Fast Convergence Blind Source Separation Using Frequency Subband Interpolation by Null Beamforming

Keiichi OSAKO, Nonmember, Yoshimitsu MORI, Student Member, Yu TAKAHASHI, Nonmember, Hiroshi SARUWATARI, Member, and Kiyohiro SHIKANO, Fellow

SUMMARY We propose a new algorithm for the blind source separation (BSS) approach in which independent component analysis (ICA) and frequency subband beamforming interpolation are combined. The slow convergence of the optimization of the separation filters is a problem in ICA. Our approach to resolving this problem is based on the relationship between ICA and null beamforming (NBF). The proposed method consists of the following three parts: (I) a frequency subband selector part for learning ICA, (II) a frequency domain ICA part with direction-of-arrivals (DOA) estimation of sound sources, and (III) an interpolation part in which null beamforming constructed with the estimated DOA is used. The results of the signal separation experiments under a reverberant condition reveal that the convergence speed is superior to that of the conventional ICA-based BSS methods.

key words: blind source separation, independent component analysis, null beamforming, DOA estimation

1. Introduction

Blind source separation (BSS) is an approach aimed at estimating original source signals using only information provided by 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, for BSS based on independent component analysis (ICA) [1], several methods, in which the unmixing matrices are calculated in the frequency domain, have been proposed to deal with the arrival lags among the elements of the microphone array system [2], [3]. However, those ICA-based approaches have the disadvantage that the nonlinear optimization convergence is poor [4].

In this paper, we propose a fast-convergence algorithm for BSS, in which the frequency subband is interpolated by beamforming. The proposed method consists of the following three parts: (I) a frequency subband selector part for learning ICA, (II) a frequency domain ICA part with direction-of-arrivals (DOA) estimation of sound sources, and (III) an interpolation part in which null beamforming (NBF) constructed with the estimated DOA is used. Inserting beamforming among the ICA iterations results in fast and high convergence optimization. In the following sections, we describe the proposed method in detail, and show that the convergence speed of the proposed algorithm is superior to those of the conventional ICA method and the conventional BSS method based on ICA and beamforming.

2. Mixing Process and Conventional ICA

In this study, the number of microphones is \( K \) and the number of multiple sound sources is \( L \); here we consider the case of \( K = L \).

Multiple mixed signals are observed at the microphone array, and these signals are converted into discrete-time series via an A/D converter. By applying the short-time discrete-time Fourier transform frame wisely, we can express the observed signals as a linear mixture in the time-frequency domain:

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

where \( f \) is an arbitrary frequency bin index, \( t \) is the frame index, \( X(f, t) = [X_1(f, t), \ldots, X_K(f, t)]^T \) is the observed signal vector, and \( S(f, t) = [S_1(f, t), \ldots, S_L(f, t)]^T \) is the source signal vector. Also, \( A(f) \) is the mixing matrix that is complex-valued because a model for dealing with the relative time delays among the microphones and room reverberations has been introduced.

Next, we perform signal separation using the complex-valued unmixing matrix \( W(f) \), so that the L-time-series output \( Y(f, t) = [Y_1(f, t), \ldots, Y_L(f, t)]^T \) becomes mutually independent; this procedure is written as

\[
Y(f, t) = W(f)X(f, t).
\]

In the conventional ICA-based BSS method, the optimal \( W(f) \) is obtained by the following iterative equation:

\[
W_{t+1}(f) = W_t(f) - \alpha \cdot \text{off-diag}(\Phi(Y_t^{(t)}(f, t)))Y_t^{(t)}(f, t)^N, \]

where \( \langle \cdot \rangle \) denotes the time-averaging operator, \( X^{(t)} \) is the value of \( X \) at the \( t \)-th step, \( \alpha \) is the step-size parameter, off-diag(\( X \)) is the operation for setting every diagonal element of matrix \( X \) to zero, and \( \Phi(\cdot) \) is the appropriate nonlinear vector function (see, for example, [3], [5]; we use [5] in this paper).

Manuscript received August 4, 2007.
Manuscript revised November 22, 2007.
The authors are with the Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630-0192 Japan.
a) E-mail: keichi-o@is.naist.jp
b) E-mail: sawatari@is.naist.jp
DOI: 10.1093/ietfec/e91-a.6.1357
3. Proposed Method

3.1 Motivation and Strategy

The conventional ICA method inherently has a significant disadvantage, that is, a slow convergence of the nonlinear optimization. In our previous study [4], an algorithm based on the temporal alternation of learning between ICA and beamforming was proposed. The unmixing matrix $W(f)$ obtained through ICA is temporally substituted by the matrix based on NBF for a temporal initialization or acceleration of the iterative optimization. This method consists of the following three parts in each iteration: (I) ICA and DOA estimation using the obtained unmixing matrix, (II) beamforming using the estimated DOA in part I, and (III) integration of parts I and II on the basis of the algorithm diversity in both iteration and frequency domains. We can see that ICA or beamforming is selected in each frequency subband from the top of Fig. 1(a); beamforming is selected in black frequency subbands, and ICA is selected in white frequency subbands. Null beamforming is used for the acceleration of learning at early times in the iterations because ICA is a rough approximation of the unmixing matrix, and ICA is used after the early part of the iterations because ICA can update the unmixing matrix more accurately.

However, with this method, a cost function for algorithm diversity (choice of ICA or NBF) is computed many times. Also, this cost function uses a coherence function, which requires a large amount of computation. Moreover, the estimated DOA is not accurate in the early part of learning. Therefore, we take the following two strategies.

Strategy 1: Select ICA or NBF before optimizing separation filter with a simple rule.

Strategy 2: Optimize selected bins not by one iteration but by more than one iteration.

By introducing strategy 1, we can reduce the computational cost of calculating diversity with the cost function. Moreover, it is not necessary to perform both learning ICA and generating NBF in whole subbands. Therefore, it can be expected that computational cost will be reduced significantly (see Fig. 1(b)). The rule for selecting ICA or beamforming is as follows: a few frequency subbands are assigned for ICA and many frequency subbands are assigned for beamforming in the early part of learning. As learning advances, we increase the subbands for ICA and decrease the subbands for beamforming. Finally, all frequency subbands are assigned for ICA. We can obtain a more accurate DOA by strategy 2 at the beginning of learning without the incrementation of computational cost. Indeed, strategy 2 requires more than one iteration. However, the total cost of computation is constant because ICA learning is only performed in a few selected frequency bins.

More specifically, the proposed method is a kind of technique to obtain better initialization of $W(f)$ with less computations. The important issue is to generate an appropriate initial $W(f)$ (NBF) based on precise DOA information, but inaccurate initialization leads to degradation of ICA's convergence (see, e.g., Fig. 6). Needless to say, it is obvious that we can construct an accurate initial $W(f)$ by NBF which makes ICA's convergence faster if we can know the true DOAs of sources in advance. However, it is difficult to use traditional high-resolution DOA estimators, e.g., MUSIC method [6] because these methods require (1) huge amount of computations, and (2) additional microphones (thus $K \geq L + 1$). Therefore, the simple combination of the traditional DOA estimator and NBF does not hold in our BSS context. Our proposed method, by contrast, can estimate the DOAs directly using directivity patterns spanned ICA filters to resolve the above-mentioned problem.

3.2 Algorithm

In the proposed algorithm, we define two terms as follows: a 'loop' represents that the number of repeats through steps 1–6, and an 'iteration' means that the number of times in ICA optimization.

[Step 0: Initialization] Set the initial $\gamma^{[0]}$ and $\delta$, the parameter of this algorithm, where $\gamma^{[i]}$ is the decimation ratio for the frequency bins of ICA, and $\delta$ is the coefficient that controls the variation of the number of bins where ICA is performed.

[Step 1: Select frequency bins] In each frequency bin, assign ICA or NBF to generate the unmixing matrix. Thus, we select which method constructs the unmixing matrix via a simple rule. This selection rule is defined as

$$f = \begin{cases} f_{\text{ICA}} : & \text{if } f \mod \gamma^{[i]} = 0 (f = 1, \ldots, N/2) \\ f_{\text{NBF}} : & \text{if } f \mod \gamma^{[i]} \neq 0 (f = 1, \ldots, N/2) \end{cases} \quad (4)$$

where $N$ is the DFT size, and $[\cdot]$ is used to express the value of the $i$-th loop in this algorithm. The selection rule is not limited within using Eq. (4), and indeed the bin selection might not be independent of the BSS accuracy. For exam-
ple, it is possible to select frequency subbands with dominant signal powers. However, that way demands more calculation cost. Therefore we use Eq. (4) in this paper because we want to reduce the total calculation cost of BSS.

**Step 2: ICA iterations** Optimize \( W(f_{\text{ICA}}) \) with update Eq. (3) \( \hat{y}[i] \) iterations in the \( i \)-th loop. The calculation cost is the same as that with the conventional full-band ICA, because selected bins are decimated to \( 1/\gamma[i] \). Therefore, \( W(f_{\text{ICA}}) \) is the new filter obtained after \( \gamma[i] \) iterations of ICA optimization.

**Step 3: DOA estimation** Estimate DOAs of the sound sources by utilizing the directivity pattern of the array system [4]. The directivity pattern for the \( i \)-th output is designated by \( F_i(f_{\text{ICA}}, \theta) \), which is generally obtained by the multiplication of the array weights and a steering vector as

\[
[F_1(f_{\text{ICA}}, \theta), \ldots, F_L(f_{\text{ICA}}, \theta)]^T = W(f_{\text{ICA}}) a(f_{\text{ICA}}, \theta),
\]

where \( a(f_{\text{ICA}}, \theta) \) is the steering vector, \( f_s \) is the sampling frequency, and \( c \) is sound velocity. In the \( i \)-th directivity pattern \( F_i(f_{\text{ICA}}, \theta) \) at the frequency of \( f_{\text{ICA}} \), at most, \( L - 1 \) directional nulls can be found. We define the set of DOAs corresponding to the directional nulls as \( \Theta_i(f_{\text{ICA}}) \), which is given by

\[
\Theta_i(f_{\text{ICA}}) = \{ \theta | |F_i(f_{\text{ICA}}, \theta) - F_i(f_{\text{ICA}}, \theta - \Delta \theta)| \leq 0; |F_i(f_{\text{ICA}}, \theta + \Delta \theta) - F_i(f_{\text{ICA}}, \theta)| > 0 \},
\]

where \( \Delta \theta \) is a positive small value, and \( \{ \theta | A; B \} \) represents a set of \( \theta \) values that satisfy conditions \( A \) and \( B \) simultaneously. \( \Theta_i(f_{\text{ICA}}) \) is evidently a good candidate of source directions. To estimate the DOAs of sources, we classify \( \Theta_i(f_{\text{ICA}}) \) with all \( f_{\text{ICA}} \) and \( i \) into \( L \) categories, and we regard the centroids as the estimated DOAs. This classification can be carried out using a Lloyd clustering algorithm as follows,

(i) Make the whole set of \( \Theta_i(f_{\text{ICA}}) \) such that it is classified as

\[
\Theta = \{ \theta_1, \theta_2, \theta_3, \ldots, \theta_Q \} = \bigcup_{i=1}^{L} \bigcup_{k=1}^{N/\gamma[i]/2} \Theta_i(f_{\text{ICA}}),
\]

where \( Q \) is the total number of detected directional nulls and at most \( Q = (L - 1) \cdot L \cdot (N/\gamma[i]/2) \).

(ii) Set initial \( L - 1 \) partitions \( \theta^{(p)}_i \), where \( p = 1, \ldots, L - 1 \) and \(-90 < \theta^{(p)}_i < \theta^{(p)}_{i+1} < \ldots < \theta^{(p)}_{L-1} < 90 \). Also, the terminal partitions \( \theta^{(p)}_0 \) and \( \theta^{(p)}_L \) are fixed at \(-90 \) and \( 90 \), respectively.

(iii) Given the partitions, calculate the \( L \) centroids \( \theta^{(c)}_i \) (\( i = 1, \ldots, L \)) as

\[
\theta^{(c)}_i = \frac{1}{Q_i} \sum_{\theta^{(p)}_j} \left( \theta^{(p)}_{i-1} \leq \theta < \theta^{(p)}_i \right),
\]

where \( Q_i \) denotes the number of \( \theta_q \) under \( \theta^{(p)}_{i-1} \leq \theta < \theta^{(p)}_i \).

(iv) Given the centroids, update the new partitions as

\[
\theta^{(p)}_i = \left( \frac{\theta^{(c)}_i + \theta^{(c)}_{i+1}}{2} \right) \text{ for } p = 1, \ldots, L - 1.
\]

(v) Go back to step (iii), and repeat the loop in steps (iii)-(v) with the appropriate number of times. If the centroids do not move, then stop the algorithm. The final centroids, \( \theta^{(c)}_i \), are regarded as the estimated DOAs \( \hat{\theta}_i \).

**Step 4: Beamforming** Construct an alternative matrix for signal separation, \( W(f_{\text{NBF}}) \), based on the null beamforming technique with the DOA information obtained in the ICA section. The unmixing matrix \( W(f_{\text{NBF}}) \) satisfies

\[
W(f_{\text{NBF}}) = \left[ w(f_{\text{NBF}}, \hat{\theta}_1), \ldots, w(f_{\text{NBF}}, \hat{\theta}_L) \right]^{-1}.
\]

**Step 5: NBF interpolation** Interpolate the \( W(f_{\text{NBF}}) \) calculated in step 4 into \( W(f_{\text{ICA}}) \). This \( W(f_{\text{ICA}}) \) is the optimized separation filter of the \( i \)-th loops.

\[
W(f_{\text{ICA}}) = \begin{cases} W(f_{\text{ICA}}) : & \text{if } f \mod \gamma[i] = 0 \\ W(f_{\text{NBF}}) : & \text{if } f \mod \gamma[i] \neq 0. \end{cases}
\]

We obtain the separated signal by using \( W(f_{\text{ICA}}) \) in Eq. (2).

**Step 6: Update ICA NBF selector** Update the \( \gamma \) value for the next \((i + 1)\)-th loops by

\[
\gamma[i+1] = \gamma[i] / \hat{\delta}.
\]

and return to step 1. If \( \gamma \) becomes equal to 1, then ICA is selected in all frequency bins.

4. Experiments in Reverberant Room

4.1 Experiment Setup

We carried out sound-separation experiments using source signals that are convolved with impulse responses recorded in the reverberant room shown in Fig. 2. The reverberation time in this room is 200 ms. A two-element array with an interelement spacing of 4.3 cm is assumed. The two speech signals are assumed to arrive from different directions, \( (\theta_1, \theta_2) = (-60^\circ, -10^\circ) \). The distance between the array and the sound source is 1.0 m. Two sentences, spoken by each of two male and two female speakers selected from the ASJ continuous speech corpus for research,
are used as the original speech samples. Using these sentences, we obtain 12 combinations with respect to speakers and source directions. The sampling frequency is 8 kHz and the length of each sound sample is limited to 3 s. The DFT size is 1024, and the frame shift length is 256. The initial value of \( W(0,f) \) corresponds to null beamforming with \((\theta_{\text{NBF}}, \phi_{\text{NBF}}) = (45^\circ, 45^\circ)\). Using the DOA information obtained in Sect. 3.2 step 3, we detect and correct the source permutation problem [7]. Then we apply the projection-back method to remove the ambiguity of amplitude [2].

In experimental results’ plots, we introduce the counting manner of the number of calculations as follows: (A) In the conventional methods, 'one iteration' means that ICA updates unmixing matrix once in all frequency bins. (B) In the proposed method, 'one iteration' is equal to 'one loop' as described in Sect. 3 because both calculation amounts are the same.

4.2 Objective Evaluation of Separated Signals

We compare the conventional BSS described in Sect. 2 and the proposed algorithm on the basis of the noise reduction rate (NRR), which is defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB. Figure 3 shows the NRR averages of all the combinations with respect to speakers. These results show that the proposed method has improved the convergence speed of the NRR. In this experimental condition, the proposed method has subtle distinctions of NRR in each pattern of \( \gamma \) and \( \delta \) parameters. Generally speaking, if the reverberation effect is moderate, we use large \( \gamma \) and \( \delta \); otherwise we should set small \( \gamma \) and \( \delta \) for stabilization of DOA estimation convergence under heavily reverberant conditions.

Next, we examine the estimated DOAs in each iteration in Fig. 4. The solid line indicates the DOA estimated by the proposed method, and the dotted line indicates the

![Fig. 3](image-url)  
**Fig. 3** Noise-reduction-rates convergence curves of proposed method and conventional BSS based on ICA for \((\theta_1, \theta_2) = (-60^\circ, -10^\circ)\).

![Fig. 5](image-url)  
**Fig. 5** Noise reduction rates convergence of proposed method and conventional BSS based on combining ICA and beamforming for \((\theta_1, \theta_2) = (-60^\circ, -10^\circ)\). Conventional method is scaled by computational cost.

![Fig. 6](image-url)  
**Fig. 6** Number of iterations necessary to reach NRR of 10 dB.
DOA estimated by the conventional method. We can see that the DOA estimation is successful from the beginning of the separation when using the proposed method.

Figure 5 shows the performance of the proposed algorithm and that of the conventional BSS based on a combination of ICA and beamforming. In this figure, we plot the NRR curve, which is scaled by the computational cost, on the x-axis, because the conventional algorithm has a computational complexity of about 2.0-fold compared with our proposed method.

Figure 6 composes the proposed algorithm and the conventional BSS method for three variations of the initialization method. The initial values of matrix are an identity matrix, principle components analysis (PCA), and null beamforming with a subjective direction, ($\theta_{\text{BF}_1}, \theta_{\text{BF}_2}$) = (-45°, +45°). The two speech signals are assumed to arrive from different directions. There are 12 patterns of combinations of DOA with fixed intervals, 50 degrees. The performance index is the number of iterations required for achieving NRR of 10 dB. The number of iterations in the proposed method does not exceed 100 times in all patterns of speech directions. Also, we can see that the average NRR is superior to those of conventional methods.

From these results, we can conclude that the proposed method converges faster than the conventional methods and is robust for various combinations of DOA of the sound sources.

5. Conclusions

In this paper, we proposed a new algorithm for the BSS approach in which ICA and frequency subband beamforming interpolation are combined. 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 its calculation cost is lower. Moreover, the proposed algorithm is also applicable to the conventional BSS approach in which ICA and beamforming are combined [4].

Acknowledgments

This work was partly supported by e-Society program, and NEDO project for strategic development of advanced robotics elemental technologies in Japan.

References