ARRAY GEOMETRY ARRANGEMENT FOR FREQUENCY DOMAIN BLIND SOURCE SEPARATION

Ryo Mukai   Hiroshi Sawada   Sébastien de la Kethulle de Ryhove   Shoko Araki   Shoji Makino

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

ABSTRACT

In this paper, we propose 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. Geometric information such as direction of arrival is helpful for solving the permutation problem, but the information becomes more uncertain as the number of source signals increases. When we use a linear microphone array, we cannot obtain reliable geometric information due to the ambiguity and sensitivity inherent to the array geometry. We propose a combination of small and large spacing microphone pairs that have various axis directions. Experimental results show that the proposed method can separate a mixture of speech signals that come from various directions, even when some come from the same direction.

1. INTRODUCTION

Blind source separation (BSS) is a technique for estimating original source signals using only observed mixtures. 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: $x_j(t) = \sum_{i=1}^{N}(h_{ji} * s_i)(t)$, $y_k(t) = \sum_{j=1}^{M}(w_{kj} * x_j)(t)$, where h_{ji} is the impulse response from source i to sensor j, w_{kj} are the separating filters, and $*$ denotes the convolution operator. 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 the frequency domain to solve a BSS problem in a reverberant environment. Using a short-time discrete Fourier transform, the model is approximated as: $\mathbf{X}(\omega, n) = \mathbf{H}(\omega)\mathbf{S}(\omega, n)$, where, ω is the angular frequency, and n represents the frame index. The separating process can be formulated in each frequency bin as: $\mathbf{Y}(\omega, n) = \mathbf{W}(\omega)\mathbf{X}(\omega, n)$, where $\mathbf{S}(\omega, n) = [S_1(\omega, n), ..., S_N(\omega, n)]^T$ is the source signal in frequency bin ω, $\mathbf{X}(\omega, n) = [X_1(\omega, n), ..., X_M(\omega, n)]^T$ denotes the observed signals, $\mathbf{Y}(\omega, n) = [Y_1(\omega, n), ..., Y_N(\omega, n)]^T$ 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 $Y_j(\omega, n)$ become mutually independent.

The ICA solution suffers scaling and permutation ambiguities. This is due to the fact that if $\mathbf{W}(\omega)$ is a solution, then $\mathbf{P}(\omega)\mathbf{D}(\omega)\mathbf{W}(\omega)$ 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 scaling and permutation problems to reconstruct separated signals in the time domain.

There is a simple and reasonable solution for the scaling problem: $\mathbf{D}(\omega) = \text{diag}(\mathbf{W}^{-1}(\omega))$, which is obtained by the minimal distortion principle (MDP) [2], and we can use it. On the other hand, the permutation problem is complicated, especially when the number of source signals is large.

2. GEOMETRIC INFORMATION FOR SOLVING PERMUTATION PROBLEM

Many methods have been proposed for solving the permutation problem, and the use of geometric information, such as direction of arrival (DOA) and beam patterns, is one effective approach [3, 4, 5, 6]. We have proposed a robust method by combining the correlation based method [7] and the DOA based method [4, 5], which almost completely solves the problem for 2-source cases [8]. However it is insufficient when the number of signals is large or when the signals come from a similar direction. In this paper, we propose a method for obtaining proper geometric information for solving the permutation problem in such cases.

2.1. Invariant in ICA solution

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

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

where $[\cdot]_{ji}$ denotes 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 utilized for solving the permutation problem.

2.2. DOA estimation with ICA solution

We can estimate the DOA of source signals by using the above invariant [9]. With a farfield model, a frequency response is formulated as:

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

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 a location of sensor j. According to this model, we have

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

where $\theta_{i,jj'}$ is the direction of source i relative to the sensor pair j and j'. By using the argument of (3) and (1), we can estimate:

$$
\begin{aligned}
\hat{\theta}_{i,jj'} &= \cos^{-1}\frac{\arg(H_{ji}/H_{j'i})}{\omega c^{-1}\|(\mathbf{p}_j-\mathbf{p}_{j'})\|} \\
&= \cos^{-1}\frac{\arg([\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i})}{\omega c^{-1}\|(\mathbf{p}_j-\mathbf{p}_{j'})\|}. \qquad (4)
\end{aligned}
$$

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

2.3. Estimation of sphere with ICA solution

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

$$H_{ji}(\omega) = \frac{1}{\|\mathbf{q}_i-\mathbf{p}_j\|}e^{j\omega c^{-1}(\|\mathbf{q}_i-\mathbf{p}_j\|)} \qquad (5)$$

where $\mathbf{q}_i$ represents the location of source i. By taking the ratio of (5) for a pair of sensors j and j' we obtain:

$$H_{ji}/H_{j'i} = \frac{\|\mathbf{q}_i-\mathbf{p}_{j'}\|}{\|\mathbf{q}_i-\mathbf{p}_j\|}e^{j\omega c^{-1}(\|\mathbf{q}_i-\mathbf{p}_j\|-\|\mathbf{q}_i-\mathbf{p}_{j'}\|)}. \qquad (6)$$

By using the modulus of (6) and (1), we have:

$$\frac{\|\mathbf{q}_i-\mathbf{p}_{j'}\|}{\|\mathbf{q}_i-\mathbf{p}_j\|} = \left|\frac{[\mathbf{W}^{-1}]_{ji}}{[\mathbf{W}^{-1}]_{j'i}}\right|. \qquad (7)$$

By solving (6) for $\mathbf{q}_i$, we have a sphere whose center $O_{i,jj'}$ and radius $R_{i,jj'}$ are given by:

$$O_{i,jj'} = \mathbf{p}_j - \frac{1}{r_{i,jj'}^2-1}(\mathbf{p}_{j'}-\mathbf{p}_j), \qquad (8)$$

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

where $r_{i,jj'} = |[\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i}|$. 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 1 shows an example of the spheres determined by (7) for various ratios $r_{i,jj'}$. This procedure is valid for sensor pairs with a large spacing.

3. SENSITIVITY AND AMBIGUITY IN SOURCE LOCATION ESTIMATION

3.1. Sensitivity of DOA estimation

The BSS performance is influenced by the source signal location. Figure 2 shows the result of a preliminary exper-
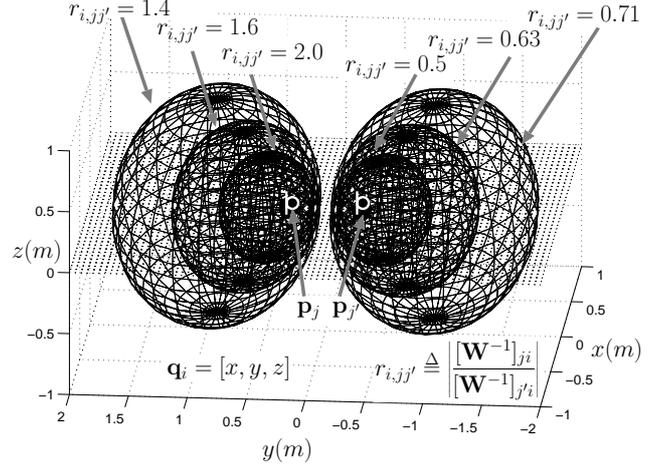

Fig. 1. Example of spheres determined by eq.(7) ($\mathbf{p}_j = [0, 0.3, 0]$, $\mathbf{p}_{j'} = [0, -0.3, 0]$)
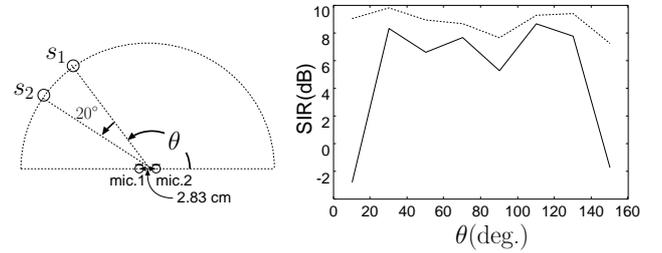

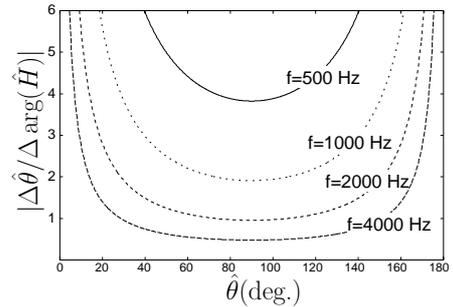**Fig. 2**. Result of preliminary experiment

Fig. 3. Sensitivity of DOA estimation

iment designed to investigate the performance for closely located signals. We used two source signals with an inter-angle of $20°$ and varied the direction of s_1 from $10°$ to $150°$ by using sets of impulse responses in the "RWCP Sound Scene Database [10]." The solid line indicates the actual performance, and the dotted line indicates the performance with optimal permutation. The actual performance deteriorates rapidly when the position of the source signals approaches $0°$ and $180°$, but the optimal permutation provides a stable performance. This means that the reason for the deterioration is the failure to solve the permutation problem.

This can be explained by the sensitivity of the DOA estimation. When we denote an error in calculated $\arg(H_{ji}/H_{j'i})$ as $\Delta\arg(\hat{H})$, and an error in $\hat{\theta}_{i,jj'}$ as $\Delta\hat{\theta}$, the ratio $|\Delta\hat{\theta}/\Delta\arg(\hat{H})|$ can be approximated by the par-

tial derivative of (4):

$$\left| \frac{\Delta\hat{\theta}}{\Delta \arg(\hat{H})} \right| \approx \left| \frac{1}{\omega c^{-1} \| \mathbf{p}_j - \mathbf{p}_{j'} \| \sin(\hat{\theta}_{i,jj'})} \right| . \quad (10)$$

Figure 3 shows examples of this value for several frequency bins. We can see that $\Delta \arg(\hat{H})$ causes a large error in the estimated DOA when the direction is near the axis of the sensor pair. Therefore, we consider the estimated DOA to be unreliable in such cases.

3.2. Ambiguity of estimated DOA

The estimation of DOA by using one sensor pair suffers from some ambiguities. Linear arrays can resolve only one angular component, and this leads to a cone of uncertainty [11]. If we assume a plane on which signals 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. When the spacing between sensors is larger than half a wavelength, spatial aliasing causes another ambiguity, but we do not consider this here.

3.3. Resolving sensitivity and ambiguity

When the number of source signals increases, BSS using ICA requires the same or a larger number of sensors than sources, and many kinds of array geometry are possible. However, if we use frequency domain BSS, we should choose an appropriate array geometry in order to obtain and utilize reliable geometrical information for solving the permutation problem.

A linear array is inappropriate when the potential source location is omnidirectional, because every sensor pair in the linear array has similar sensitivity for DOA estimation, and even when the estimated DOA is reliable, the cone of uncertainty remains. A nonlinear arrangement of sensors is suitable for resolving both sensitivity and ambiguity. Thus, we propose a combination of small and large spacing sensor pairs that have various axis directions.

By using the DOA estimation described in Sec.2.2 with the small spacing sensor pairs that have different axis directions, we can estimate cones which have various vertex angles for one source direction. Because of the sensitivity explained in Sec.3.1, we assume that obtuse cones are reliable, and acute cones are unreliable. Then, we can determine a bearing line pointing to a source direction by using the reliable cones (Fig. 4).

Even when some signals come from the same or a similar direction, we can distinguish between them by using the information provided by the large spacing sensor pair described in Sec.2.3. The source locations can be estimated by combining the estimated direction and spheres (Fig. 5). Then, we can classify separated signals in the frequency domain according to the estimated source locations.

4. EXPERIMENTS

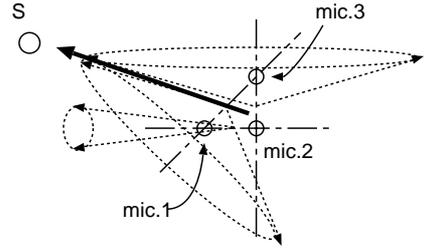We carried out experiments for 6 sources and 8 microphones using speech signals convolved with impulse re-

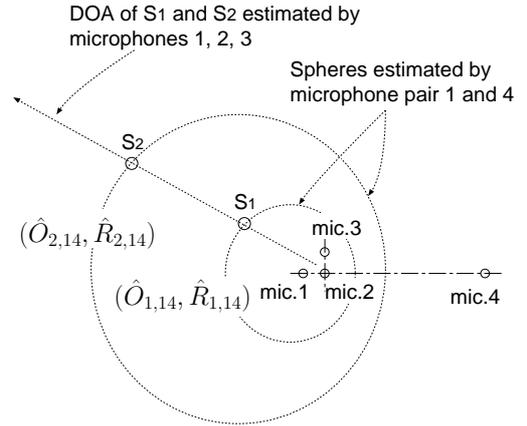**Fig. 4**. Combination of small spacing sensor pairs with different axes

Fig. 5. Combination of small spacing sensor pairs and a large spacing sensor pair

sponses measured in a room. The room layout is shown in Fig. 6. Other conditions are summarized in Table 1. 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. Then we estimate DOAs by using the rows of $\mathbf{W}^{-1}(\omega)$ corresponding to the small spacing microphone pairs (1-3, 2-4, 1-2 and 2-3). Figure 7 shows a histogram of estimated DOAs. We can find five clusters in this histogram, and one cluster is twice the size of the others. This implies that these are six source signals, and two of them come from the same direction (about $150°$). Then, we apply the estimation of spheres to the signals that belong to the large cluster by using the rows of $\mathbf{W}^{-1}(\omega)$ corresponding to the large spacing microphone pairs (7-5, 7-8, 6-5, 6-8). Finally, we can classify the signals into six clusters.

Unfortunately, the classification by 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 [8] for the rest. The correlation based method solves the permutation so that the inter-frequency correlation for neighboring or harmonic frequency bins becomes maximized. In addition, we use the spectral smoothing method proposed in [12] to make separating filters in the
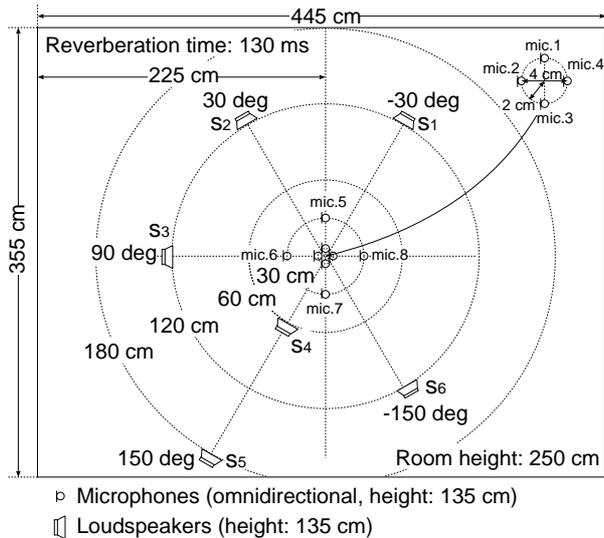
Fig. 6. Room layout

Fig. 7. Histogram of estimated DOAs obtained by using small spacing microphone pairs

Table 1. Experimental conditions

Sampling rate	8 kHz
Data length	6 s
Frame length	2048 point (256 ms)
Frame shift	512 point (64 ms)
ICA algorithm	Infomax (complex valued)

Table 2. Experimental results (dB)

	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	4.5	10.8	14.4	4.5	5.4	8.8	8.1
D+S+C	12.3	5.6	14.5	7.6	8.9	10.8	10.0

time domain from the result of ICA $\mathbf{W}(\omega)$.

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:

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

We measured SIRs for three permutation solving strategies: only the correlation based method ("C"), the estimated DOAs and correlation ("D+C"), and the 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 2.

Our proposed method succeeded in separating six speech signals. We can see 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 the combination of small and large spacing microphone pairs with various axis directions in order to obtain proper geometrical information for solving the permutation problem in frequency domain BSS. In experiments, our method succeeded in the separation of six speech signals, even when two come from the same direction. The computation time was about 1 min. for 6 s. data.
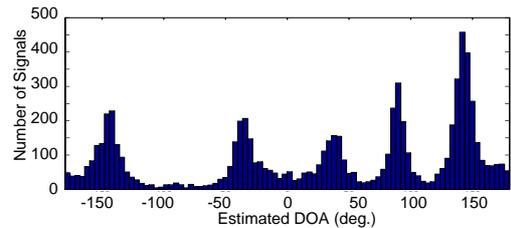
6. REFERENCES

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

[2] 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.

[3] 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.

[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] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "A combined approach of array processing and independent component analysis for blind separation of acoustic signals," in *Proc. ICASSP 2001*, May 2001, pp. 2729–2732.

[8] 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.

[9] 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.

[10] Real World Computing Partnership, "RWCP sound scene database in real acoustic environments," http://tosa.mri.co.jp/sounddb/indexe.htm.

[11] Harry L. Van Trees, *Optimum array processing, Part IV of Detection, Estimation, and Modulation Theory*, John Wiley & Sons, 2002.

[12] H. Sawada, R. Mukai, S. de la Kethulle de Ryhove, S. Araki, and S. Makinol, "Spectral smoothing for frequency-domain blind source separation," in *Proc. of IWAENC 2003*, 2003, to be published.