BLIND SOURCE SEPARATION
BASED ON SUBBAND ICA AND BEAMFORMING

Hiroshi Sawatari*  Suzuki Kawai†, Kazuya Takeda†, Funitada Itakura‡, Kiyohiro Shikano*

*Graduate School of Information Science, Nara Institute of Science and Technology, JAPAN
†Nagoya University/CLAIR, JAPAN
E-mail: sawatari@ast-nara.ac.jp

ABSTRACT

This paper describes a new blind source separation (BSS) method on microphone array using the subband independent component analysis (ICA) and beamforming. The proposed array system consists of the following three sections: (1) subband-ICA based BSS section, (2) null beamforming section, and (3) integration of (1) and (2) based on the algorithm diversity. Using this technique, we can resolve the low-convergence problem on optimization in ICA. Signal separation and speech recognition experiments clarify that the noise reduction rate (NRR) of about 18 dB is obtained under the reverberation condition, and NRRs of 8 dB and 6 dB are obtained in the case that the reverberation times are 150 ms and 300 ms. These performances are superior to those of both simple ICA-based BSS and simple beamforming method. Also, the improvements of the proposed method in word recognition rates are superior to those of the conventional ICA-based BSS method under all reverberant conditions.

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 the 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 in each frequency independently, the following problems arise in these methods: (1) permutation of each source source, (2) arbitrariness of each source gain. To resolve these problems, a priori assumption of similarity among the envelopes of source signal waveforms must be required [2].

In this paper, a new BSS method on microphone array using the subband ICA and beamforming is proposed. The proposed array system consists of the following three sections (see Fig. 1 as 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 the 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. Next, based on the DOA estimated in above-mentioned ICA section, we construct a null beamformer, in which the directional null is steered to the direction of undesired sound source, in parallel with the ICA-based BSS. This approach for signal separation carries the advantage that there is no problem 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 two 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 performances 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 \(d_k (k = 1, \ldots, K)\), and the directions of arrival of multiple sound sources are designated as \(\theta_i (i = 1, \ldots, L)\) 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:

\[ X = AS. \]  \hspace{1cm} (1)
where $X = [X_1(f), \ldots, X_n(f)]^T$ is the observed signal vector, and $S = [S_1(f), \ldots, S_l(f)]^T$ is the source signal vector. $A$ is the mixing matrix which is assumed to be complex-valued because we introduce the model to deal with the arriving bags among each element of the microphone array.

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

$$W_{t+1} = \Psi_t(\Phi(Y)Y^*) - \Phi(Y)Y^*)[W_t]^{-1},$$

(2)

where $\Psi_t$ denotes the averaging operator, $l$ is used to express the value of the $t$-th step in the iterations, and $\eta$ is the step size parameter. Also, we define the nonlinear vector function $\Phi(\cdot)$ as

$$\Phi(Y) = i\{\exp(-jY^{(1)}) + j\cdot i\{1 + \exp(-jY^{(1)})\}\},$$

(3)

where $Y^{(1)}$ and $Y^{(1)}$ are the real and the imaginary parts of $Y$, respectively.

Since the above-mentioned calculations are carried out in each frequency independently, problems about the source permutation and scaling indeterminacy arise in 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, $F(f, \theta)$, which is is given by

$$F(f, \theta) = \sum_{k=1}^{K} W_k(f) \exp[-j2\pi f d_k \sin \theta/c],$$

(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 only in two particular directions. Accordingly, by taking statistics with respect to directions of the nulls in all frequency bins, we can estimate the DOAs of the sound sources. The DOA of the $l$-th sound source, $\theta_l$, can be estimated as

$$\theta_l = \frac{2}{N} \sum_{n=1}^{N/2} \theta_l(f_n),$$

(5)

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

$$\theta_l(f_n) = \min_{\mu} \{g[i] - g[j] : \arg \max_{i,j} \left\{ F_l(f_n, \theta) \right\} \}. \quad \text{(6)}$$

$$\theta_l(f_n) = \max_{\mu} \{g[i] - g[j] : \arg \max_{i,j} \left\{ F_l(f_n, \theta) \right\} \}. \quad \text{(7)}$$

where $\min[x, y]$ $(\max[x, y])$ is defined as a function to obtain the smaller (larger) value among $x$ and $y$. Based on these DOA infromations, 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 $\theta_1$ and directional null is steered to $\theta_2$, the elements of the unmixing matrix are given as

$$W^{(1)}_{12}(f_n) = \exp[-j2\pi f_n d_1 \sin \theta_1/c]$$

$$\times \{ \exp[j2\pi f_n d_2 (\sin \theta_1 - \sin \theta_2)/c] - \exp[j2\pi f_n d_2 (\sin \theta_1 + \sin \theta_2)/c] \}, \quad \text{(8)}$$

$$W^{(1)}_{21}(f_n) = -\exp[-j2\pi f_n d_2 (\sin \theta_1 - \sin \theta_2)/c]$$

$$\times \{ \exp[j2\pi f_n d_1 (\sin \theta_1 + \sin \theta_2)/c] - \exp[j2\pi f_n d_1 (\sin \theta_1 - \sin \theta_2)/c] \}, \quad \text{(9)}$$

Also in the case that the look direction is $\theta_2$ and directional null is steered to $\theta_1$, the elements of the unmixing matrix are given as

$$W^{(1)}_{21}(f_n) = \exp[-j2\pi f_n d_2 (\sin \theta_2/c)]$$

$$\times \{ \exp[j2\pi f_n d_1 (\sin \theta_2 - \sin \theta_1)/c] - \exp[j2\pi f_n d_1 (\sin \theta_2 + \sin \theta_1)/c] \}, \quad \text{(10)}$$

$$W^{(1)}_{12}(f_n) = -\exp[-j2\pi f_n d_1 (\sin \theta_2 - \sin \theta_1)/c]$$

$$\times \{ \exp[j2\pi f_n d_2 (\sin \theta_2 + \sin \theta_1)/c] - \exp[j2\pi f_n d_2 (\sin \theta_2 - \sin \theta_1)/c] \}, \quad \text{(11)}$$

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 the 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 proper estimated DOA of undesired sound source, we use the unmixing matrix obtained by the subband ICA, $W^{(SC)}_{k,l}(f)$. (2) If the directional null departs from the estimated DOA, we use the unmixing matrix obtained by the null beamforming, $W_{k,l}^{(NC)}(f)$, in preference to that of the subband ICA. The above strategy yields the following algorithm:

$$W_k(f) = \begin{cases} W^{(SC)}_{k,l}(f), & (|\theta_k(f) - \theta_l| < h \cdot \sigma_l) \\ W^{(NC)}_{k,l}(f), & (|\theta_k(f) - \theta_l| \geq h \cdot \sigma_l) \end{cases} \quad \text{(12)}$$
where $h$ is a magnification parameter of the threshold, and $\sigma_t$ represents the deviation with respect to the estimated DOA of the $l$th sound source; it can be given as

$$\sigma_t = \sqrt{\frac{2}{N} \sum_{m=1}^{N/2} (\hat{\theta}_m - \hat{\theta})^2}, \quad (13)$$

Using the algorithm with an adequate value of $h$, we can recover the unmixing matrix trapped on a local minimizer of optimization procedure in ICA. Also, by changing the parameter $h$, we can construct various 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, $-30^\circ$ and $30^\circ$. 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 the two environments specified by the different reverberation times (RTs), 150 m sec, and 300 m sec. The analysis conditions in these experiments are summarized in Table 1.

![Graph](image)

Table 1: Analysis Conditions in Signal Separation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling Frequency</td>
<td>8 kHz</td>
</tr>
<tr>
<td>Frame Length</td>
<td>32 m sec</td>
</tr>
<tr>
<td>Frame Shift</td>
<td>16 m sec</td>
</tr>
<tr>
<td>Window</td>
<td>Rectangular window</td>
</tr>
<tr>
<td>Number of Iterations</td>
<td>500</td>
</tr>
<tr>
<td>Step Size Parameter $\eta$</td>
<td>$1.0 \times 10^{-1}$</td>
</tr>
</tbody>
</table>

![Graph](image)

Figure 3: Noise reduction rates for different threshold parameter $h$. The reverberation time is 150 m sec.

![Graph](image)

Figure 4: Comparison of the noise reduction rates obtained by proposed method ($h = 2$) and Murata's method in the case that the learning duration on ICA is (a) 5 sec, (b) 3 sec, and (c) 1 sec.

3.2. Objective Evaluation

In order to illustrate the behavior of the proposed array for the different values of $h$, the noise reduction rate (NRR), defined as output signal-to-noise ratio (SNR) in dB minus input SNR in dB, is shown in Fig. 3 on the typical reverberant tests. These values are taken the average of the whole combinations with respect to speakers and source directions.

From Fig. 3, it is shown that (1) the NRR monotonically increase 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 best performances by setting the appropriate value of $h$, e.g., $h = 2$, in the case that the learning duration is 3 and 5 sec. We can summarize form these results that the proposed combination algorithm of ICA and null beamforming is effective for the improvement of the signal separation performance.

In order to compare with the conventional BSS method, we also perform the same BSS experiments using Murata's method [2]. Fig. 4 (a) shows the results obtained by the proposed method and Murata's method where the observed signals of 5 sec duration are used to learn the unmixing matrix, Fig. 4 (b) shows those of 3 sec duration, and Fig. 4 (c) shows those of 1 sec duration. In these ord
From Figs. 4 (a) (c), in both an reverberant 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 the conventional Murata's method. In particular, from Fig. 4 (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 heavy degradations on the proposed method compared with those of Murata's method.

### 3.3. Word Recognition Test

The HMM continuous speech recognition (CSR) experiment is performed in a speaker-dependent manner. For the CSR experiment, 10 sentences spoken by one speaker are used as test data, and the monophone HMM model is trained using 140 phonetically balanced sentences. Both test and training sets are selected from the ASJ continuous speech corpus for research. The remaining conditions are summarized in Table 2.

Figure 5 shows the results of the word recognition rates under different reverberation conditions. Compared with the results of Murata’s BSS method, it is evident that the improvements of the proposed method are superior to those of the conventional ICA-based BSS method under all reverberation conditions. These results indicate that the proposed method is applicable to the speech recognition system, especially when confronted with some interfering speech.

### 4. CONCLUSION

In this paper, a new blind source separation (BSS) method using the subband independent component analysis (ICA) and beamforming was described. In order to evaluate its effectiveness, the signal separation and speech recognition experiments were performed under various reverberant conditions. From the signal separation experiments, 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. From the speech recognition experiments, it was evident that the improvements of the proposed method are superior to those of the conventional Murata's BSS method under all reverberant conditions.

### 5. ACKNOWLEDGEMENT

This work was partly supported by Grant-in-Aid for COE Research (No. 11CE2005) and CREST (Core Research for Evolutional Science and Technology) in Japan.

### 6. REFERENCES


