

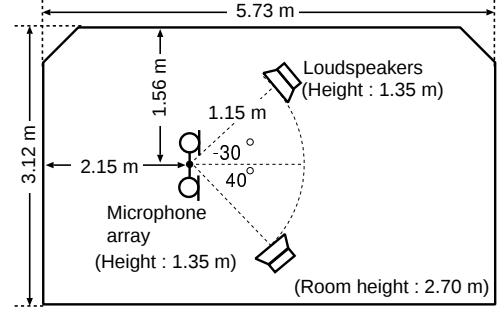
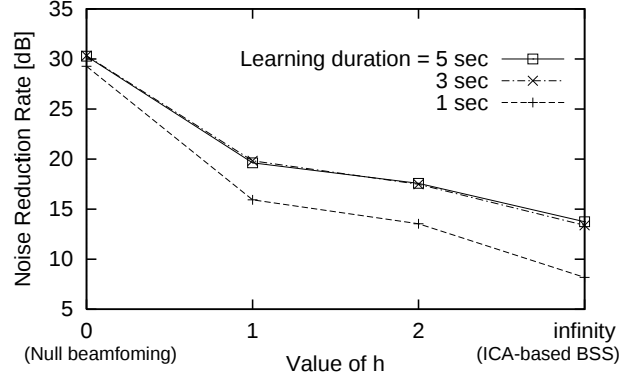
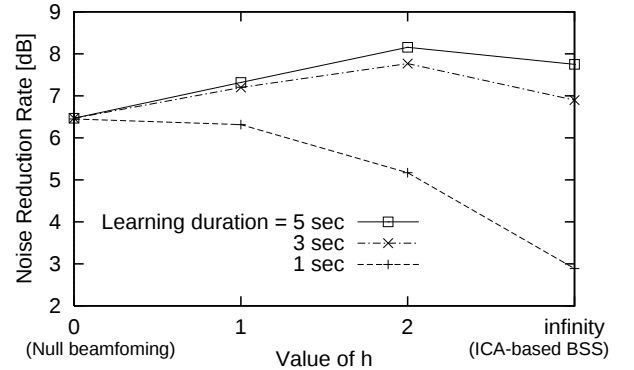
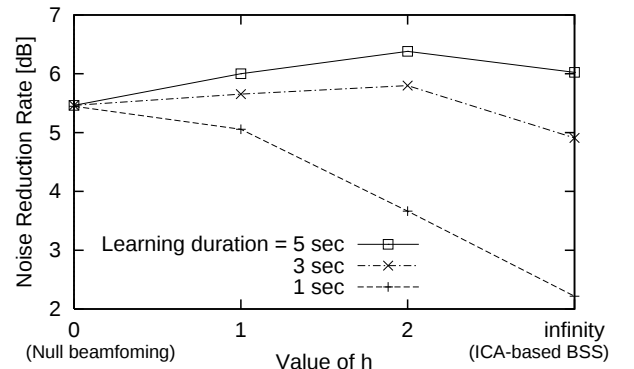


Figure 3: Layout of reverberant room used in experiments.

Figure 4: Noise reduction rates for different values of threshold parameter h . Reverberation time is 0 msec.Figure 5: Noise reduction rates for different values of threshold parameter h . Reverberation time is 150 msec.Figure 6: Noise reduction rates for different values of threshold parameter h . Reverberation time is 300 msec.

3.3. Results 2: Comparison with Conventional BSS Method

In order to perform a comparison with the conventional BSS method, we also perform the same BSS experiments using Murata's method [2]. Figure 7 (a) shows the results obtained using the proposed method and Murata's method where the observed signals of 5 sec duration are used to learn the unmixing matrix, Fig. 7 (b) shows those of 3 sec duration, and Fig. 7 (c) shows those of 1 sec duration. In these experiments, the parameter h in the proposed method is set to be 2.

From Figs. 7 (a)–(c), in both nonreverberant and reverberant tests, it can be seen that the BSS performances obtained by using the proposed method are the same as or superior to those of Murata's conventional method. In particular, from Fig. 7 (c), it is evident that the NRRs of Murata's method degrade remarkably in the case that the learning duration is 1 sec; however, there are no significant degradations in the case of the proposed method compared with those of Murata's method. We can summarize the main reasons for the degradations in Murata's method by looking at the similarity (e.g., *cosine distance*) among the source signals of different lengths as follows (see Fig. 8). (1) The envelopes of the original source speech become more similar to each other as the duration of the speech shortens. (2) The separated signals' envelopes at the same frequency are similar to each other since the inaccurate unmixing matrix is estimated to have many components of cross talk. Therefore, the recovery of the permutation tends to fail in Murata's method. In contrast, our method did not fail to recover the source permutation because we did not use any information of signal waveforms but rather, used only the directivity patterns.

4. CONCLUSION

In this paper, a new blind source separation (BSS) method using subband independent component analysis (ICA) and beamforming was described. In order to evaluate its effectiveness, the signal separation experiments were performed under various reverberant conditions. From the results, it was shown that the noise reduction rate (NRR) of about 18 dB is obtained under the nonreverberant condition, and NRRs of 8 dB and 6 dB are obtained in the case that the reverberation times are 150 msec and 300 msec. These performances were superior to those of both simple ICA-based BSS and simple beamforming technique.

5. ACKNOWLEDGEMENT

The authors are grateful to Prof. Fumitada Itakura of Nagoya University for his suggestions and discussions on this work. This work was partly supported by a Grant-in-Aid for COE Research (No. 11CE2005) and CREST (Core Research for Evolutional Science and Technology) in Japan.

6. REFERENCES

- [1] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol.7, pp.1129–1159, 1995.
- [2] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. of 1998 International Symposium on Nonlinear Theory and Its Application (NOLTA98)*, pp.923–926, 1998.
- [3] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol.22, pp.21–34, 1998.

- [4] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. ICASSP2000*, pp.3140–3143, 2000.
- [5] Y. Karasawa, T. Sekiguchi, and T. Inoue, "The software antenna: a new concept of kaleidoscopic antenna in multimedia radio and mobile computing era," *IEICE Trans. Commun.*, vol.E80-B, no.8, pp.1214–1217, 1997.
- [6] A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identification and blind separation of sources," *IEEE Trans. Circuits and Systems I*, vol.43, no.11, pp.894–906, 1996.

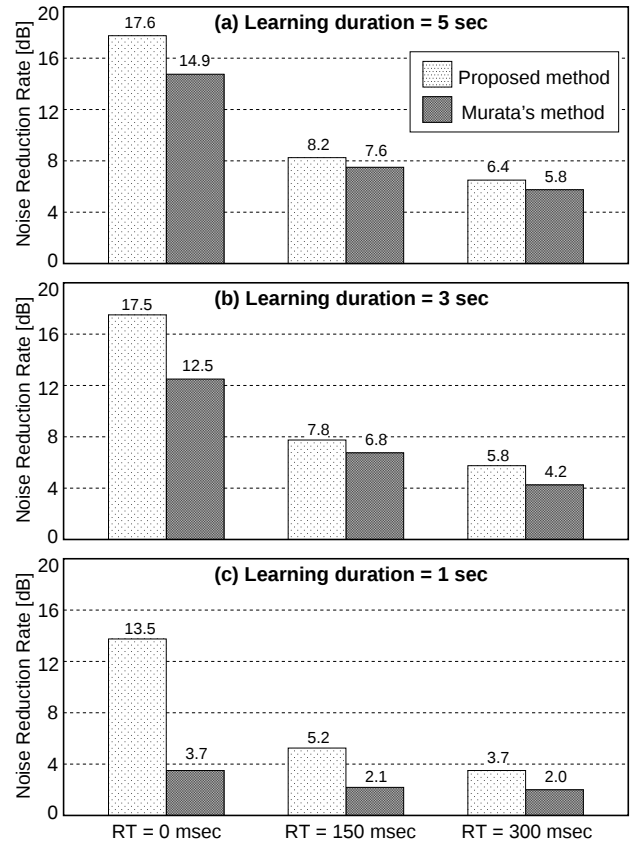


Figure 7: Comparison of noise reduction rates obtained by the proposed method ($h = 2$) and Murata's method in the case that the learning duration for ICA is (a) 5 sec, (b) 3 sec, and (c) 1 sec.

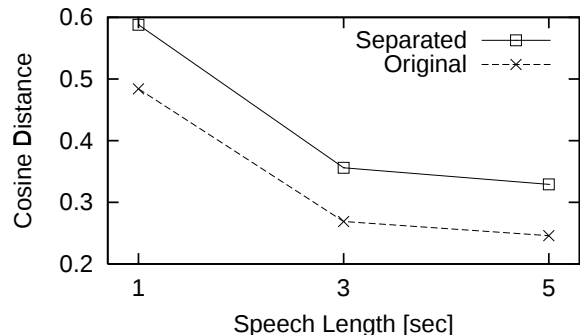


Figure 8: Cosine distances for different speech lengths. These values are the average of the all of the frequency bins.