Frequency Domain Blind Source Separation for Many Speech Signals

Ryo Mukai, Hiroshi Sawada, Shoko Araki, and 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 the number of source signals is large, and the potential source locations are omnidirectional. We propose a combination of small and large spacing sensor pairs with various axis directions in order to obtain proper geometric information for solving the permutation problem. Experimental results in a room (reverberation time T_R=130 ms) with eight microphones show that the proposed method can separate a mixture of six speech signals that come from various directions, even when two of them come from the same direction.

1 Introduction

Independent component analysis (ICA) is one of the major statistical methods for blind source separation (BSS). It is theoretically possible to solve the BSS problem with a large number of sources by ICA if we assume that the number of observed signals is equal to or greater than the number of source signals. However, there are many practical difficulties, and although a large number of studies have been undertaken on audio BSS in a reverberant environment, only a few studies have dealt with more than two source signals.

In a reverberant environment, the signals are mixed in a convolutive manner with reverberations, and the unmixing system that we have to estimate is a matrix of filters, not just a matrix of scalars. There are two major approaches to solving the convolutive BSS problem. The first is the time domain approach, where ICA is applied directly to the convolutive mixture model. Matsuoka et al. have proved that time domain ICA can solve the convolutive BSS problem of eight sources with eight microphones in a real environment [1]. Unfortunately, the time domain approach incurs considerable computation cost, and it is difficult to obtain a solution in a practical time.

The other approach is frequency domain BSS, where ICA is applied to multiple instantaneous mixtures in the frequency domain. This approach takes much less computation time than time domain BSS. However, it poses another problem in that we need to align the output signal order for every frequency bin so that a separated signal in the time domain contains frequency components from one source signal. This problem is known as the permutation problem.

C.G. Puntonet and A. Prieto (Eds.): ICA 2004, LNCS 3195, pp. 461–469, 2004.

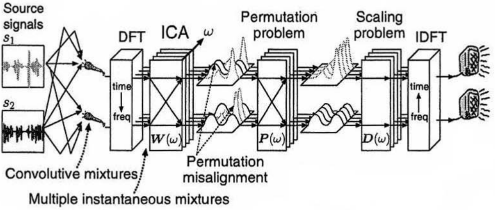


Fig. 1. Flow of frequency domain BSS

Many methods have been proposed for solving the permutation problem, and the use of geometric information, such as beam patterns [2–4], direction of arrival (DOA) and source locations [5], is an effective approach. We have proposed a robust method that combines the DOA based method [2, 3] and the correlation based method [6], which almost completely solves the problem for 2-source cases [7]. However it is insufficient when the number of signals is large or when the signals come from the same or similar direction. In this paper, we propose a method for obtaining proper geometric information for solving the permutation problem in such cases.

2 Frequency Domain BSS Using ICA

When the source signals are $s_i(t)(i = 1, ..., N)$, the signals observed by sensor 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: $x_j(t) = \sum_{i=1}^{N}(h_{ji} * s_i)(t)$, $y_k(t) = \sum_{j=1}^{M}(w_{kj} * x_j)(t)$, where h_{ji} is the impulse response from source i to sensor j, w_{kj} are the separating filters, and $*$ denotes the convolution operator. Figure 1 shows the flow of BSS 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 independent component analysis (ICA) algorithm [8] in the frequency domain to solve a BSS problem in a reverberant environment. Using a short-time discrete Fourier transform, the model is approximated as: $\mathbf{X}(\omega, m) = \mathbf{H}(\omega)\mathbf{S}(\omega, m)$, where, ω is the angular frequency, and n represents the frame index. The separating process can be formulated in each frequency bin as: $\mathbf{Y}(\omega, m) = \mathbf{W}(\omega)\mathbf{X}(\omega, m)$, where $\mathbf{S}(\omega, m) = [S_1(\omega, m), ..., S_N(\omega, m)]^T$ is the source signal in frequency bin ω, $\mathbf{X}(\omega, m) = [X_1(\omega, m), ..., X_M(\omega, m)]^T$ denotes the observed signals, $\mathbf{Y}(\omega, m) = [Y_1(\omega, m), ..., Y_N(\omega, m)]^T$ is the estimated source signal, and $\mathbf{W}(\omega)$ represents the separating matrix. $\mathbf{W}(\omega)$ is determined so that $Y_i(\omega, m)$ and $Y_j(\omega, m)$ 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) [9], and we can use it. On the other hand, the permutation problem is complicated, especially when the number of source signals is large.

3 Geometric Information for Solving Permutation Problem

3.1 Invariant in ICA Solution

If a separating matrix $\mathbf{W}(\omega)$ is calculated successfully and it extracts source signals with scaling ambiguity, $\mathbf{D}(\omega)\mathbf{W}(\omega)\mathbf{H}(\omega) = \mathbf{I}$ holds (except for singular frequency bins). Because of the scaling ambiguity, we cannot obtain $\mathbf{H}(\omega)$ simply from the ICA solution. However, the ratio of elements in the same column $H_{ji}/H_{j'i}$ is invariable in relation to $\mathbf{D}(\omega)$, and given by

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

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 is used to solve the permutation problem.

If we have more sensors than sources ($N < M$), principal component analysis (PCA) is performed as a preprocessing of ICA [10] so that the N dimensional subspace spanned by the row vectors of $\mathbf{W}(\omega)$ is almost identical to the signal subspace, and the Moore-Penrose pseudo-inverse $\mathbf{W}^{+} \triangleq \mathbf{W}^T(\mathbf{W}\mathbf{W}^T)^{-1}$ is used instead of $\mathbf{W}^{-1}$.

3.2 DOA Estimation with ICA Solution

We can estimate the DOA of source signals by using the above invariant $H_{ji}/H_{j'i}$ [7]. With a farfield model, a frequency response is formulated as:

$$H_{ji}(\omega) = e^{\jmath\omega c^{-1}\mathbf{a}_i^T\mathbf{p}_j}, \tag{2}$$

where c is the speed of wave propagation, $\mathbf{a}_i$ is a unit vector that points to the direction of source i, and $\mathbf{p}_j$ represents the location of sensor j. According to this model, we have

$$\begin{align} H_{ji}/H_{j'i} &= e^{\jmath\omega c^{-1}\mathbf{a}_i^T(\mathbf{p}_j-\mathbf{p}_{j'})} \tag{3}\\ &= e^{\jmath\omega c^{-1}\|\mathbf{p}_j-\mathbf{p}_{j'}\|\cos\theta_{i,jj'}}, \tag{4}\end{align}$$

where $\theta_{i,jj'}$ is the direction of source i relative to the sensor pair j and j'. By using the argument of (4) and (1), we can estimate:

$$\begin{align} \hat{\theta}_{i,jj'} &= \arccos\frac{\arg(H_{ji}/H_{j'i})}{\omega c^{-1}\|(\mathbf{p}_j-\mathbf{p}_{j'})\|} \nonumber\\ &= \arccos\frac{\arg([\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i})}{\omega c^{-1}\|(\mathbf{p}_j-\mathbf{p}_{j'})\|}. \tag{5}\end{align}$$

This procedure is valid for sensor pairs with a small spacing.

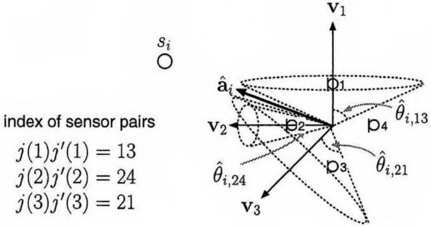


Fig. 2. Solving ambiguity of estimated DOAs

3.3 Ambiguity of DOA Estimation

DOA estimation involves some ambiguities. When we use only one pair of sensors or a linear array, the estimated $\hat{\theta}_{i,jj'}$ determines a cone rather than a direction. If we assume a horizontal plane on which sources exist, the cone is reduced to two half-lines. However, the ambiguity of two directions that are symmetrical with respect to the axis of the sensor pair still remains. This is a fatal problem when the source locations are omnidirectional.

When the spacing between sensors is larger than half a wavelength, spatial aliasing causes another ambiguity, but we do not consider this here.

3.4 Solving Ambiguity of DOA Estimation

The ambiguity can be solved by using multiple sensor pairs. If we use sensor pairs that have different axis directions, we can estimate cones with various vertex angles for one source direction. If the *relative* DOA $\hat{\theta}_{i,jj'}$ is estimated without any error, the *absolute* direction of the source signal $\mathbf{a}_i$ satisfies:

$$\frac{(\mathbf{p}_j - \mathbf{p}_{j'})^T \mathbf{a}_i}{\|\mathbf{p}_j - \mathbf{p}_{j'}\|} = \cos \hat{\theta}_{i,jj'}. \tag{6}$$

When we use L sensor pairs whose indexes are $j(l)j'(l)(1 \leq l \leq L)$, $\mathbf{a}_i$ is given by the solution of the following equation:

$$\mathbf{V}\mathbf{a}_i = \mathbf{c}_i, \tag{7}$$

where $\mathbf{v}_l \triangleq \frac{\mathbf{p}_{j(l)} - \mathbf{p}_{j'(l)}}{\|\mathbf{p}_{j(l)} - \mathbf{p}_{j'(l)}\|}$ is a normalized axis, $\mathbf{V} \triangleq (\mathbf{v}_1, ..., \mathbf{v}_L)^T$, and $\mathbf{c}_i \triangleq [\cos(\hat{\theta}_{i,j(1)j'(1)}), ..., \cos(\hat{\theta}_{i,j(L)j'(L)})]^T$. Sensor pairs should be selected so that $\text{rank}(\mathbf{V}) \geq 3$ if potential source locations are three-dimensional, or $\text{rank}(\mathbf{V}) \geq 2$ if we assume a plane on which sources exist.

Actually, $\hat{\theta}_{i,j(l)j'(l)}$ has an estimation error, and (7) has no solution. Thus we adopt an optimal solution by employing certain criteria such as:

$$\hat{\mathbf{a}}_i = \underset{\mathbf{a}}{\text{argmin}} \, ||\mathbf{V}\mathbf{a} - \mathbf{c}_i|| \quad (\text{subject to } \; ||\mathbf{a}|| = 1) \tag{8}$$

This can be solved approximately by using the Moore-Penrose pseudo-inverse $\mathbf{V}^+ \triangleq (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$, and we have:

$$\hat{\mathbf{a}}_i \approx \frac{\mathbf{V}^+\mathbf{c}_i}{||\mathbf{V}^+\mathbf{c}_i||}. \tag{9}$$

Accordingly, we can determine a unit vector $\hat{\mathbf{a}}_i$ pointing to the direction of source s_i (Fig. 2).

3.5 Estimation of Sphere with ICA Solution

The interpretation of the ICA solution with a nearfield model yields other geometric information [11]. When we adopt the nearfield model, including the attenuation of the wave, $H_{ji}(\omega)$ is formulated as:

$$H_{ji}(\omega) = \frac{1}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{\jmath\omega c^{-1}(\|\mathbf{q}_i - \mathbf{p}_j\|)} \tag{10}$$

where $\mathbf{q}_i$ represents the location of source i. By taking the ratio of (10) for a pair of sensors j and j' we obtain:

$$H_{ji}/H_{j'i} = \frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{\jmath\omega c^{-1}(\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i - \mathbf{p}_{j'}\|)}. \tag{11}$$

By using the modulus of (11) and (1), we have:

$$\frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} = \left|\frac{[\mathbf{W}^{-1}]_{ji}}{[\mathbf{W}^{-1}]_{j'i}}\right|. \tag{12}$$

By solving (11) for $\mathbf{q}_i$, we have a sphere whose center $O_{i,jj'}$ and radius $R_{i,jj'}$ are given by:

$$O_{i,jj'} = \mathbf{p}_j - \frac{1}{r_{i,jj'}^2 - 1}(\mathbf{p}_{j'} - \mathbf{p}_j), \tag{13}$$

$$R_{i,jj'} = \|\frac{r_{i,jj'}}{r_{i,jj'}^2 - 1}(\mathbf{p}_{j'} - \mathbf{p}_j)\|, \tag{14}$$

where $r_{i,jj'} \triangleq |[\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i}|$. Thus, we can estimate a sphere $(\hat{O}_{i,jj'}, \hat{R}_{i,jj'})$ on which $\mathbf{q}_i$ exists by using the result of ICA $\mathbf{W}$ and the locations of the sensors $\mathbf{p}_j$ and $\mathbf{p}_{j'}$. Figure 3 shows an example of the spheres determined by (12) for various ratios $r_{i,jj'}$. This procedure is valid for sensor pairs with a large spacing.

3.6 Solving Permutation Problem

We solve the permutation problem by classification using the geometric information together with a correlation based method. This is similar to our previously reported proposal [7].

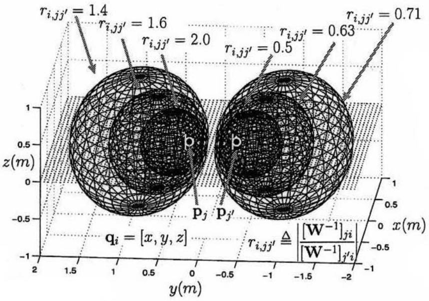


Fig. 3. Example of spheres determined by (12) ($\mathbf{p}_j = [0, 0.3, 0]$, $\mathbf{p}_{j'} = [0, -0.3, 0]$)

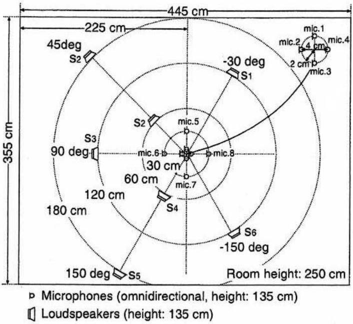
The models (2) and (10) are simple approximations without multi-path propagation and reverberation, however we can use them to obtain information for classifying signals. Even when some signals come from the same or a similar direction, we can distinguish between them by using the information obtained by the method described in Sec.3.5. The source locations can be estimated by combining the estimated direction and spheres. Then, we can classify separated signals in the frequency domain according to the estimated source locations.

Unfortunately, classification on the basis of the estimated location tends to be inconsistent especially in a reverberant environment. In many frequency bins, several signals are assigned to the same cluster, and such classification is inconsistent. We solve the permutation only for frequency bins with a consistent classification, and we employ a correlation based method for the rest. The correlation based method solves the permutation so that the inter-frequency correlation for neighboring or harmonic frequency bins is maximized.

4 Experiments

We carried out experiments with 6 sources and 8 microphones using speech signals convolved with impulse responses measured in a room with reverberation time of 130 ms. The room layout and other experimental conditions are shown in Fig. 4. We assume that the number of source signals $N = 6$ is known. The experimental procedure is as follows.

First, we apply ICA to $x_j(t) (j = 1, ..., 8)$, and calculate separating matrix $\mathbf{W}(\omega)$ for each frequency bin. The initial value of $\mathbf{W}(\omega)$ is calculated by PCA. Then we estimate DOAs by using the rows of $\mathbf{W}^{+}(\omega)$ (pseudo-inverse) corresponding to the small spacing microphone pairs (1-3, 2-4, 1-2 and 2-3). Figure 5 shows a histogram of the estimated DOAs. We can find five clusters in this histogram, and one cluster is twice the size of the others. This implies that two



Reverberation time	$T_R = 130$ ms
Sampling rate	8 kHz
Data length	6 s
Window	hanning
Frame length	2048 points (256 ms)
Frame shift	512 points (64 ms)
ICA algorithm	Infomax (complex valued)
Number of iterations	100

Fig. 4. Room layout and experimental conditions

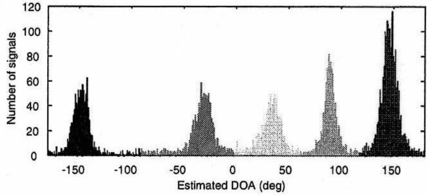


Fig. 5. Histogram of estimated DOAs obtained by using small spacing microphone pairs

signals come from the same direction (about 150°). We can solve the permutation problem for other four sources by using this DOA information (Fig. 6(a)).

Then, we apply the estimation of spheres to the signals that belong to the large cluster by using the rows of $\mathbf{W}^{+}(\omega)$ corresponding to the large spacing microphone pairs (7-5, 7-8, 6-5 and 6-8). Figure 6(b) shows estimated radiuses for S_4 and S_5 for the microphone pair 7-5. Although the radius estimation includes a large error, it provides sufficient information to distinguish two signals. Finally, we can classify the signals into six clusters. We determine the permutation only for frequency bins with a consistent classification, and we employ a correlation based method for the rest. In addition, we use the spectral smoothing method proposed in [12] to construct separating filters in the time domain from the ICA result in the frequency domain.

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

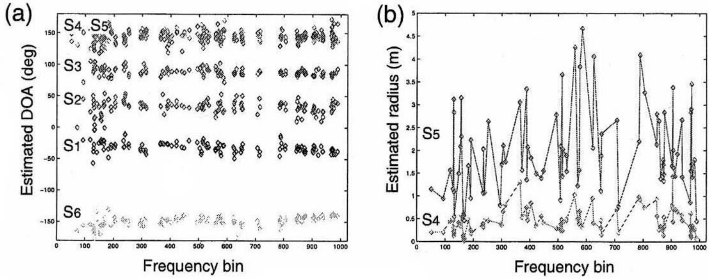


Fig. 6. Permutation solved by using (a) DOAs and (b) estimated radiuses

Table 1. Experimental results (dB), T_R=130 ms

	SIR_1	SIR_2	SIR_3	SIR_4	SIR_5	SIR_6	ave.
Input SIR	-8.3	-6.8	-7.8	-7.7	-6.7	-5.2	-7.1
C	4.4	2.6	4.0	9.2	3.6	-2.0	3.7
D+C	9.6	9.3	14.7	2.7	6.5	14.0	9.4
D+S+C	10.8	10.4	14.5	7.0	11.0	12.2	11.0

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

We measured SIRs for three permutation solving strategies: the correlation based method ("C"), estimated DOAs and correlation ("D+C"), and a combination of estimated DOAs, spheres and correlation ("D+S+C", proposed method). We also measured input SIRs by using the mixture observed by microphone 1 for the reference ("Input SIR"). The results are summarized in Table 1.

Our proposed method succeeded in separating six speech signals. It can be seen that the discrimination obtained by using estimated spheres is effective in improving the separation performance for signals coming from the same direction.

5 Conclusion

We proposed using a combination of small and large spacing microphone pairs with various axis directions to obtain proper geometric information for solving the permutation problem in frequency domain BSS. In experiments (T_R=130 ms), our method succeeded in the separation of six speech signals, even when two came from the same direction. The computation time was about 1 minute for 6 seconds of data. Some sound examples can be found on our web site [13].

References

1. Matsuoka, K., Ohba, Y., Toyota, Y., Nakashima, S.: Blind separation for convolutive mixture of many voices. In: Proc. IWAENC 2003. (2003) 279–282
2. Kurita, S., Saruwatari, H., Kajita, S., Takeda, K., Itakura, F.: Evaluation of blind signal separation method using directivity pattern under reverberant conditions. In: Proc. ICASSP 2000. (2000) 3140–3143
3. Ikram, M.Z., Morgan, D.R.: A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation. In: Proc. ICASSP 2002. (2002) 881–884
4. Parra, L.C., Alvino, C.V.: Geometric source separation: Merging convolutive source separation with geometric beamforming. IEEE Trans. Speech Audio Processing **10** (2002) 352–362
5. Soon, V.C., Tong, L., Huang, Y.F., Liu, R.: A robust method for wideband signal separation. In: Proc. ISCAS '93. (1993) 703–706
6. Asano, F., Ikeda, S., Ogawa, M., Asoh, H., Kitawaki, N.: A combined approach of array processing and independent component analysis for blind separation of acoustic signals. In: Proc. ICASSP 2001. (2001) 2729–2732
7. Sawada, H., Muaki, R., Araki, S., Makino, S.: A robust and precise method for solving the permutation problem of frequency-domain blind source separation. IEEE Trans. Speech Audio Processing **12** (2004)
8. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)
9. Matsuoka, K., Nakashima, S.: Minimal distortion principle for blind source separation. In: Proc. ICA 2001. (2001) 722–727
10. Winter, S., Sawada, H., Makino, S.: Geometrical understanding of the PCA subspace method for overdetermined blind source separation. In: Proc. ICASSP 2003. Volume 5. (2003) 769–772
11. Mukai, R., Sawada, H., Araki, S., Makino, S.: Near-field frequency domain blind source separation for convolutive mixtures. In: Proc. ICASSP 2004. (2004)
12. Sawada, H., Mukai, R., de la Kethulle, S., Araki, S., Makino, S.: Spectral smoothing for frequency-domain blind source separation. In: Proc. IWAENC 2003. (2003) 311–314
13. http://www.kecl.ntt.co.jp/icl/signal/mukai/demo/ica2004/