EVALUATION OF BLIND SIGNAL SEPARATION METHOD USING DIRECTIVITY PATTERN UNDER REVERBERANT CONDITIONS

Satoshi KURITA¹, Hiroshi SARUWATARI¹, Shoji KAJITA¹, Kazuya TAKEDA¹ and Fumitada ITAKURA¹

¹Graduate School of Engineering/CIAIR, ²Center for Information Media Studies/CIAIR, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8603, JAPAN
E-mail: kurita@itakura,nuee,nagoya-u,ac,jp

ABSTRACT

This paper describes a new blind signal separation method using the directivity patterns of the microphone array. In this method, to deal with the arriving lags among each microphone, the inverses of the mixing matrices are calculated in the frequency domain so that the separated signals are mutually independent. Since the calculations are carried out in each frequency independently, the following problems arise: (1) permutation of each sound source, (2) arbitrariness of each source gain. In this paper, we propose a new solution that directivity patterns are explicitly used to estimate the each sound source direction. As the results of signal separation experiments, it is shown that the proposed method improves the SNR of degraded speech by about 16 dB under non-reverberant condition. Also, the proposed method improves the SNR by 8.7 dB when the reverberation time is 184 ms, and by 5.1 dB when the reverberation time is 322 ms.

1. INTRODUCTION

Blind signal 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 for 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 [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, 4]. Since the calculations are carried out in 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 must be required [2, 3].

In this paper, we propose a new solution in which the directivity patterns of the microphone array system are explicitly used to estimate the each direction of the sound sources [5]. Using this method, we can resolve both permutation and arbitrariness problems simultaneously without the assumption for the source signal waveforms. In addition, we also show that the proposed method didn't cause heavy degradations of the separation performance compared with those of the previous method [2, 3] under reverberant conditions.

The rest of this paper is organized as follows. In the next section, the formulation for the general BSS problems and the principle of the proposed method is explained. In Section 3, the signal separation experiments are performed. Following discussion on the results of the experiments, we give conclusions in Section 4.

2. ALGORITHM

2.1. Blind Signal Separation on Microphone Array

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 \( \theta_l \) \((l = 1, \ldots, L)\) (see Fig. 1).

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

\[
X = AS,
\]

where \( X \) is the observed signal vector, \( S \) is the source signal vector, and \( A \) is the mixing matrix; these are given as

\[
X = \begin{bmatrix} X_1(f) \\ \vdots \\ X_K(f) \end{bmatrix},
\]

\[
S = \begin{bmatrix} S_1(f) \\ \vdots \\ S_L(f) \end{bmatrix},
\]

\[
A = \begin{bmatrix} A_{11}(f) & \cdots & A_{1L}(f) \\ \vdots & \ddots & \vdots \\ A_{K1}(f) & \cdots & A_{KL}(f) \end{bmatrix}.
\]

We introduce the model to deal with the arriving lags among each element of the microphone array. In this case, \( A_{kl}(f) \) is assumed to be complex-valued. Hereafter, for convenience, we only consider the relative lags among each element with respect to the arrival time of the wavefront of the each sound source, and neglect the pure delay between the microphone and sound source. Also, \( S \) is identically regarded as the source signals observed at the origin. By ignoring the effect of the room reverberation, we can rewrite the elements in the mixing matrix (Eq. (4)) as the following simple expression:

\[
A_{kl}(f) = \exp(j2\pi f \tau_{kl}), \quad (\tau_{kl} \equiv (1/c) d_k \sin \theta_l),
\]

where \( \tau_{kl} \) is the arriving lag with respect to \( l \) th source signal from the direction of \( \theta_l \), observed at \( k \) th microphone at the coordinate of \( d_k \). Also, \( c \) is the velocity of sound.

In this study, we perform the signal separation procedure as described below (see Fig. 2), where we deal with the case that the number of sound sources \( L \) equals to that of microphones \( K \), i.e., \( K = L \). First, the short-time analysis of observed signals is conducted by using DFT frame by frame. By plotting the spectral values in a frequency bin of one microphone input framewisely, we consider them as the time series. Also, the other inputs at the same frequency bin are dealt with the same manner. Next, we perform the signal separation by using the complex-valued inverse...
of the mixing matrix, \( W \), so that \( L \) time series output \( Y \) become mutually independent; this procedure can be given as

\[
Y = WX. \tag{6}
\]

We perform this procedure with respect to all frequency bins. Finally, by applying the inverse DFT and the overlap-add technique to the separated time series \( Y \), we reconstruct the resultant source signals in the time domain.

As for the calculation of the inverse of the mixing matrix, \( W \), we use the optimization algorithm based on the minimization of the Kullback-Leibler divergence; this algorithm has been introduced by Murata and Ikeda for an on-line learning [2, 3] and modified by the authors for an off-line learning with the stable convergence [5]. The optimal \( W \) is obtained by using the following iterative equation:

\[
W_{i+1} = \eta \left( \text{diag} \left( \Phi(Y)^H \right) - \Phi(Y)^H \right) \left( W_i^H \right)^{-1} + W_i, \tag{7}
\]

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

\[
\Phi(Y) = \frac{1}{1 + \exp(-Y^{(R)})} + j \frac{1}{1 + \exp(-Y^{(I)})}, \tag{8}
\]

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

### 2.2. Signal Reconstruction

In this section, we describe the problems which arise after signal separation shown in Section 2.1, and the solution for these problems are newly proposed. Hereafter, we assume the two-channel model without loss of generality, i.e., \( K = L = 2 \).

#### 2.2.1. Problems of Source Permutation and Gain Arbitrariness

We assume that the following separation has been completed in a frequency bin \( f \):

\[
\begin{bmatrix}
\hat{S}_1(f) \\
\hat{S}_2(f)
\end{bmatrix} = \begin{bmatrix}
W_{11}(f) & W_{12}(f) \\
W_{21}(f) & W_{22}(f)
\end{bmatrix} \begin{bmatrix}
X_1(f) \\
X_2(f)
\end{bmatrix}, \tag{9}
\]

where \( X_1(f) \) and \( X_2(f) \) are the components of the observed signals at the frequency bin \( f \), \( \hat{S}_1(f) \) and \( \hat{S}_2(f) \) are the components of the estimated source signals, and \( W_{ik}(f) \) represents the element of the unmixing matrix \( W \). Since the above calculations are carried out in each frequency independently, the following two problems arise (see Fig. 3).

### Problem 1: The permutation of the source signals \( \hat{S}_1(f) \) and \( \hat{S}_2(f) \) arises. That is, the separated signal components are possible to be permuted every frequency bin, e.g., in a frequency bin of \( f = f_1 \), \( \hat{S}_1(f_1) = S_1(f_1) \) and \( \hat{S}_2(f_1) = S_2(f_1) \), and in another frequency bin of \( f = f_2 \), \( \hat{S}_1(f_2) = S_2(f_2) \) and \( \hat{S}_2(f_2) = S_1(f_2) \).

### Problem 2: The gains of \( \hat{S}_1(f) \) and \( \hat{S}_2(f) \) are arbitrary. That is, different gains are obtained at the different frequency bins \( f = f_1 \) and \( f = f_2 \).
of \( \hat{S}_1 (\hat{S}_2) \) on the right (left) hand side of this figure. From this constraint, we replace \( F_1 (f_2, \theta) \) with \( F_2 (f_2, \theta) \) at the frequency bin of \( f = f_2 \) (see Fig. 4). By performing this procedure, we can resolve the Problem 1.

Step 4: Problem 2 is resolved by normalizing the directivity patterns by the gain in each source direction after the classification (see Fig. 4). In Fig. 4, \( \alpha_1 \) and \( \alpha_2 \) are the constants which normalize the gain in the direction of \( \hat{S}_1 \) and \( \beta_1 \) and \( \beta_2 \) are the constants which normalize the gain in the direction of \( \hat{S}_2 \). Without the permutations of the sources, \( \alpha_m \) and \( \beta_m \) are described as

\[
\alpha_m = 1 / F_1 (f_m, \hat{\theta}_m), \quad \beta_m = 1 / F_2 (f_m, \hat{\theta}_m), \quad (12)
\]

where \( \hat{\theta}_m \) is the estimated direction of the \( m \)th source. If the source permutation was detected in the previous Step 3, we change \( \alpha_m \) with \( \beta_m \).

By substituting \( W \) after the above-mentioned modification for Eq (9) and applying inverse DFT to the outputs \( \hat{S}_1 (f) \) and \( \hat{S}_2 (f) \), we can obtain the source signals correctly.

3. EXPERIMENTS

3.1. Conditions for Experiments

Signal separation experiments were conducted using the sound data convolved with the impulse responses recorded in the six environments specified by the different reverberation times (RTs). In these experiments, we investigated the performance of separation under the different reverberant conditions.

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 \(+40^\circ\). Two sentences spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research are used as the original speech. Using these sentences, we obtain twelve 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, (2) the original speech convolved with the impulse responses recorded in the six environments specified by the different reverberation times. Hereafter, we designate the experiments using the signals described in (1) as the non-reverberant tests, and those of (2) as the reverberant tests. The impulse responses are recorded in the room shown in Fig. 5. The reverberation times of the impulse responses recorded in the Room 1 and Room 2 are 184 ms and 322 ms, respectively. Also the impulse responses whose reverberation times are 198 ms, 218 ms, 248 ms and 264 ms are recorded in the Room 1 each other. The remaining conditions of the rooms are summarized in Table 1. The analysis conditions in these experiments are shown in Table 2.

3.2. Alternative Method for Comparison

In order to compare with the proposed method, we also performed the BSS experiment using the alternative method proposed by Murata et al. [2, 3] with the modification for an off-line learning.

Our proposed method is based on the utilization of directivity patterns, in contrast, Murata's method is based on the utilization of \( W^{-1} \) for the normalization of gains, and a priori assumption of similarity among the envelopes of source signal waveforms for the recovery of the source permutation. In this method, the following operations are performed:

\[
Z = [Z_1 (f), \ldots, Z_L (f)]^T = WX, \quad (13)
\]

\[
\hat{S} (f, l) = W^{-1} (0, \ldots, 0, Z_i (f), 0, \ldots, 0)^T, \quad (14)
\]

where \( \hat{S} (f, l) \) denotes the component of the \( l \)th estimated source signal in the frequency bin of \( f \). By using both \( W \) and \( W^{-1} \), the gain arbitrariness vanishes in the separation procedure. Also, the source permutation can be detected and recovered by measuring the similarity among the envelopes of \( \hat{S} (f, l) \) between the different frequency bins.

3.3. Results and Discussion

In order to illustrate the behavior of the proposed array for the different RTs, the noise reduction rate (NRR), defined as output
signal-to-noise ratio (SNR) in dB minus input SNR in dB, is shown in Fig. 6. These values are taken the average of the whole combinations with respect to speakers and source sentences. SNRs correspond to the objective evaluation score in the case that the suppressed signal is regarded as the noise. In this figure, the keys, "-30" and "40", represent the NRRs of the proposed method for the directions of -30° and 40°, respectively. Also, the key "(Simulation)" represents those under non-reverberant conditions.

From Fig. 6, in the non-reverberant tests, it can be seen that the NRRs of about 16 dB are obtained using the proposed method. This indicates that the directions of the sources are estimated correctly in the proposed method. However, in the reverberant tests, NRRs decrease as the reverberation time increases. Especially, in the direction of -30°, the NRR is 8.4 dB in the case that the RT is 184 ms, and the NRR is 4.5 dB in the case that the RT is 322 ms. The main reason for this phenomenon is that since a large number of artificial sound sources are produced under the reverberation condition, it is hard to suppress the signals which are independent of the target signal.

Figure 7 shows the comparison of the NRRs of the proposed method with those of the conventional Murata's method under the typical reverberant conditions. These values are taken the average of the both direction of sound sources. From this figure, it is shown that the noise reduction rate is slightly inferior by 0.8 dB to Murata's method in the non-reverberant tests, however, there are no heavy degradations on the proposed method in the reverberant tests, compared with those of Murata's method. The main reason for the degradations in Murata's method is that the output envelopes in the same frequency are similar each other since the inaccurate unmixing matrix is estimated with many components of the cross talk because of the reverberation. Therefore, the recovery of the permutation tends to fail in Murata's method. In contrast, our method didn't fail to recover the source permutation because we did not use any informations of signal waveforms but use the directivity patterns only.

4. CONCLUSION

In this paper, a new blind signal separation method using the directivity patterns was described. In order to evaluate its effectiveness, the signal separation experiments were performed under reverberant conditions. As the result, it was shown that the noise reduction rate (NRR) of about 16 dB is obtained under the non-reverberant condition, and NRR of 8.7 dB is obtained in the case that the reverberation time is 184 ms. Also, it was shown that NRR decreases as the reverberation time increases, and NRR is 5.1 dB in the case that the reverberation time is 322 ms, however these performances are superior to those of the previous Murata's method.

5. ACKNOWLEDGEMENT

This work was supported by Grant-in-Aid for COE Research (No. 11CE2005).

6. REFERENCES


