RCES REVIEW OF COMPUTER ENGINEERING STUDIES

ISSN: 2369-0755 (Print), 2369-0763 (Online)
Vol. 4, No. 4, December, 2017, pp. 103-107
DOI: 10.18280/rces.040401

A publication of IIETA

IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering
http://www.iieta.org/Journals/RCES

Blind source separation algorithm for convolution mixed signals

Chunli Wang*, Quanyu Wang, Yuping Cao

College of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

Email: wcl@mail.lzjtu.cn

ABSTRACT

In the actual speech enhancement application, a large number of observation data need longer filters. The time domain algorithm has the disadvantages of large computation amount and slow processing speed. Transforming the time domain convolution operation into the frequency domain product operation can not only avoid the complicated convolution operation, but also reduce the calculation amount to a large extent, and improve the effectiveness of the blind source separation algorithm. Simulation experiment results show that the blind deconvolution algorithm in the frequency domain can improve the intelligibility and articulation of separated speech.

Keywords: Speech Enhancement, Frequency Domain, Convolution, Blind Source Separation, Effectiveness.

1. INTRODUCTION

The classical case of convolution blind source separation is "cocktail party effect" [1]. In a multi-person simultaneous speaking or noisy music environment, the microphone receives the component weighted signals after mixing and delay, that's, the convolution signals, and the required speech signal can be distinguished by the "Blind deconvolution" of the human ear. In practical applications, most of the observed signals to be separated are convolutionally mixed, and their statistical properties change with time [2]. The traditional instantaneous blind source separation algorithm has some drawbacks in dealing with convolution problems.

2. MATHEMATICAL MODEL OF CONVOLUTION MIXING

In the actual environment, the signal received by the microphone is mixed with pure speech signal, room reverberation, noise and some other interferences, with its model shown in Figure. 1 [3].

A convolution mixed signal with m source signals and n mixed signals is represented by Equation (1).

$$x_j(k)=\sum_{i=1}^{m}\sum_{p=0}^{P-1}h_{ji}(p)s_i(k-p),\ j=1,2,\cdots,n \tag{1}$$

where, $x_j(k)$ is the j mixed signal, s_i is the i source signal, $h_{ji}(p)$ is the transfer function from the i source signal to the j mixed signal, and P is the order of this transfer function. The above equation is rewritten as a matrix as shown in Equation (2).

$$x(k)=\sum_{p=0}^{P-1}H(p)s(k-p) \tag{2}$$

where, $H(p)$ is a mixed matrix of $n\times m$ order, each element of which is a filter of P order [4]. Then, the blind separation task of the convolution mixed signals is to solve a Q order separation filter matrix, so that the Equation (3) holds.

$$y(k)=\sum_{q=0}^{Q-1}W(q)x(k-q) \tag{3}$$

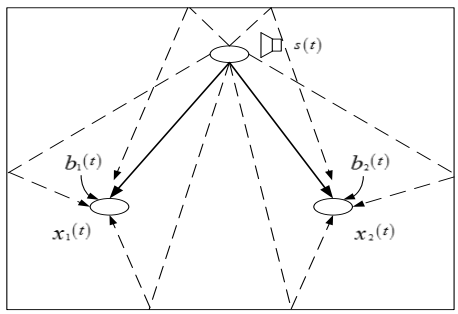


Figure 1. Model of speech signal transmission in actual environment

where, $y(k)$ is the estimation of the source signal $x(k)$ and $W(k)$

is the separation matrix of $m \times n$, and each element therein is a filter of Q order. If $y_i(k)=a_j(s_j(k))$, $a_j(\cdot)$ is an unknown transfer function, Equation (3) may also be written as Equation (4).

$$y_i(k) = \sum_{j=1}^{n} \sum_{q=0}^{Q-1} w_{ij}(q) x_j(k-q), i = 1, 2, \cdots, m \tag{4}$$

When Equations (2) and (3) are transformed into the frequency domain, Equation (5) holds.

$$\begin{aligned} x(w) &= H(w) s(w) \\ y(w) &= W(w) x(w) \end{aligned} \tag{5}$$

As can be seen from Equation (5), it's blind source separation of the instantaneous mixed signal to convert the time-domain convolution mixed signal to the frequency domain.

3 FREQUENCY DOMAIN BLIND DECONVOLUTION ALGORITHM

In practical application, a large amount of observation data needs a long filter, and the time domain algorithm has the disadvantages of long time consuming and large calculation amount. In this case, it is necessary to transform convolution operation in the time domain into product operation in the frequency domain, to avoid complex convolution operation and reduce the calculation amount [5].

3.1 ICA blind deconvolution algorithm in frequency domain

Frequency domain algorithm is the most commonly used method in speech signal blind separation at present. The main idea is to transform the time domain mixed signals into frequency domain product by short-time Fourier transform, then perform instantaneous mixed blind source separation at each frequency point (generally called independent component analysis ICA algorithm [6]). Because of the scaling and arrangement in frequency domain algorithm, it is necessary to descale and de-arrange the frequency domain signals, and then carry out inverse short-time Fourier transform of signals in the separated frequency domain to recover the original signals.

The time domain signal x_j is converted into a frequency domain time series signal $X_j(w, k)$ by T-point windowed discretized short-time Fourier transform, as shown in Equation (6).

$$X_j(\omega, k) = \sum_{\tau=0}^{T-1} x_j(\tau + kR) w(\tau) e^{-j\omega\tau} \tag{6}$$

where, $w(\tau)$ is a window function, with Hanning window, Hamming window and others avaible, k is the position in a window of T width, R is the time interval of window movement and ω is frequency, $\omega = 0, \frac{2\pi}{T}, \cdots, \frac{(T-1)2\pi}{T}$.

For each frequency point, Equation (7) holds.

$$X(\omega, k) = H(\omega) S(\omega, k) \tag{7}$$

where, $H(\omega)$ is mixed matrix. So, separated signals can be obtained by the separation matrix $W(\omega)$ of each frequency point, as shown in Equation (8).

$$Y(\omega, k) = W(\omega) X(\omega, k) \tag{8}$$

In Equations (7) and (8), the following relationship is satisfied:

$$\begin{aligned} S(\omega, k) &= \left[S_1(\omega, k), \cdots, S_M(\omega, k) \right]^T \\ X(\omega, k) &= \left[X_1(\omega, k), \cdots, X_N(\omega, k) \right]^T \\ Y(\omega, k) &= \left[Y_1(\omega, k), \cdots, Y_M(\omega, k) \right]^T \end{aligned} \tag{9}$$

where, $S_j(\omega, k)$ and $Y_j(\omega, k)$ are discrete short-time Fourier transform of source and separated signals. $W(\omega)$ can be iterated and updated at each frequency point until the components in $Y(\omega, k)$are independent of each other.

3.2 Second-order statistical method

The idea of the method: for each element in the time block k(k=0, 1, …, K-1), it is necessary to find $W(\omega)$ to diagonalize the covariance matrix $R_{yy}(\omega, k)$ [7], and the process is as shown in Equation (10).

$$\begin{aligned} R_{yy}(\omega, k) &= W(\omega) R_{xx}(\omega, k) W^H(\omega) \\ &= W(\omega) H(\omega) R_s(\omega, k) W^H(\omega) \\ &= R_c(\omega, K) \end{aligned} \tag{10}$$

where, $R_s(\omega, k)$ represents covariance matrix of the source signals, and varies with k, $R_c(\omega, k)$ is an arbitrary diagonal matrix and $R_{xx}(\omega, k)$ is covariance matrix of $X(\omega)$, and can obtain via Equation (11).

$$R_{xx}(\omega, k) = \frac{1}{M} \sum_{m=o}^{M-1} X(\omega, Mk+m) X^H(\omega, Mk+m) \tag{11}$$

For each frequency, the separation filter $W(\omega)$ can be estimated via Equation (12).

$$\hat{W}(\omega) = \arg\min \sum_{k} \left\| V(\omega, k) \right\|^2 \tag{12}$$

where, $\| \|^2$ is square of the norm Frobenius and Equation (13) holds.

$$\begin{aligned} V(\omega, k) = W(\omega) R_{xx}(\omega, k) W^H(\omega) - \\ diag \left| W(\omega) R_{xx}(\omega, k) W^H(w) \right| \end{aligned} \tag{13}$$

The constraint conditions are as shown in Equation (14).

$$\sum_{k} diag \left\| W(\omega) R_{xx}(\omega, k) W^H(k) \right\|^2 \neq 0 \tag{14}$$

The iterative equation (15) of the second-order statistical

algorithm for blind deconvolution in frequency domain can be obtained with the steepest descent method.

$$W^{(l+1)}(\omega) = W^{l}(\omega) - \mu(\omega)\frac{\partial}{W^{(l)H}(\omega)}\left\{\sum_{k}\left\|V^{(l)}(\omega,k)\right\|^{2}\right\} \tag{15}$$

where, $\mu(\omega)$ is the step size at each frequency point.

3.3 Improved natural gradient blind deconvolution algorithm based on KL divergence

(1) Natural gradient algorithm

Set coefficient space: $S=\{w\in R^{n}\}$, where, $w=(w_1,w_2,\ldots,w_n)^{T}$. An objective function $J(w)$ is defined in S, the standard gradient of $J(w)$ is defined as shown in Equation (16) [8].

$$\nabla J(w) = \left[\frac{\partial J}{\partial w_1},\frac{\partial J}{\partial w_2},\cdots,\frac{\partial J}{\partial w_n}\right]^{T} \tag{16}$$

The objective function $J(w)$ is obtained by the standard gradient method, as shown in Equation (17).

$$w(t+1) = w(t) - \mu\nabla J(w(t)) \tag{17}$$

The assumption conditions of the standard descent method: the parameter space S is Euclidean space with an orthogonal coordinate system, and the length of the coefficient vector increment dw can be obtained via Equation (18).

$$\|dw\| = \sqrt{\sum_{i=1}^{n}(dw_i)^{2}} \tag{18}$$

If the coordinate system is not orthogonal, the length of dw can be obtained via Equation (19).

$$\|dw\| = \sqrt{\sum_{i=1}^{n} g_{ij}(w)\,dw_i dw_j} \tag{19}$$

(2) Selection of nonlinear functions

Since the performance of the algorithm depends largely on the similarity between the selected nonlinear function and the probability density function of the source signals, and the probability density function of the actual signals is generally unknown, the selection of the nonlinear function is closely related to the performance of the algorithm.

Assuming that the probability density function of the source signal S can be approximately estimated as $q_i(y_i)$, the definition of the nonlinear function $f_i(y_i)$ is as shown in Equation (20).

$$f_i(y_i) = \frac{d\log q_i(y_i)}{dy_i} = -\frac{dq_i(y_i)/dy_i}{q_i(y_i)} = -\frac{q_i'(y_i)}{q_i(y_i)} \tag{20}$$

The above equation contains the probability density function $q(y)$ of the source signals, which is an unknown quantity, and in the reference [9], the selected nonlinear function is generally fixed, such as for sub-Gaussian signals, $f_i(y_i)=y_i^{3}$ is generally selected; for super-Gaussian signals, $f_i(y_i)=\tanh(y_i)$ is generally selected.

When the distribution of the source signals is inconsistent and includes both sub-Gaussian and super-Gaussian signals, the adjustable nonlinear function must be selected to adapt to the probability density function of different source signals by adjusting its parameters.

(3) Natural gradient blind deconvolution algorithm based on KL divergence

The model of convolution mixing can be expressed as Equation (21).

$$x(k) = \sum_{p=-\infty}^{\infty} H_p s(k-p) \tag{21}$$

where, for source originals: $s(k)=[s_1(k), s_2(k), \ldots, s_m(k)]^{T}$, when the filter matrix at p of the delay is H_p, the mixed signals: $x(k)=[x_1(k), x_2(k), \ldots, x_n(k)]^{T}$.

Assuming that $y(k)$ is the estimation of the source signal $s(k)$ after the mixed signal $x(k)$ is separated by the filter W, as shown in Equation (22).

$$y(k) = \sum_{p=-\infty}^{\infty} W_p(k)\,x(k-p) \tag{22}$$

For frequency domain blind source separation of convolution mixed signals, the expression of the objective function based on KL divergence is as shown in Equation (23).

$$J(W_p(k)) = -\frac{1}{j2\pi}\oint \log\left|\det(W(z,k))\right| z^{-1} - \sum_{i=1}^{M}\log p_i(y_i(k)) \tag{23}$$

where, $W(z,k)=\sum_{p=-\infty}^{\infty} W_p(k)z^{-p}$, $y_i(k)$ is the i element of $y(k)$, and $p_i(y_i)$ is the probability density function of the i element. By minimizing the objective function using the natural gradient method with the steepest descent, a blind natural gradient deconvolution algorithm is obtained as shown in Equation (24).

$$W_p(l+1) = W_p(l) + \mu(l) \cdot\left(W_p(l) - f\left(y^{(l)}(k-l)\right)u^{(l)H}(k-p)\right) \tag{24}$$

where, $y(k)=\sum_{p=0}^{L} W_p(k)x(k-p)$, $u(k)=\sum_{q=0}^{L} W_{L-q}^{H}(k)y(k-q)$, L is the length of the separation filter and $(\cdot)^{H}$ is conjugate transpose.

In reference [10], the good convergence property of the Equation (24) algorithm has been proven. Another property of the algorithm is the equivariant property, that's, its separation effect is not related to the characteristics of the transmission channel. The objective funtion of Equation (24) is $f_i(y_i) = -\frac{d\log[p_i(y_i)]}{dy_i}$, where $p_i(y_i)$ is the probability density function of the i signal yi in the estimated signal y. however, $p_i(y_i)$ is unknown in advance, and a nonlinear function is needed to replace the unknown objective function. However, this method also has limitations because the selection of source signal

objective functions with different statistical characteristics will also be different. Therefore, it is very difficult to select the objective function when the statistical characteristics of the source signals are unknown. In this paper, Equation (23) is used as the objective function. For the super-Gaussian speech signal, the improved algorithm can perform blind deconvolution effectively.

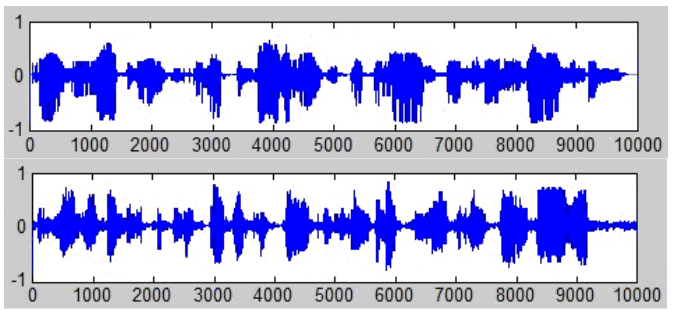
4. SIMULATION EXPERIMENT AND ANALYSIS OF RESULTS

In order to verify the validity of the algorithm, two sets of MATLAB simulation experiments are carried out. The first set is the simple time delay mixed speech signals, and the second set is the convolution mixed speech signals.

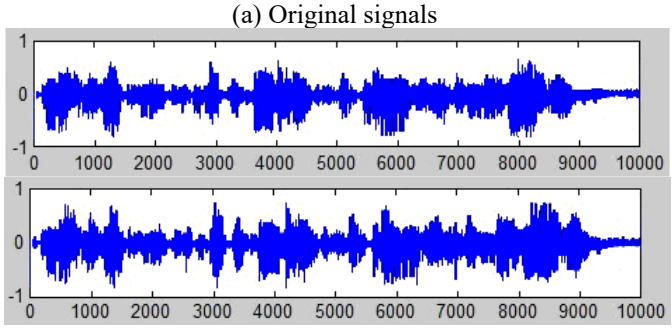
The original speech signal used in Experiment 1 is provided by the ICALAB database [11], the selected delay coefficient is τ_{12}=1 and τ_{21}=2, and the two observed signals are respectively:

$$x_1(t) = s_1(t) + 0.9s_2(t - \tau_{12})$$
$$x_2(t) = s_2(t) + 0.65s_1(t - \tau_{21}) \quad (25)$$

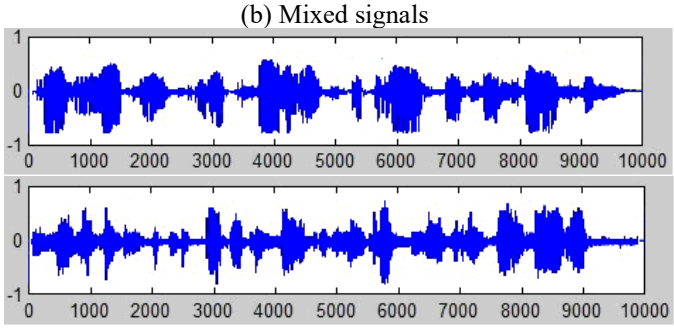
Using the frequency domain method in the experiment, the waveforms of the original signals, the mixed signals and the separated signals are shown in Figure 2.



(a) Original signals



(b) Mixed signals

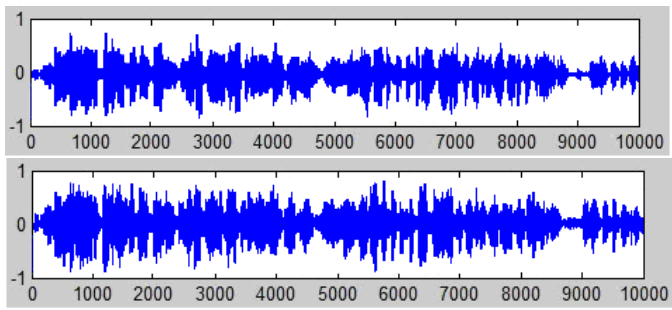


(c) Separated signals

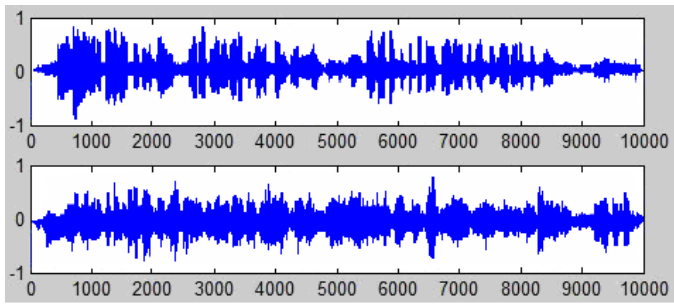
Figure 2. Waveforms of the original signals, the mixed signals and the separated signals

Comparing the diagrams (a), (b) and (c) in Figure 2, it can be seen that the waveform of separated signals is very close to that of original signals, and the part of the mixed signal in which the two voices are mixed is difficult to be observed in the separated signal diagram, achieving ideal separation effect.

The speech test data used in experiment 2 is provided by Lucas Parra [11], which is a mixed signals of speaker and TV sound. The frequency domain method is adopt in the experiment, and the observed signals of the microphone and the separated signals obtained by the algorithm are shown in Figure 3.



(a) Observed signals



(b) Separated signals

Figure 3. Waveforms of observed signals and separated signals

Since the original signals of the two sounds cannot be obtained directly, the difference between the source signals and the separated signals cannot be seen intuitively from the above figure, but from the definition of the output signals, the separated signals are obviously better than the observed signals, indicating that the signal quality is improved to a certain extent to achieve the ideal separation effect.

5. CONCLUSIONS

In this paper, the convolution-mixed speech signal enhancement processing method is studied. The mixed signal is closer to the actual environment, and the time-domain blind deconvolution algorithm is complicated with longer time delay, so the frequency-domain algorithm is adopted to reduce the algorithm complexity and improve the blind source separation efficiency. On the basis of previous researches, this paper introduces a blind source separation algorithm of natural gradient convolution mixed signals based on KL divergence. Through frequency domain blind deconvolution simulation of two sets of mixed speech signals with different characteristics, the effectiveness of the improved algorithm is verified, and the intelligibility and articulation of the separated signals are improved.

ACKNOWLEDGEMENTS

Supported by the Youth Fundation of Lanzhou Jiaotong University Project NO.2014003; Supported by the National Nature Fundation Project of China NO.61461024; Supported by the Graduate educational reform project of Lanzhou Jiaotong University (Training and practice in the ability of innovation and experiment based on the course of the modern electronic technology for graduate students).

REFERENCES

[1] Mitianoudis N., Stathaki T. (2007). Batch and online underdetermined source separation using alpaca mixture models, *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 15, No. 6, pp. 1818-1832.

[2] Pedersen M.S., Wang D.L., Larsen J., Kjems U. (2008). Two- microphones separation of speech mixtures, *IEEE Transactions on Neural Networks*, Vol. 19, No. 3, pp. 475-492. DOI: 10.1109/TNN.2007.911740

[3] Hiroshi S., Shoko A., Shoji M. (2011). Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 19, No. 3, pp. 516-527. DOI: 10.1109/TASL.2010.2051355

[4] Rivet B., Girin L., Jutten C. (2007). Mixing audiovisual speech processing and blind source separation for the extraction of speech signals from convolutive mixtures, *IEEE Trans on Audio, Speech and Language Processing*, Vol. 15, No. 1, pp. 96-108. DOI: 10.1109/TASL.2006.872619

[5] Kirei B.S., Topa M., Muresan I., Homana I., Toma N. (2011). Blind source separation for convolutive mixtures with neural networks, *Advances in Electrical and Computer Engineering*, Vol. 11, pp. 63-68. DOI: 10.4316/AECE.2011.01010

[6] Prasad R., Saruwatari H., Shikano K. (2009). Enhancement of speech signals separated from their convolutive mixture by FDICA algorithm, *Digital Signal Processing*, Vol. 19, pp. 127-133. DOI: 10.1016/j.dsp.2008.01.007

[7] Wang L., Ding H., Yin F. (2010). An improved method for permutation correction in convolutive blind source separation, *Archives of Acoustics*, Vol. 35, No. 4, pp. 493-504. DOI: 10.2478/v10168-010-0038-9

[8] Guo W., Yu F.Q. (2015). Improved speech music signal separation based on negative entropy maximization, *Computer Engineering and Application*, Vol. 51, No. 4, pp. 209-212. DOI: 10.3778/j.issn.1002-8331.1306-0039

[9] Zhang Y.Y., Xin J.H., Liu G.B. (2016). Applications of combined with cumulant slice joint diagonalization of blind source separation, *Journal of Huazhong University of Science and Technology (Natural Science)*, Vol. 44, No. 7, pp. 86-90. DOI: 10.13245/j.hust.160717

[10] Zhou J. (2016). Research of underdetermined source estimation and blind extraction method for mechanical fault signals, *Doctoral Dissertation of Kunming University*, pp. 38-45.

[11] Yang J.M., Qi H.Y. (2015). Improved nonlinear blind source separation algorithm based on the minimization of mutual information, *Electric Measurement and Instrument*, Vol. 52, No. 9, pp. 66-69. DOI: 10.3969/j.issn.1001-1390.2015.09.013