Acoust. Sci. & Tech. **25**, 4 (2004)

ACOUSTICAL LETTER

Convolutive blind source separation for more than two sources in the frequency domain

Hiroshi Sawada, Ryo Mukai, Shoko Araki and Shoji Makino

NTT Communication Science Laboratories, NTT Corporation,
2–4 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619–0237 Japan

(*Received 12 February 2004, Accepted for publication 8 March 2004*)

Keywords: Blind source separation, Independent component analysis, Convolutive mixtures
PACS number: 43.60.Fg, 43.60.Np [DOI: 10.1250/ast.25.296]

1. Introduction

Blind source separation (BSS) is a technique for recovering source signals solely from their mixtures [1]. Various kinds of BSS methods have been proposed for separating audio signals mixed in a reverberant condition. Frequency-domain BSS, where independent component analysis (ICA) [2,3] is performed separately in each frequency bin, has been widely considered for its simplicity and efficiency [4–6]. However, most of the results were for separating only two sources. This is mainly because the permutation problem of frequency-domain BSS is difficult to solve when there are many sources.

We have recently proposed a method for solving the permutation problem [7], which is even effective for more than two sources. However, just solving the permutation problem does not provide good separation performance if there are many sources. We need to solve another problem, namely the circularity problem, which originates with the circularity of discrete frequency representation [8]. This problem has not been carefully considered because it is not serious in a two-source case but it becomes serious as the number of sources increases. By solving these two problems, we succeeded in separating many sources from their convolutive mixtures in the frequency domain. This letter reports the basic structure of the BSS method, and also experimental results for the separation of up to four sources.

2. BSS for convolutive mixtures

Suppose that N source signals $s_k(t)$ are mixed and observed at M sensors

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

where $h_{jk}(l)$ represents the impulse response from source k to sensor j. If we have enough sensors for the number of sources ($N \leq M$), a set of FIR filters $w_{ij}(l)$ of length L is typically used to obtain the separated signals

$$y_i(t) = \sum_{j=1}^{M} \sum_{l=0}^{L-1} w_{ij}(l) x_j(t-l). \tag{2}$$

ICA is a statistical tool for calculating the filters $w_{ij}(l)$ without any information on the mixing system $h_{jk}(l)$ and the sources $s_k(t)$. We can classify BSS methods into two categories based on how we apply ICA for convolutive mixtures.

The first is time-domain BSS, where ICA is applied directly to the convolutive mixture model [9]. It provides good separation once the algorithm converges, and is easy to extend to more than two sources. However, ICA for convolutive mixtures is not as simple as ICA for instantaneous mixtures, and computationally expensive for long filters.

The other approach is frequency-domain BSS, where complex-valued ICA for instantaneous mixtures is applied in each frequency bin. 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, there are two problems to be solved as discussed in the Introduction.

3. Frequency-domain BSS

This section explains the flow (Fig. 1) of frequency-domain BSS, and also our methods for solving the two problems. In this flow, time-domain filters $w_{ij}(l) = [\mathbf{w}(l)]_{ij}$ of length L are obtained by the inverse discrete Fourier transform of frequency responses $W_{ij}(f) = [\mathbf{W}(f)]_{ij}$ calculated by ICA and the following several processes.

Time-domain signals $x_j(t)$ at sensors are first 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 frame shift and f is sampled at L discrete frequencies. Then, to obtain the frequency responses $W_{ij}(f)$, complex-valued ICA

$$\mathbf{Y}(f,t) = \mathbf{W}(f)\mathbf{X}(f,t) \tag{3}$$

is solved, where $\mathbf{X}(f,t) = [X_1(f,t), \ldots, X_M(f,t)]^{\mathrm{T}}$, $\mathbf{Y}(f,t) = [Y_1(f,t), \ldots, Y_N(f,t)]^{\mathrm{T}}$ and $\mathbf{W}(f)$ is an $N \times M$ separation matrix whose elements are $W_{ij}(f)$. 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 corresponds to the same source $s_i(t)$. This is the permutation problem of frequency-domain BSS. Our method for the problem [7] is based on direction of arrival (DOA) estimation [6] and the inter-frequency correlation of output signal envelopes [5]. Since the DOA estimation method proposed in [7] is applicable for more than two sources, the method has become practical for more than two sources.

The scaling ambiguity of an ICA solution should be taken

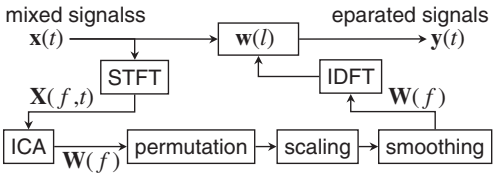

Fig. 1 Flow of frequency-domain BSS.

care of not only in frequency-domain BSS but also in time-domain BSS. The minimal distortion principle (MDP) [9] resolves this ambiguity by making $y_i(t)$ as close to $\sum_l h_{ii}(l)s_i(t-l)$ as possible. The same principle can be considered in the frequency domain [5], and can be realized by a simple operation $\mathbf{W}(f) \leftarrow \mathrm{diag}[\mathbf{W}^{-1}(f)]\mathbf{W}(f)$.

The second problem of frequency-domain BSS comes from the circularity of discrete frequency representation. The circularity refers to the fact that frequency responses $W_{ij}(f)$ sampled at L points with an interval f_s/L (f_s: sampling frequency) represent a periodical time-domain filter whose period is L/f_s. If the required time-domain filter length is less than L, the effect of the circularity is not so serious. However, if the required length is more than L, the time-domain filters have an overlap with another period, and may cause a problem as shown in the upper half of Fig. 3. The required length is generally not as long in a two-source case, but become longer as the number of sources increases.

Our approach to this problem involves smoothing the frequency responses $W_{ij}(f)$ so that the corresponding time-domain filter $w_{ij}(l)$ fits length L and has a 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. 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 that degrades the separation performance. Therefore, we minimize the error by adjusting the scaling ambiguity of the ICA solution before windowing. See [8] for the details of the error and its minimization.

4. Experimental results

We performed experiments to separate speech signals in an environment whose conditions are summarized in Fig. 2. 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.

The signal-to-interference ratio (SIR) and signal-to-distortion ratio (SDR) were calculated to evaluate how well the sources are separated, and the degree to which the filter operation distorts the signals, respectively. To calculate these values for each output, a separated signal $y_i(t)$ is first decomposed into a target signal $\sum_l u_{ii}(l)s_i(t-l)$ and an interference signal $\sum_{k\neq i}\sum_l u_{ik}(l)s_k(t-l)$, where $u_{ik}(l)$ is the impulse responses from a source $s_k(t)$ to the separated signal $y_i(t)$ defined as

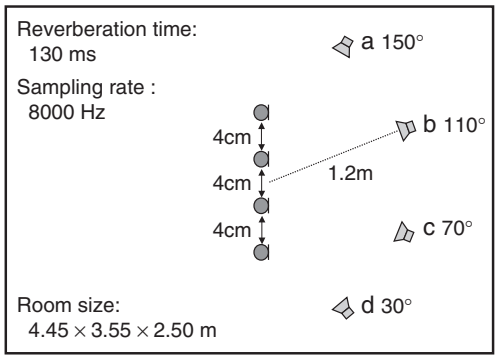

Fig. 2 Experimental conditions.

Table 1 Batch processing results.

#sources	2		3		4	
position	a b		a b d		a b c d	
smoothing	no	yes	no	yes	no	yes
SIR (dB)	19.3	20.3	13.7	16.9	9.3	13.2
SDR (dB)	18.0	19.3	13.9	15.7	10.8	11.3
exec. time	9.9 s		18.7 s		28.3 s	

$$u_{ik}(l) = \sum_{j=1}^{M}\sum_{\tau=0}^{L-1} w_{ij}(\tau)h_{jk}(l-\tau). \tag{4}$$

The SIR is defined as the power ratio of the target and the interference. Then, the target signal is decomposed into a scaled version of a reference $r_i(t)$ and a distortion $e_i(t)$. We selected $r_i(t) = \sum_l h_{ii}(l)s_i(t-l)$ as the reference following the MDP [9]. Thus, the target signal is decomposed as $\sum_l u_{ii}(l)s_i(t-l) = \alpha_i \cdot r_i(t) + e_i(t)$, where α_i is a real-valued scalar that minimizes the distortion $e_i(t)$. The SDR is defined as the power ratio of $\alpha_i \cdot r_i(t)$ and $e_i(t)$.

Table 1 shows the result of batch processing for 7 second observations. We see that the smoothing discussed in the previous section improved the SIR and SDR in every setup, especially with three and four sources. We used filters of length $L = 2048$ because this length performed the best under the conditions. The ICA algorithm we used was FastICA [3] followed by InfoMax combined with the natural gradient [2] of 50 iterations to obtain further improvement. Good separation was achieved with a practical execution time.

Figure 3 shows the impulse responses $u_{ik}(l)$ when there were three sources. Those on the left $u_{11}(l)$ correspond to the extraction of a target signal, and those on the right $u_{13}(l)$ correspond to the suppression of an interference signal. The upper responses were obtained without considering the circularity problem, and the lower ones obtained with smoothing. We see that the smoothing effectively eliminates the spikes caused by the circularity of the discrete Fourier transform.

The BSS system was capable of operating in real time for three sources when we used shorter filters $L = 1024$ and decreased the number of iterations to 35 in the natural

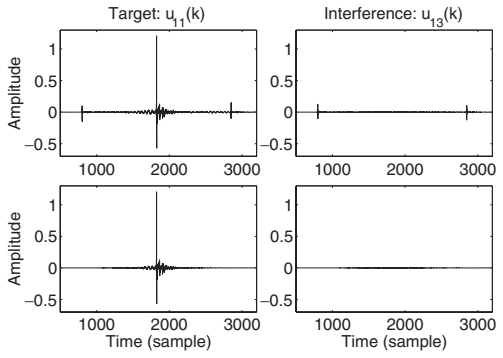

Fig. 3 Impulse responses $u_{ik}(l)$ obtained without considering the circularity problem (above) and with smoothing (below).

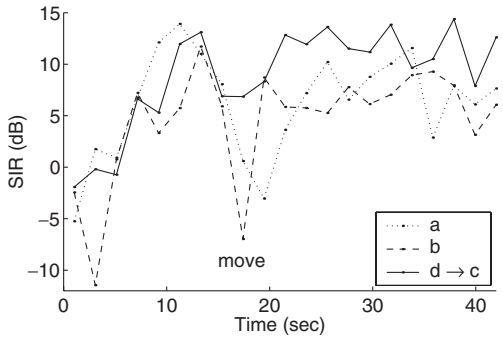

Fig. 4 Separation performance of real-time processing for moving sources.

gradient. We used the same system structure as that used for two sources described previously [10]. Figure 4 shows the SIR for each source, where the source at position "d" started to move to position "c" at a time of 15 seconds. Since the filter coefficients were updated every 2 seconds, the system tracked the movement and recovered the SIRs. We see that the SIR of the moving source ("d"→"c") is not degraded as much as the other sources when it moves. An interpretation of this kind of phenomenon is discussed in [10].

5. Conclusion

We have presented a frequency-domain BSS method that is practically applicable for more than two sources by overcoming the permutation and circularity problems. The experimental results show the effectiveness and efficiency of the BSS method. We also reported results for the separation of six sources with a planar array of eight sensors [11].

References

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

[2] T. W. Lee, *Independent Component Analysis — Theory and Applications* (Kluwer Academic Publishers, Boston, 1998).

[3] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis* (John Wiley & Sons, New York, 2001).

[4] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, **22**, 21–34 (1998).

[5] N. Murata, S. Ikeda and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, **41**, 1–24 (2001).

[6] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. Int. Conf. Acoust., Speech, Signal Process.* (*ICASSP 2000*), pp. 3140–3143 (2000).

[7] 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," *Proc. Int. Symp. Independent Component Analysis and Blind Signal Separation* (*ICA 2003*), pp. 505–510 (2003).

[8] H. Sawada, R. Mukai, S. de la Kethulle, S. Araki and S. Makino, "Spectral smoothing for frequency-domain blind source separation," *Proc. Int. Workshop Acoustic Echo and Noise Control* (*IWAENC 2003*), pp. 311–314 (2003).

[9] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. Int. Conf. Independent Component Analysis and Blind Signal Separation* (*ICA 2001*), pp. 722–727 (2001).

[10] R. Mukai, H. Sawada, S. Araki and S. Makino, "Robust real-time blind source separation for moving speakers in a room," *Proc. Int. Conf. Acoust., Speech, Signal Process.* (*ICASSP 2003*), pp. 469–472 (2003).

[11] R. Mukai, H. Sawada, S. de la Kethulle, S. Araki and S. Makino, "Array geometry arrangement for frequency domain blind source separation," *Proc. Int. Workshop Acoustic Echo and Noise Control* (*IWAENC 2003*), pp. 219–222 (2003).