Blind Source Separation for Speech Based on Fast-Convergence Algorithm with ICA and Beamforming

Hiroshi SARUWATARI, Toshiya KAWAMURA, and Kiyohiro SHIKANO

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

Abstract

We propose a new algorithm for blind source separation (BSS), in which independent component analysis (ICA) and beamforming are combined to resolve the low-convergence problem through optimization in ICA. The proposed method consists of the following three parts: (1) frequency-domain ICA with direction-of-arrival (DOA) estimation, (2) null beamforming based on the estimated DOA, and (3) integration of (1) and (2) based on the algorithm diversity in both iteration and frequency domain. The inverse of the mixing matrix obtained by ICA is temporally substituted by the matrix based on null beamforming through iterative optimization, and the temporal alternation between ICA and beamforming can realize fast- and high-convergence optimization. The results of the signal separation experiments reveal that the signal separation performance of the proposed algorithm is superior to that of the conventional ICA-based BSS method, even under reverberant conditions.

2. Data model and conventional BSS method

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), where we deal with the case of \( K = L = 2 \).

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

\[
X(f) = A(f)S(f),
\]

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

\[
X(f) = [X_1(f), \ldots, X_K(f)]^T, \quad S(f) = [S_1(f), \ldots, S_L(f)]^T, \quad A(f) = \begin{bmatrix} A_{11}(f) & \cdots & A_{1L}(f) \\ \vdots & \ddots & \vdots \\ A_{K1}(f) & \cdots & A_{KL}(f) \end{bmatrix}.
\]

\( A(f) \) is assumed to be complex-valued because we introduce a model to deal with the arrival lags among each of the elements of the microphone array and room reverberations.

In the frequency-domain ICA, first, the short-time analysis of observed signals is conducted by frame-by-frame discrete Fourier transform (DFT) (see Fig. 2). By plotting the spectral values in a frequency bin of each microphone input frame by frame, we consider them as \( L \) time series. Hereafter, we designate the time series as \( X(f,t) = [X_1(f,t), \ldots, X_K(f,t)]^T \). Next, we perform signal separation using the complex-valued inverse of the mixing matrix, \( \mathbf{W}(f) \), so that the \( L \) time-series output \( Y(f,t) \) becomes mutually independent; this procedure...
can be given as
\[ Y(f, t) = W(f)X(f, t), \]  
where
\[ Y(f, t) = [Y_1(f, t), \ldots, Y_L(f, t)]^T, \]  
\[ W(f) = \begin{bmatrix} W_{11}(f) & \cdots & W_{1K}(f) \\ \vdots & \ddots & \vdots \\ W_{L1}(f) & \cdots & W_{LK}(f) \end{bmatrix}. \]

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 \( Y(f, t) \), we reconstruct the resultant source signals in the time domain.

In the conventional ICA-based BSS method, the optimal \( W(f) \) is obtained by the following iterative equation [3, 7]:
\[ W_{i+1}(f) = \eta \left[ \text{diag} \left( \Phi(Y(f, t))) Y^H(f, t) \right) \right] W_i(f) + W_i(f), \]  
where \( \text{diag} \) denotes the time-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 vector function \( \Phi(\cdot) \) as
\[ \Phi(Y(f, t)) = [\Phi(Y_1(f, t)), \ldots, \Phi(Y_L(f, t))]^T, \]  
\[ \Phi(Y_1(f, t)) = \left[ 1 + \exp(-Y_{1R}(f, t)) \right]^{-1} + j \left[ 1 + \exp(-Y_{1I}(f, t)) \right]^{-1}, \]
where \( Y_{1R}(f, t) \) and \( Y_{1I}(f, t) \) are the real and imaginary parts of \( Y_1(f, t) \), respectively.

### 3. Proposed algorithm

The conventional ICA method inherently has a significant disadvantage which is due to low convergence through nonlinear optimization in ICA. In order to resolve the problem, we propose an algorithm based on the temporal alternation of learning between ICA and beamforming: the inverse of the mixing matrix, \( W(f) \), obtained through ICA is temporally substituted by the matrix based on null beamforming for a temporal initialization or acceleration of the iterative optimization. The proposed algorithm is conducted by the following steps with respect to all frequency bins in parallel (see Fig. 3).

#### [Step 1: Initialization]
Set the initial \( W_0(f) \), i.e., \( W_0(f) \), to an arbitrary value, where the subscripts \( i \) is set to be 0.

#### [Step 2: 1-time ICA iteration]
Optimize \( W_i(f) \) using the following 1-time ICA iteration:
\[ W_{i+1}(f) = \eta \left[ \text{diag} \left( \Phi(Y(f, t))) Y^H(f, t) \right) \right] W_i(f) + W_i(f), \]  
where the superscript "(ICA)" is used to express that the inverse of the mixing matrix is obtained by ICA.

#### [Step 3: DOA estimation]
Estimate DOAs of the sound sources by utilizing the directivity pattern of the array system, \( F(f, \theta) \), which is given by
\[ F(f, \theta) = \sum_{k=1}^{K} W_{ik}^{(ICA)}(f) \exp[j2\pi fd_k \sin \theta/c], \]  
where \( W_{ik}^{(ICA)}(f) \) is the element of \( W_{i+1}^{(ICA)}(f) \). 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 can be estimated as
\[ \theta_l(f_m) = \min[\arg\min_{\theta} |F_l(f_m, \theta)|, \arg\max_{\theta} |F_l(f_m, \theta)|], \]  
\[ \theta_l(f_m) = \max[\arg\min_{\theta} |F_l(f_m, \theta)|, \arg\max_{\theta} |F_l(f_m, \theta)|], \]  
where \( \min[x, y] \) (\( \max[x, y] \)) is defined as a function in order to obtain the smaller (larger) value among \( x \) and \( y \).

#### [Step 4: Beamforming]
Construct an alternative matrix for signal separation, \( W_{i+1}^{(BF)}(f) \), based on the null-beamforming technique where the DOA results obtained in the previous step...
and the gain inconsistency obtained in the Increment of otherwise go back to [Step 4].

Thus, an observation of the conditions yields the following algorithm:

\[
W(f) = \begin{cases} 
W^{(ICA)}(f), & \text{if } J^{(ICA)}(f) \leq J^{(BF)}(f) \\
W^{(BF)}(f), & \text{if } J^{(ICA)}(f) > J^{(BF)}(f)
\end{cases}
\]

If the \((i + 1)\)th iteration was the final iteration, go to [Step 6]. Otherwise, go back to [Step 2] and repeat the ICA iteration inserting \(W_i(f)\) in Eq. (11) into \(W_i(f)\).

[Step 6: Ordering and scaling] Using the DOA information obtained in [Step 3], we detect and correct the source permutation and the gain inconsistency [8].

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

(15)

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

(16)

[Step 5: Diversity with cost function] Select the most suitable unmixing matrix in each frequency bin and each iteration point, i.e., algorithm diversity in both iteration and frequency domain. As a cost function used to achieve the diversity, we calculate two kinds of cosine distances between the separated signals which are obtained by ICA and beamforming. These are given by

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

(17)

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

(18)

Objective evaluation of separated signals

In order to compare the performance of the proposed algorithm with that of the conventional BSS described in Sect. 2 for different iteration points in ICA, the noise reduction rate (NRR), defined as the output signal-to-noise ratio (SNR) in dB minus input SNR in dB, is shown in Fig. 5. These values were averages of all of the combinations with respect to speakers and source directions. As for the proposed algorithm, we also plot the NRR which is rescaled by the computational cost (see dotted lines) because the proposed algorithm has a computational complexity of about 1.9-fold compared with the conventional ICA.

In Fig. 5, it is evident that the separation performances of the proposed algorithm are superior to those of the conventional ICA-based BSS method at every iteration point, even considering the additional computational cost of the proposed algorithm. For example, compared with the conventional method, the proposed method can improve the NRR of about 4.6 dB at the 50-iteration point in the conventional ICA when the RT is 150 msec. Also, when the RT is 300 msec, the proposed method can improve the NRR of about 1.5 dB.

Figure 6 shows a result of alternation between ICA and null beamforming through iterative optimization by the proposed algorithm when the RT is 300 msec. In this figure, the symbol "-" represents that the null beamforming is used in the iteration point and frequency bin. As shown in Fig. 6, the proposed algorithm can work automatically as follows: (1) null beam-
forming is used for the acceleration of learning at early times in the iterations because $W^{(BF)}(f)$ is a rough approximation of the inverse of the mixing matrix $A(f)$. (2) ICA is used after the early part of the iterations because ICA can update the inverse of the mixing matrix more accurately, and (3) the inverse of the mixing matrix obtained by ICA is substituted by the matrix based on null beamforming through whole iteration points at particular frequency bins where the independence between the sources is low. From these results, although null beamforming is not suitable for signal separation under the condition that the direct sounds and their reflections exist, we can confirm that the temporal utilization of null beamforming for algorithm diversity through ICA iterations is effective for improving the separation performance and convergence.

5. Conclusion

In this paper, we described a fast- and high-convergence algorithm for BSS where null beamforming is used for temporal algorithm diversity through ICA iterations. The results of the signal separation experiments reveal that the signal separation performance of the proposed algorithm is superior to that of the conventional ICA-based BSS method, and the utilization of null beamforming in ICA is effective for improving the separation performance and convergence, even under reverberant conditions.

6. Acknowledgement

The authors are grateful to Dr. Shoji Makino and Dr. Ryo Mukai of NTT. CO., LTD. for their suggestions and discussions on this work. This work was partly supported by CREST (Core Research for Evolutional Science and Technology) in Japan.

7. References