4-1-2008

A Fast and Efficient Frequency-Domain Method for Convolutive Blind Source Separation

Peng Xie

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

A Fast and Efficient Frequency-Domain Method for Convolutive Blind Source Separation*

Peng Xie, Steven L. Grant†
Dept. of Electrical and Computer Engineering,
Missouri University of Science and Technology‡,
Rolla, MO 65409, USA

Abstract-**In this paper, the problem of blind separation of a convolutive mixture of audio signals is considered. A fast and efficient frequency-domain Blind Source Separation (BSS) method using Independent Component Analysis (ICA) is investigated. The main difficulties of this approach lie in the so called permutation and amplitude problems. In order to solve the permutation ambiguity, the final value of the ICA derived separation matrix of one frequency bin, is used to initialize the ICA iterations in the next frequency bin. The amplitude problem is addressed by utilizing the elements in the inverse of the separation matrix. Experimental results demonstrate that successful separation is achieved and compared with conventional frequency-domain BSS methods, it is less computationally complex and has faster convergence.**

I. Introduction

Blind Source Separation (BSS) is a statistical approach for determining a number of independent random or deterministic 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 speech recognition systems, telecommunications, and medical signal processing.

Within BSS research there are two important issues that are generally considered: instantaneous BSS and convolutive BSS. The distinction between them is based primarily on the nature of the signal mixing process. Instantaneous BSS separates signals that are mixed without introducing time delays. It is where the development of BSS as a research field began. Convolutive BSS is an extension of instantaneous BSS which can achieve separation when time delays are involved. Recently convolutive BSS is drawing much of researchers' attention, because in many real-world applications such as communication and acoustics, the signals are mixed in a convolutive manner.

The major approaches to separate the convolutive mixtures can be divided into time-domain and frequency-domain methods. The time-domain method suffers from the high computational complexity. In order to overcome the shortcoming of the time-domain methods, people are moving to the frequency domain, where the problem of convolutive mixing simplifies to instantaneous mixing allowing standard instantaneous ICA algorithms to be employed. Frequency-domain BSS takes much less computation time than time-domain BSS. However, it encounters problems, namely permutation and amplitude ambiguity.

During the last few years several frequency-domain methods have been reported to address the permutation and amplitude indeterminacy [1, 2, 3, 4, and 5]. Among these methods people are utilizing geometric information or spectrum characteristics to solve the permutation problem and using methods like directivity patterns or reference sensors to address the scaling problem. Although most of them cost much less computation time than time-domain methods, they are still time consuming for real-time processing. In this paper, we propose a frequency-domain separating system which runs faster than those methods without lowering the efficiency.

II. BSS in the Frequency Domain

A. Mixing and Separation Model

In this section, we introduce the basic model of the convolutive mixtures. It is believed that a linear mixture of source signals weighted by filters is a sufficient model to describe the mixture. Assume M source signals are recorded by N sensors in a reverberant environment. In this model, the observed signals $\mathrm{x}_1(t)$, …, $\mathrm{x}_N(t)$ are obtained as the sum of convolutions of the source signals $\mathrm{s}_1(t)$, …, $\mathrm{s}_M(t)$ and the room impulse response:

$$\mathrm{x}_i(t) = \sum_j \sum_\tau \mathrm{a}_{ij}(\tau)\mathrm{s}_j(t-\tau) \tag{1}$$

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

The objective of BSS is to design a causal, stable separation filter $\mathrm{b}_{ji}(\tau)$ to obtain the estimation of original source signals, which is denoted by $\mathrm{y}_1(t)$, …, $\mathrm{y}_M(t)$:

* This work is sponsored by the Wilkens Missouri Endowment
† Formerly known as Steven L. Gay
‡ Formerly known as University of Missouri Rolla

978-1-4244-2077-3/08/$25.00 ©2008 IEEE.

$$y_j(t) = \sum_i \sum_\tau b_{ji}(\tau) x_i(t-\tau) \quad (2)$$

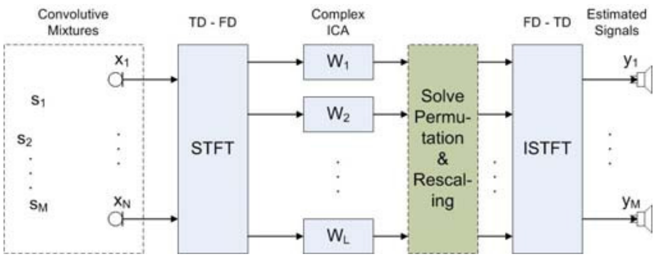


Fig. 1. Diagram of frequency-domain BSS

The flow of frequency-domain BSS is shown in Fig. 1. Using short-time Fourier transform (STFT), the time-domain observed signals are transformed into frequency-domain signals:

$$X_i(\omega,t) = \sum_{k=0}^{K-1} x_i(t+k) w(k) e^{-j\omega k/K} \quad (3)$$

where $w(k)$ denotes a window function. Then the BSS model is converted into the frequency domain:

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

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

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

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

$$y_j(t) = \frac{1}{K} \sum_{k=0}^{K-1} Y_j(\omega,t) e^{j\omega k/K} \quad (6)$$

B. Independent Component Analysis

In each frequency bin, the instantaneously mixed frequency-domain signals are separated. Independent Component Analysis (ICA) is the most widely used approach to attack this problem. ICA 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, good separation is achieved.

There have been lots of existing ICA methods such as InfoMax [6], JADE [7] and FastICA [8]. 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 and is formulated as follows [9].

$$\begin{aligned} \tilde{B}_n &= E\left\{\mathbf{X}(B_n\mathbf{X})^* h(|B_n\mathbf{X}|^2)\right\} \\ &\quad - E\left\{h(|B_n\mathbf{X}|^2) + |B_n\mathbf{X}|^2 h'(|B_n\mathbf{X}|^2)\right\} B_n \quad (7) \\ B_n &= \tilde{B}_n \big/ \|\tilde{B}_n\| \end{aligned}$$

where B_n is a demixing weight vector, which forms the n-th row of the demixing matrix $\mathbf{B}$. $h(\cdot)$ is a nonlinear function and $h'(\cdot)$ denotes its differential.

C. Permutation and Amplitude Ambiguity

Even though the ICA algorithm for instantaneous mixtures precisely estimates the demixing matrix at each frequency bin, it still has 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:

$$\mathbf{B}(\omega)\mathbf{G}(\omega)\mathbf{A}(\omega) = \mathbf{P}(\omega)\mathbf{D}(\omega) \quad (8)$$

where $\mathbf{G}$ represents the whitening matrix used in ICA process $\mathbf{P}$ is a permutation matrix and $\mathbf{D}$ is a diagonal matrix, of which the elements denote the scaling factors.

If the permutation matrix $\mathbf{P}$ 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. The scaling ambiguity 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, those above indeterminacy problems must be essentially solved before making an inverse STFT from the frequency domain to the time domain.

III. Proposed Method

A. Solving Permutation

In order to overcome the permutation problem, a fact needs to be noticed. When the sensor signals are converted into the frequency domain, their spectrums change gradually along the frequency axis, which means if the frequency bins are narrow enough, the spectrums between neighboring bins are highly correlated. Therefore, we can expect that the separation matrices obtained by FastICA in adjacent frequency bins will not have great changes in their coefficients, as well as their permutation orders.

Due to this fact, we can employ the final solution of the separation matrix in the previous frequency bin as the initial value of the FastICA iteration in the current frequency bin. If the resolution in the frequency domain is high enough, the separation matrices in consecutive frequency bins will tend to converge in the same permutation order, which means the step of solving permutation can be avoided.

Fig. 2 shows the flow of the iteration to compute the separation matrices. The iteration can be performed in two equally efficient ways: 1. Start from the lowest frequency; 2. Start from the highest. Method 2 runs slightly faster than method 1, as it converges faster in the lowest frequency bins.

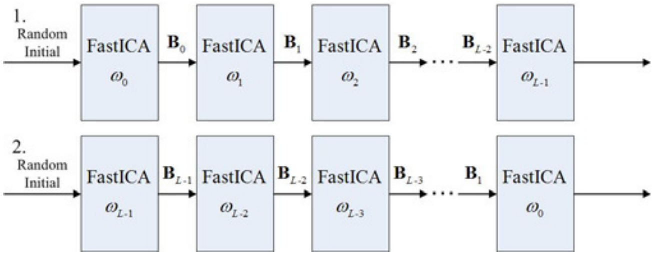


Fig. 2. Iteration flow across all frequency bins

B. Rescaling

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 separation matrix **B** is successfully calculated, and **C** is its inverse (or pseudo inverse). Matrix **C** can be denoted by:

$$\mathbf{C}(\omega) = \mathbf{B}^{-1}(\omega) = \begin{bmatrix} c_{11}(\omega) & \cdots & c_{1M}(\omega) \\ \vdots & \ddots & \vdots \\ c_{M1}(\omega) & \cdots & c_{MM}(\omega) \end{bmatrix} \quad (9)$$

Let the rescaling matrix **R** be the diagonal elements of **C**:

$$\mathbf{R}(\omega) = diag\{\mathbf{C}(\omega)\} \quad (10)$$

Then (5) changes into:

$$\mathbf{Y}(\omega,t) = \mathbf{R}(\omega)\mathbf{B}(\omega)\mathbf{X}(\omega,t) \quad (11)$$

With (11), good amplitude rescaling can be achieved.

IV. 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 [10] and performed on a Dell laptop with a Pentium M CPU at 1.7GHz. A typical reverberant room was simulated, as shown in Fig. 3.

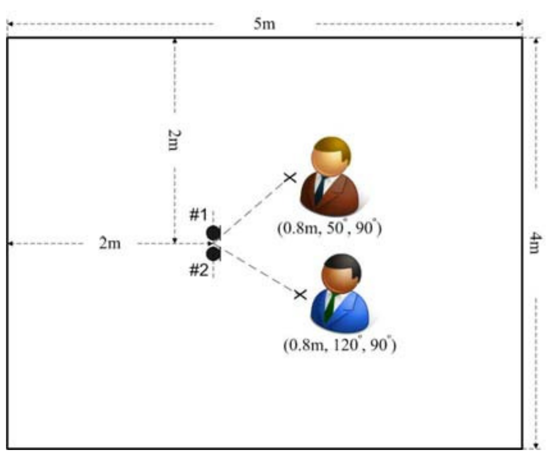


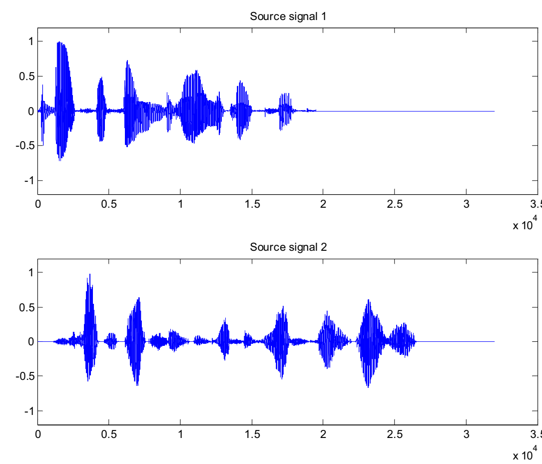
Fig. 3. Experiment room setup

A 2-input and 2-output noise free case was considered. And the experiment parameters and conditions are shown in the following table.

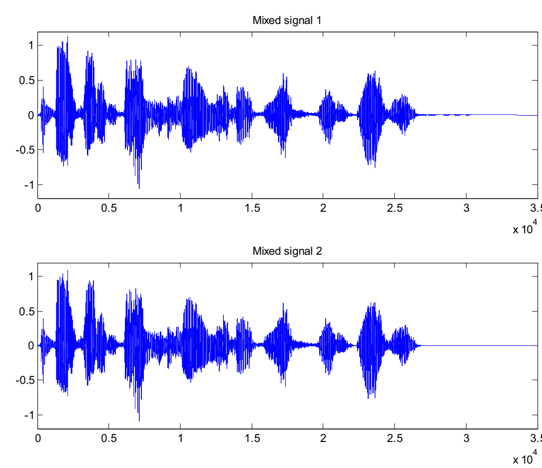
TABLE I
PARAMETERS AND CONDITIONS

Room dimension	L: 5m, W: 4m, H: 3m
Reverberation time ($T_{60} \approx$)	100ms/150ms/200ms/300ms
Number of sensors	2
Distance between sensors	2cm
Direction of Arrivals (DOA)	50° and 120°
Distance of sources	0.8m
Source signals	2 male speeches of 4s
Sample rate	8000Hz
Frame length of STFT	1024
Number of frequency bins	513
Nonlinear function in ICA	$G_2(u) = \log(0.1+u)$

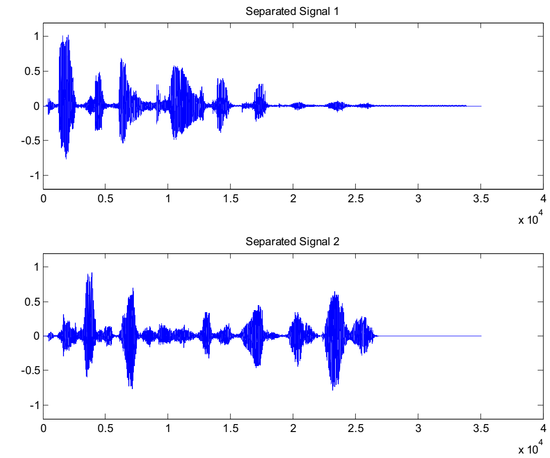
Fig. 4 shows the separation results when room reverberation time is 150ms and the performance was compared with the method by Kurita [2], as shown in Fig. 5.



(a)



(b)



(c)

Fig. 4. Separation results when $T_{60} \approx$ 150ms: (a) Source signals; (b) Mixed signals recorded by Sensor #1 and #2 respectively; (c) Separated signals

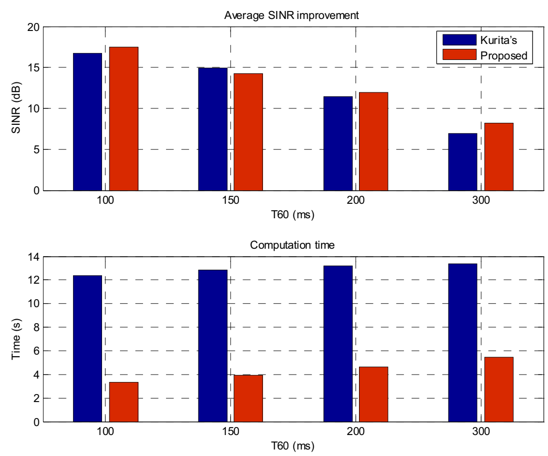


Fig. 5. The performance evaluations of the Kurita's method and proposed method

From Fig. 5, we can see that the proposed method achieves the same level on the average SINR improvement as Kurita's method. The average SINR is improved by 14.3dB when $T_{60} \approx$ 150ms. Saving time is the biggest advantage of the proposed method over conventional BSS techniques. In the second case, it reduces the computation time in ICA stage from 10 seconds in Kurita's method to around 3.5 seconds. Also, it jumps over the permutation solving step and saves more than 2.5 seconds in this stage.

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, there are a decrease in the SINR improvement and an increase in the computation time, when the T_{60} gets higher. 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 approach, the proposed method is equally efficient but runs much faster.

V. Conclusion

A new approach for blind separation of convolutive mixtures has been presented. It is based on taking advantage of the separation matrix obtained by ICA at each frequency bin, where the current separation matrix is used to initialize the ICA process in next frequency bin and the diagonal elements of it are exploited as the scaling factors. In contrast to other frequency-domain BSS algorithms, this method does not suffer from permutation indeterminacy across frequency bins and achieves faster convergence and easier rescaling.

Acknowledgment

The authors would like to thank Aapo Hyvärinen and Erkki Oja, Helsinki University of Technology, Finland, for making their valuable FastICA code available.

References

[1] V. C. Soon, L. Tong, Y. F. Huang, and R. Liu, "A wideband blind identification approach to speech acquisition using a microphone array," in *ICASSP'92*, vol. 1, pp. 293–296, 1992.

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

[3] M. Ikram and D. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation," in *ICASSP'02*, pp. 881-884, 2002.

[4] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Blind source separation of many signals in the frequency domain," in *ICASSP'06*, pp. 969-972, 2006.

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

[6] 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.

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

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

[9] 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.

[10] 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.