# FREQUENCY-DOMAIN BLIND SOURCE SEPARATION WITHOUT ARRAY GEOMETRY INFORMATION

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

NTT Communication Science Laboratories, NTT Corporation, Kyoto 619-0237, Japan
Email: {sawada,ryo,shoko,maki}@cslab.kecl.ntt.co.jp

## Introduction

Blind source separation (BSS) for convolutive mixtures is an active research area. Its audio applications include speech enhancement for distant talk. In the frequency-domain approach to BSS, independent component analysis (ICA) is employed in each frequency bin to separate the mixtures. This approach has the advantage that ICA is simple and fast because it is for instantaneous mixtures. The price we have to pay instead is that permutation ambiguities in ICA solutions should be aligned so that a time-domain separated signal contains components of the same source signal. We have proposed a method for solving the permutation problem that utilizes information on source locations estimated from the ICA solutions [1, 2]. It is beneficial to have the source locations estimated during the BSS process. However, some might say that such a technique is not completely *blind*, since the BSS system needs the array geometry information to estimate the source locations.

To respond to such misgivings, we propose a new method for solving the permutation problem that does not need the array geometry information, but just the upper bound of the distances between a sensor and any other sensor. The new method clusters basis vectors that are normalized with the operation described below. It has two main advantages over the previously reported source-localization based method. 1) It makes it easy to use a non-uniform arrangement of sensors, and also there is no need for position calibration. 2) The use of all the information obtained from the basis vectors solves the permutation problem more accurately and therefore improves the BSS performance.

## Blind source separation in frequency domain

Table 1 shows equations related to a mixing model and ICA. Convolutive mixtures in the time domain can be approximated as multiple instantaneous mixtures in the frequency domain. Equation (1) shows the mixing model. The numbers of sources and sensors are denoted by $N$ and $M$, respectively. A vector $\mathbf{h}_k = [h_{1k}, \ldots, h_{Mk}]^T$ represents frequency responses from source $s_k$ to all $M$ sensors. A sensor vector $\mathbf{x}(f,\tau) = [x_1(f,\tau), \ldots, x_M(f,\tau)]^T$ at frequency $f$ and time $\tau$ is modeled as a mixture of the frequency response vectors $\mathbf{h}_k(f)$ each of which is excited by a source $s_k(f,\tau)$.

To separate the mixture $\mathbf{x}$, ICA is employed in each frequency bin (2), where $\mathbf{y} = [y_1, \ldots, y_N]^T$ is a vector of the separated signals and $\mathbf{W}$ is an $N \times M$ separation matrix. By calculating the Moore-Penrose pseudoinverse $\mathbf{W}^+$ of the separation matrix (3), we have the decomposition (4) of $\mathbf{x}$, where $\mathbf{a}_i$ is a basis vector. If ICA works well, each of the terms $\mathbf{a}_1 y_1, \ldots, \mathbf{a}_N y_N$ of (4) is close to each of the terms $\mathbf{h}_1 s_1, \ldots, \mathbf{h}_N s_N$ of (1). However, the correspondence of these terms is not clear because of the permutation ambiguity of an ICA solution. This permutation ambiguity must be aligned so that the $\mathbf{a}_i(f) y_i(f,\tau)$ terms of all frequencies $f$ are grouped together when they correspond to the same source.

Table 1: Mixing model and ICA solution

$$\mathbf{x}(f,\tau) = \textstyle\sum_{k=1}^{N} \mathbf{h}_k(f) s_k(f,\tau) \quad (1)$$

$$\mathbf{y}(f,\tau) = \mathbf{W}(f)\mathbf{x}(f,\tau) \quad (2)$$

$$[\mathbf{a}_1, \cdots, \mathbf{a}_N] = \mathbf{W}^+, \ \mathbf{a}_i = [a_{1i}, \ldots, a_{Mi}]^T \quad (3)$$

$$\mathbf{x}(f,\tau) = \textstyle\sum_{i=1}^{N} \mathbf{a}_i(f) y_i(f,\tau) \quad (4)$$

Table 2: Normalizing and clustering basis vectors

$$\bar{a}_{ji}(f) \leftarrow |a_{ji}(f)| \exp\left[\jmath \frac{\arg[a_{ji}(f)/a_{Ji}(f)]}{4fc^{-1}d}\right] \quad (5)$$

$$\bar{\mathbf{a}}_i(f) \leftarrow \bar{\mathbf{a}}_i(f) \,/\, ||\bar{\mathbf{a}}_i(f)|| \quad (6)$$

$$\mathbf{c}_k \leftarrow \textstyle\sum_{\bar{\mathbf{a}} \in C_k} \bar{\mathbf{a}}/|C_k|, \ \ \mathbf{c}_k \leftarrow \mathbf{c}_k/||\mathbf{c}_k|| \quad (7)$$

$$\mathcal{J} = \textstyle\sum_{k=1}^{M} \mathcal{J}_k, \ \mathcal{J}_k = \sum_{\bar{\mathbf{a}} \in C_k} ||\bar{\mathbf{a}} - \mathbf{c}_k||^2. \quad (8)$$

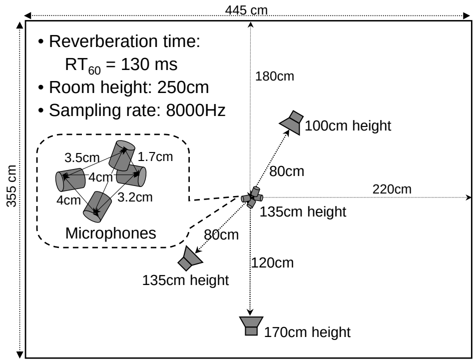


Figure 1: Experimental conditions

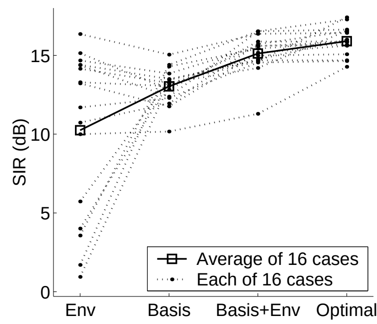


Figure 2: Separation performance

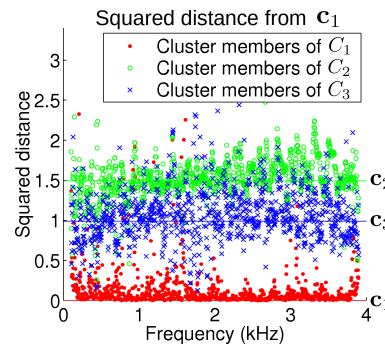


Figure 3: Clustering result

**New method for permutation problem**

The previously reported method estimates geometric information about the sources (direction [1, 2] and distance [2]) from basis vectors $\mathbf{a}_i(f)$ and sensor array geometry, and then clusters the estimations to solve the permutation problem. In contrast, the new method normalizes all the basis vectors $\mathbf{a}_i(f)$, $i = 1, \ldots, M$, for all frequencies $f$ such that they form clusters, each of which corresponds to an individual source. Table (2) shows the flow of the normalization and clustering with the new proposed method. The normalization is performed by selecting a reference sensor $J$ and calculating $\bar{a}_{ji}(f)$ by (5), where $c$ is the propagation velocity and $d$ is the upper bound of the distances between the reference sensor $J$ and any sensor $^\forall j \in \{1, \ldots, M\}$. Then, we apply unit-norm normalization (6) to $\bar{\mathbf{a}}_i(f) = [\bar{a}_{1i}(f), \ldots, \bar{a}_{Mi}(f)]^T$. The reason for the normalized basis vectors $\bar{\mathbf{a}}_i(f)$ forming a cluster for a source can be understood [3] if we assume a direct-path (near-field) model $h_{jk}(f) = (1/d_{jk}) \exp\left[\jmath\, 2\pi f c^{-1}(d_{jk} - d_{Jk})\right]$, where $d_{jk} > 0$ is the distance between source $k$ and sensor $j$.

The next step is to find clusters $C_1, \ldots, C_N$ formed by the normalized vectors $\bar{\mathbf{a}}_i(f)$. The centroid $\mathbf{c}_k$ of a cluster $C_k$ is calculated by (7), where $|C_k|$ is the number of vectors in $C_k$. The criterion of clustering is to minimize the total sum (8) of the squared distances between cluster members and their corresponding centroid. This minimization can be performed efficiently with the k-means clustering algorithm [4]. After we have found $N$ cluster centroids $\mathbf{c}_1, \ldots, \mathbf{c}_N$, we decide the permutation $\Pi_f$ for each frequency $f$ by $\Pi_f = \mathrm{argmin}_\Pi \sum_{k=1}^{N} ||\bar{\mathbf{a}}_{\Pi(k)}(f) - \mathbf{c}_k||^2$.

**Experimental results**

We performed experiments to separate three English speeches, each 3 seconds long, mixed in the condition shown in Fig. 1. We used a non-uniform arrangement of four sensors while knowing only $d = 4$ cm. Figure 2 shows the results for 16 combinations of three speeches measured by using the signal-to-interference ratio (SIR). We compared four methods for solving the permutation problem. Env uses only signal envelope information $|y_i(f, \tau)|$. Basis is the new proposed method using basis vector information $\mathbf{a}_i(f)$. Basis+Env integrates both types of information to obtain better performance [1]. Optimal is the optimal solution obtained by utilizing $s_k$ and $\mathbf{h}_k$ information. We would like to stress that Basis provided satisfactory results even without the integration of +Env. This is an improvement over the former method [1, 2] where the geometric information alone did not provide satisfactory results. Figure 3 shows a clustering result. The vertical axis shows the squared distance from centroid $\mathbf{c}_1$ to cluster members $\bar{\mathbf{a}}$ or to the other centroids $\mathbf{c}_2, \mathbf{c}_3$. The fact that most of the members of $C_1$ were centered close to the centroid $\mathbf{c}_1$ demonstrates the validity of basis vector normalization.

**References**


[1] 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," *IEEE Trans. Speech Audio Processing*, vol. 12, pp. 530–538, Sept. 2004.

[2] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Frequency domain blind source separation using small and large spacing sensor pairs," in *Proc. ISCAS 2004*, May 2004, vol. V, pp. 1–4.

[3] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind extraction of a dominant source signal from mixtures of many sources," in *Proc. ICASSP 2005*, Mar. 2005, accepted.

[4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley Interscience, 2nd edition, 2000.