NEAR-FIELD FREQUENCY DOMAIN BLIND SOURCE SEPARATION FOR CONVOLUTIVE MIXTURES

Ryo Mukai *Hiroshi Sawada* *Shoko Araki* *Shoji Makino*

NTT Communication Science Laboratories, NTT Corporation
2–4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619–0237, Japan
{ryo,sawada,shoko,maki}@cslab.kecl.ntt.co.jp

ABSTRACT

This paper presents a method for solving the permutation problem of frequency domain blind source separation (BSS) when source signals come from the same or similar directions. Geometric information such as the direction of arrival (DOA) is helpful for solving the permutation problem, and a combination of the DOA based and correlation based methods provides a robust and precise solution. However when signals come from similar directions, the DOA based approach fails, and we have to use only the correlation based method whose performance is unstable. In this paper, we show that an interpretation of the ICA solution by a near-field model yields information about spheres on which source signals exist, which can be used as an alternative to the DOA. Experimental results show that the proposed method can robustly separate a mixture of signals arriving from the same direction.

1. INTRODUCTION

Independent Component Analysis (ICA) [1] in the frequency domain is one of the most important and practical methods for Blind Source Separation (BSS) of convolutive mixtures. A convolutive mixture in the time domain is converted into multiple instantaneous mixtures in the frequency domain, and a complex-valued ICA algorithm for instantaneous mixtures can be applied for each frequency. This approach has the advantage of fast convergence, which enables realtime processing [2]. However, a permutation ambiguity in the ICA solutions is a serious problem. We need to align separated signals so that a separated signal in the time domain contains frequency components from the same source signal. This problem is known as the permutation problem of frequency domain BSS.

The other approach for separating convolutive mixtures is time domain BSS, which does not suffer the permutation problem, however it takes much more computational time than frequency domain BSS [3]. Therefore we adopt the frequency domain approach.

Many methods have been proposed for solving the permutation problem. The use of geometric information, such as beam patterns [4, 5, 6], direction of arrival (DOA) and source locations [7], is one of the effective approaches. Another approach is based on the inter-frequency correlations [8, 9], which uses the inter-frequency correlations of output signal envelopes to align the permutations. However the correlation based method is not robust since a misalignment at one frequency bin causes consecutive misalignments.

We have proposed a robust and precise method by combining the DOA based method and the correlation based method, which almost completely solves the permutation problem for two sources that come from different directions [10]. However the DOA based method fails in the first stage when the signals come from the same or similar directions. In such cases we have to rely on the correlation based method that is unstable.

In this paper, we show that the interpretation of the ICA solution by a near-field model yields information about spheres on which source signals exist. This information can substitute for the DOA when signals come from the same direction.

2. FREQUENCY DOMAIN BSS

When the source signals are $s_i(t)(i = 1, ..., N)$, the signals observed by microphone j are $x_j(t)(j = 1, ..., M)$, and the separated signals are $y_k(t)(k = 1, ..., N)$, the BSS model can be described as:

$$x_j(t) = \textstyle\sum_{i=1}^{N}(h_{ji} * s_i)(t), \tag{1}$$

$$y_k(t) = \textstyle\sum_{j=1}^{M}(w_{kj} * x_j)(t), \tag{2}$$

where h_{ji} is the impulse response from source i to microphone j, w_{kj} are the separating filters, and $*$ denotes the convolution operator.

Figure 1 shows the BSS flow in the frequency domain. A convolutive mixture in the time domain is converted into multiple instantaneous mixtures in the frequency domain. Therefore, we can apply an ordinary ICA algorithm [1] in

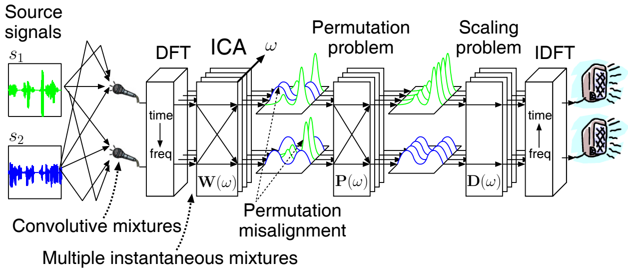

Fig. 1. Flow of frequency domain blind source separation

the frequency domain to solve a BSS problem in a reverberant environment. Using a short-time discrete Fourier transform, the model is approximated as:

$$\mathbf{X}(\omega, n) = \mathbf{H}(\omega)\mathbf{S}(\omega, n), \tag{3}$$

where ω is the angular frequency, n represents the frame index, $\mathbf{H}(\omega)$ is the mixing system in the frequency domain, $\mathbf{S}(\omega, n) = [S_1(\omega, n), ..., S_N(\omega, n)]^T$ is the source signal, and $\mathbf{X}(\omega, n) = [X_1(\omega, n), ..., X_M(\omega, n)]^T$ denotes the observed signals. The separating process can be formulated in each frequency bin as:

$$\mathbf{Y}(\omega, n) = \mathbf{W}(\omega)\mathbf{X}(\omega, n), \tag{4}$$

where $\mathbf{Y}(\omega, n) = [Y_1(\omega, n), ..., Y_N(\omega, n)]^T$ is the estimated source signal, and $\mathbf{W}(\omega)$ represents the separating matrix. $\mathbf{W}(\omega)$ is determined so that $Y_i(\omega, n)$ and $Y_j(\omega, n)$ $(i \neq j)$ become mutually independent.

The ICA solution suffers permutation and scaling ambiguities. This is due to the fact that if $\mathbf{W}(\omega)$ is a solution, then $\mathbf{D}(\omega)\mathbf{P}(\omega)\mathbf{W}(\omega)$ is also a solution, where $\mathbf{D}(\omega)$ is a diagonal complex valued scaling matrix, and $\mathbf{P}(\omega)$ is an arbitrary permutation matrix. We thus have to solve the permutation and scaling problems to reconstruct separated signals in the time domain.

There is a simple and reasonable solution for the scaling problem: $\mathbf{D}(\omega) = \mathrm{diag}\{[\mathbf{P}(\omega)\mathbf{W}(\omega)]^{-1}\}$, which is obtained by the minimal distortion principle (MDP) [11], and we can use it. In contrast, the permutation problem is complicated, and many solutions have been proposed.

3. PROPOSED METHOD

3.1. Invariant in ICA solution

If a separating matrix $\mathbf{W}(\omega)$ is calculated successfully and it extracts source signals with scaling ambiguity, there is a diagonal matrix $\mathbf{D}(\omega)$, and $\mathbf{D}(\omega)\mathbf{W}(\omega)\mathbf{H}(\omega) = \mathbf{I}$ holds. Because the scaling matrix determined by the MDP does not satisfy this equation, and $\mathbf{D}(\omega)$ cannot be determined in general, we cannot obtain $\mathbf{H}(\omega)$ simply from the ICA solution. However, the ratio of elements in the same column $H_{ji}(\omega)/H_{j'i}(\omega)$ is invariable in relation to $\mathbf{D}(\omega)$, and given

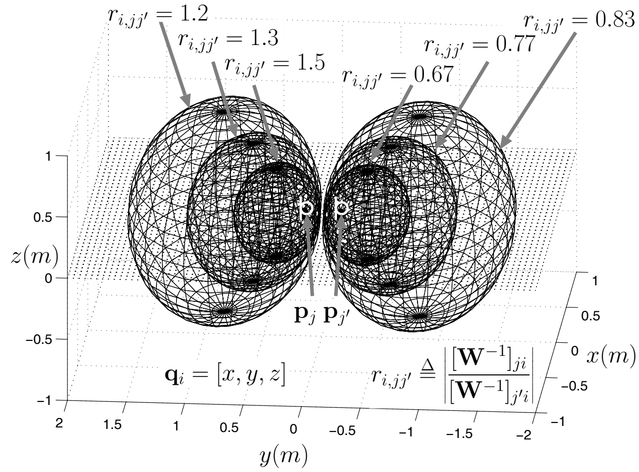

Fig. 2. Example of spheres determined by eq.(8) ($\mathbf{p}_j = [0, 0.15, 0]$, $\mathbf{p}_{j'} = [0, -0.15, 0]$)

by

$$\frac{H_{ji}(\omega)}{H_{j'i}(\omega)} = \frac{[\mathbf{W}^{-1}(\omega)\mathbf{D}^{-1}(\omega)]_{ji}}{[\mathbf{W}^{-1}(\omega)\mathbf{D}^{-1}(\omega)]_{j'i}} = \frac{[\mathbf{W}^{-1}(\omega)]_{ji}}{[\mathbf{W}^{-1}(\omega)]_{j'i}}, \tag{5}$$

where $[\cdot]_{ji}$ denotes the ji-th element of the matrix. We can estimate several types of geometric information related to source signals by using this invariant. The estimated information can be used for solving the permutation problem.

Previously we have shown that the DOA of source signals can be estimated by comparing (5) with a far-field model [10, 12]. The interpretation of the ICA solution by a near-field model yields other geometric information as described below.

3.2. Estimation of sphere with ICA solution

When we adopt the near-field model, including the attenuation of the wave, $H_{ji}(\omega)$ is formulated as:

$$H_{ji}(\omega) = \frac{1}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{\jmath\omega c^{-1}(\|\mathbf{q}_i - \mathbf{p}_j\|)}, \tag{6}$$

where $\mathbf{p}_i$ represents the location of microphone j, $\mathbf{q}_i$ is the location of source i, and c is the speed of wave propagation. By taking the ratio of (6) for a pair of microphones j and j' we obtain:

$$\frac{H_{ji}(\omega)}{H_{j'i}(\omega)} = \frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{\jmath\omega c^{-1}(\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i - \mathbf{p}_{j'}\|)}. \tag{7}$$

By using the modulus of (7) and (5), we have:

$$\frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} = \left|\frac{[\mathbf{W}^{-1}(\omega)]_{ji}}{[\mathbf{W}^{-1}(\omega)]_{j'i}}\right|. \tag{8}$$

By solving (8) for $\mathbf{q}_i$, we have a sphere whose center $O_{i,jj'}$ and radius $R_{i,jj'}$ are given by:

$$O_{i,jj'} = \mathbf{p}_j - \frac{1}{r_{i,jj'}^2 - 1}(\mathbf{p}_{j'} - \mathbf{p}_j), \tag{9}$$

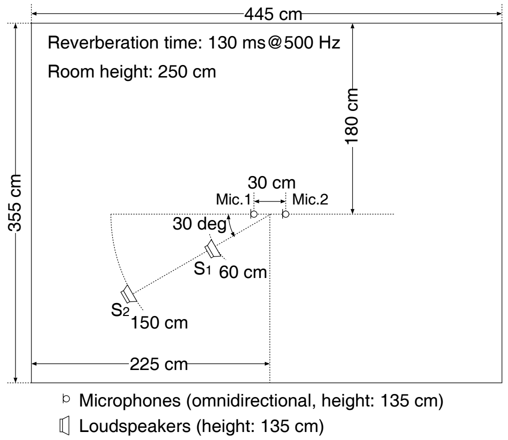

Fig. 3. Room layout

Table 1. Experimental conditions

Sampling rate	8 kHz
Data length	2 s
Window	Hanning
Frame length	1024 point (128 ms)
Frame shift	256 point (32 ms)
ICA algorithm	Infomax (complex valued)

$$R_{i,jj'} = \|\frac{r_{i,jj'}}{r_{i,jj'}^2 - 1}(\mathbf{p}_{j'} - \mathbf{p}_j)\|, \tag{10}$$

where $r_{i,jj'} = |[\mathbf{W}^{-1}(\omega)]_{ji}/[\mathbf{W}^{-1}(\omega)]_{j'i}|$. Therefore, we can estimate a sphere $(\hat{O}_{i,jj'}, \hat{R}_{i,jj'})$ on which $\mathbf{q}_i$ exists by using the separation matrix $\mathbf{W}(\omega)$ obtained by ICA and the locations of the microphones $\mathbf{p}_j$ and $\mathbf{p}_{j'}$. Figure 2 shows an example of spheres determined by (8) for various ratios $r_{i,jj'}$.

3.3. Solving permutation problem

The estimated spheres can be utilized for sorting or classifying the output signals of ICA, thus we can solve the permutation problem even when signals come from the same or similar directions. However, geometric information such as the DOA or the sphere tends to have a large error in a reverberant environment, especially for lower frequency bins. Accordingly, we use the geometric information based solution only for frequency bins whose geometric information is highly reliable. Then we apply the correlation based method to the rest of frequency bins. The correlation based method solves the permutation problem by maximizing the inter-frequency correlation for neighboring or harmonic frequency bins [10].

4. EXPERIMENTS

We carried out experiments with 2 sources and 2 microphones using speech signals convolved with impulse responses measured in a room. The room layout is shown

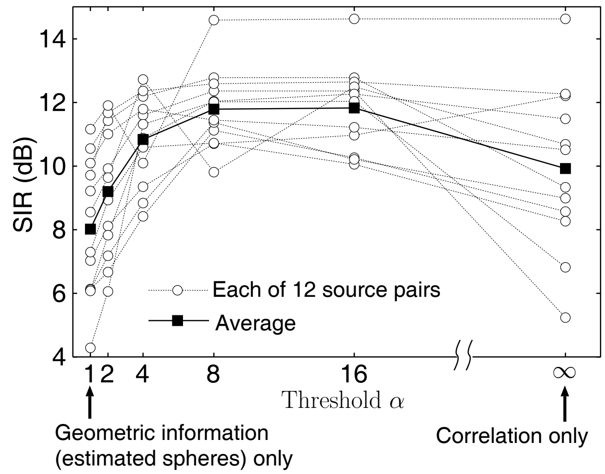

Fig. 4. Experimental results. SIRs are evaluated for 12 combinations of source signals with various values for threshold parameter α.

in Fig. 3. The sources are located in the same direction for the microphone pair. The reverberation time of the room was 130 ms at 500 Hz. Other conditions are summarized in Table 1. The experimental procedure is as follows.

First, we apply ICA to observed signals $x_j(t) (j = 1, 2)$, and calculate separating matrix $\mathbf{W}(\omega)$ for each frequency bin. Then we estimate spheres by using $\mathbf{W}^{-1}(\omega)$. We use the radiuses of two spheres $\hat{R}_{1,12}$ and $\hat{R}_{2,12}$, and the permutation is aligned so that $\hat{R}_{2,12} \geq \hat{R}_{1,12}$. In order to evaluate the reliability of the solution provided by the estimated spheres, we introduce a threshold parameter $\alpha \geq 1$, and we accept solutions only for frequency bins that satisfy the condition $\hat{R}_{2,12}/\hat{R}_{1,12} \geq \alpha$. We then apply the correlation based method to the rest of frequency bins. The permutation problem is solved simply by using the geometric information when $\alpha = 1$, and simply by using the correlation when $\alpha = \infty$.

The performance is measured by the signal-to-inference ratio (SIR). The portion of $y_k(t)$ that comes from $s_i(t)$ is calculated by $y_{ki}(t) = \sum_{j=1}^{2}(w_{kj} * h_{ji} * s_i)(t)$, and the output SIR for $y_k(t)$ is defined as:

$$\mathrm{SIR}_{Ok} = 10 \log[\textstyle\sum_t y_{kk}(t)^2 / \sum_t (\sum_{i \neq k} y_{ki}(t))^2] \text{ (dB)}.$$

We define the SIR as an average of SIR_{O1} and SIR_{O2} in order to cancel out the effect of input SIR. We measured SIRs for 12 combinations of source signals using two male and two female speakers and varying the threshold parameter α.

5. RESULTS AND DISCUSSION

Figure 4 shows the experimental results. When we solve the permutation problem using only the estimated spheres ($\alpha = 1$), the performance is insufficient. In contrast, the performance we obtain using only the correlation ($\alpha = \infty$)

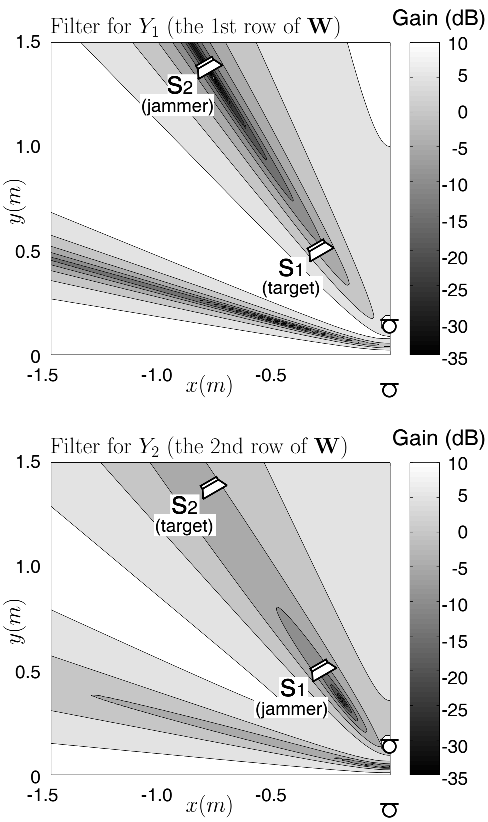

Fig. 5. Example of spatial gain patterns of separating filters (f = 1000 Hz)

is unstable. The combination of both methods yields good and stable performance. These tendencies are similar to the results we obtain when we use DOAs as geometric information [10].

We obtained good performance when the threshold parameter α was relatively large. When α was 8 to 16, the portion of frequency bins whose permutation was determined by the geometric information was about 1/5 to 1/10.

Finally, we show the gain patterns of the separating system. Figure 5 shows the spatial gain patterns of separating filters in one frequency bin (f = 1000 Hz). Microphone 1 is used as a reference, *i.e.*, when a signal observed by microphone 1 is directly outputted, the gain is 0 dB. We can see that the separating filter forms a spot null beam focusing on the jammer signal.

6. CONCLUSION

The interpretation of the ICA solution by a near-field model yields information about spheres on which source signals exist. This information can be used as an alternative to the DOA when signals come from the same or similar directions. Experimental results showed that the proposed method can robustly separate a mixture of signals originating from the same direction. Some sound examples can be found on our web site [13].

The proposed method is valid for a microphone pair with a large spacing. When we can use many microphone pairs and the DOAs are available through the use of microphone pairs with a small spacing, the source locations can be estimated in more detail. We have succeeded in separating a mixture of six speech signals arriving from various directions, even when two of them come from the same direction, by using microphone pairs with small and large spacings [14, 15].

7. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[2] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Robust real-time blind source separation for moving speakers in a room," in *Proc. of ICASSP'03*, 2003, vol. 5, pp. 469–472.

[3] K. Matsuoka, Y. Ohba, Y. Toyota, and S. Nakashima, "Blind separation for convolutive mixture of many voices," in *Proc. IWAENC 2003*, 2003, pp. 279–282.

[4] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. of ICASSP'00*, 2000, pp. 3140–3143.

[5] M. Z. Ikram and D. R. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation," in *Proc. of ICASSP'02*, 2002, vol. 1, pp. 881–884.

[6] L. C. Parra and C. V. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 6, pp. 352–362, Sep. 2002.

[7] V. C. Soon, L. Tong, Y. F. Huang, and R. Liu, "A robust method for wideband signal separation," in *Proc. of ISCAS '93*, 1993, pp. 703–706.

[8] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *ICA 2000*, 2000, pp. 215–220.

[9] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, 2001.

[10] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," in *Proc. Intl. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, 2003, pp. 505–510.

[11] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. of Intl. Workshop on Independent Component Analysis and Blind Signal Separation (ICA'01)*, 2001, pp. 722–727.

[12] H. Sawada, R. Mukai, and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," in *Proc. of ISSPA'03*, 2003, vol. 2, pp. 411–414.

[13] http://www.kecl.ntt.co.jp/icl/signal/mukai/demo/icassp2004/

[14] R. Mukai, H. Sawada, S. de la Kethulle de Ryhove, S. Araki, and S. Makino, "Array geometry arrangement for frequency domain blind source separation," in *Proc. IWAENC 2003*, 2003, pp. 219–222.

[15] http://www.kecl.ntt.co.jp/icl/signal/mukai/demo/iwaenc2003/