A SOLUTION FOR THE PERMUTATION PROBLEM OF OVERDETERMINED SOURCE SEPARATION USING SUBSPACE METHOD

[1]*Shunichiro Hirayanagi and* [2]*Nozomu Hamada*

[1]hirayanagi@hamada.sd.keio.ac.jp
[12]Signal Processing Lab., School of Integrated Design Engineering, Keio University
Hiyoshi 3-14-1, Kohoku-ku, Yokohama, 223-8522 Japan

ABSTRACT

This paper proposes a new method to solve the permutation problem in the frequency domain source separation based on Independent Component Analysis (ICA), for overdetermined case, namely the case where the number of microphones is larger than the number of sources. The method is based on the directivity patterns of separation filters and the subspace method to estimate the source directions. The source directions estimation method uses a result of Principal Component Analysis (PCA) as preprocess of ICA. The separation experiments in a real environment show that the proposed method gives good performance for the permutation problem in the overdetermined case.

1. INTRODUCTION

Source separation is a technique that separates mixed sound signals into each source signals. In recent years, methods based on ICA have been actively researched. In these methods, signals are separated depending on the statistical independency of source signals.

The separation problem is formulated as follows. Assume that N source signals are observed by M microphones under the assumption $M > N$. The mixing model is generally a convolutive mixture, but we can treat this as instantaneous mixtures in the frequency domain by applying Short Time discrete Fourier Transform (STFT). The mixing model is described as

$$\mathbf{x}(\omega, n) = \mathbf{A}(\omega)\mathbf{s}(\omega, n) + \mathbf{n}(\omega, n) \tag{1}$$

where ω is the integer denoting the frequency bin, n is the frame index, $\mathbf{x}(\omega, n) = [x_1(\omega, n), \ldots, x_M(\omega, n)]^T$ denotes the observed signal vector, $\mathbf{s}(\omega, n) = [s_1(\omega, n), \ldots, s_N(\omega, n)]^T$ denotes the source signal vector, $\mathbf{A}(\omega)$ denotes a $M \times N$ mixing matrix and $\mathbf{n}(\omega, n)$ is Gaussian noise vector. The separated output signal vector $\mathbf{y}(\omega, n) = [y_1(\omega, n), \ldots, y_N(\omega, n)]^T$ is obtained as

$$\mathbf{y}(\omega, n) = \mathbf{W}(\omega)\mathbf{x}(\omega, n) \tag{2}$$

where $\mathbf{W}(\omega)$ is a $N \times M$ separation matrix. In the ICA framework, $\mathbf{W}(\omega)$ is determined adaptively so that the output signals are mutually independent.

In the overdetermined source separation problem, namely the case where the number of microphones is larger than the number of sources, the more microphones generally improve the separation performance. There are a lot of methods for solving the overdetermined problem such as methods using subarray [1] [2]. The typical one uses the subspace method such as PCA for the preprocess of ICA [3] [4] [5]. Followings are the advantages of using PCA as the preprocessing of ICA. 1) Since the observed signals are pre-whitened at first, the convergence speed of ICA is improved. (This advantage is not restricted to the overdetermined case.) 2) The dimension of ICA input is reduced from M to N. 3) The influence of the noise is more suppressed as M is larger than N. We adopt PCA not only for these purposes but also for source directions estimation described later. The detail of PCA is discussed in Sect.2.1.

If separation matrix $\mathbf{W}(\omega)$ is obtained by ICA, $\mathbf{P}(\omega)\mathbf{W}(\omega)$ gives another solution for any permutation matrix $\mathbf{P}(\omega)$. The permutation matrix $\mathbf{P}(\omega)$ exchanges the row vectors of $\mathbf{W}(\omega)$. In the frequency-domain ICA, this causes the remixture of the output signals in the time domain even though the signals are separated in each frequency bin. Many methods solving the permutation problem have been proposed.

One of them is the method using the envelope correlation of output signals [6]. In this method, one mis-permutation at certain frequency bin causes mis-permutation in other frequency bins. The method using directivity patterns[7] is another effective method. This method estimates the source directions by searching peaks of the directivity patterns of separation filters. Then we can judge on which source each filter is focusing by the estimated source directions and the directivity patterns so that the permutation problem can be solved. The detail of this method is described in Sect.2.3.

In this paper, the source directions are estimated by using the direction of arrival (DOA) estimation method based on the orthogonality of the noise subspace and the signal subspace as in MUSIC method [8]. DOA estimation is achieved by using a result of PCA processes. The proposed method is described in Sect.3.

2. SOURCE SEPARATION

2.1. Principal component analysis

The PCA uses the spatial correlation matrix $\mathbf{R}(\omega)$ of $\mathbf{x}(\omega, n)$ defined as

$$\mathbf{R}(\omega) = E[\mathbf{x}(\omega, n)\mathbf{x}^H(\omega, n)] \tag{3}$$

The eigenvalues of $\mathbf{R}(\omega)$ are denoted as $\lambda_1(\omega), \ldots, \lambda_M(\omega)$ ($\lambda_1(\omega) \geq, \cdots, \geq \lambda_M(\omega)$) and the corresponding eigenvectors are denoted as $\mathbf{e}_1(\omega), \ldots, \mathbf{e}_M(\omega)$. When the number of active sources is N and the number of microphones is M, it is generally satisfied that

$$\lambda_1(\omega), \cdots, \lambda_N(\omega) \gg \lambda_{N+1}(\omega), \cdots, \lambda_M(\omega) \tag{4}$$

The vectors $\mathbf{e}_1(\omega), \ldots, \mathbf{e}_N(\omega)$ are the basis vectors spanned the source signal subspace, and $\mathbf{e}_{N+1}(\omega), \ldots, \mathbf{e}_M(\omega)$ are the basis vectors spanned the noise subspace. The PCA matrix is determined as

$$\mathbf{W}_{PCA}(\omega) = \mathbf{\Lambda}^{-\frac{1}{2}}(\omega)\mathbf{E}^H(\omega) \tag{5}$$

where $\mathbf{\Lambda}(\omega) = \text{diag}(\lambda_1(\omega), \ldots, \lambda_N(\omega))$, $\mathbf{E}(\omega) = [\mathbf{e}_1(\omega), \ldots, \mathbf{e}_N(\omega)]$. We obtain the output signals as follows

$$\mathbf{y}_{PCA}(\omega, n) = \mathbf{W}_{PCA}(\omega)\mathbf{x}(\omega, n) \tag{6}$$

2.2. Independent component analysis

The ICA algorithm used in this paper is the Infomax algorithm [3]. Its learning rule is described as

$$\mathbf{W}_{i+1}(\omega) = \mathbf{W}_i(\omega) + \eta[\mathbf{I} - \phi(\mathbf{y}(\omega, n))\mathbf{y}^H(\omega, n)]\mathbf{W}_i(\omega) \tag{7}$$

where η is the step size, $[\cdot]^H$ denotes the Hermitian transposition and $\phi(\cdot)$ denotes the score function. The score function in this paper is defined as

$$\phi(y) = 2\tanh(G \cdot \text{Re}(y)) + 2j\,\tanh(G \cdot \text{Im}(y)) \tag{8}$$

where G is the gain constant. We finally obtain the converged matrix $\mathbf{W}_i(\omega)$ denoted by $\mathbf{W}_{ICA}(\omega)$.

The entire separation matrix $\mathbf{W}(\omega)$ is given as follows

$$\mathbf{W}(\omega) = \mathbf{W}_{ICA}(\omega)\mathbf{W}_{PCA}(\omega) \tag{9}$$

Thus, the output signals are obtained as

$$\mathbf{y}(\omega, n) = \mathbf{W}(\omega)\mathbf{x}(\omega, n) \tag{10}$$

Each row vector of $\mathbf{W}(\omega)$ is the separation filter focusing on each source.

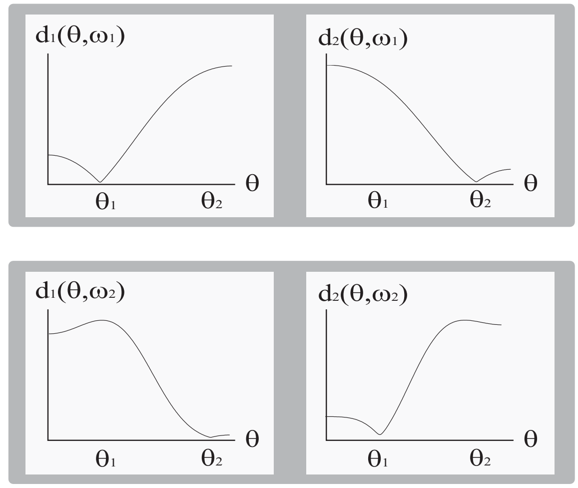

Figure 1: Examples of directivity pattern

2.3. Conventional solution

In [7], the method for permutation using the directivity patterns of separation filters is proposed. Assuming that a steering vector is $\mathbf{v}(\theta, \omega)$, which is determined uniquely by the array arrangement, the directivity patterns are defined as

$$\mathbf{d}(\theta, \omega) = |\mathbf{W}(\omega)\mathbf{v}(\theta, \omega)| \tag{11}$$

,and $d_n(\theta, \omega)$, which is the nth element of $\mathbf{d}(\theta, \omega)$, denotes the directivity pattern of nth separation filter. ICA-based separation filter and the adaptive beamformer have a similar goal to suppress signals located in certain directions. So analyzing the directivity patterns of the separation filters is useful to judge on which source each filter is focusing. For example, Fig.1 shows the directivity patterns in the case of two sources and two microphones at different ω. Searching the directions giving nulls, we can estimate the source directions θ_1 and θ_2. Then we exchange the channels of filter at each frequency bin in order that nth output channel focuses on the nth source throughout all frequency bin. This is the basic idea to solve the permutation problem in [7].

In the overdetermined problem, the method of [7] is not always effective, because the directions giving the nulls of the directivity patterns of $\mathbf{W}(\omega)$ do not always correspond to the source directions. Fig.2 shows an example of directivity patterns of $\mathbf{W}(\omega)$ in the overdetermined case. Since $\mathbf{W}(\omega)$ steers the nulls toward not only the source directions, therefore directivity pattern of $\mathbf{W}(\omega)$ fails to estimate the source directions. In the next section, we adopt the subspace method to estimate the source directions.

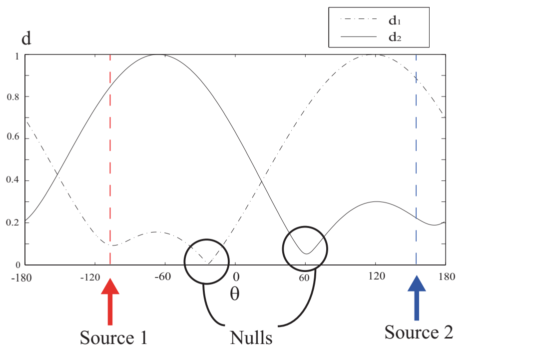

Figure 2: Directivity pattern of $\mathbf{W}(\omega)$ in an overdetermined case

3. PROPOSED METHOD

In this section, we consider the case of $N = 2$, $M = 3$. As the result of PCA, the eigenvectors of $\mathbf{R}(\omega)$ denoted by $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are obtained. The vector $\mathbf{e}_3$ is the basis of the noise subspace. Thus, $\mathbf{e}_3^H$ is used as a filter that suppresses all source signals. This means that the directivity pattern of $\mathbf{e}_3^H$ is applicable to estimate the source directions as shown in Fig.3. The directions giving nulls correspond to the source directions accurately. This DOA estimation method is same as the MUSIC method [8]. The process is as follows. At first we calculate the directivity patterns of $\mathbf{e}_3^H(\omega_n)$ for all ω_n.

$$\mathbf{d}(\theta, \omega_n) = |\mathbf{e}_3^H(\omega_n)\mathbf{v}(\theta, \omega_n)| \tag{12}$$

Next we search the directions $\theta_1(\omega_n)$,$\theta_2(\omega_n)$ ($\theta_1(\omega_n) < \theta_2(\omega_n)$) giving nulls. Then we can estimate the source directions $\hat{\theta}_1$,$\hat{\theta}_2$ as the average of $\theta_1(\omega_n)$,$\theta_2(\omega_n)$ with respect to ω_n.

$$\hat{\theta}_1 = \text{average}[\theta_1(\omega_n)] \tag{13}$$

$$\hat{\theta}_2 = \text{average}[\theta_2(\omega_n)] \tag{14}$$

Using the estimated source directions $\hat{\theta}_1$ and $\hat{\theta}_2$, we compare $d_1(\hat{\theta}_1, \omega_n)$ with $d_2(\hat{\theta}_1, \omega_n)$. If $d_1(\hat{\theta}_1, \omega_n) > d_2(\hat{\theta}_1, \omega_n)$, the channel 1 filter is focusing on the source 1 at ω_n. And if $d_1(\hat{\theta}_1, \omega_n) < d_2(\hat{\theta}_1, \omega_n)$, the channel 2 filter is focusing on the source 1 at ω_n. Then the channel exchange process follows. The flow of the proposed system is shown in Fig.4.

4. EXPERIMENT RESULT

Separation experiments are performed in a real environment. The experimental situation is as follows. Three microphones are located in the shape of a triangle and source signals are about 3 sec long voice signals in Japanese provided by ASJ (Acoustical Society of Japan) Continuous

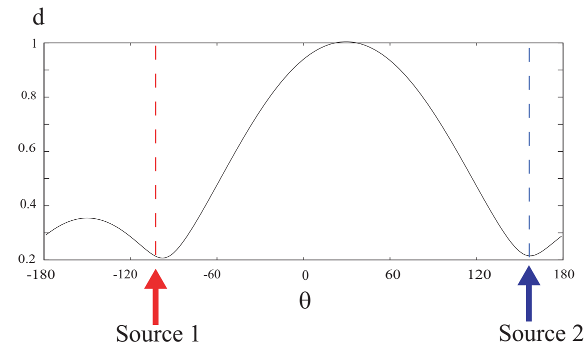

Figure 3: Directivity pattern of $\mathbf{e}_3^H(\omega)$

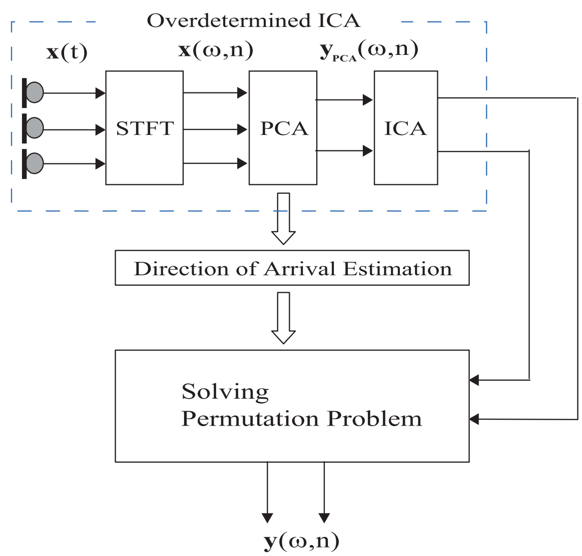

Figure 4: Flow of the entire system

Speech Corpus for Research Database. We use two sources and each source is located on the every $15°$ position, $0°, 15°, \ldots, 360°$. It is assumed that the array arrangement and the number of sources are known. The physical parameters of the recording room are shown in Fig.5. The parameters are shown in Tab.1.

Fig.6 (a),(b) show the rate of incorrect permutation in the case of using conventional method [7] and using the proposed method respectively. For the purpose of the evaluation, we derive “correct permutation” from the cross correlation between the output spectrogram and the source spectrogram (provided in the experiment) [3]. Fig.6 (b) shows that the permutation problem is almost solved by the proposed method except in the case that two sources are closely located.

Fig.7 shows the average of output Signal to Interference Ratio (SIR). SIR is defined as the ratio between the power of the desired signal and the power of the interference signal. The proposed method provides improved

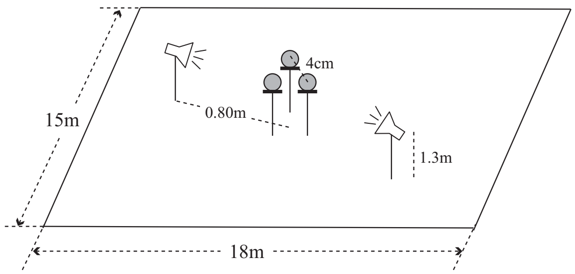

Figure 5: Recording environment

Table 1: Parameters

sampling frequency	8000Hz
window length of STFT	256
window shift of STFT	32
window function	hamming
step size η	0.001
G in score function ϕ	100

SIR, especially when source direction difference is within $[60°, 90°]$.

5. CONCLUSION

A method to solve the permutation problem is proposed by combining the method using directivity pattern and the subspace DOA estimation method. Experiments show that the proposed method gives higher rate of correct permutation.

6. REFERENCES

[1] A. Koutras, E. Dermatas and G.Kokkinakis, "Improving simultaneous speech recognition in real room environments using overdetermined blind source separation," *Eurospeech2001*, pp.1009-1012, Sept. 2001.

[2] T. Nishikawa, H. Abe, H. Saruwatari, K. Shikano, A. Kaminuma, "Overdetermined Blind Separation for Real Convolutive Mixtures of Speech Based on Multistage ICA Using Subarray Processing," IEICE Trans. Fundamentals, Vol.E87-A, No.8, pp. 1924–1932, August 2004.

[3] F. Asano, S. Ikeda, M. Ogawa, H. Aso, and N. Kitawaki, "Combined approach of array processing and independent component analysis for blind separation of acoustic signal," IEEE Trans. on Speech and Audio processing, pp.204-215, May.2003

[4] Joho M., Mathis H., Lambert R. H.,"Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture", Proceedings of ICA 2000, pp. 81-86, Helsinki, Finland, June 19-22, 2000.

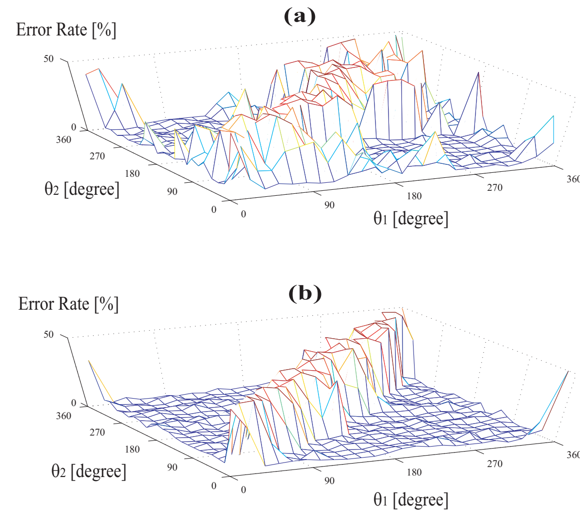

Figure 6: Error rate of the permutation at each source direction: (a) conventional method (b) proposed method

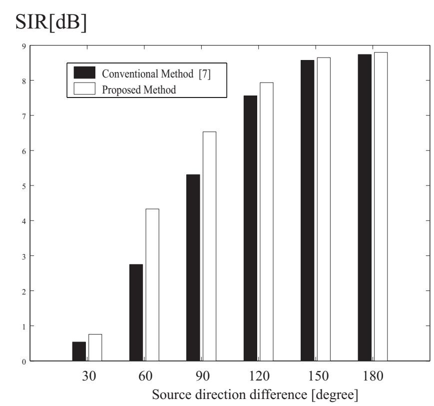

Figure 7: Average of SIR

[5] F.Asano, Y.Motomura, H.Asoh, T.Matsui, "Effect of PCA filter in blind source separation", Proc. of of Int. conf. on Independent Component Analysis (ICA2000), 2000.6.19

[6] 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, August 2001.

[7] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakkura, "Evaluation of Blind Signal Separation Method Using Directivity Pattern Under Reverberant Conditions ," *ICASSP2000*, pp.3140-3143, 2000.

[8] H. Krim, M. Viberg. "Two Decades of Array Signal Processing Research: The Parametric Approach", IEEE Signal Processing Magazine, 13(4):67-94, July 1996.