

Research Article

A Fetal Electrocardiogram Signal Extraction Algorithm Based on the Temporal Structure and the Non-Gaussianity

Yibing Li,¹ Wei Nie,¹ Fang Ye,¹ and Ao Li²

¹College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

²School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China

Correspondence should be addressed to Fang Ye; yefang0815@sina.cn

Received 9 November 2015; Revised 24 December 2015; Accepted 28 December 2015

Academic Editor: Luca Faes

Copyright © 2016 Yibing Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fetal electrocardiogram (FECG) extraction is an important issue in biomedical signal processing. In this paper, we develop an objective function for extraction of FECG. The objective function is based on the non-Gaussianity and the temporal structure of source signals. Maximizing the objective function, we can extract the desired FECG. Combining with the solution vector obtained by maximizing the objective function, we further improve the accuracy of the extracted FECG. In addition, the feasibility of the innovative methods is analyzed by mathematical derivation theoretically and the efficiency of the proposed approaches is illustrated with the computer simulations experimentally.

1. Introduction

In biomedical signal processing, fetal electrocardiogram (FECG) extraction is full of challenges. FECG provides important information about the health of the fetus. However, FECG is always buried in various interferences and noises. Among these interferences and noises, maternal electrocardiogram (MECG) is the strongest one with high amplitude. Besides, breathing artifact, electrode artifact, and other noises also affect the desired FECG. Therefore, it is a difficult task to extract accurate FECG.

Traditional methods cannot get satisfactory results, such as multireference adaptive noise cancellation [1] and singular value decomposition [2]. Recently, the blind source separation (BSS) model [3–6] is introduced to solve the extraction problem of the desired FECG and shows satisfactory results. BSS algorithms can separate all sources from mixtures without any priori knowledge. In BSS algorithms, the ICA model [7–9] utilizes the non-Gaussianity [10] of signals to separate all source signals. This model is suitable for biomedical signal processing [11] and non-Gaussianity becomes an important tool to process these kinds of signals. In [12], traditional BSS algorithms and joint BSS algorithms were used to separate the maternal signal and the fetal signal. Authors [13] proposed

two methods based on hybrid BSS to get FECG signals. These methods get better results than traditional BSS algorithms. However, the extraction of the desired FECG with the BSS algorithms is not the best choice. On one hand, separating all sources is a waste of time and not necessary if only the desired FECG is needed. On the other hand, the prior knowledge about the desired FECG can be utilized by us. According to this situation, blind source extraction (BSE) rises in response to the proper time and conditions and becomes a better choice than BSS. BSE only separates one source signal every time, so it has higher efficiency when it extracts the desired FECG. FECG is a special signal with the temporal structure. As prior knowledge, this structure can help us recover source signals. Barros and Cichocki [14] proposed a BSE algorithm called BCBSE. This algorithm can extract the desired FECG by utilizing the temporal structure. However, the algorithm is very sensitive to the estimation error of the optimal time delay. To overcome the drawback of BCBSE, Shi and Zhang [15] developed another BSE algorithm called SemiBSE based on the non-Gaussianity and the mean squared error function that is described in [14]. This method improves the performance of BCBSE. It is more robust for the estimation error of the optimal time delay. However, the performance of this algorithm is up to the choice of

the parameters. If we initialize the parameters randomly, the performance of the algorithm becomes weaker. A BSE method called MACBSE [16] was proposed based on several time-delay autocorrelations of primary sources. MACBSE adopts a fixed-point learning algorithm for extraction of the desired source signal without choosing the learning step size. Authors [17] proposed a BSE algorithm called GABSE based on the generalized linear or nonlinear autocorrelations of the sources. This method has fast convergence speed and good stability if any two sources are uncorrelated with each other and have different temporal structures. A fast and robust fixed-point algorithm based on the nonlinear autocorrelation was proposed in [18]. Its convergence speed is shown to be at least quadratic. Li and Liao [19] proposed an algorithm based on the eigenvalue decomposition of the cross-correlation of whitened source signals at a given time tag. This algorithm functions without iterations. Wang et al. [20] proposed a robust separation algorithm to recover original sources through a joint diagonalizer of several average delayed covariance matrices at positions of the optimal time delay and its integers.

In order to settle the above problem better, we put forward the method based on the non-Gaussianity and the temporal structure of source signals. This paper consists of two main parts. First, we design an objective function to get the weight vector for extraction of the desired FECG. Then, the weight vector obtained by maximizing the objective function is combined with the FastICA algorithm to further improve the performance of the algorithm.

This paper is organized as follows. Section 2 introduces the basic theory of our algorithms. As the principle part, we highlight the algorithms and the analysis about the algorithms in Section 3. Simulation results are presented in Section 4, and conclusions are made in Section 5.

2. The Basic Model

The linear instantaneous mixed model of BSS problems can be denoted as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$ is the n -dimension mixture vector, $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m]$ is an unknown mixing matrix, $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_m(t)]^T$ is the m -dimension source vector, and $\mathbf{n}(t)$ is the noise vector. If the noise is ignored, the noiseless model of BSS problems can be expressed as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \quad (2)$$

Preprocessing for the mixtures is usually necessary before the BSS algorithms. First, the mixtures are made to own zero mean through the following formula:

$$\hat{\mathbf{x}}(t) = \mathbf{x}(t) - E[\mathbf{x}(t)], \quad (3)$$

where $\hat{\mathbf{x}}(t)$ is the new mixture vector. Then, the prewhitening process makes any two variables of the new mixtures orthogonal and gets the whitened mixtures

$$\tilde{\mathbf{x}}(t) = \mathbf{V}\hat{\mathbf{x}}(t), \quad (4)$$

where \mathbf{V} is the prewhitening matrix. For a BSE algorithm, only a source signal can be separated at a time. If the unmixing matrix is $\mathbf{w} = (w_1, w_2, \dots, w_n)^T$, the estimated source signal can be denoted as

$$y(t) = \mathbf{w}^T \tilde{\mathbf{x}}(t). \quad (5)$$

Meanwhile, the delayed estimated source signal can be expressed as

$$y(t - \tau) = \mathbf{w}^T \tilde{\mathbf{x}}(t - \tau), \quad (6)$$

where τ is the time delay that is some lag constant. The estimation error of the optimal time delay may have bad effect on the performance of the algorithms, so the reasonable estimation is important. The specific estimation method of the optimal time delay refers to [14].

3. The Proposed Algorithms

3.1. The Proposed Algorithm 1

3.1.1. Objective Function. In order to avoid the blindness of choosing parameters, we combine the non-Gaussianity and the temporal structure to design the following constrained maximization problem:

$$\begin{aligned} \max_{\|\mathbf{w}\|=1} \psi(\mathbf{w}) &= E\{G[y(t)]G[y(t)y(t-\tau)]\} \\ &= E\{G[\mathbf{w}^T \tilde{\mathbf{x}}(t)]G[\mathbf{w}^T \tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t-\tau)^T \mathbf{w}]\}, \end{aligned} \quad (7)$$

where G is a differentiable nonlinear function. This function utilizes the non-Gaussianity, the temporal structure, and the nonlinear correlation.

3.1.2. Learning Algorithm. With respect to \mathbf{w} , the gradient of $E\{G[y(t)]G[y(t)y(t-\tau)]\}$ can be obtained as

$$\begin{aligned} \frac{\partial E\{G[y(t)]G[y(t)y(t-\tau)]\}}{\partial \mathbf{w}} &= E\{\tilde{\mathbf{x}}(t)g[y(t)]G[y(t)y(t-\tau)]\} \\ &\quad + E\{\tilde{\mathbf{x}}(t)y(t-\tau)g[y(t)y(t-\tau)]G[y(t)]\} \\ &\quad + E\{\tilde{\mathbf{x}}(t-\tau)y(t)g[y(t)y(t-\tau)]G[y(t)]\}, \end{aligned} \quad (8)$$

where g is the derivative of G . According to the gradient ascent learning rule, a gradient method can be derived as follows:

$$\begin{aligned} \mathbf{w} \leftarrow \mathbf{w} + \mu \frac{\partial E\{G[y(t)y(t-\tau)]\}}{\partial \mathbf{w}} &= \mathbf{w} \\ &\quad + \mu(E\{\tilde{\mathbf{x}}(t)g[y(t)]G[y(t)y(t-\tau)]\} \\ &\quad + E\{\tilde{\mathbf{x}}(t)y(t-\tau)g[y(t)y(t-\tau)]G[y(t)]\} \\ &\quad + E\{\tilde{\mathbf{x}}(t-\tau)y(t)g[y(t)y(t-\tau)]G[y(t)]\}), \end{aligned} \quad (9)$$

$$\mathbf{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|},$$

where μ is a learning rate. The desired vector can be obtained through this method, but the corresponding cost is that the convergence speed of the algorithm is slow and the performance of the algorithm is dependent on the proper choice of the learning rate. If the learning rate is chosen improperly, the convergence property will be destroyed. Therefore, how to find a new approach to radically improve the convergence speed and the reliability is an important issue that needs to be solved. The fixed-point algorithm is a choice to solve this issue. For getting a more efficient fixed-point algorithm, we note that the gradient must point in the direction of \mathbf{w} at a stable point of the gradient algorithm. It means the gradient must be equal to \mathbf{w} multiplied by some scalar constant. In such a case, adding the gradient to \mathbf{w} does not change its direction and the convergence can be obtained. Through normalization to the unit norm, the value of \mathbf{w} is not changed except by changing its sign. Therefore, the following formula can be obtained:

$$\begin{aligned} \mathbf{w} \propto & E \{ \bar{\mathbf{x}}(t) g[y(t)] G[y(t) y(t-\tau)] \} \\ & + E \{ \bar{\mathbf{x}}(t) y(t-\tau) g[y(t) y(t-\tau)] G[y(t)] \} \\ & + E \{ \bar{\mathbf{x}}(t-\tau) y(t) g[y(t) y(t-\tau)] G[y(t)] \}. \end{aligned} \quad (10)$$

Through the above formula, the fixed-point algorithm can be updated as

$$\begin{aligned} \mathbf{w} \leftarrow & E \{ \bar{\mathbf{x}}(t) g[y(t)] G[y(t) y(t-\tau)] \} \\ & + E \{ \bar{\mathbf{x}}(t) y(t-\tau) g[y(t) y(t-\tau)] G[y(t)] \} \\ & + E \{ \bar{\mathbf{x}}(t-\tau) y(t) g[y(t) y(t-\tau)] G[y(t)] \}, \end{aligned} \quad (11)$$

$$\mathbf{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|}.$$

Utilizing the above fixed-point algorithm, we can extract the desired signal.

3.1.3. Stability Analysis. In this part, we analyze the stability of the proposed algorithm.

Theorem 1. Assume that the input data meet the model as (2). The data are prewhitened through equation $\bar{\mathbf{x}} = \mathbf{V}\mathbf{A}\mathbf{s}$ and G is a quite smooth even function. Furthermore, we assume that $\delta s_i = s_i(t)s_i(t-\tau)$ ($i = 1, 2, \dots, n$) are mutually independent. To simplify it, $s_i(t)s_i(t-\tau)$ is replaced by $s_i s_{i\tau}$. Under the constraint $\|\mathbf{w}\| = 1$, the local maxima of $E\{G[y(t)]G[y(t-\tau)]\}$ include one row of the inverse of the matrix $\mathbf{V}\mathbf{A}$ if the corresponding desired signal s_i satisfies

$$\begin{aligned} & E \left[g'(s_i) G(s_i s_{i\tau}) + s_{i\tau} g(s_i) g(s_i s_{i\tau}) \right. \\ & + 2s_i^2 g'(s_i s_{i\tau}) G(s_i) + 2s_i s_{i\tau} s_j s_{j\tau} g'(s_i s_{i\tau}) G(s_i) \\ & + 3s_i s_j s_{j\tau} g(s_i) g(s_i s_{i\tau}) - s_i g(s_i) G(s_i s_{i\tau}) \\ & \left. - 2s_i s_{i\tau} g(s_i s_{i\tau}) G(s_i) \right] < 0 \quad (\forall i \neq j), \end{aligned} \quad (12)$$

where g' is the derivative of g .

Proof. According to the above conditions, we make the orthogonal transform of coordinates $\mathbf{p} = \mathbf{A}^T \mathbf{V}^T \mathbf{w}$ and obtain

$$H(\mathbf{p}) = E \{ G(\mathbf{p}^T \mathbf{s} \mathbf{s}^T \mathbf{p}) \}. \quad (13)$$

We analyze the stability at the point $\mathbf{p} = \mathbf{e}_1 = (1, 0, 0, 0, \dots)^T$. The independency assumptions are utilized to evaluate the gradient and the Hessian matrix of $H(\mathbf{p})$ at the point $\mathbf{p} = \mathbf{e}_1$. Then, through making a small perturbation $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$, where ε_1 and ε_2 are the elements of ε , we get

$$\begin{aligned} H(\mathbf{e}_1 + \varepsilon) = & H(\mathbf{e}_1) + \varepsilon^T \frac{\partial H(\mathbf{e}_1)}{\partial \mathbf{p}} + \frac{1}{2} \varepsilon^T \frac{\partial^2 H(\mathbf{e}_1)}{\partial \mathbf{p}^2} \varepsilon \\ & + o(\|\varepsilon\|^2), \end{aligned} \quad (14)$$

where $\partial H(\mathbf{e}_1)/\partial \mathbf{p}$ and $\partial^2 H(\mathbf{e}_1)/\partial \mathbf{p}^2$ are expressed as

$$\begin{aligned} \frac{\partial H(\mathbf{e}_1)}{\partial \mathbf{p}} = & \mathbf{e}_1 E \{ s_1 g(s_1) G(s_1 s_{1\tau}) + 2s_1 s_{1\tau} g(s_1 s_{1\tau}) \\ & \cdot G(s_1) \}, \\ \frac{\partial^2 H(\mathbf{e}_1)}{\partial \mathbf{p}^2} = & \text{diag} \left(E \{ s_1^2 g'(s_1) G(s_1 s_{1\tau}) \right. \\ & + 4s_1^2 s_{1\tau} g(s_1) g(s_1 s_{1\tau}) \\ & + 4s_1^2 s_{1\tau}^2 g'(s_1 s_{1\tau}) G(s_1) \}, E \{ g'(s_1) G(s_1 s_{1\tau}) \\ & + s_{1\tau} g(s_1) g(s_1 s_{1\tau}) + 2s_1^2 g'(s_1 s_{1\tau}) G(s_1) \\ & + 2s_1 s_{1\tau} s_2 s_{2\tau} g'(s_1 s_{1\tau}) G(s_1) \\ & + 3s_1 s_2 s_{2\tau} g(s_1) g(s_1 s_{1\tau}) \}, \dots, E \{ g'(s_1) G(s_1 s_{1\tau}) \\ & + s_{1\tau} g(s_1) g(s_1 s_{1\tau}) + 2s_1^2 g'(s_1 s_{1\tau}) G(s_1) \\ & + 2s_1 s_{1\tau} s_j s_{j\tau} g'(s_1 s_{1\tau}) G(s_1) \\ & \left. + 3s_1 s_j s_{j\tau} g(s_1) g(s_1 s_{1\tau}) \} \right). \end{aligned} \quad (15)$$

Because of the constraint $\|\mathbf{w}\| = 1$, we can know $\varepsilon_1 = \sqrt{1 - \varepsilon_2^2 - \dots} - 1$. Meanwhile, $\sqrt{1 - \gamma} = 1 - \gamma/2 + o(\gamma)$ is known. The order of ε_1^2 is $o(\|\varepsilon\|^2)$, so the higher order terms can be neglected. Through the first-order approximation of ε_1 that is described above, we know $\varepsilon_1 = -\sum_{i>1} \varepsilon_i^2/2 + o(\|\varepsilon\|^2)$ and obtain

$$\begin{aligned} H(\mathbf{e}_1 + \varepsilon) = & H(\mathbf{e}_1) + \frac{1}{2} E \{ g'(s_1) G(s_1 s_{1\tau}) \\ & + s_{1\tau} g(s_1) g(s_1 s_{1\tau}) + 2s_1^2 g'(s_1 s_{1\tau}) G(s_1) \\ & + 2s_1 s_{1\tau} s_j s_{j\tau} g'(s_1 s_{1\tau}) G(s_1) \\ & + 3s_1 s_j s_{j\tau} g(s_1) g(s_1 s_{1\tau}) - s_1 g(s_1) G(s_1 s_{1\tau}) \\ & - 2s_1 s_{1\tau} g(s_1 s_{1\tau}) G(s_1) \} \sum_{j>1} \varepsilon_j^2 + o(\|\varepsilon\|^2). \end{aligned} \quad (16)$$

$\mathbf{p} = \mathbf{e}_i$ is an extremum that is implied in the condition of Theorem 1 if the following condition is satisfied:

$$\begin{aligned} & E \left[g'(s_1) G(s_1 s_{1\tau}) + s_{1\tau} g(s_1) g(s_1 s_{1\tau}) \right. \\ & \quad + 2s_1^2 g'(s_1 s_{1\tau}) G(s_1) \\ & \quad + 2s_1 s_{1\tau} s_j s_{j\tau} g'(s_1 s_{1\tau}) G(s_1) \\ & \quad + 3s_1 s_j s_{j\tau} g(s_1) g(s_1 s_{1\tau}) - s_1 g(s_1) G(s_1 s_{1\tau}) \\ & \quad \left. - 2s_1 s_{1\tau} g(s_1 s_{1\tau}) G(s_1) \right] < 0. \end{aligned} \quad (17)$$

Based on the above descriptions, this condition can be expanded as

$$\begin{aligned} & E \left[g'(s_i) G(s_i s_{i\tau}) + s_{i\tau} g(s_i) g(s_i s_{i\tau}) \right. \\ & \quad + 2s_i^2 g'(s_i s_{i\tau}) G(s_i) + 2s_i s_{i\tau} s_j s_{j\tau} g'(s_i s_{i\tau}) G(s_i) \\ & \quad + 3s_i s_j s_{j\tau} g(s_i) g(s_i s_{i\tau}) - s_i g(s_i) G(s_i s_{i\tau}) \\ & \quad \left. - 2s_i s_{i\tau} g(s_i s_{i\tau}) G(s_i) \right] < 0 \quad (\forall i \neq j). \end{aligned} \quad (18)$$

□

3.1.4. Convergence Analysis. Convergence is also important like stability. In this part, we analyze the convergence of the algorithm.

Theorem 2. *If the following two conditions are satisfied, the algorithm described in formula (11) can reach convergence:*

- (1) $\{s_i, s_{i\tau}\}$ and $\{s_j, s_{j\tau}\}$ are mutually independent.
- (2) $E\{s_i g(s_i) G(s_i s_{i\tau}) + 2s_i s_{i\tau} g(s_i s_{i\tau}) G(s_i)\} \neq 0$.

Proof. Based on the orthogonal coordinate transformation $\mathbf{p} = \mathbf{A}^T \mathbf{V}^T \mathbf{w}$, formula (11) is changed to

$$\begin{aligned} \hat{\mathbf{p}}(k) &= E \left\{ \mathbf{s} g \left[\mathbf{p}^T(k) \mathbf{s} \right] G \left[\mathbf{p}^T(k) \mathbf{s} \mathbf{s}_\tau^T \mathbf{p}(k) \right] \right\} \\ &+ E \left\{ \mathbf{s} \left[\mathbf{p}^T(k) \mathbf{s}_\tau \right] g \left[\mathbf{p}^T(k) \mathbf{s} \mathbf{s}_\tau^T \mathbf{p}(k) \right] G \left[\mathbf{p}^T(k) \mathbf{s} \right] \right\} \\ &+ E \left\{ \mathbf{s}_\tau \left[\mathbf{p}^T(k) \mathbf{s} \right] g \left[\mathbf{p}^T(k) \mathbf{s} \mathbf{s}_\tau^T \mathbf{p}(k) \right] G \left[\mathbf{p}^T(k) \mathbf{s} \right] \right\}, \end{aligned} \quad (19)$$

$$\mathbf{p}(k+1) = \frac{\hat{\mathbf{p}}(k)}{\|\hat{\mathbf{p}}(k)\|},$$

where k is the number of iterations. Using a Taylor approximation for G and g , we have

$$\begin{aligned} G(\mathbf{p}^T(k) \mathbf{s}) &= G(p_i s_i) + g(p_i s_i) \mathbf{p}_{-i}^T \mathbf{s}_{-i} \\ &\quad + o(\|\mathbf{p}_{-i}\|^2), \\ g(\mathbf{p}^T(k) \mathbf{s}) &= g(p_i s_i) + g'(p_i s_i) \mathbf{p}_{-i}^T \mathbf{s}_{-i} \\ &\quad + o(\|\mathbf{p}_{-i}\|^2), \end{aligned}$$

$$\begin{aligned} G[\mathbf{p}^T(k) \mathbf{s} \mathbf{s}_\tau^T \mathbf{p}(k)] &= G(p_i s_i p_i s_{i\tau}) \\ &\quad + g(p_i s_i p_i s_{i\tau}) \mathbf{p}_{-i}^T \mathbf{s}_{-i} \mathbf{s}_{-i\tau}^T \mathbf{p}_{-i} \\ &\quad + o(\|\mathbf{p}_{-i}\|^2), \end{aligned}$$

$$\begin{aligned} g[\mathbf{p}^T(k) \mathbf{s} \mathbf{s}_\tau^T \mathbf{p}(k)] &= g(p_i s_i p_i s_{i\tau}) \\ &\quad + g'(p_i s_i p_i s_{i\tau}) \mathbf{p}_{-i}^T \mathbf{s}_{-i} \mathbf{s}_{-i\tau}^T \mathbf{p}_{-i} \\ &\quad + o(\|\mathbf{p}_{-i}\|^2), \end{aligned} \quad (20)$$

where \mathbf{p}_{-i} is the vector \mathbf{p} without its i th component, \mathbf{s}_{-i} is the vector \mathbf{s} without its i th component, and $\mathbf{s}_{-i\tau}$ is the vector \mathbf{s}_τ without its i th component.

Similarly, we analyze the convergence at the point $\mathbf{p} = \mathbf{e}$, where e_i is 1 and e_j is 0 ($\forall j \neq i$). Based on the above conditions, we get

$$\begin{aligned} \hat{p}_i &= E \left\{ s_i g(s_i) G(s_i s_{i\tau}) + 2s_i s_{i\tau} g(s_i s_{i\tau}) G(s_i) \right\} \\ &\quad + o(\|\mathbf{p}_{-i}\|^2), \\ \hat{p}_j &= E \left\{ g'(s_i) G(s_i s_{i\tau}) + s_{i\tau} g(s_i) g(s_i s_{i\tau}) \right. \\ &\quad \left. + s_i s_j s_{j\tau} g(s_i) g(s_i s_{i\tau}) \right\} p_j + o(\|\mathbf{p}_{-i}\|^2). \end{aligned} \quad (21)$$

The above equations clearly show that the algorithm converges to such a vector \mathbf{e} in which e_i is 1 and e_j is 0 if condition (2) is satisfied. It means that $\mathbf{w} = ((\mathbf{V}\mathbf{A})^T)^{-1} \mathbf{q}$ can converge without computing the sign. This shows that \mathbf{w} converges as Theorem 2. □

3.2. The Proposed Algorithm 2. On the basis of the above algorithm, we propose another improved algorithm by utilizing the FastICA algorithm. The FastICA algorithm is a typical ICA algorithm. The source signal can be estimated through maximizing the non-Gaussianity function of signals. One of the representative non-Gaussianity functions is negentropy. However, this algorithm is based on the theory of projection pursuit, so any signal may be first extracted. According to this problem, we propose an improved algorithm by presetting the weight vector. The method of projection pursuit searches for a projection output to maximize the objective function. Based on the theory of ICA, a local extremum in the solution space of the projection direction is obtained if the negentropy is the objective function. The output of the corresponding data for this local extremum is an independent component. If the mixtures consist of m source signals, the parameter space formed by ICA will have $2m$ local extrema that correspond to signed solutions of all source signals. Meanwhile, a local optimal solution owns its attraction region for a greedy optimization algorithm that is based on projection pursuit. In other words, the initial point in this region converges to this local optimal solution with this algorithm. The FastICA algorithm may not extract the desired signal if the initial weight matrix is arbitrary. However, if the initial weight matrix is near the local optimal solution that corresponds to

the desired signal, the desired source signal can be extracted. The solution vector of the proposed algorithm 1 can be the initial weight matrix of the FastICA algorithm and help the algorithm extract the desired source signal. The solution vector of the proposed algorithm 1 is denoted as

$$\mathbf{w}_0 = [w_1, \dots, w_i, \dots, w_j, \dots, w_n]^T. \quad (22)$$

It is the initial weight matrix of the proposed algorithm 2. Then, we can extract the desired signal with the FastICA algorithm based on the negentropy. The approximate negentropy in literatures can be denoted as

$$J(y) \propto \{E[G(y)] - E[G(v)]\}^2. \quad (23)$$

The gradient of $J(y)$ is written as

$$\nabla J(y) \propto \gamma E\{\bar{\mathbf{x}}(t) g[y(t)]\}, \quad (24)$$

where $\gamma = E[G(y)] - E[G(v)]$; the learning algorithm is obtained as follows:

$$\begin{aligned} \Delta \mathbf{w} &= \{E[G(y)] - E[G(v)]\} E\{\bar{\mathbf{x}}(t) g[y(t)]\}, \\ \mathbf{w} &\leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}. \end{aligned} \quad (25)$$

The corresponding fixed-point algorithm is denoted as

$$\begin{aligned} \mathbf{w} &\leftarrow E\{\bar{\mathbf{x}}(t) g[y(t)]\}, \\ \mathbf{w} &\leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}. \end{aligned} \quad (26)$$

However, the convergence property of the FastICA algorithm based on the negentropy is unsatisfactory because the non-polynomial moment does not own good algebra property. Therefore, the Newton iteration algorithm is usually utilized to improve the process. If the maximization of $E\{G[\mathbf{w}^T \bar{\mathbf{x}}(t)]\}$ under the constraint $\|\mathbf{w}\| = 1$ is considered, the optimization problem with the Lagrangian multiplier method can be changed to

$$F(\mathbf{w}) = E\{G[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} + \beta \mathbf{w}^T \mathbf{w}. \quad (27)$$

The first-order derivative and the second-order derivative can be denoted as

$$\begin{aligned} \frac{\partial F(\mathbf{w})}{\partial \mathbf{w}} &= E\{\bar{\mathbf{x}}(t) g[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} + 2\beta \mathbf{w}, \\ \frac{\partial^2 F(\mathbf{w})}{\partial \mathbf{w}^2} &= E\{\bar{\mathbf{x}}(t) \bar{\mathbf{x}}(t)^T g'[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} + 2\beta \mathbf{I} \\ &\approx E\{\bar{\mathbf{x}}(t) \bar{\mathbf{x}}(t)^T\} E\{g'[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} + 2\beta \mathbf{I} \\ &= E\{g'[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} \mathbf{I} + 2\beta \mathbf{I}. \end{aligned} \quad (28)$$

In order to avoid calculating the inverse of the matrix, the simple approximation is made to get the following Newton iteration algorithm:

$$\begin{aligned} \mathbf{w} &\leftarrow \mathbf{w} - \frac{\partial F(\mathbf{w}) / \partial \mathbf{w}}{\partial^2 F(\mathbf{w}) / \partial \mathbf{w}^2} \\ &= \mathbf{w} - E\{\bar{\mathbf{x}}(t) g[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} \\ &\quad + \frac{2\beta \mathbf{w}}{(E\{g'[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} + 2\beta)}. \end{aligned} \quad (29)$$

Both sides of the above equation are multiplied by $E\{g'[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} + 2\beta$ and then the learning algorithm after the algebraic simplification is denoted as

$$\begin{aligned} \mathbf{w} &\leftarrow E\{\bar{\mathbf{x}}(t) g[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} - E\{g'[\mathbf{w}^T \bar{\mathbf{x}}(t)]\} \mathbf{w}, \\ \mathbf{w} &\leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}. \end{aligned} \quad (30)$$

Formula (30) is the basic formula of the fixed-point FastICA algorithm.

Here, we analyze the condition that the initial weight vector must meet in order to get the desired signal. According to formula (30), we get

$$\begin{aligned} \hat{\mathbf{w}}(k) &\leftarrow E\{\bar{\mathbf{x}}(t) g[\mathbf{w}(k)^T \bar{\mathbf{x}}(t)]\} \\ &\quad - E\{g'[\mathbf{w}(k)^T \bar{\mathbf{x}}(t)]\} \mathbf{w}(k), \\ \mathbf{w}(k+1) &= \frac{\hat{\mathbf{w}}(k)}{\|\hat{\mathbf{w}}(k)\|}. \end{aligned} \quad (31)$$

Based on formula (31) and $\mathbf{q} = \mathbf{A}^T \mathbf{V}^T \mathbf{w}$, the following formula is obtained:

$$\begin{aligned} \hat{\mathbf{q}}(k) &\leftarrow E\{\mathbf{s}(t) g[\mathbf{q}(k)^T \mathbf{s}(t)]\} \\ &\quad - E\{g'[\mathbf{q}(k)^T \mathbf{s}(t)]\} \mathbf{q}(k), \\ \mathbf{q}(k+1) &= \frac{\hat{\mathbf{q}}(k)}{\|\hat{\mathbf{q}}(k)\|}. \end{aligned} \quad (32)$$

Assuming $a_k = 1/\|\hat{\mathbf{q}}(k)\|$, we obtain

$$\begin{aligned} \mathbf{q}(k+1) &= a_k E\{\mathbf{s}(t) g[\mathbf{q}(k)^T \mathbf{s}(t)]\} \\ &\quad - a_k E\{g'[\mathbf{q}(k)^T \mathbf{s}(t)]\} \mathbf{q}(k). \end{aligned} \quad (33)$$

To analyze conveniently, we choose $G = y^4/4$ to illustrate. Therefore, $g = y^3$ and $g' = 3y^2$ can be obtained. Formula (33) can be simplified as

$$\begin{aligned}
\mathbf{q}(k+1) &= a_k \left[E \left\{ s_1 \left(\sum_{m=1}^M q_m(k) s_m \right)^3 \right\} - 3q_1(k) \right. \\
&\quad \cdot E \left\{ \left(\sum_{m=1}^M q_m(k) s_m \right)^2 \right\}, \dots, \\
&\quad E \left\{ s_i \left(\sum_{m=1}^M q_m(k) s_m \right)^3 \right\} - 3q_i(k) \\
&\quad \cdot E \left\{ \left(\sum_{m=1}^M q_m(k) s_m \right)^2 \right\}, \dots, \\
&\quad E \left\{ s_M \left(\sum_{m=1}^M q_m(k) s_m \right)^3 \right\} - 3q_M(k) \\
&\quad \left. \cdot E \left\{ \left(\sum_{m=1}^M q_m(k) s_m \right)^2 \right\} \right]^T.
\end{aligned} \tag{34}$$

Next, the following formula can be obtained:

$$\begin{aligned}
&q_i(k+1) \\
&= a_k E \left\{ s_i \left(\sum_{m=1}^M q_m(k) s_m \right)^3 \right\} \\
&\quad - 3a_k q_i(k) E \left\{ \left(\sum_{m=1}^M q_m(k) s_m \right)^2 \right\} \\
&= a_k E \left\{ s_i \left[q_i(k) s_i + \sum_{m=1, m \neq i}^M q_m(k) s_m \right]^3 \right\} \\
&\quad - 3a_k q_i(k) E \left\{ \left[q_i(k) s_i + \sum_{m=1, m \neq i}^M q_m(k) s_m \right]^2 \right\} \\
&= a_k E \left\{ q_i^3(k) s_i^4 + 3q_i(k) s_i^2 \left[\sum_{m=1, m \neq i}^M q_m(k) s_m \right]^2 \right\} \\
&\quad - 3a_k q_i(k) E \left\{ q_i^2(k) s_i^2 + \left[\sum_{m=1, m \neq i}^M q_m(k) s_m \right]^2 \right\} \\
&= a_k q_i^3(k) E \{ s_i^4 \} - 3a_k q_i^3(k) E \{ s_i^2 \} \\
&= a_k q_i^3(k) (E \{ s_i^4 \} - 3E \{ s_i^2 \}) = a_k q_i^3(k) k_4(s_i),
\end{aligned} \tag{35}$$

where $k_4(\cdot)$ denotes the kurtosis of the signal. Similarly, we obtain

$$q_j(k+1) = a_k q_j^3(k) k_4(s_j). \tag{36}$$

The following formula is assumed:

$$\begin{aligned}
b(k+1) &= \frac{q_j(k+1)}{q_i(k+1)} = \frac{q_j^3(k) k_4(s_j)}{q_i^3(k) k_4(s_i)} \\
&= b^3(k) \frac{k_4(s_j)}{k_4(s_i)}.
\end{aligned} \tag{37}$$

Then, $b(k)$ can be denoted as

$$\begin{aligned}
b(k) &= \frac{q_j(k)}{q_i(k)} \\
&= \left(\sqrt{\left| \frac{k_4(s_j)}{k_4(s_i)} \right|} \right)^{-1} \left[\frac{q_j(1)}{q_i(1)} \sqrt{\left| \frac{k_4(s_j)}{k_4(s_i)} \right|} \right]^{3^{k-1}}.
\end{aligned} \tag{38}$$

If we want to get the desired signal, $q_j(k)$ must be close to 0 and $q_i(k)$ must be close to 1. Therefore, the following formula must be met:

$$\left| \frac{q_j(1)}{q_i(1)} \sqrt{\left| \frac{k_4(s_j)}{k_4(s_i)} \right|} \right| < 1. \tag{39}$$

The above formula (39) can be simplified as

$$q_i^2(1) > q_j^2(1) \left| \frac{k_4(s_j)}{k_4(s_i)} \right|. \tag{40}$$

Combining the solution vector \mathbf{w}_0 described in formula (22), we get

$$\mathbf{q}_0 = \mathbf{A}^T \mathbf{V}^T \mathbf{w}_0 = [q_{01}, \dots, q_{0i}, \dots, q_{0j}, \dots, q_{0n}]^T. \tag{41}$$

In conclusion, the desired signal can be obtained if the following condition is satisfied:

$$q_{0i}^2 > q_{0j}^2 \left| \frac{k_4(s_j)}{k_4(s_i)} \right|. \tag{42}$$

Similarly, if another function G is chosen, there will be another condition that is similar to formula (42) helping us get the desired source signal.

4. Simulation Results and Analysis

4.1. Experiments on Artificial Data. In order to verify the performance of the above algorithms, simulation experiments are performed with the data that are shown in Figure 1. Source signals include maternal electrocardiogram (MECG), fetal electrocardiogram (FECG), breathing artifact, electrode artifact, and two kinds of noises. The above source signals are

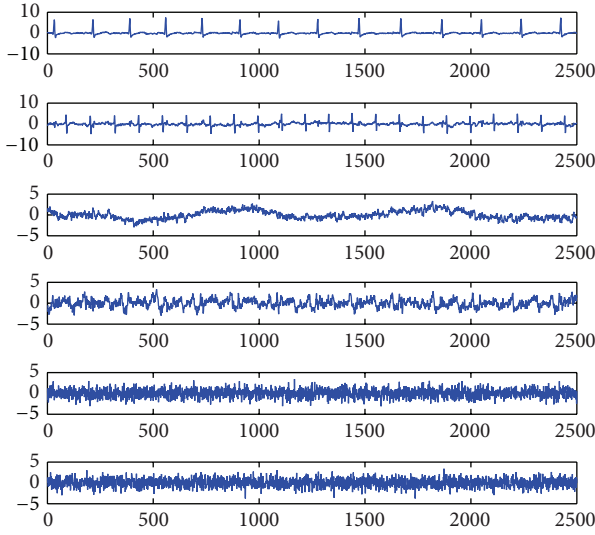


FIGURE 1: Source signals data.

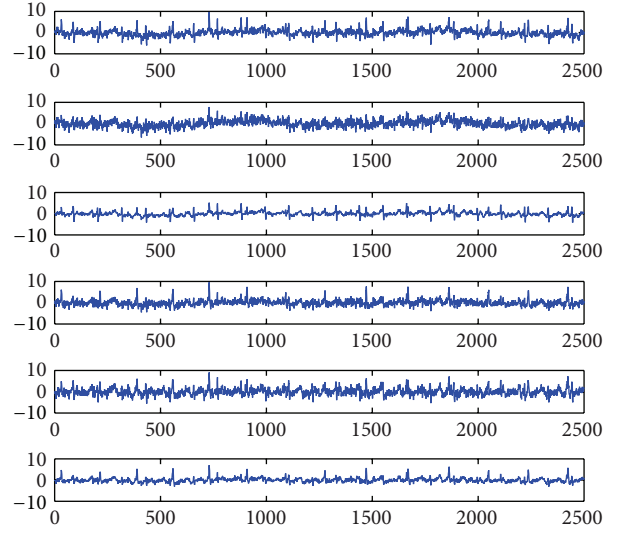


FIGURE 2: Mixed signals data.

mixed with a matrix to get the mixtures that are shown in Figure 2. The mixing matrix in this paper is written as

$$\begin{bmatrix} 0.8246 & 0.9469 & 0.7384 & 0.1109 & 0.5211 & 0.5746 \\ 0.4530 & 0.5101 & 0.9764 & 0.3752 & 0.7743 & 0.8639 \\ 0.3806 & 0.7919 & 0.5233 & 0.3299 & 0.1203 & 0.1986 \\ 0.9259 & 0.4522 & 0.4299 & 0.3421 & 0.6255 & 0.6725 \\ 0.7408 & 0.8492 & 0.2072 & 0.8171 & 0.3466 & 0.9018 \\ 0.7376 & 0.3904 & 0.3234 & 0.5317 & 0.3346 & 0.1992 \end{bmatrix} \cdot \quad (43)$$

Corresponding simulations are needed to compare the performance of the proposed algorithms and other algorithms. The performance of algorithms can be measured with the performance index (PI) that is denoted as

$$PI = \sum_{j=1}^n \frac{|p_j|}{\max_k |p_k|} - 1, \quad k = 1, \dots, n, \quad (44)$$

where p_j denotes the element of the global vector $\mathbf{p} = \mathbf{w}^T \mathbf{V} \mathbf{A}$. If the desired signal is extracted perfectly, the value of PI is 0. In other words, the value of PI is lower if the performance is better. PI is always used to be the measurement index for the artificial data.

In the comparison algorithms, the MACBSE algorithm, the GABSE algorithm, the SemiBSE algorithm, and the FastNA algorithm are the algorithms that need iterations. Therefore, we compare these algorithms with the proposed algorithm through the relationship of the average performance index and the iterations. In this paper, the function $G(u) = \log(\cosh(u))$ is chosen. The parameters of other algorithms are selected based on the references. In the GABSE algorithm and the FastNA algorithm, $G(u) = \log(\cosh(u))$ is also chosen. In the SemiBSE algorithm, the nonlinearity is chosen as $g(u) = \text{sign}(u)$ and the constant coefficient is initialized to 0.3. The learning rates are set to 0.1 and 0.0001. All

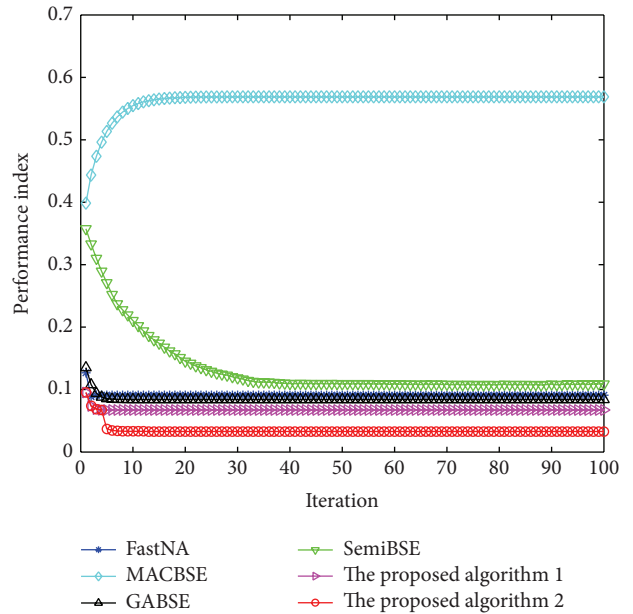


FIGURE 3: The performance indexes against iteration numbers by the six algorithms.

the algorithms described above adopt the same initial weight vector $[-0.3331, 0.3768, 0.2715, 0.6498, 0.3576, 0.3510]^T$ and the optimal time delay 112. The average performance indexes over 100 independent trials against iteration numbers by the six algorithms are shown in Figure 3.

It is obviously verified in Figure 3 that our algorithms have better performance than other algorithms when the algorithms achieve the convergence.

Meanwhile, we compare our algorithms with two noniterative algorithms which include Li's algorithm and the LAJD algorithm. In the LAJD algorithm, L and K are set to 3 and 2,

TABLE 1: The PI values of the algorithms.

Li's algorithm	The LAJD algorithm	Our algorithm 1	Our algorithm 2
0.2405	0.195	0.0673	0.033

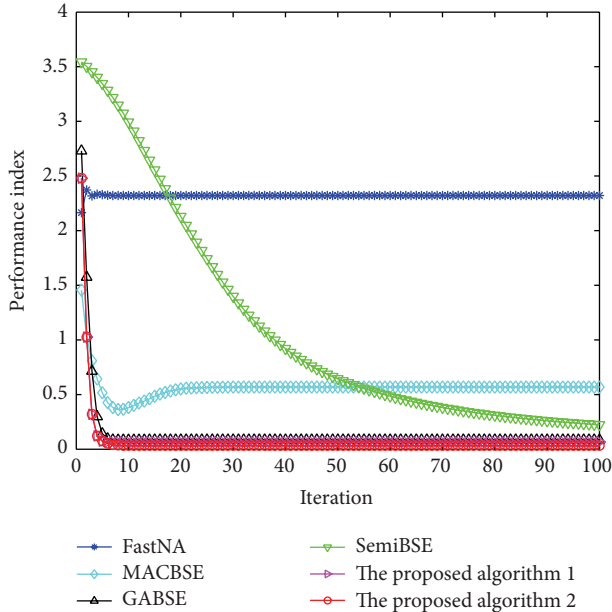


FIGURE 4: The average performance indexes in 100 independent trials against iteration numbers by the six algorithms when the mixing matrix is random.

respectively. The performance indexes of the algorithms after the convergence are shown in Table 1.

From Table 1, it is easy to know that our algorithms have lower performance indexes than Li's algorithm and the LAJD algorithm.

The above comparison is done with the same mixing matrix. However, the mixing matrix is unknown and changing in practice. Therefore, Monte Carlo simulations need to be carried out with different random mixing matrices. We set the initial weight vector as $[0, 0, 0, 0, 0, 1]^T$ and compare different algorithms. The average performance indexes over 100 independent trials against iteration numbers by the six iterative algorithms are shown in Figure 4. In 100 independent trials, the mixing matrices are different and random.

When the mixing matrix is random, the average performance indexes of our algorithms and two noniterative algorithms in 100 independent trials are shown in Table 2.

As shown in Figure 4 and Table 2, our algorithms still own better performance when the mixing matrix is random.

Based on the comparisons of the proposed algorithms and all other algorithms, we can know that our algorithms own good performance for processing the artificial data.

4.2. Experiments on Real-World Data. In order to further demonstrate the practicability of our algorithm, the simulations of real-world EEG data which are shown in Figure 5

TABLE 2: The average PI values of the algorithms when the mixing matrix is random.

Li's algorithm	The LAJD algorithm	Our algorithm 1	Our algorithm 2
0.2405	0.195	0.0672	0.033

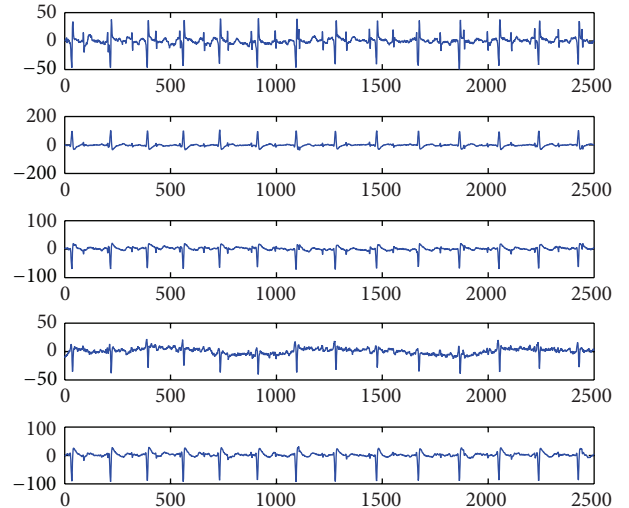


FIGURE 5: The real-world EEG data.

need to be performed. The well-known ECG data are measured from a pregnant woman and shared by De Moor [21]. The ECG data are mixed with maternal electrocardiogram (MECG), fetal electrocardiogram (FECG), breathing artifact, electrode artifact, and some other noises. When extracting FECG from the real-world ECG signals with various kinds of algorithms, we cannot measure the performance of the algorithms with PI because the mixing matrix is unknown. However, the robustness of the algorithms is also an important index. It is always used to be the measurement index for the real-world data. Because there may be estimation error of the optimal time delay in practice, the robustness of the algorithms becomes more important. The FECG signals extracted by all algorithms at the optimal delay 112 are shown in Figure 6. The FECG signals extracted by all algorithms at delays 106 and 120 are shown in Figures 7 and 8. From Figure 6, we can find that all algorithms can extract the FECG signal at the optimal delay 112. However, if the estimation error of the delay is too large, the FECG may not be extracted by the algorithms. According to Figures 7 and 8, we can know that only the proposed algorithms can extract the FECG signal. Our algorithms function when the delay ranges from 106 to 120, which directly illustrates that our algorithms are more robust than other algorithms. In conclusion, our algorithm is applicable to real-world data and owns better robustness than other algorithms.

5. Conclusion

We have proposed novel BSE algorithms for extraction of FECG. First, we design an objective function based on

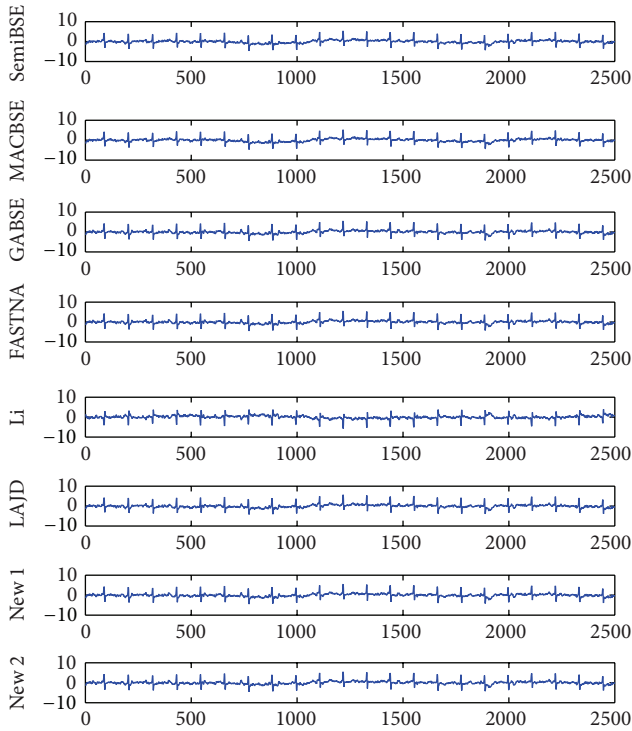


FIGURE 6: The FECG signals extracted by all algorithms at the optimal delay 112.

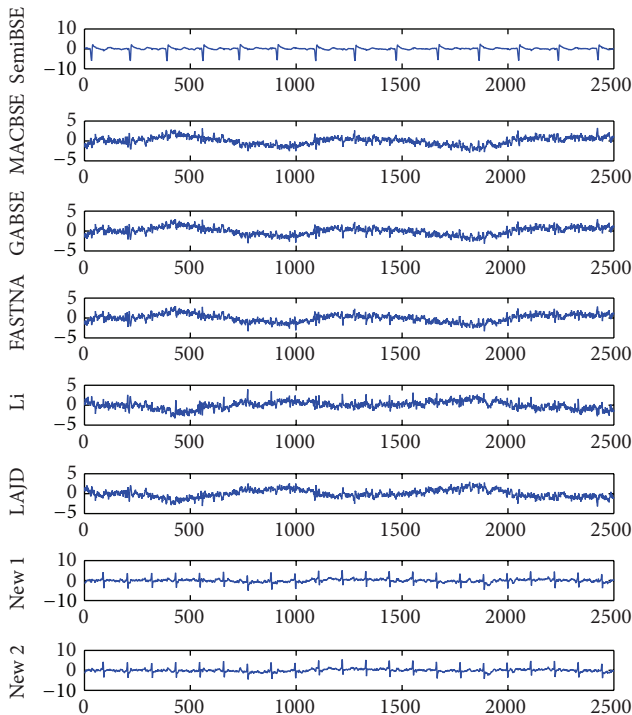


FIGURE 7: The FECG signals extracted by all algorithms at the optimal delay 106.

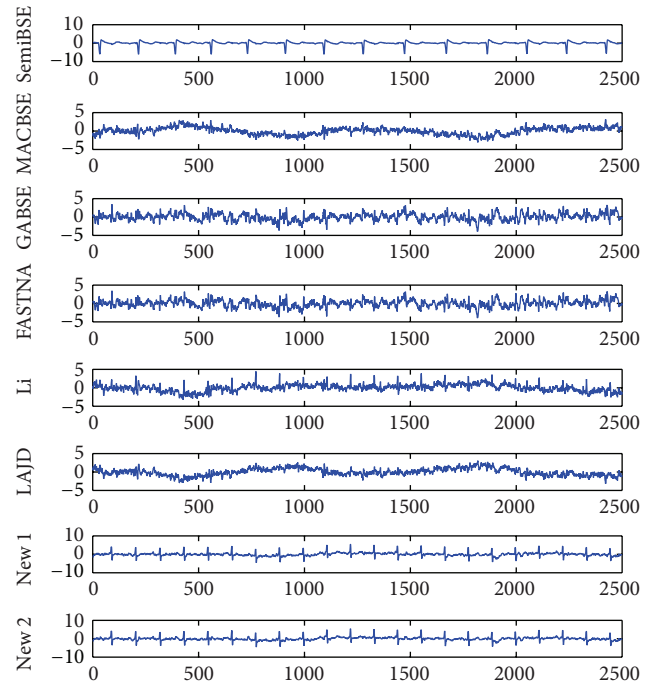


FIGURE 8: The FECG signals extracted by all algorithms at the optimal delay 120.

the non-Gaussianity and the temporal structure of the desired FECG. Secondly, a clearer FECG can be obtained by the learning algorithm. Furthermore, the proposed algorithm is further improved through combining the vector obtained by the objective function with the FastICA algorithm. The algorithms of this paper are effective and robust. They can save more time and get more satisfactory results than traditional BSS methods through utilizing the prior information. Experimental results certify the effectiveness of the proposed algorithms. Due to the generality of the presented BSE algorithms, we believe that they can also be extended to other signal processing applications. With our algorithms, it is easy to design suitable functions to solve various kinds of problems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by National Natural Science Foundation of China (no. 51509049 and no. 61501147), Heilongjiang Province Natural Science Foundation (F201345), and the Fundamental Research Funds for the Central Universities of China (no. HEUCF150812/150810).

References

[1] V. Zarzoso and A. K. Nandi, “Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise

- cancellation,” *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 1, pp. 12–18, 2001.
- [2] P. P. Kanjilal, S. Palit, and G. Saha, “Fetal ECG extraction from single-channel maternal ECG using singular value decomposition,” *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 1, pp. 51–59, 1997.
 - [3] S. Horihata, Z. Zhang, T. Imamura, T. Miyake, H. Toda, and Y. Yasuda, “Biological signal analysis by independent component analysis using complex wavelet transform,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 8, no. 4, pp. 595–608, 2010.
 - [4] X. Wang, Z. Huang, and Y. Zhou, “Semi-blind signal extraction for communication signals by combining independent component analysis and spatial constraints,” *Sensors*, vol. 12, no. 7, pp. 9024–9045, 2012.
 - [5] H. Wang, R. Li, G. Tang, H. Yuan, Q. Zhao, and X. Cao, “A Compound fault diagnosis for rolling bearings method based on blind source separation and ensemble empirical mode decomposition,” *PLoS ONE*, vol. 9, no. 10, Article ID e109166, 2014.
 - [6] Y. Li, W. Nie, and F. Ye, “A complex mixing matrix estimation algorithm based on single source points,” *Circuits, Systems, and Signal Processing*, vol. 34, no. 11, pp. 3709–3723, 2015.
 - [7] G. R. Naik and D. K. Kumar, “Improving the quality of the audio sources using Gaussianity reduction technique,” *International Journal of Electronics*, vol. 98, no. 7, pp. 949–959, 2011.
 - [8] G. R. Naik, “Measure of quality of source separation for sub- and super-Gaussian audio mixtures,” *Informatika*, vol. 23, no. 4, pp. 581–599, 2012.
 - [9] G. R. Naik and W. Wang, “Audio analysis of statistically instantaneous signals with mixed Gaussian probability distributions,” *International Journal of Electronics*, vol. 99, no. 10, pp. 1333–1350, 2012.
 - [10] K. Mohanaprasad and P. Arulmozhivarman, “Wavelet based ICA using maximisation of non-Gaussianity for acoustic echo cancellation during double talk situation,” *Applied Acoustics*, vol. 97, pp. 37–45, 2015.
 - [11] G. R. Naik, D. K. Kumar, and M. Palaniswami, “Signal processing evaluation of myoelectric sensor placement in low-level gestures: sensitivity analysis using independent component analysis,” *Expert Systems*, vol. 31, no. 1, pp. 91–99, 2014.
 - [12] D. Sugumar, P. T. Vanathi, and S. Mohan, “Joint blind source separation algorithms in the separation of non-invasive maternal and fetal ECG,” in *Proceedings of the International Conference on Electronics and Communication Systems (ICECS '14)*, pp. 1–6, IEEE, Coimbatore, India, February 2014.
 - [13] L. O. Sarmiento-Álvarez, J. Millet-Roig, and A. González-Salvador, “Fetal electrocardiogram extraction using hybrid BSS technique: COMBI and MULTICOMBI algorithms,” *Iteckne*, vol. 11, no. 2, pp. 121–128, 2014.
 - [14] A. K. Barros and A. Cichocki, “Extraction of specific signals with temporal structure,” *Neural Computation*, vol. 13, no. 9, pp. 1995–2003, 2001.
 - [15] Z. Shi and C. Zhang, “Semi-blind source extraction for fetal electrocardiogram extraction by combining non-Gaussianity and time-correlation,” *Neurocomputing*, vol. 70, no. 7–9, pp. 1574–1581, 2007.
 - [16] Z. Shi, D. Zhang, and C. Zhang, “MACBSE: extracting signals with linear autocorrelations,” *Neurocomputing*, vol. 71, no. 4–6, pp. 1082–1091, 2008.
 - [17] Z. Shi and C. Zhang, “Blind source extraction using generalized autocorrelations,” *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1516–1524, 2007.
 - [18] Z. Shi and C. Zhang, “Fast nonlinear autocorrelation algorithm for source separation,” *Pattern Recognition*, vol. 42, no. 9, pp. 1732–1741, 2009.
 - [19] C. Li and G. Liao, “Notes on two temporal structure-based methods for blind extraction of fetal electrocardiogram,” *Neural Computing and Applications*, vol. 21, no. 7, pp. 1647–1650, 2012.
 - [20] J. Wang, Y. Shen, H. Li, H. Wei, and S. Xu, “A separation algorithm for sources with temporal structure only using second-order statistics,” *Radioengineering*, vol. 22, no. 3, pp. 861–865, 2013.
 - [21] D. De Moor, Ed., *Daisy: Database for the Identification of Systems*, 1997, <http://www.esat.kuleuven.ac.be/sista/daisy>.