Frequency-Domain Blind Source Separation with Permutation Control

Peng Xie

Steven L. Grant
Missouri University of Science and Technology, sgrant@mst.edu

FREQUENCY-DOMAIN BLIND SOURCE SEPARATION WITH PERMUTATION CONTROL*

Peng Xie, Steven L. Grant†

Dept. of Electrical and Computer Engineering,
Missouri University of Science and Technology,
Rolla, MO 65409, USA

ABSTRACT

This paper explores the problem of frequency-domain Blind Source Separation (BSS) of convolutive mixtures. The main difficulties of this approach lie in the so called permutation and amplitude problems. In order to solve the permutation ambiguity, a new hybrid approach is proposed, in which the Independent Component Analysis (ICA) processes across all frequency bins are concatenated and each of them is embedded with a permutation control unit. In each frequency bin, when the separation matrix is obtained by the ICA process, the control unit detects the possible permutation and aligns the matrix only if the permutation is confirmed. Then the final value of separation matrix is used to initialize the ICA iterations in the next frequency bin. The amplitude problem is addressed by utilizing the elements in estimated mixing matrix. The method is compared with conventional frequency-domain BSS approaches and the experimental results demonstrate superior performances of the proposed method.

Index Terms— convolutive mixture, frequency-domain, Blind Source Separation, permutation control

1. INTRODUCTION

Blind Source Separation (BSS) is a statistical technique for recovering a number of original source signals when only their linear mixtures are available for observation. With the understanding that both source signals and mixing procedure are unknown, the process is termed "blind" and this blindness enables the technique to be used in a wide variety of situations. These include noise-robust speech recognition, hands-free telecommunication, and medical signal processing.

Based on the nature of the signal mixing process, there are two important issues in BSS research that are generally investigated: instantaneous BSS and convolutive BSS. Recently convolutive BSS is drawing much of researchers' attention, because in many real-world applications, the signals are mixed in a convolutive manner. The major approaches to separate convolutive mixtures can be divided into time-domain and frequency-domain methods. Time-domain BSS suffers from high computational complexity to compute convolution of long filters and update filter coefficients. Frequency-domain BSS can overcome this shortcoming by simplifying convolutive mixing to instantaneous mixing which allows standard instantaneous ICA algorithms to be employed. However, it encounters problems, namely permutation and amplitude ambiguity.

During the last few years, to solve the permutation problem, approaches that employ geometric beamforming [1, 2], filter consistency [3] and spectrum continuity [4], etc, have been investigated. Some of them can sufficiently solve the permutation, but they are still time consuming for real-time processing. Some other methods are computationally efficient but come with a relatively low degree of accuracy. It is still open to a satisfying solution. In this paper, we propose a novel hybrid separating framework which achieves efficiency both on computation and performance.

2. FREQUENCY-DOMAIN BSS

In this section, we briefly review the general model of convolutive mixtures and frequency-domain BSS.

2.1. Mixing and Separation Model

It is widely believed that a linear mixture of source signals weighted by filters is a sufficient model to describe the convolutive mixture. Assume P source signals are recorded by Q sensors in a reverberant environment. In this model, the observed signals $\mathrm{x}_j(k)$, $j = 1, \ldots, Q$ are obtained as the sum of linear convolutions of the source signals $\mathrm{s}_i(k)$, $i = 1, \ldots, P$ and the room impulse response:

$$\mathrm{x}_j(k) = \sum_i \sum_\tau \mathrm{a}_{ji}(\tau)\mathrm{s}_i(k-\tau) \quad (1)$$

where $\mathrm{a}_{ji}(\tau)$ denotes the impulse response from source i to sensor j. The additive noise is not considered because it is sufficient to evaluate this model in a noise free situation.

* This work is sponsored by the Wilkens Missouri Endowment

† Formerly known as Steven L. Gay

The objective of BSS is to design a causal, stable separation filter $b_{ij}(\tau)$ to obtain the estimation of original source signals, which is denoted by $y_i(k)$, $i = 1, \ldots, P$:

$$y_i(k) = \sum_j \sum_\tau b_{ij}(\tau) x_j(k-\tau) \quad (2)$$

Using an L-point short-time Fourier transform (STFT), the time-domain observed signals are transformed into frequency-domain signals:

$$X_j(\omega,k) = \sum_{l=0}^{L-1} x_j(k+l) g(l) e^{-j\omega l} \quad (3)$$

where $g(l)$ represents a window function. Then the BSS model is converted into the frequency domain:

$$\mathbf{X}(\omega,k) = \mathbf{A}(\omega)\mathbf{S}(\omega,k) \quad (4)$$

where $\mathbf{A}(\omega)$ is the mixing matrix in the frequency bin ω. $\mathbf{X}=[X_1,\ldots,X_Q]^T$ and $\mathbf{S}=[S_1,\ldots,S_P]^T$ are time-frequency representations of the observed signals and source signals, respectively. And the estimated signals are turned into:

$$\mathbf{Y}(\omega,k) = \mathbf{B}(\omega)\mathbf{X}(\omega,k) \quad (5)$$

where $\mathbf{B}(\omega)$ is the demixing matrix in the frequency bin ω and $\mathbf{Y}=[Y_1,\ldots,Y_P]^T$. At the last step, the time-domain signals are reconstructed using the inverse STFT:

$$y_i(k+l) = \frac{1}{L}\sum_{l=0}^{L-1} Y_i(\omega,k) e^{j\omega l} \quad (6)$$

2.2. Independent Component Analysis

In each frequency bin, the instantaneously mixed frequency-domain signals are separated. ICA is the most widely used approach to attack this problem. It exploits the statistical independence between the original source signals in order to separate them from the observed mixtures, attempting to make the signals as independent as possible. When the source signals are non-Gaussian and mutually independent, sufficient separation can be achieved.

There have been lots of ICA methods such as InfoMax [5], JADE [6] and FastICA [7], etc. In the proposed method, the well-known FastICA algorithm by Hyvärinen is implemented. According to the complex data value in the frequency domain, the algorithm is complex-valued [8].

2.3. Permutation and Amplitude Ambiguity

Even though the ICA algorithm for instantaneous mixtures accurately estimates the demixing matrix in each frequency bin, it still encounters indeterminacy of permutation and scaling, because ICA does not take into account the order and gain in which the original sources are recovered. Each ICA solution satisfies:

$$\mathbf{W}(\omega)\mathbf{A}(\omega) = \mathbf{\Pi}(\omega)\mathbf{\Gamma}(\omega) \quad (7)$$

where $\mathbf{\Pi}$ represents a permutation matrix and $\mathbf{\Gamma}$ is a diagonal matrix, of which the elements denote the scaling factors.

If the permutation matrix $\mathbf{\Pi}$ is not consistent across all frequencies then contributions from different sources will be combined into a single channel when converting the signal back to the time domain. It is the biggest challenge in the context of frequency-domain BSS. The scaling ambiguity $\mathbf{\Gamma}$ at each frequency bin results in a filtering effect on the sources in the time domain. In order to perfectly recover the sources in the time domain, these indeterminacy problems must be essentially solved before making an inverse STFT.

3. PROPOSED METHOD

In this approach, the permutation control process is embedded into the ICA iterations. In a single frequency bin, the separation matrix obtained by the ICA process is immediately fed into the permutation control unit, where the possible permutation is checked and then corrected if necessary. In the next frequency bin, the ICA step is initialized with the final value of the separation matrix in the previous bin, which does not have permutation. After all the ICA iterations are completed, the separation matrices are sent to the rescaling stage, where the amplitude ambiguity is solved. The flow of proposed method is shown in Fig. 1.

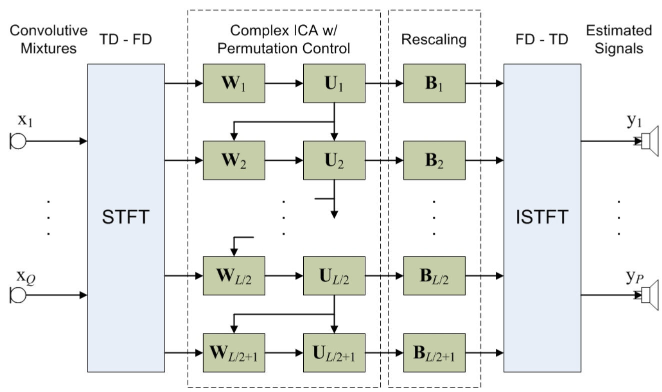


Fig. 1. Diagram of proposed frequency-domain BSS

3.1. Permutation Control

In most of the conventional frequency-domain methods, permutation is corrected in all frequency bins, which is time consuming, especially when the number of sources and sensors is large. In this letter, instead of spending time on all bins, we focus on those bins where permutation could possibly happen. A permutation control unit is proposed to accomplish this job, which is shown in Fig. 2.

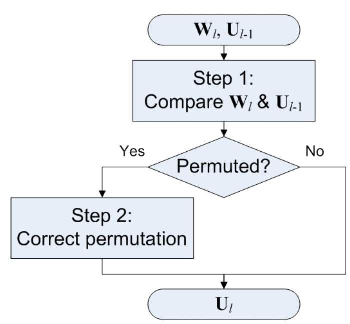


Fig. 2. Flowchart of the permutation control unit

In the first step, criteria must be set to determine whether a possible permutation exists. In order to set the criteria, a fact needs to be noticed. When the sensor signals are converted into the frequency domain, their spectrums change gradually along the frequency axis. Therefore, if the frequency bins are narrow enough, we can also expect that the separation matrices obtained by ICA process in adjacent bins will not have great changes in their coefficients.

Due to this fact, we can employ the distance between adjacent matrices as criteria, which is calculated as

$$D = \sum_{i,j} \left| \mathbf{W}_l(i,j) - \mathbf{U}_{l-1}(i,j) \right| \begin{cases} < \varepsilon \Rightarrow \text{Not permuted.} \\ > \varepsilon \Rightarrow \text{Permuted.} \end{cases} \quad (8)$$

where ε is the threshold. If the distance is under the threshold, no changes will be made. But if it is above the threshold, the permutation will be corrected in the second step. The method utilized to correct the permutation is the beamforming alignment method proposed by Kurita [1]. Then we get the demixing matrix with correct order:

$$\mathbf{U}(\omega) = \mathbf{\Pi}^{-1}(\omega)\mathbf{W}(\omega) \quad (9)$$

Under this two-stage framework, most of the existing permutation solving methods could be exploited. The combination could vary but the method in the step 1 must be computationally efficient, because it is implemented in all frequency bins. The distance method we used is not sufficient to correct all permutations, but good enough to find them. Accuracy and robustness are the key issues for the method in step 2, where the possible permutation will be corrected if necessary. Focusing only on the permuted frequency bins allows more complicated permutation correcting methods to be employed without greatly increasing the overall computational load.

The output of the permutation control module is used as the initial value of the ICA iteration in the next frequency bin. This effort can not only increase the convergence speed, but also significantly control the permutation. In some cases, if the resolution in the frequency domain is high enough, the separation matrices in neighboring frequency bins will even tend to converge in the same order, which means the step of correcting permutation can be avoided.

3.2. Rescaling

The separation matrices obtained in the previous stage needs to be rescaled to remove the amplitude ambiguity. For simplicity, we assume the number of sources and the number of sensors are equal, which means $N = M$, in the following discussions. Assume at the frequency bin ω, the demixing matrix $\mathbf{U}$ is successfully calculated. Then the rescaling matrix $\mathbf{\Gamma}^{-1}$ can be obtained by:

$$\mathbf{\Gamma}^{-1}(\omega) = \mathrm{diag}\left\{\mathbf{U}^{-1}(\omega)\right\} \quad (10)$$

Then (5) turns into:

$$\mathbf{Y}(\omega,k) = \mathbf{\Gamma}^{-1}(\omega)\mathbf{U}(\omega)\mathbf{X}(\omega,k) \quad (11)$$

With (11), the ambiguity of scaling can be resolved.

4. EXPERIMENTAL RESULTS

In this section we present the results of experiments carried out to test the performance of the proposed method. The experiments were conducted using the Image Model [9] and performed on a laptop with a 1.7GHz Pentium M CPU.

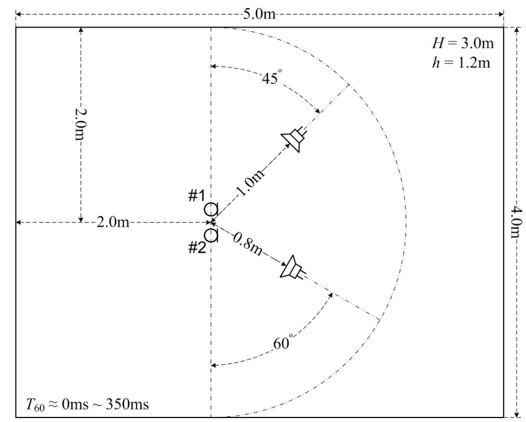


Fig. 3. Experiment room setup

A typical reverberant room was simulated and a 2-input and 2-output noise free case was considered, as shown in Fig. 3. The experiment parameters and conditions are shown in the following table.

Table 1
Parameters and Conditions

Room dimension	L: 5m, W: 4m, H: 3m
Reverberation time ($T_{60} \approx$)	0ms ~ 350ms
Distance between sensors	4cm
Direction of Arrivals (DOA)	45° and 120°
Distance of sources	1.0m and 0.8m
Source signals	2 female speeches of 16seconds
Sample rate	8000Hz
Permutation threshold (ε)	0.2

Fig. 4 shows the results when reverberation time T_{60} is about 220ms. The performance was compared with Kurita's method [1] (using FastICA), as shown in Fig. 5.

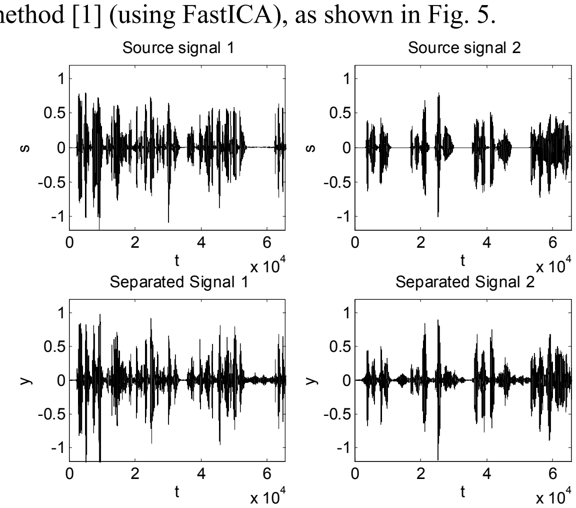
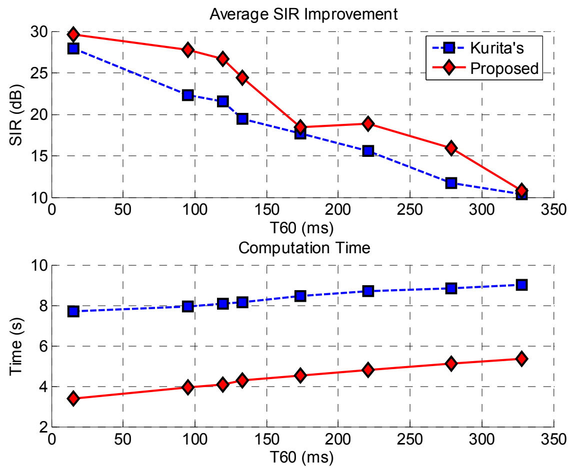
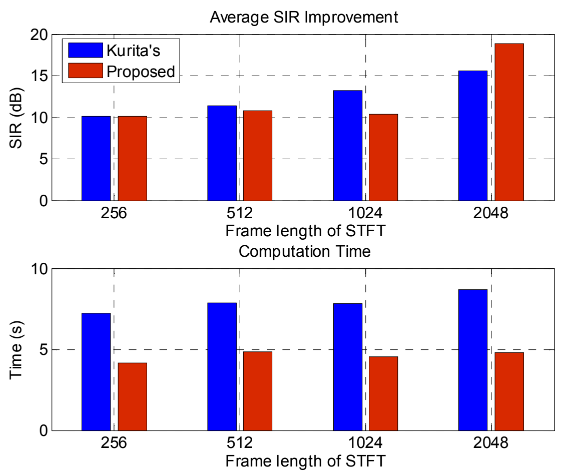


Fig. 4. Results when Frame length = 2048 and $T_{60} \approx$ 220ms



(a) Frame length = 2048, T_{60} varies



(b) $T_{60} \approx$ 220ms, frame length varies

Fig. 5. Performance comparison on average SIR improvement & computation time between Kurita's method and proposed method.

From Fig. 5 (a), it can be seen that the proposed method achieves a higher level on the average SIR improvement than Kurita's method does. When frame length = 2048, the average SIR is improved by 30dB, 19dB and 11dB when $T_{60} \approx$ 0ms, 220ms and 330ms, respectively. Computational efficiency is another advantage of the proposed method over conventional BSS approaches. As we can see from Fig. 5 (b), when $T_{60} \approx$ 220ms, the time saving reaches at least 3 seconds and is gradually increasing as we increase the frame length of STFT. In the fourth case, by using inherited initial values and concentrating on only a few frequency bins which could possibly have permutations, the proposed method reduces the overall computation time of Kurita's method from approximately 9 seconds to 5 seconds, and meanwhile, it achieves a nearly 4dB higher average SIR improvement.

A bigger permutation threshold ε could make the proposed method run even faster, but it could also result in the missing of some permuted frequency bins. In addition, a significantly increase of time saving can be expected when the number of sources to be recovered gets large. By increasing the frame length or improving the ICA algorithm, more decent separation results could be achieved.

One shortcoming of this method is that it may not be competent for the job separating quickly moving sources, as well as most of the existing BSS methods. Furthermore, when the T_{60} gets higher, there are a decrease in the SIR improvement and a slightly increase in the computation time. This is a general problem in frequency-domain BSS, which is caused by the degradation of convergence, and it is even worse in time-domain BSS.

To sum up, in the BSS for static sources, compared with the Kurita's conventional approach, the proposed method is more efficient on computation and performance.

5. CONCLUSION

A new approach for blind separation of convolutive mixtures has been presented. It is based on embedding a two-stage permutation control unit into the ICA process and taking advantage of the separation matrix obtained at each frequency bin to initialize the ICA in next frequency bin. In contrast to conventional frequency-domain BSS algorithms, this method aligns permutation only in a small amount of frequency bins and achieves less computational complexity without lowering the efficiency.

6. REFERENCES

[1] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of frequency-domain blind signal separation using directivity pattern under reverberant conditions," *ICASSP'00*, pp. 3140-3143, 2000.

[2] W. Baumann, D. Kolossa and R. Orglmeister, "Beamforming-based convolutive source separation," *ICASSP'03*, pp. 357-360, 2003.

[3] W. Baumann, B.-U. Köhler, D. Kolossa, and R. Orglmeister, "Real time separation of convolutive mixtures," *ICA'01*, pp. 65-69, 2001.

[4] K. Rahbar and J. P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 5, pp. 832-844, 2005.

[5] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, Vol. 7, pp. 1129-1159, 1995.

[6] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceeding Part F*, Vol. 140, No. 6, pp. 362-370, 1993.

[7] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, Vol. 9, pp. 1482-1492, 1997.

[8] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex-valued signals," *Int. J. Neural Systems*, 10(1):1, pp. 1-8, 2000.

[9] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, Vol. 65, No. 4, pp. 943-950, 1979.