*International Workshop on Acoustic Echo and Noise Control (IWAENC2003), Sept. 2003, Kyoto, Japan*

# SPECTRAL SMOOTHING FOR FREQUENCY-DOMAIN BLIND SOURCE SEPARATION

*Hiroshi Sawada   Ryo Mukai   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
{sawada,ryo,delaketh,shoko,maki}@cslab.kecl.ntt.co.jp

## ABSTRACT

This paper describes the circularity problem of frequency-domain blind source separation (BSS), and presents a new method for solving it. Frequency-domain BSS performs independent component analysis (ICA) in each frequency bin. It is more efficient than time-domain BSS where ICA is applied to convolutive mixtures. However, frequency-domain BSS has two problems. The first is the permutation problem, for which we have recently proposed a method. It provides a robust and precise solution for the permutation problem and reveals the influence of the second problem, namely the circularity problem. Our solution for this second problem is based on spectral smoothing by windowing. However, the direct application of windowing changes the frequency responses for separation obtained by ICA and causes an error. Therefore, we adjust the frequency responses before windowing so that the error is minimized. The effectiveness of the method is shown by experimental results for the separation of up to four sources.

## 1. INTRODUCTION

We consider the problem of blind source separation (BSS) for convolutive mixtures [1]. Suppose that $N$ source signals $s_p(t)$ are mixed and observed at $M$ sensors $x_q(t) = \sum_{p=1}^{N}\sum_l h_{qp}(l)s_p(t-l)$, where $h_{qp}(l)$ represents the impulse response from source $p$ to sensor $q$. The goal is to separate the mixtures $x_q(t)$ and to obtain a filtered version of a source $s_p(t)$ at each output $y_r(t)$. The separation system typically consists of a set of FIR filters $w_{rq}(l)$ of length $L$ to produce separated signals $y_r(t) = \sum_{q=1}^{M}\sum_{l=0}^{L-1} w_{rq}(l)$ $x_q(t-l)$. In a practical situation, separation should be performed without knowing $s_p(t)$ or $h_{qp}(l)$. Independent component analysis (ICA) [2] is one of the major statistical tools for solving this problem. We can classify BSS methods into two approaches based on how we apply ICA.

The first approach is time-domain BSS, where ICA is applied directly to the convolutive mixture model. The ICA algorithm correctly evaluates the independence of separated signals $y_r(t)$. Thus, the approach achieves good separation once the algorithm converges. However, if the algorithm starts from a solution far from the final one, it takes many iterations and much time to converge. This is because filter coefficients $w_{rq}(l)$ depend on each other in the algorithm. We need thousands of filter taps to separate acoustic signals in a room. The approach is impractical for such cases unless starting from a good initial solution.

The other approach is frequency-domain BSS, where complex-valued ICA for an instantaneous mixture is applied in each frequency bin [3–8]. 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. The difficulty lies in solving the permutation problem and the circularity problem. The permutation problem is well known to be a difficult problem. Recently, we have proposed a method [6] that provides a robust and precise solution and also enables the separation of more than two sources. As a consequence, the influence of the circularity problem is highlighted. It deteriorates separation performance as explained in Sec. 3.

One well-known aspect of the circularity problem is that a multiplication in the frequency domain corresponds to a circular convolution in the time domain, which is different from a linear one. A widely used technique for simulating a linear convolution by frequency-domain multiplication involves constraining frequency responses such that the corresponding time-domain filter contains enough consecutive zeros at the end [9]. To maintain this constraint strictly in BSS, the gradient of a filter in the ICA algorithm should also be constrained to contain enough consecutive zeros. If this constraint is followed in frequency-domain BSS, it becomes a frequency-domain implementation of time-domain convolutive ICA [10, 11], whose convergence is slow as in the case of time-domain BSS. Some BSS methods interleave this constraint with a frequency-domain ICA algorithm [7, 8]. With these methods, the constraint conflicts with the ICA solution, and these two different operations should be applied repeatedly until convergence. In both cases, the algorithm moves back and forth between the two domains, and loses the attractive characteristic of frequency-domain BSS whereby ICA can be applied separately in each frequency bin.

Our recognition of the circularity problem is more general than the circular convolution problem. The circularity problem arises from discrete frequency representation as explained in Sec. 3. Our technique for solving the problem involves spectral smoothing by a window that tapers smoothly to zero at each end, and forcing a filter to fit a specific length $L$. The direct application of windowing, however, changes

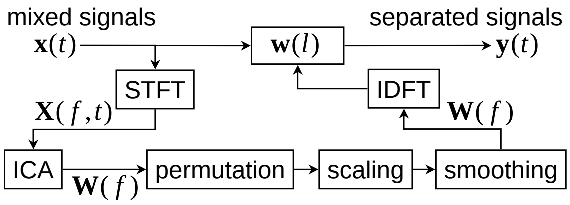


**Fig. 1**. Flow of frequency-domain BSS

the ICA solutions and causes an error. Therefore, we adjust the ICA solutions within their scaling ambiguity before windowing so that the error is minimized in the least-squares sense. These techniques are explained in Sec. 4.

## 2. FREQUENCY-DOMAIN BSS

This section describes frequency-domain BSS where ICA is applied separately in each frequency bin [3–6]. Figure 1 shows the flow. First, time-domain signals $x_q(t)$ are converted into frequency-domain time-series signals $X_q(f,t)$ by short-time Fourier transform (STFT), where $t$ is now down-sampled with the distance of the frame shift. Then, to obtain the frequency responses $W_{rq}(f)$ of filters $w_{rq}(l)$, complex-valued ICA $\mathbf{Y}(f,t) = \mathbf{W}(f)\mathbf{X}(f,t)$ is solved, where $\mathbf{X}(f,t) = [X_1(f,t), \ldots, X_M(f,t)]^T$, $\mathbf{Y}(f,t) = [Y_1(f,t), \ldots, Y_N(f,t)]^T$, and $\mathbf{W}(f)$ is a separation matrix whose elements are $W_{rq}(f)$. Note that any ICA algorithm can be used in this scheme. The ICA solution in each frequency has scaling and permutation ambiguity. We perform permutation alignment $\mathbf{W}(f) \leftarrow \mathbf{P}(f)\mathbf{W}(f)$, where $\mathbf{P}(f)$ is a permutation matrix obtained by the method [6]. The scaling ambiguity is solved by the minimal distortion principle, $\mathbf{W}(f) \leftarrow \text{diag}(\mathbf{W}(f)^{-1})\mathbf{W}(f)$, to make $Y_r(f,t)$ as close to $X_r(f,t)$ as possible [4, 12]. Then, we solve the circularity problem by spectral smoothing as described in Sec. 4. Finally, separation filters $w_{rq}(l)$ are obtained by applying inverse DFT to $W_{rq}(f)$.

## 3. THE CIRCULARITY PROBLEM

The frequency-domain BSS described in the previous section is influenced by the circularity of the 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 time-domain signal whose period is $L/f_s$. Figure 2 shows two time-domain filters. The upper one is a periodical infinite-length filter represented by frequency responses $W_{rq}(f)$ calculated by ICA at $L$ points. Since this filter is unrealistic, we usually use the one-period realization of the filter shown in the lower part.

However, one-period filters may cause a problem. Figure 3 shows impulse responses from a source $s_p(t)$ to a separated signal $y_r(t)$: $u_{rp}(l) = \sum_{q=1}^{M} \sum_{\tau=0}^{L-1} w_{rq}(\tau) h_{qp}(l-\tau)$. 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 are obtained by the infinite-length filters, and the lower ones are by the one-period filters. We see that the one-period filters create spikes, which distort a target signal and degrade the separation performance. Since these spikes does not exist in the infinite-length case, both adjacent periods are needed to eliminate a spike. Here, we consider two reasons for these spikes. One is that the frequency responses are undersampled and the corresponding time domain filter has an overlap with another period. This problem does not occur if the required filter length is less 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.

The ICA algorithm in the frequency domain does not know the length $L$ of the time-domain filters, and ICA solutions in various frequency bins might require the length of the time-domain filter to be longer than $L$ and generally infinite. Therefore, we need to control the frequency responses so that the corresponding time-domain filter fits length $L$ and has small amplitude around the ends.

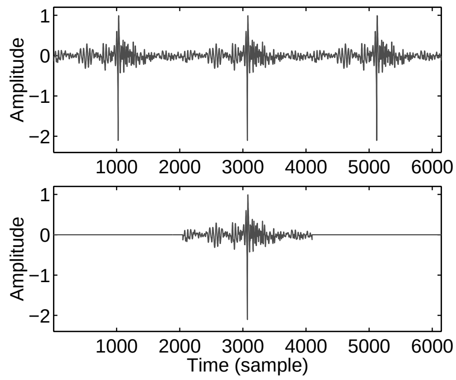


**Fig. 2**. Periodical time-domain filter represented by frequency responses sampled at $L = 2048$ points (above) and its one-period realization (below).

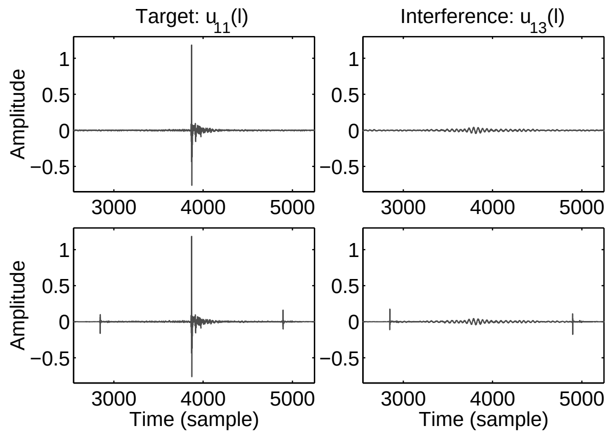


**Fig. 3**. Impulse responses $u_{rp}(l)$ obtained by the periodical filter (above) and by the finite-length filter (below).

## 4. SPECTRAL SMOOTHING

### 4.1. By Windowing

We propose a method for changing the frequency responses to meet the requirement considered in the previous section. The basic idea is spectral smoothing by windowing. We apply a window $g(l)$ to a filter $w_{rq}(l)$ to obtain a new filter $w_{rq}(l) \cdot g(l)$. To ensure that the new filter have small amplitude around both ends, we use a window that tapers smoothly to zero at each end, such as a Hanning window $g(l) = \frac{1}{2}(1 + \cos\frac{2\pi l}{L})$. By 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

$$\mathbf{W}(f) \leftarrow [\mathbf{W}(f-\Delta f) + 2\mathbf{W}(f) + \mathbf{W}(f+\Delta f)]/4$$

since the frequency responses $G(f)$ of a Hanning window are $G(0) = 1/2$, $G(\Delta f) = G(f_s - \Delta f) = 1/4$, and zero for the other frequency bins.

Although the windowing eliminates the spikes, it changes the frequency response obtained by ICA and causes an error. The error can be calculated for each row $\mathbf{w}_r(f) = [W_{r1}(f), \ldots, W_{rM}(f)]$ of $\mathbf{W}(f)$. Let $\mathbf{b}_r(f)$ be the smoothed frequency responses, and $\alpha_r$ be a complex-valued scalar representing the scaling ambiguity of the ICA solution. The error is $\mathbf{e}_r(f) = min_{\alpha_r}[\mathbf{b}_r(f) - \alpha_r\mathbf{w}_r(f)]$, and its least-squares solution is

$$\mathbf{e}_r(f) = \mathbf{b}_r(f) - \frac{\mathbf{b}_r(f)\mathbf{w}_r(f)^H}{||\mathbf{w}_r(f)||^2}\mathbf{w}_r(f).$$

If we use a Hanning window, the smoothed frequency response is $\mathbf{b}_r(f) = [\mathbf{w}_r(f-\Delta f)+2\mathbf{w}_r(f)+\mathbf{w}_r(f+\Delta f)]/4$. Thus, the error can be represented as

$$\begin{aligned}
\mathbf{e}_r(f) &= [\mathbf{c}_r^-(f) + \mathbf{c}_r^+(f)] \;/\; 4, \text{ where} \\
\mathbf{c}_r^-(f) &= \mathbf{w}_r(f-\Delta f) - \frac{\mathbf{w}_r(f-\Delta f)\mathbf{w}_r(f)^H}{||\mathbf{w}_r(f)||^2}\mathbf{w}_r(f), \\
\mathbf{c}_r^+(f) &= \mathbf{w}_r(f+\Delta f) - \frac{\mathbf{w}_r(f+\Delta f)\mathbf{w}_r(f)^H}{||\mathbf{w}_r(f)||^2}\mathbf{w}_r(f).
\end{aligned}$$

This $\mathbf{c}_r^-$ (or $\mathbf{c}_r^+$) represents the difference of two vectors $\mathbf{w}_r(f)$ and $\mathbf{w}_r(f-\Delta f)$ (or $\mathbf{w}_r(f+\Delta f)$). These differences are usually not very large, therefore the error does not seriously affect the separation.

### 4.2. Pre-scaling

Even if the error caused by the windowing is not very large, we can improve the separation by minimizing the error. We minimize the error by adjusting the scaling of the ICA solution before windowing. Let $d_r(f)$ be a complex-valued scalar and $\mathbf{v}_r(f) = d_r(f)\mathbf{w}_r(f)$ be a new row vector of a separation matrix. We want to find $d_r(f)$ such that the error

$$\mathbf{e}_r(f) = \mathbf{b}_r(f) - \frac{\mathbf{b}_r(f)\mathbf{v}_r(f)^H}{||\mathbf{v}_r(f)||^2}\mathbf{v}_r(f)$$

is minimized, where $\mathbf{b}_r(f)$ is the smoothed frequency responses. A scalar $d_r(f)$ should be close to 1 to avoid any great change in the predetermined scaling. Thus, the total cost to be minimized is $J = \sum_f J_r(f)$, where

$$J_r(f) = ||\mathbf{e}_r(f)||^2 \;/\; ||\mathbf{w}_r(f)||^2 + \beta|d_r(f) - 1|^2,$$

and $\beta$ is a parameter indicating the importance of maintaining the predetermined scaling. The minimization can be performed iteratively by $d_r(f) = d_r(f) - \mu\frac{\partial J}{\partial d_r(f)}$ with a small step-size $\mu$. If a Hanning window is used, the smoothed frequency responses are

$$\mathbf{b}_r(f) = [\mathbf{v}_r(f-\Delta f) + 2\mathbf{v}_r(f) + \mathbf{v}_r(f+\Delta f)] \;/\; 4,$$

and the error can be represented as

$$\mathbf{e}_r(f) = [d_r(f-\Delta f)\mathbf{c}_r^-(f) + d_r(f+\Delta f)\mathbf{c}_r^+(f)] \;/\; 4.$$

using previously defined $\mathbf{c}_r^-$ and $\mathbf{c}_r^+$. Thus, the gradient is

$$\begin{aligned}
\frac{\partial J}{\partial d_r(f)} &= \frac{\partial J_r(f-\Delta f)}{\partial d_r(f)} + \frac{\partial J_r(f+\Delta f)}{\partial d_r(f)} + \frac{\partial J_r(f)}{\partial d_r(f)} \\
&= [\mathbf{e}_r(f-\Delta f)\mathbf{c}_r^+(f-\Delta f)^H + \\
&\quad\; \mathbf{e}_r(f+\Delta f)\mathbf{c}_r^-(f+\Delta f)^H] \;/\; (8 \cdot ||\mathbf{w}_r(f)||^2) \\
&\quad +2\beta(d_r(f) - 1).
\end{aligned}$$

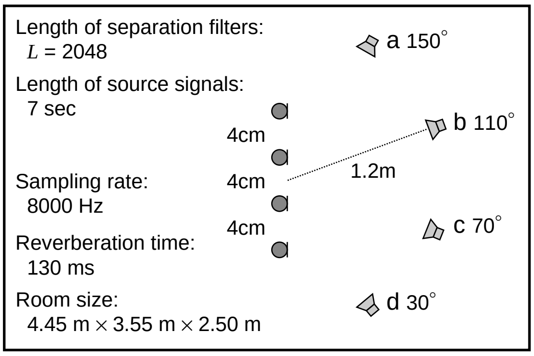


**Fig. 4**. Experimental conditions

## 5. EXPERIMENTAL RESULTS

We performed experiments under the conditions shown in Fig. 4. As the ICA algorithm, we used FastICA [2] followed by Infomax + Natural gradient [1] with 50 loops to improve the performance. Table 1 summarizes the separation results. To show the effectiveness of the proposed method, we compare cases where the spectral smoothing was applied differently: no smoothing ("no"), simply multiplying a Hanning window ("win") and pre-scaling before multiplying a Hanning window ("pre + win"). In the "pre + win" case, we tested two different values for $\beta$.

We evaluated the separation using signal-to-interference ratio (SIR) and signal-to-distortion ratio (SDR). SIR is calculated by the ratio of the power of a target component $u_{rr}(l)$ and an interference component $\sum_{p\neq r} u_{rp}(l)$. To calculate SDR, we decompose the target component $u_{rr}(l)$ into a scaled version of a reference and a distortion $e_r(t)$. We selected $h_{rr}(l)$ as the reference following the minimal distortion principle [12]. Thus, the target component is decomposed as $u_{rr}(l) = \alpha_r \cdot h_{rr}(t) + e_r(t)$, where $\alpha_r$ is a

**Table 1**. Experimental results

| #sources / position | 2 / a b | | | | 3 / a b d | | | | 4 / a b c d | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| smoothing | no | win | pre + win | | no | win | pre + win | | no | win | pre + win | |
| $\beta$ | | | 0.1 | 0.01 | | | 0.1 | 0.01 | | | 0.1 | 0.01 |
| SIR (dB) | 20.6 | 21.6 | 21.8 | 22.1 | 13.4 | 17.2 | 17.8 | 19.1 | 11.8 | 15.6 | 17.0 | 19.0 |
| SDR (dB) | 20.5 | 21.1 | 20.8 | 19.6 | 14.9 | 17.0 | 17.6 | 17.0 | 15.1 | 17.9 | 15.8 | 12.2 |
| Execution time (s) | 9.7 | 9.7 | 9.8 | 10.1 | 17.9 | 18.0 | 18.2 | 18.7 | 27.0 | 27.0 | 27.3 | 27.9 |

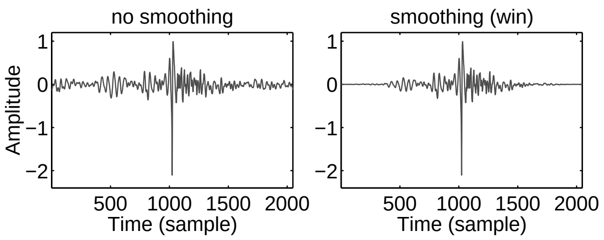


**Fig. 5**. Separation filters $w_{11}(l)$

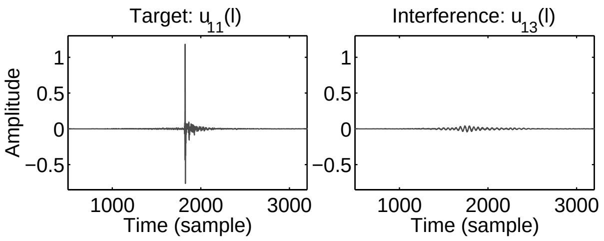


**Fig. 6**. Impulse responses $u_{rp}(l)$ obtained by filters, one of which is shown in "smoothing (win)" in Fig. 5.

real-valued scalar to make the distortion $e_r(t)$ minimum. Finally, SDR is calculated by the ratio of the power of $\alpha_r \cdot h_{rr}(t)$ and $e_r(t)$.

We found that spectral smoothing improved the SIR in all cases. With spectral smoothing, both ends of a separation filter converge to zero as shown in Fig. 5. This eliminates the spikes caused by the circularity as shown in Fig. 6. However, spectral smoothing alone changes the ICA solutions and causes an error. The pre-scaling minimizes the error and improves SIR further. If we use a smaller $\beta$, a better SIR is obtained at the cost of a worse SDR. An appropriate $\beta$ can be chosen to balance SIR and SDR.

## 6. CONCLUSION

We have observed the circularity problem, and have proposed a method for solving the problem based on spectral smoothing. This method provides good separation in combination with our previously reported solution to the permutation problem [6]. We have succeeded in separating many sources with a practical execution time. The experiments 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 [13].

## 7. ACKNOWLEDGEMENT

We thank Prof. Scott C. Douglas for valuable discussions.

## 8. REFERENCES

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

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

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

[4] S. Ikeda and N. Murata, "A method of ICA in time–frequency domain," in *Proc. ICA '99*, Jan. 1999, pp. 365–370.

[5] H. Saruwatari, S. Kurita, and K. Takeda, "Blind source separation combining frequency-domain ICA and beamforming," in *Proc. ICASSP 2001*, May 2001, MULT-P2.2.

[6] 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.

[7] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[8] L. Schobben and W. Sommen, "A frequency domain blind signal separation method based on decorrelation," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 1855–1865, Aug. 2002.

[9] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[10] M. Joho and P. Schniter, "Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient," in *Proc. ICA2003*, Apr. 2003, pp. 543–548.

[11] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutive mixtures," in *Proc. ICA2003*, Apr. 2003, pp. 945–950.

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

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