Solving the Permutation and Circularity Problems of Frequency-Domain Blind Source Separation

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

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

Abstract

Blind source separation (BSS) for convolutive mixtures can be performed efficiently in the frequency domain, where independent component analysis (ICA) is applied separately in each frequency bin. However, frequency-domain BSS involves two major problems that must be solved. The first is the permutation problem: the permutation ambiguity of ICA should be aligned so that a separated signal in the time-domain contains the frequency components of the same source signal. The second problem is the circularity problem: the frequency responses obtained separately by ICA should be constrained so that the corresponding time-domain filter does not rely on the circularity effect of discrete frequency representation. This paper discusses these two problems and presents our methods for solving them. The effectiveness of the BSS method is shown by experimental results for the separation of up to four sources in a reverberant environment.

1. Introduction

Blind source separation (BSS) [1, 2] is a technique for estimating original source signals solely from their mixtures at sensors. Its potential audio signal applications include teleconferences, voice control and hearing aids. In such applications, signals are mixed in a convolutive manner with reverberations. This makes the BSS problem much more difficult to solve than the instantaneous mixture problem. Let us formulate the convolutive BSS problem. Suppose that N source signals $s_k(t)$ are mixed and observed at M sensors

$$x_j(t) = \sum_{k=1}^{N} \sum_l h_{jk}(l) s_k(t-l),$$

where $h_{jk}(l)$ represents the impulse response from source k to sensor j. The goal is to obtain N output signals $y_i(t)$, each of which is a filtered version of a source $s_k(t)$. If we have enough sensors ($M \geq N$), a set of FIR filters $w_{ij}(l)$ of length L is typically used to produce separated signals

$$y_i(t) = \sum_{j=1}^{M} \sum_{l=0}^{L-1} w_{ij}(l) x_j(t-l)$$

at the outputs, and independent component analysis (ICA) [3, 4] is generally used to obtain the FIR filters $w_{ij}(l)$. We can classify the BSS methods into two types based on how we apply ICA to convolutive mixtures.

The first is time-domain BSS, where ICA is applied directly to the convolutive mixture model [5, 6]. It provides good separation once the algorithm converges, and is easy to extend to more than two sources. However, if the algorithm starts from an initial solution far from the final one, it takes many iterations and much time to converge because filter coefficients $w_{ij}(l)$ are interdependent in the algorithm.

The other approach is frequency-domain BSS, where complex-valued ICA for an instantaneous mixture is applied in each frequency bin [7–11]. The merit of this approach is that the ICA algorithm can be performed separately at each frequency, and the convergence of each ICA is fast. However, frequency-domain BSS involves two major problems that must be solved. The first is the well-known permutation problem. Although various methods have been proposed for overcoming the permutation problem, most of them are applicable only to two sources or their performance deteriorates as the number of sources increases. Section 3 of this paper presents a method for solving the permutation problem robustly and precisely. It is based on direction of arrival (DOA) estimation and also the inter-frequency correlation of signal envelopes. The method performs well even when there are more than two sources.

However, just solving the permutation problem does not provide good separation performance. We need to solve the second problem, namely the circularity problem, which originates with the circularity effect of discrete frequency representation. This problem is not as well known as the permutation problem. We discuss the influence and the reason for this problem and present an approach for its solution in Sec. 4. By solving these two problems, the frequency-domain BSS effectively separates many sources in a reverberant environment with low computational cost. The effectiveness of the presented methods is shown by experimental results for up to four sources in Sec. 5.

2. Frequency-domain BSS

This section describes frequency-domain BSS whose flow is shown in Fig. 1. First, time-domain signals $x_j(t)$ at sensors are converted into frequency-domain time-series signals $X_j(f,t)$ by short-time Fourier transform (STFT), where t is now down-sampled with the distance of the

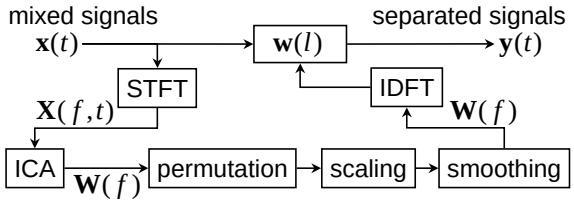

Figure 1: Flow of frequency-domain BSS

frame shift. Then, the frequency responses $W_{ij}(f)$ of filters $w_{ij}(l)$ are obtained by complex-valued ICA

$$\mathbf{Y}(f,t) = \mathbf{W}(f)\mathbf{X}(f,t),$$

where $\mathbf{W}(f)$ is a separation matrix whose elements are $W_{ij}(f)$, $\mathbf{X}(f,t) = [X_1(f,t), \ldots, X_M(f,t)]^T$ and $\mathbf{Y}(f,t) = [Y_1(f,t), \ldots, Y_N(f,t)]^T$. Any complex-valued ICA algorithm can be used in this scheme.

The ICA solution in each frequency bin has permutation and scaling ambiguity: even if we permute the rows of $\mathbf{W}(f)$ or multiply a row by a constant, it is still an ICA solution. The permutation ambiguity should be fixed so that $Y_i(f,t)$ at all frequencies correspond to the same source $s_i(t)$. Thus, the rows of $\mathbf{W}(f)$ are permuted by a permutation Π_f: $\{1,\ldots,N\} \rightarrow \{1,\ldots,N\}$ obtained by a method, such as those discussed in Sec. 3. The scaling ambiguity is solved by the frequency-domain version of the minimal distortion principle, $\mathbf{W}(f) \leftarrow \mathrm{diag}[\mathbf{W}(f)^{-1}]\mathbf{W}(f)$, to make $Y_i(f,t)$ as close to $X_i(f,t)$ as possible [5, 9]. Then, we solve the circularity problem by the spectral smoothing described in Sec. 4. Finally, time-domain separation filters $w_{ij}(l)$ are obtained by applying inverse DFT to $W_{ij}(f)$.

3. The permutation problem

Various methods have been proposed for solving the permutation problem. Let us begin with the direction of arrival (DOA) approach, where the DOAs of source signals are estimated to align permutations. The methods described in [10, 11] plot the directivity patterns formed by a separation matrix, and estimate the direction of a source as the minimum of a directivity pattern. In practice, the methods only work for two sources since the directivity patterns become too complicated to analyze for more than two sources.

We have proposed another way of estimating directions that works for any number of sources [12]. It first calculates the inverse $\mathbf{H}(f) = \mathbf{W}(f)^{-1}$ of the separation matrix $\mathbf{W}(f)$ obtained by ICA. Then, the direction θ_i of a source corresponding to the i-th row of $\mathbf{W}(f)$ is calculated by

$$\theta_i = \arccos \frac{\arg(H_{ji}/H_{j'i})}{2\pi f c^{-1}(d_j - d_{j'})}, \quad (1)$$

where H_{ji} is the element of the j-th row and i-th column of $\mathbf{H}(f)$, c is the propagation velocity, and d_j is the position of sensor j. The scaling ambiguity of the ICA solution is eliminated by taking the ratio $H_{ji}/H_{j'i}$ of two

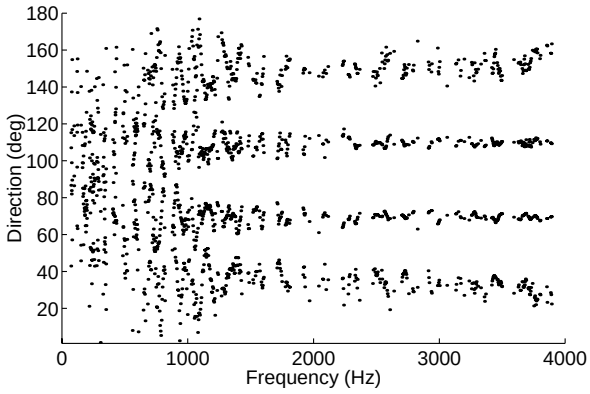

Figure 2: DOA estimations for four sources using ICA

elements from the same column. Figure 2 shows DOA estimations for mixtures of four sources obtained with (1). We see that directions are well estimated and permutation can be aligned by sorting the estimated directions at each frequency. However, at some frequencies (especially low frequencies), estimations are not obtained or are inaccurate. Therefore, the DOA approach alone does not provide a highly precise solution as shown at "D" in Fig. 6.

We also employ the correlation approach [8, 9] to align permutations more precisely. We use the envelope $v_i^f(t) = |Y_i(f,t)|$ of a separated signal $Y_i(f,t)$ to measure correlation. The correlation between two signals $x(t)$ and $y(t)$ is defined as $\mathrm{cor}(x,y) = (\mu_{x\cdot y} - \mu_x \cdot \mu_y)/(\sigma_x \cdot \sigma_y)$, where μ_x is the mean and σ_x is the standard deviation of x. Envelopes have high correlation at neighboring frequencies if separated signals correspond to the same signal. A simple criterion for deciding the permutation Π_f of frequency f is to maximize the sum of the correlations between neighboring frequencies within distance δ:

$$\Pi_f = \mathrm{argmax}_{\Pi} \sum_{|g-f|\le\delta} \sum_{i=1}^{N} \mathrm{cor}(v^f_{\Pi(i)}, v^g_{\Pi_g(i)}), \quad (2)$$

where Π_g is the permutation at frequency g. This criterion is based on local information and has a drawback in that mistakes in a narrow range of frequencies may lead to the complete misalignment of the frequencies beyond the range. As shown at "C" in Fig. 6, the correlation approach alone does not provide a robust solution.

Our method effectively integrates these two approaches to solve the permutation problem robustly and precisely [12]. First, we decide permutations for frequency bins where the confidence of the DOA estimation is sufficiently high. Let $\mathcal{F}$ be the set of frequency bins where the permutation is already decided. Then, we apply (2) to frequency bins that are close neighbors with $f \in \mathcal{F}$. This procedure can avoid a consecutive misalignment. However, the permutations at low frequencies are not usually decided at this stage because the DOA estimations are unreliable as shown in Fig. 2. To decide permutations for these frequencies, we utilize the harmonic structure of a signal. If the signals are speech, there is a strong correlation be-

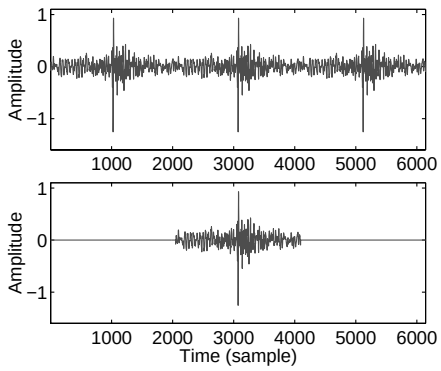

Figure 3: Periodical time-domain filter represented by frequency responses sampled at $L = 2048$ points (above) and its one-period realization (below).

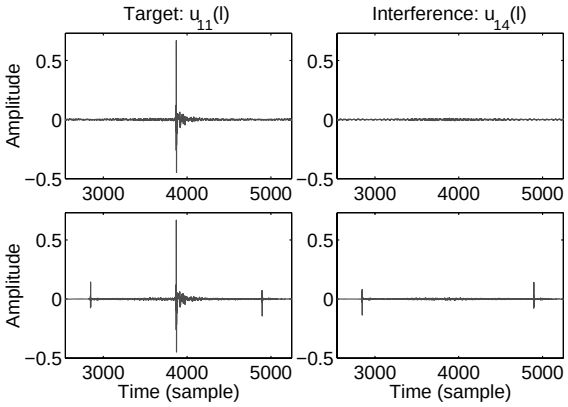

Figure 4: Impulse responses $u_{ik}(l)$ obtained with the periodical filter (above) and with its one-period realization (below).

tween the envelopes of a frequency f and its harmonics $2f, 3f$ and so forth. Thus, we decide the permutation at frequency f with high confidence, if the sum shown below can be clearly maximized:

$$\Pi_f = \text{argmax}_\Pi \sum_{g=2f,3f,\ldots} \sum_{i=1}^{N} \text{cor}(v^f_{\Pi(i)}, v^g_{\Pi_g(i)}).$$

Finally, we apply (2) again for frequencies where the permutation is not yet decided.

4. The circularity problem

The frequency-domain BSS described in Sec. 2 is influenced by the circularity of discrete frequency representation. The circularity refers to the fact that frequency responses sampled at L points with an interval f_s/L (f_s: sampling frequency) represent a periodical time-domain signal whose period is L/f_s. Figure 3 shows two time-domain filters. The upper one is a periodical infinite-length filter represented by frequency responses $W_{ij}(f)$ calculated by ICA at L points. Since this filter is unrealistic, we usually use its one-period realization shown in the lower part.

However, such one-period filters may cause a problem. Figure 4 shows impulse responses from a source $s_k(t)$ to an output $y_i(t)$:

$$u_{ik}(l) = \sum_{j=1}^{M} \sum_{\tau=0}^{L-1} w_{ij}(\tau) h_{jk}(l-\tau).$$

Those on the left $u_{11}(l)$ correspond to the extraction of a target signal, and those on the right $u_{14}(l)$ correspond to the suppression of an interference signal. The upper responses are obtained with the infinite-length filters, and the lower ones with the one-period filters. We see that the one-period filters create spikes, which distort the target signal and degrade the separation performance.

Here, we consider two reasons for these spikes. One is that the frequency responses are under-sampled and the corresponding time-domain filter has an overlap with another period. ICA solutions separately obtained in frequency bins generally require the time-domain filters to be longer than L. The other reason is that adjacent periods work together to perform some filtering even if the first problem is solved. The effect of the second problem can be mitigated if the amplitude of the filter coefficients around both ends is small. It might be thought that a sufficiently large L would solve these problems. However, an excessively long STFT frame results in fewer samples at each frequency and worse ICA solutions [13].

Our approach to this problem involves controlling the frequency responses $W_{ij}(f)$ so that the corresponding time-domain filter $w_{ij}(l)$ fits length L and has small amplitude around the ends. This is carried out by windowing $w_{ij}(l) \cdot g(l)$ with a window $g(l)$ that tapers smoothly to zero at each end, such as a Hanning window. With this operation, frequency responses $\mathbf{W}(f)$ obtained by ICA are smoothed as $\mathbf{W}(f) \leftarrow \sum_{\phi=0}^{f_s-\Delta f} G(\phi)\mathbf{W}(f-\phi)$, where $G(f)$ is the frequency response of $g(l)$ and $\Delta f = f_s/L$. If a Hanning window is used, the frequency responses are smoothed as $\mathbf{W}(f) \leftarrow [\mathbf{W}(f-\Delta f)+2\mathbf{W}(f) + \mathbf{W}(f+\Delta f)]/4$. The windowing successfully eliminates the spikes. However, it changes the frequency response obtained by ICA and causes an error. Thus, we minimize the error by adjusting the scaling of the ICA solution before windowing. See [14] for the details of the error and how to minimize it.

5. Experimental results

We performed experiments to separate speech signals in an environment whose conditions are summarized in Fig. 5. We tested cases of two, three and four sources whose positions are indicated in Table 1. The sensors were arranged linearly, and the number of sensors used was the same as the number of sources. We used filters of length $L = 2048$ because this length performed the best under the conditions. The ICA algorithm we used was the complex-valued version of FastICA [4].

The results shown in Table 1 are the average of eight combinations of 7-second speeches. The signal to interference ratio (SIR) at output i is calculated as the ratio of the power of a target component $\sum_l u_{ii}(l)s_i(t-l)$ and in-

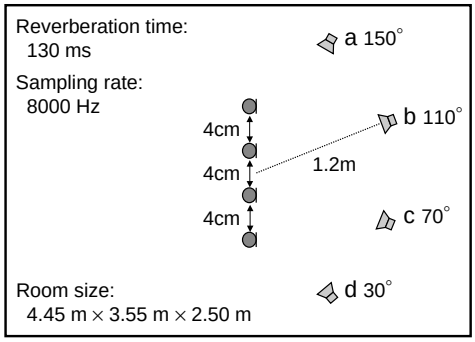

Figure 5: Experimental conditions

Table 1: Overall results

#sources / position	2 / a c		3 / a b d		4 / a b c d	
Spectral smoothing	no	yes	no	yes	no	yes
SIR (dB)	18.4	21.2	13.3	16.5	9.4	11.6
Execution time (s)	4.1	4.1	7.5	7.6	12.3	12.5

terference components $\sum_{k\neq i}\sum_l u_{ik}(l)s_k(t-l)$. We see that the spectral smoothing discussed in Sec. 4 improves the average SIR with every setup. The short execution time, as shown in Table 1, enables the BSS system to perform in real-time if the number of source signals is not very large.

Figure 6 shows SIRs for three and four sources with the different methods for solving the permutation problem discussed in Sec. 3: "D" is the DOA approach alone, "C" is the correlation approach alone, "D+C+Ha" is the proposed method, and "Optimal" is the optimal solution obtained by utilizing the information of $s_k(t)$ and $h_{jk}(l)$. The performance of "D" was stable but insufficient. The performance of "C" was unstable and very poor for four sources. The proposed method "D+C+Ha" performed very well and was close to "Optimal" even when the number of sources was more than two.

6. Conclusion

This paper presented effective methods for overcoming the two major problems of frequency domain BSS. We succeeded in separating many sources mixed in a real environment with a short execution time. The results shown here were for up to four sources with linearly arranged sensors. We have also separated six sources with a planar array of eight sensors based on similar techniques [15].

7. References

[1] S. Haykin, Ed., *Unsupervised Adaptive Filtering (Volume I: Blind Source Separation)*, John Wiley & Sons, 2000.

[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, 2002.

[3] T. W. Lee, *Independent Component Analysis - Theory and Applications*, Kluwer Academic Publishers, 1998.

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

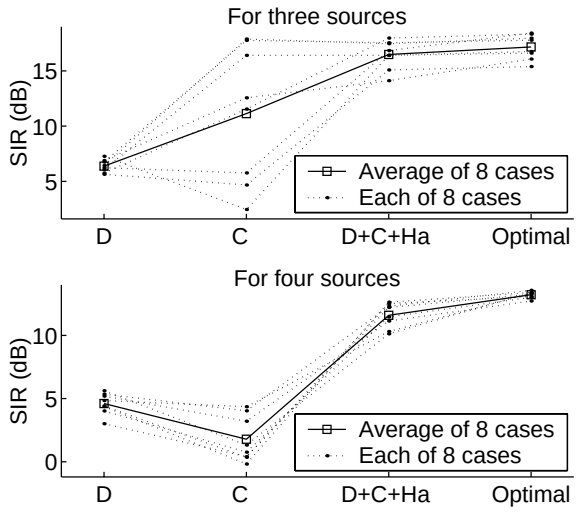

Figure 6: Separation performance with different methods for solving the permutation problem

[5] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. ICA 2001*, Dec. 2001, pp. 722–727.

[6] S. C. Douglas and X. Sun, "Convolutive blind separation of speech mixtures using the natural gradient," *Speech Communication*, vol. 39, pp. 65–78, 2003.

[7] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[8] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. ICA 2000*, June 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, Oct. 2001.

[10] 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. ICASSP 2000*, June 2000, pp. 3140–3143.

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

[12] 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. ICA2003*, Apr. 2003, pp. 505–510.

[13] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 2, pp. 109–116, 2003.

[14] H. Sawada, R. Mukai, S. de la Kethulle, S. Araki, and S. Makino, "Spectral smoothing for frequency-domain blind source separation," in *Proc. IWAENC2003*, Sept. 2003.

[15] R. Mukai, H. Sawada, S. de la Kethulle, S. Araki, and S. Makino, "Array geometry arrangement for frequency domain blind source separation," in *Proc. IWAENC2003*, Sept. 2003.