Proceedings of the 5th WSEAS Int. Conf. on SIGNAL, SPEECH and IMAGE PROCESSING, Corfu, Greece, August 17-19, 2005 (pp68-73)

Blind Source Separation Using Variable Step-Size Adaptive Algorithm in Frequency Domain

HYUN-TAE KIM, *JANG-SIK PARK
Department of Multimedia Engineering, Dongeui University
Gaya-Dong, San 24, Busanjin-Gu, Busan, 614-714
KOREA(SOUTH)
*Department of Visual Technologies, Dongeui Institute of Technology
Yangjung-Dong, San 72, Busanjin-Gu, Busan, 614-715
KOREA(SOUTH)

Abstract: This paper introduces a variable step-size adaptive algorithm for blind source separation. From frequency characteristics of mixed input signals, we need to adjust the convergence speed regularly in each frequency bin. This algorithm varies a step-size according to the magnitude of input at each frequency bin. This guarantee of the regular convergence in each frequency bin would become more efficient in separation performances than conventional fixed step-size frequency domain ICA. Computer simulation results show the improvement of about 5 dB in signal to interference ratio(SIR) and the better separation quality.

Key-Words: Blind source separation, Independent component analysis, Adaptive algorithm, Variable step-size, Frequency domain ICA

1 Introduction

Recently, blind source separation(BSS) is a technique for estimating original source signals using only observed mixtures. The adjective “blind” stresses the fact that firstly sources are not known and secondly no information is available about the mixing information. A typical modeling is to record two people talking at the same time using two microphones. The recorded signals would then of course consist of a mixture of the two speech sounds. The applied algorithm then tries to estimate the inverse channel and force the recorded signals to be independent of each other in order to separate the signals[1]. BSS based on independent component analysis(ICA) technique has been found effective in signal separation comparing other BSS methods. ICA is a statistical method that was originally introduced in the context of neural network modeling [2].

Methods for constructing separation filters can be classified into two approaches. The first one is a time-domain approach(TDICA)[3], where the coefficients of the separating filters are calculated directly in the convolved mixture model. It has an advantage in that ICA is applied to instantaneous mixtures, which are easier to solve than convolved mixture in the time domain. The other is a frequency-domain approach(FDICA)[4,5], where the frequency responses of the separating filters are first calculated, and then the time-domain representations of the filters are obtained by applying an inverse discrete Fourier transform(DFT) to them. FDICA for convolutive mixtures can be performed efficiently, where the analysis is applied separately in each frequency bin independently. Computationally, it may be lighter to move to the frequency domain, as convolutions in the time domain become efficient multiplications in the frequency domain.

This paper deals with the frequency-domain approach. Frequency-domain approach transforms the observed signals into the each frequency component bin by short-time DFT frame by frame. Then optimize the inverse of the mixing component in each frequency bin. Finally, the optimized weights at each bin can reconstruct the full-band separated signals in time-domain. In FDICA, we have to consider complex data in general. For this purpose Smaragdis[4] proposed a complex-valued ICA algorithm, which was an extension of infomax algorithm[1]. The nonlinear function use in the extension was based on the Cartesian coordinates of a complex number. The nonlinear function is applied separately in the real and imaginary parts. This type of nonlinear function has been widely used by other researchers[6]. However, there are disadvantages for converging to the optimal solution. The separation performance is saturated before reaching a sufficient performance because the independence assumption collapses in each frequency bin. Secondly, the permutation among source signals and indeterminacy of each source gain each bin. As for these disadvantages, various solutions have been already proposed[7,8]. The separation is performed independently bin-to-bin, the convergence of the separation matrix is non-uniform at each bin.

As discussed above, the non-uniform convergence at each frequency causes pre-saturation or permutation. Hence, in order to resolve FDICA problems, we propose a new algorithm in which the variable step-size is used. Proposed method is to adopt a variable step-size which is normalized by the input signal in each bin. This modified version is only modest increase in computation about 15%[9] over the conventional ICA algorithm, while convergence time is reduced in some instances by about a factor of 2. This uniform convergence at each bin reduces the effects of a permutation and pre-saturation problem. A good control of step-size is for faster convergence and better separation quality.

With the results of simulations on separating speech signals in a convolved mixture, we compare the behaviors of proposed algorithm with conventional ICA. Then this paper discusses the performance about the efficient separation.

2 Frequency Domain ICA

In this study, the number of microphone is K and the number of multiple sound sources is L. When the multiple sound sources are linearly mixed, the observed signals are expressed as

$$\mathbf{x}(t) = \sum_{n=0}^{N-1} \mathbf{a}(n)\mathbf{s}(t-n) = \mathbf{A}(z)\mathbf{s}(t) \tag{1}$$

where $\mathbf{s}(t) = [s_1(t), \ldots, s_L(t)]^T$ is the source signal vector, and $\mathbf{x}(t) = [x_1(t), \ldots, x_K(t)]^T$ is the observed signal vector. Also, $\mathbf{a}(n) = [a_{kl}(n)]_{kl}$ is the mixing filter matrix with the length of N, $\mathbf{A}(z) = \left[\sum_{n=0}^{N-1} a_{kl}(n)z^{-n}\right]_{kl}$ is the z-transform of $\mathbf{a}(n)$, where z^{-1} is used as the unit-delay operator, i.e., z^{-n}. $x(t) = x(t-n)$, $a_{kl}(n)$ is the impulse response between the k-th microphone and the l-th sound source, and $[X]_{ij}$ denotes the matrix which includes the element X in the i-th row and the j-th column. Hereafter, we only deal with the case of $K = L$ in this paper.

A conventional mixture in the time domain corresponds to instantaneous mixtures in the frequency domain. Hereafter, the convolutive BSS problem is considered in the frequency domain unless stated otherwise. Note that digital signal processing in the time and frequency domains are essentially identical, and all discussions here in the frequency domain are also essentially true for the time-domain convolutive BSS problem. Therefore, we can apply an ordinary ICA algorithm in the frequency domain to solve BSS problem in a reverberant environment. Smaragdis[4] exploited the transform of convolved mixing into simple multiplicative operation and proposed the application of a short-time discrete Fourier transform(STDFT) for (1), and then to separate independent components in every frequency bin.

Thus, in the frequency domain, the entire process of convolved signal separation is transformed into the computation of the separation matrix in each frequency bin for each source.

Applying the model in the frequency domain introduces a new problem: the frequency bins are treated as being mutually independent. As a result, the estimated source signal components are recovered with a different order in the frequency bins. And in FDICA, the scaling problem also become nontrivial, i.e., the estimated source signal components are recovered with a different gain in the different frequency bins.

Also it is easy to converge to the separation filter in an iterative ICA learning with a high stability. However, the separation performance is saturated before reaching a sufficient performance because the independent assumption collapses in each narrowband [8]. This is because we transform the full-band signals into narrow band signals especially when the umber of sub-band is large. This is a serious and inherent problem, and this prevent us from applying FDICA in a real acoustic environment with a long reverberation.

The signal model in the frequency domain is the following form.

$$\mathbf{X}(w,t) = \mathbf{H}(w)\mathbf{S}(w,t) \tag{2}$$

where, w is the angular frequency, and t represents the frame index. The separating process can be formulated in each frequency bin as :

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

where $\mathbf{S}(w,t) = [S_1(w,t), \cdots, S_L(w,t)]^T$ is the source signal in frequency bin w, $\mathbf{X}(w,t) = [X_1(w,t), \ldots, X_K(w,t)]^T$ denotes the observed signals.

Next, $\mathbf{Y}(w,t) = [Y_1(w,t), \ldots, Y_L(w,t)]^T$ is the estimated source signal vector, and $\mathbf{W}(w)$ represents the separating matrix. $\mathbf{W}(w)$ is determined so that $Y_i(w,t)$ and $Y_j(w,t)$ become mutually independent. For the simple notation, we will annihilate the terms w and t.

To calculate the separating matrix $\mathbf{W}(w)$, we use an optimization algorithm based on the minimization of the mutual information of mixed signals. Different theories, such as informax approach, maximum likelihood, negentropy maximization nonlinear principal component analysis (PCA) and Bussgang cost function based algorithm, for ICA lead to the same iterative learning rule for BSS [7].

To deal with complex signals in ICA at each frequency, the separating matrix was updated using the following learning rules.

$$\mathbf{W}_{\mathbf{i+1}}(w) = \mathbf{W}_{\mathbf{i}}(w) + m \cdot \Delta\mathbf{W}_{\mathbf{i}}(w) \tag{4}$$

$$\textit{where } \Delta\mathbf{W} = m\left[\mathbf{I} - \Phi(\mathbf{Y})\mathbf{Y}^H\right]\mathbf{W},$$
$$\Phi(\mathbf{Y}) = \tanh\left[re(\mathbf{Y})\right] + j\tanh\left[im(\mathbf{Y})\right].$$

where sub $\mathbf{i}$ means the iteration number and w is frequency bin. $\mathbf{Y}^H$ represents the conjugate transpose of $\mathbf{Y}$, and $re[\mathbf{Y}]$ and $im[\mathbf{Y}]$ are the real and imaginary parts of $\mathbf{Y}$, respectively. In the nonlinear functional $\Phi(Y_i)$, $\tanh(\cdot)$ is applied separately in the real and imaginary parts. The matrix $\mathbf{I}$ is an identity matrix. The constant m is termed the learning rate or step-size. Fig. 1 represents the total processing of the FDICA algorithm.

Generally, fixed value of m is used in conventional ICA algorithm. However, in general, we can observe that the characteristics of the speech signals have a large energy in low frequency band. Thus the convergence is irregular with each frequency bin in adapting the FDICA algorithm. This irregularity would cause the problem of the FDICA. If adjusting this converging rate regularly, we would expect better result. It would be the solution to the problems of the FDICA.

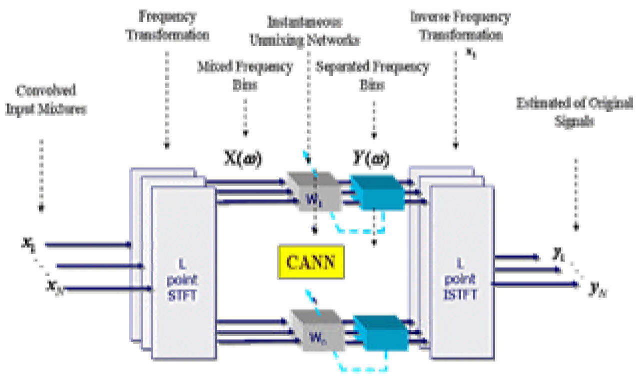

Fig. 1. Block diagram of the frequency domain algorithm

3 Proposed Variable Step-size ICA

This paper proposes a variable step-size algorithm at each frequency bin component for improving the separation performance based on FDICA. The step-sizes are various to each input component differently. In equation (4), m is (fixed) step-size parameter that ranged with $0 < m < 1$. Generally, this parameter controls the speed of convergence. Since the convergence time is inversely proportional to m, a large m is selected for fast convergence in applications with non-stationary input signals. This selection, however, results in increased excess mean squared error (MSE). And small m causes a slowness of convergence to the weak input signals.

We need to normalize the step-size to the input signals. Especially, in FDICA method, a different input at each frequency component could have exhibited irregular convergence speeds among the different frequency bins. This irregularity could have caused to mislead to the local minimums.

A modified version of the ICA update algorithm will be proposed and we will show the characteristic of the input speech signal in frequency domain. The step-size parameter is variable to the input magnitude by normalizing the input signals. The step-size changes at each frame as well as at each bin. The variable step-size algorithm is shown in (5).

$$\mathbf{W}_{\mathbf{i+1}}(w) = \mathbf{W}_{\mathbf{i}}(w) + m_{\mathbf{i}}(w) \cdot \Delta\mathbf{W}_{\mathbf{i}}(w) \tag{5}$$

where $m_{\mathrm{i}}(w) = m / \max[\,\|X_1(w)\|, \|X_2(w)\|\,]$.

The term $m_{\mathrm{i}}(w)$ is time-varying step-size, this term $m_{\mathrm{i}}(w)$ is regularized for the updating equation at different bin.

There are various methods to normalize the step-size. A method using the norm of the input signals frame by frame may have large perturbation in case of radical changes of inter-frame frequency bin. So we divide the magnitude of the input signal by ten levels as shown in Fig. 2. The Fig. 2 represents a normalized histogram of the 1024-tap frequency components of the mixing data. We can observe that the lower frequency component below than about 400$^{\text{th}}$ bin has a more dominant energy than the higher one.

Then we endow the step-size with inversely proportioned values between 0.1 and 1. This 10-step quantizing method copes with the perturbation of separating matrix by the small changes of input. Moreover, this technique has an effect as like the adopting of low pass filter which help the separating matrix for stable convergence, too.

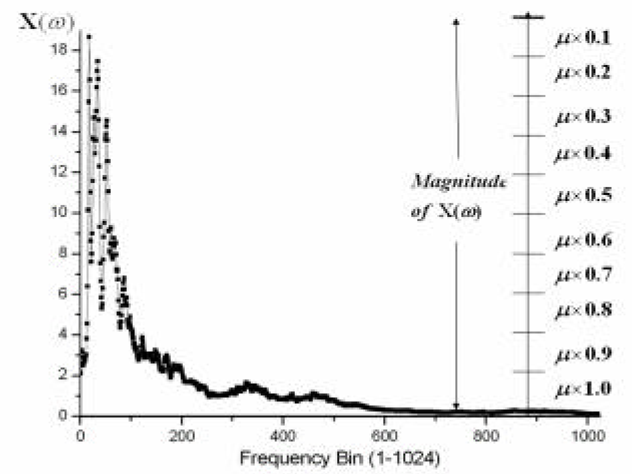

Fig. 2. Histogram of input data and step-size normalization

4 Simulation Results

To examine the effectiveness of the proposed method, we carried out computer simulations using dry speech signals with 8 kHz sampling rate. It is assumed that two omni-directional microphones with an inter-element spacing of 4-cm. The step size parameter m in (4) affects the separation performance of BSS when the convolved mixing channel changes. We chose m to optimize the performance for each mixing channel.

To test the algorithms, we used mixing systems as follows in z-transform domain,

$$A_1(z) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \tag{6}$$

The mixing system of (6) shows an instantaneous (nonreverberant) mixing case.

$$A_2(z) = \begin{bmatrix} 0.9 + 0.5z^{-1} + 0.3z^{-2} & 0.5z^{-6} + 0.3z^{-7} + 0.2z^{-8} \\ -0.7z^{-9} - 0.3z^{-10} - 0.2z^{-11} & 0.8 - 0.1z^{-1} \end{bmatrix} \tag{7}$$

The mixing system of (7) is a convolutive (reverberant) mixture, and has a minimum phase with all its zeros inside the unit circle used in [10].

We assumed the straight component $y_{ii}(t)$ as a signal, and the cross-channel component $y_{ij(i\neq j)}(t)$ as interference.

We define the output signal-to-interference ratio (SIR_{Oi}) for $y_i(t)$ as :

$$SIR_{Oi} \equiv 10\log \frac{\sum_t |y_{ii}(t)|^2}{\sum_t \left(\sum_{i\neq j} y_{ij}(t) \right)^2} (dB) \tag{8}$$

SIR is used as an average of SIR_{O1} and SIR_{O2} in order to measure the performance. This measurement is consistent with the performance evaluation of BSS in which the crosstalk component assumed as interference. We measured SIR with six combinations of source signals using two males and two females speakers, and averaged them.

Table 1. Comparison of the SIR results

Algorithm	Inst. case	Conv. case	Max. SIR	Min. SIR
FDICA(dB)	25	15	32	12
Proposed FDICA(dB)	25	21	32	17

Table 1 shows the mean SIR values. It is obvious that

Proceedings of the 5th WSEAS Int. Conf. on SIGNAL, SPEECH and IMAGE PROCESSING, Corfu, Greece, August 17-19, 2005 (pp68-73)

SIR improvements for instantaneous mixing cases are almost the same for both fixed step-size algorithm and variable step-size algorithm. However, for reverberant condition, proposed method gives good result in the performance about 4-6 dB. Furthermore, this table shows a maximum SIR and minimum SIR. In the comparing of minimum SIR' s, ours outperformed to conventional ICA about 5 dB above.

And second measure is a performance matrix. After the data pass the estimate unmixing FIR matrix was multiplied with the corresponding mixing matrix to obtain a performance matrix. This matrix is an indication of how well the inputs were separated, and has to be close to the unit FIR matrix (a matrix where the diagonal elements are delta function and the rest is zero) to denote success. The performance matrix is shown in (9).

$$\mathbf{M} = \mathbf{WA} \approx \mathbf{I} \tag{9}$$

Now, we compare the separation qualities by Fig 3.and Fig. 4. These two figures show the performance matrix. The Fig. 3 represents the result of the conventional ICA and proposed algorithm result is in Fig. 4. In this Fig. 4, the mixture was separated with more conservative adaptation parameter selection and by using the influence factor described in the previous. Resulting separation of Fig. 4 was almost inaudible in this case.

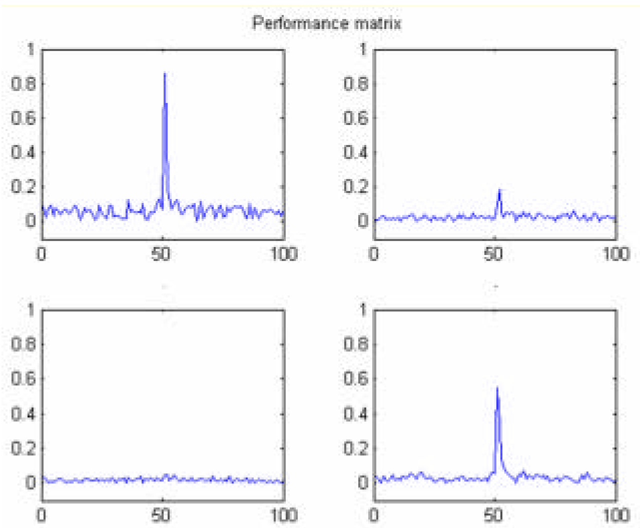

Fig. 3. Performance matrix for FDICA

As shown in Fig. 2, it is clear that separation will be better especially in low frequency range below 2 *kHz*. Speech signals do not have significant high frequency content so training of the high frequency network is usually bad. This is evident in the plots where the interfering signals are seen as high-pass filters, while the cleaned signals are more impulse-like.

This can be seen better in Fig. 5 which are the frequency domain representations of the performance matrix using proposed method, i.e., $\mathbf{M}(f)$. The dashed lines are interfering signal ($\mathbf{M}_{12}(f)$) and the solid lines are the desired signal ($\mathbf{M}_{11}(f)$). We can note the seemingly poor performance at the high frequency regions as discussed in [4]. This is however not a problem given that there is no excitation at these levels. Excluding permuting problems for the frequency range of the inputs (10 Hz – 800 Hz) the algorithm works fine.

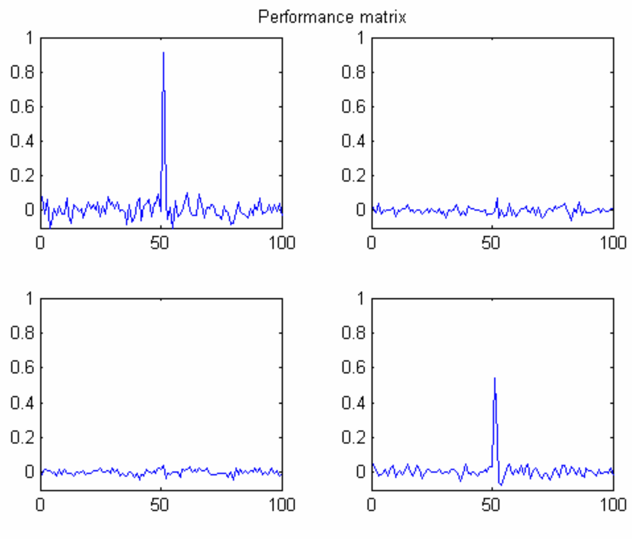

Fig. 4. Performance matrix for proposed ICA

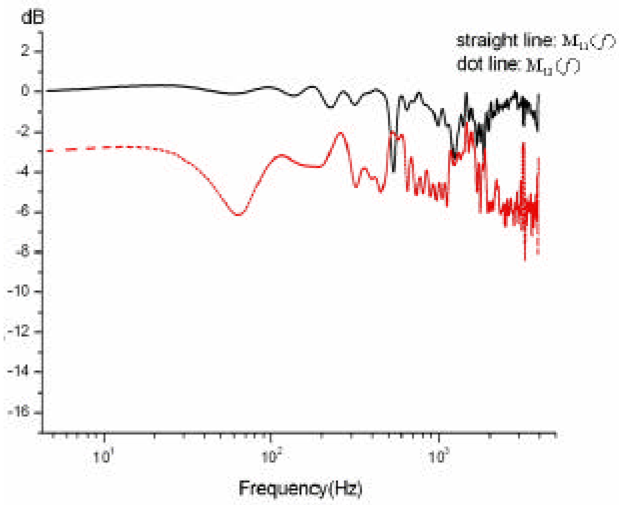

Fig. 5. Frequency domain representation of the performance matrix

5 Conclusion

Our work is aimed at developing the BSS algorithm based on FDICA to the convolved mixing case. This paper proposes a variable step-size algorithm at each frequency bin component for improving the separation performance based on FSICA. The different magnitude of each frequency component causes a bad result for successful convergence. The regular weights are updated by normalizing the step-size with the magnitude of input signal at each frequency bin. The performance of this method has been verified by subjective listening tests and by quantitative measurements.

References:

[1]A. J. Bell and T. J. Sejnowski, *An information-maximization approach to blind separation and blind deconvolution*, Neural Computation, Vol. 7, No. 6, pp. 1129-1159, 1995.

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

[3] S. I. Amari, A. Cichocki and H. H. Yang, *A new learning algorithm for blind signal separation*, Advances in neural information Processing systems *8*. MIT Press, Cambridge, MA.

[4] P. Smaragdis, *Blind separation of convolved mixtures in the frequency domain*, Nerucomputing, Vol. 22, pp. 21-34, 1998.

[5] H. Saruwatari, T. Kawamura and K. Shikano, *Blind source separation for speech based on fast-convergence algorithm with ICA and beamforming*, Proc. Eurospeech 2001, pp. 2603-2606, Sep. 2001.

[6] R. Prasad, H. Saruwatari, A. Lee and K. Shikano, "A fixed- point algorithm for convoluted speech signal separation," *4th international symposium on ICA 2003*, pp. 579-584, Apr. 2003.

[7] S. Araki, S. Makino, T. Nishikawa and H. and Saruwatari, *Fundamental limitation of frequency domain blind separation for convolutive of speech*, Proc. ICASSP 2001, pp. 2737-2740, May. 2001.

[8] S. Kurita, H. Saruwatari, S. Kajita, K. Ikeda and F. Itakura, *Evaluation of blind signal separation method using directivity pattern under reverberant conditions, Proc.* ICASSP 2000, pp. 3140-3143, Jun. 2000.

[9] S. Haykin, *Adaptive Filter Theory*, Third Edition, Prentice Hall Inc., NJ 1996.

[10]T. W. Lee, A. Bell and R. Lambert, *Blind separation of delayed and convolved sources*, Advances in neural information processing systems 9, MIT Press, 1999.