# BLIND SEPARATION OF BALLISTOCARDIOGRAM FROM EEG VIA SHORT-AND-LONG-TERM LINEAR PREDICTION FILTERING

*Saideh Ferdowsi*, Vahid Abolghasemi†, and Saeid Sanei**

*Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK
†School of Engineering and Design, Brunel University, Uxbridge, UK
{s.ferdowsi, s.sanei}@surrey.ac.uk, vahid.abolghasemi@brunel.ac.uk

## ABSTRACT

In this paper the problem of removing ballistocardiogram (BCG) artifact from EEG signals is addressed. This kind of artifact appears in simultaneous EEG-fMRI recordings. We propose a new Blind source extraction method based on linear prediction technique. The proposed method is a joint short-and-long-term prediction (SLTP) strategy to extract the BCG sources. The main reason of using this technique is to jointly model the temporal structure (short-term prediction) of sources and exploiting the prior information of BCG sources (long-term prediction). The results of extensive experiments on both synthetic and real data confirm the strength of the proposed technique to effectively remove the BCG artifact.

***Index Terms*—** Blind source extraction, ballistocardiogram, linear prediction.

## 1. INTRODUCTION

Simultaneous recording of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) is a useful technique for the study of brain function. Although both EEG and fMRI provide valuable information about brain functionality, the former suffers from low spatial resolution and the latter from low temporal resolution. Fusion of EEG and fMRI can therefore compensate for the existing shortcoming of each technique. However simultaneous recording of EEG and fMRI poses some difficulties. One of the major limitations is the effect of artifacts introduced in the EEG. Recorded EEG signals in the magnetic field are distorted by two major artifacts, gradient artifact and ballistocardiogram (BCG). Changes of magnetic field in MRI scanner during image acquisition cause gradient artifact [1]. Since this artifact does not show significant variability over time it can be subtracted from EEG signal using an average template approach [2]. The EEG signal after removing the gradient artifact suffers from another severe artifact known as BCG. This artifact is generated because of EEG electrode movements resulting from heart pulsation inside the magnetic field. Cardiac pulsation causes a small movement in each electrode and as a result a voltage will be induced into each electrode which obscure the EEG signal mainly at alpha frequencies (8–13 Hz) and below [3].

In order to have a successful EEG-fMRI integration an efficient removal of BCG artifact is needed. Average artifact subtraction (AAS) is an effective algorithm which has been proposed for cancelation of the BCG artifact [1]. In this method first of all, the QRS events of ECG signal are detected. Then, a number of EEG signal slices segmented based on detected QRS are averaged to build a template for artifact in each channel. The BCG can be removed by the subsequent template subtraction from each channel. One of the major drawbacks of this widely used method is incompatibility with artifact changes over time. Since the shape of BCG artifact is affected by both variation of heart beat and subject movements, the assumption of occurring similar artifacts in all the trials is not always valid. In order to deal with heart beat timing variation, Kruggel et al. [4] improved the AAS using a dynamic template with adaptive amplitude calculated by sliding and averaging. Niazy et al. [5] developed an algorithm using principal component analysis (PCA) known as optimal basis set (OBS). In their method, first a mixture of artifact trials is formed for each channel, then, the principal components of this mixture are calculated. In the third step, few of the first principal components are selected as the basis set. In the last step a template is created using the selected basis set and subtracted from each BCG trial. This method does not lead to good results when there are additional artifacts in the EEG data due to subject movement.

Another class of artifact removal methods is based on blind source separation (BSS). Independent component analysis (ICA) is a well-known BSS technique and a powerful statistical algorithm to remove the BCG in EEG [6]. The BSS approaches are useful when no ECG signal is available for template matching. Methods using ICA assume that the recorded EEG signals can be represented by a linear mixture of neural activity inside the brain, and artifacts caused by muscles and noise. On the other hand, ICA decomposes the EEG signals into a set of independent components (ICs). Removing ICs containing BCG artifact and backprojecting the remaining ICs achieves the clean EEG signals. In several studies ICA has been used for removing the BCG artifact [7]. Nakamura et al. [8] evaluated the performance of different ICA algorithms for removing BCG from EEG data. The most important issue in BCG removal using ICA, is choosing the correct number of extracted components that should be deflated. Different numbers for BCG components among the ICs have been reported in different researches. For example, in [8] three independent components are selected as BCG artifact. In contrast, Mantini et al. [7] show that ICs should be removed from the extracted sources comprises of three to six independent components. Some methods use the correlation between the estimated ICs and the ECG channel as a criterion to select the BCG components [7].

This research is supported by the Leverhulme Trust.

Selecting and removing only a small number of sources as BCG may leave some artifacts in the signals. In contrast, selecting and removing a large number of sources may eliminate some of the useful information from the EEG signals. Although ICA algorithms have broad application in BCG removal, their performance may be limited as it does not use *a priori* information about the BCG artifact. In a recent method, Ghaderi et al. [9] proposed a blind source extraction (BSE) technique with cyclostationarity constraint to extract the BCG sources. In another attempt, Leclercq et al. [10] proposed a constrained independent component analysis (cICA) to remove the BCG artifact.

In this paper, we propose a BSE method by incorporating a linear predictor (LP) [11]. In contrast to BSS algorithms in which all the sources are recovered simultaneously, in BSE algorithms only one or a selected number of sources are extracted at any time. In BSE methods, it is desired to extract the sources with particular statistical features or properties such as smoothness, predictability, or temporal correlation. In this work predictability is used as the criterion for source extraction. The main reason of using this criterion is to exploit the temporal structure of BCG sources as *a priori*. Although LP-based BSE algorithms extract the sources with specific temporal structure, they are only able to model autoregressive (AR) structure of the sources. On the other hand, these methods employ merely a model that describes short-term prediction. Since the BCG sources have more temporal (particularly long-term) correlation than any other extracted sources, a technique based on short-term prediction is not enough to extract this artifact completely. In this work an algorithm which jointly models both short-term and long-term prediction is proposed to extract the BCG sources.

In the next section the LP-based source extraction methods are described. In section 3, we present the mathematical details of the proposed method. The simulation results are presented in section 4. Finally, the paper is concluded in section 5.

## 2. PROBLEM FORMULATION

Assume that $m$ observed signals $x_i(t)$ are linear combinations of $n$ unknown source signals $s_j(t)$, with $(m \geq n)$. Source signals are either statistically independent or have different temporal structures. The BSS problem can be formulated as follows.

$$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t) \qquad i = 1, 2, \ldots, m, \tag{1}$$

or in matrix notation

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t) \tag{2}$$

where $\mathbf{x}(t) = [x_1(t), ..., x_m(t)]^T$ is the $m \times 1$ observation vector, $\mathbf{s}(t) = [s_1(t), ..., s_n(t)]^T$ is the vector of $n$ source signals at time $t$ which are assumed to be zero mean, $\mathbf{A}$ is an $m \times n$ unknown full column rank mixing matrix and $\mathbf{v}(t)$ is $m \times 1$ noise vector. In BSS-based methods, source signals can be retrieved by determining an $n \times m$ separation matrix $\mathbf{W}$. In general, $\mathbf{W}$ can be represented by product of $\mathbf{A}^{-1}$ and scaling and permutation matrices. After obtaining the separation matrix the source signals are calculated by:

$$y_j(t) = \sum_{i=1}^{m} w_{ji} x_i(t) \qquad j = 1, 2, \ldots, n, \tag{3}$$

or in matrix notation

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t). \tag{4}$$

where $\mathbf{y}(t) = [y_1(t), ..., y_n(t)]^T$ is the $n \times 1$ estimated source signals at time $t$.

The problem of estimating the sources of interest in BSE is equivalent to identification of the corresponding columns of mixing matrix, $\mathbf{a}_j$, or their pseudo-inverse, $\mathbf{w}_j$, which are rows of separation matrix $\mathbf{W} = \mathbf{A}^{\dagger}$.

A BSE algorithm may exhibit weaker convergence than simultaneous blind separation techniques. The main reason is some ill-conditioned problems due to accumulation of error during deflation procedures. In order to eliminate this problem a pre-withening is applied to the observation signals $\mathbf{x}$ in the form:

$$\tilde{\mathbf{x}}(t) = \mathbf{Q}\mathbf{x}(t) \tag{5}$$

where $\mathbf{Q} \in \mathbb{R}^{n \times m}$ is a decorrelation matrix ensuring that the correlation matrix $\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} = E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{I}_n$ is an identity matrix.

It is shown that a blind source extraction using linear predictor is capable of extracting temporally correlated sources [12]. Since this is in the scope of this paper we describe linear prediction in detail in the following section.

## 3. ALGORITHM

### 3.1. Short-term prediction

A generic linear prediction problem is expressed as follows:

$$e_s(t) = y(t) - \sum_{p=1}^{K_1} b_p y(t-p) = \mathbf{w}^T\tilde{\mathbf{x}}(t) - \mathbf{b}^T\bar{\mathbf{y}}(t) \tag{6}$$

where $\mathbf{w} = [w_1, w_2, \ldots, w_m]^T$, $\mathbf{b} = [b_1, b_2, \ldots, b_{K_1}]^T$, $\bar{\mathbf{y}}(t) = [y(t-1), y(t-2), \ldots, y(t-K_1)]^T$ and $K_1$ is the prediction order.

It is desired to estimate the optimal value of $\mathbf{w}$ and $\mathbf{b}$ leading to a successfully extracted $\mathbf{y}$. For this purpose, the following cost function should be minimized:

$$\mathcal{J}_s(\mathbf{w}, \mathbf{b}) = \frac{1}{2}E\{e_s^2(t)\} \tag{7}$$

where $E\{.\}$ denotes the expected value.

Applying such a cost function allows us to model the sources of interest using an AR model [13]. However, the performance could be significantly improved by imposing *a priori* about the temporal structure of those sources. This can play a crucial role in applications concerning artifact removal from biomedical signals especially EEG signals.

### 3.2. Long-term prediction

The purpose of a long-term linear predictor is to model a signal based on samples that do not immediately precede the current sample. The resulting error of long-term linear predictor is formulated as follows:

$$e_l(t) = y(t) - \sum_{p=1}^{K_2} d_p y(t-\tau-p) = \mathbf{w}^T\tilde{\mathbf{x}}(t) - \mathbf{d}^T\bar{\mathbf{y}}_l(t) \tag{8}$$

where $\mathbf{w} = [w_1, w_2, \ldots, w_m]^T$, $\mathbf{d} = [d_1, d_2, \ldots, d_{K_2}]^T$, $\bar{\mathbf{y}}_l(t) = [y(t-\tau-1), y(t-\tau-2), \ldots, y(t-\tau-K_2)]^T$, $\tau$ is the prediction distance referring to the cycle duration of periodic source signal and $K_2$ is the prediction order.

### 3.3. Proposed method

In the proposed algorithm a joint model based on short-term and long-term predictors is defined as the cost function:

$$\mathcal{J}(\mathbf{w},\mathbf{b},\mathbf{d})=\frac{1}{2}E\{e_s^2\}+\frac{1}{2}E\{e_l^2\}=\mathcal{J}_s(\mathbf{w},\mathbf{b})+\mathcal{J}_l(\mathbf{w},\mathbf{d}) \quad (9)$$

where $e_s$ and $e_l$ are short-term and long-term prediction errors respectively. The minimization of the proposed cost function leads to extracting the sources which are temporally correlated and their period is equal to $\tau$ defined in long-term prediction error.

A consequence of applying such a joint model for BCG removal application, is exploiting the temporal structure of BCG artifact as *a priori* for source separation. A normalized autocorrelation of ECG channel or EEG signals containing BCG artifact is used to estimate the periodicity length $\tau$ in long-term prediction.

We use the standard stochastic gradient descent algorithm [11] to estimate vectors $\mathbf{w}$, $\mathbf{b}$ and $\mathbf{d}$ which minimize the proposed cost function. The partial derivatives of $\mathcal{J}(\mathbf{w},\mathbf{b},\mathbf{d})$ with respect to $\mathbf{w}$, $\mathbf{b}$ and $\mathbf{d}$ are calculated as follows:

$$\frac{\partial\mathcal{J}(\mathbf{w},\mathbf{b},\mathbf{d})}{\partial\mathbf{w}}=\frac{\partial\mathcal{J}_s(\mathbf{w},\mathbf{b})}{\partial\mathbf{w}}+\frac{\partial\mathcal{J}_l(\mathbf{w},\mathbf{d})}{\partial\mathbf{w}}, \quad (10)$$

$$\begin{aligned}\frac{\partial\mathcal{J}_s(\mathbf{w},\mathbf{b})}{\partial\mathbf{w}}&=E\{e_s(t)(\tilde{\mathbf{x}}(t)-\sum_{p=1}^{K_1}b_p\tilde{\mathbf{x}}(t-p))\} \quad (11)\\&=\mathbf{w}^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(0)-2\mathbf{w}^T\sum_{p=1}^{K_1}b_p\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p)\\&\quad+\mathbf{w}^T\sum_{p=1}^{K_1}\sum_{q=1}^{K_1}b_pb_q\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p,q),\end{aligned}$$

$$\begin{aligned}\frac{\partial\mathcal{J}_l(\mathbf{w},\mathbf{d})}{\partial\mathbf{w}}&=E\{e_l(t)(\tilde{\mathbf{x}}(t)-\sum_{p=1}^{K_2}d_p\tilde{\mathbf{x}}(t-\tau-p))\} \quad (12)\\&=\mathbf{w}^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(0)-2\mathbf{w}^T\sum_{p=1}^{K_2}d_p\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p+\tau)\\&\quad+\mathbf{w}^T\sum_{p=1}^{K_2}\sum_{q=1}^{K_2}d_pd_q\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p+\tau,q+\tau),\end{aligned}$$

$$\begin{aligned}&\forall\ k=1,\dots K_1: \quad (13)\\&\frac{\partial\mathcal{J}(\mathbf{w},\mathbf{b},\mathbf{d})}{\partial b_k}=E\{e_s(t)y(t-k)\}\\&\quad=\mathbf{w}^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(k)\mathbf{w}\\&\quad-\sum_{p=1}^{K_1}b_p\mathbf{w}^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p,k)\mathbf{w},\end{aligned}$$

$$\begin{aligned}&\forall\ k=1,\dots K_2: \quad (14)\\&\frac{\partial\mathcal{J}(\mathbf{w},\mathbf{b},\mathbf{d})}{\partial d_k}=E\{e_l(t)y(t-\tau-k)\}\\&\quad=\mathbf{w}^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(\tau+k)\mathbf{w}\\&\quad-\sum_{p=1}^{K_2}d_p\mathbf{w}^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p,\tau+k)\mathbf{w},\end{aligned}$$

where $b_k$ and $d_k$ are the $k$th element of vectors $\mathbf{b}$ and $\mathbf{d}$ respectively.

The learning rules for $\mathbf{w}$, $\mathbf{b}$ and $\mathbf{d}$ are derived by substituting the derivatives obtained in (11-14) into the stochastic gradient descent approach.

$$\begin{aligned}\mathbf{w}^{(r+1)}&=\mathbf{w}^{(r)}-\eta_{\mathbf{w}}[2\mathbf{w}^{(r)^T}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(0) \quad (15)\\&-2\mathbf{w}^{(r)^T}\sum_{p=1}^{K_1}b_p^{(r)}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p)\\&-2\mathbf{w}^{(r)^T}\sum_{p=1}^{K_2}d_p^{(r)}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p+\tau)\\&+\mathbf{w}^{(r)^T}\sum_{p=1}^{K_1}\sum_{q=1}^{K_1}b_p^{(r)}b_q^{(r)}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p,q)\\&+\mathbf{w}^{(r)^T}\sum_{p=1}^{K_2}\sum_{q=1}^{K_2}d_p^{(r)}d_q^{(r)}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p+\tau,q+\tau)],\end{aligned}$$

$$\begin{aligned}&\forall\ k=1,\dots K_1: \quad (16)\\&b_k^{(r+1)}=b_k^{(r)}-\eta_{\mathbf{b}}[\mathbf{w}^{(r)^T}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(k)\mathbf{w}^{(r)}\\&\quad-\sum_{p=1}^{K_1}b_p^{(r)}\mathbf{w}^{(r)^T}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p,k)\mathbf{w}^{(r)}],\end{aligned}$$

$$\begin{aligned}&\forall\ k=1,\dots K_2: \quad (17)\\&d_k^{(r+1)}=d_k^{(r)}-\eta_{\mathbf{d}}[\mathbf{w}^{(r)^T}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(k+\tau)\mathbf{w}^{(r)}\\&\quad-\sum_{p=1}^{K_2}d_p^{(r)}\mathbf{w}^{(r)^T}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(p,\tau+k)\mathbf{w}^{(r)}],\end{aligned}$$

where $\eta_{\mathbf{w}}$, $\eta_{\mathbf{b}}$ and $\eta_{\mathbf{d}}$ are the step sizes and $\mathbf{w}^{(r+1)}$, $b_k^{(r+1)}$ and $d_k^{(r+1)}$ are the updated values of parameters of the proposed cost function.

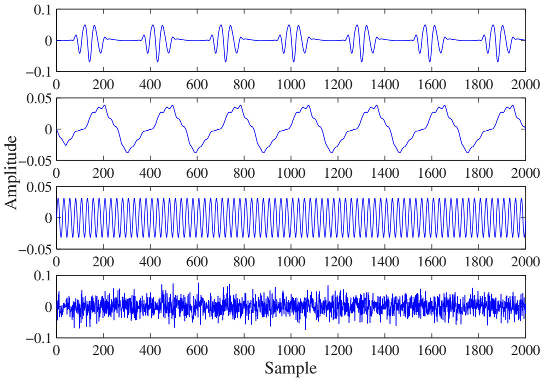
The above update rules are followed for all the sources from 1 to $n$ and the procedure of updating the values of $\mathbf{w}$, $\mathbf{b}$ and $\mathbf{d}$ is repeated to minimize (9). It is important to note that normalization of $\mathbf{w}$ is necessary after each iteration to preserve the column norm of separation matrix.

After estimating the separation matrix, the sources which are related to the BCG artifact can be determined visually. Preferably, the normalized autocorrelation of the extracted sources can be used to cluster the BCG sources automatically. In the last step, the sources which are selected as BCG are set to zero and the remaining sources are projected back to the sensor space.

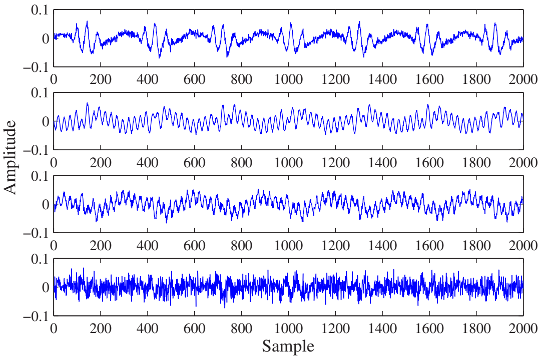
## 4. EXPRIMENTS

In this section, the performance of the proposed method for BCG removal is evaluated. First, it is compared with different ICA algorithms using a set of synthetic signals. Then, it is applied to a set of real-world EEG data and the results are compared with the results of commonly used methods for BCG removal.

Results of the proposed algorithm for synthetic data are compared with those of well-known BSS algorithms such as Infomax, FastICA and JADE. Infomax was proposed by Bell et al. [14] and is




(a)




(b)

**Fig. 1**. Simulated EEG data; (a) sources, and (b) mixtures.

working based on the maximization of the estimated entropy. FastICA is another popular algorithm for blind source extraction which has been proposed by Hyvarinen et al. [15]. Maximization of non-Gaussianity as a measure of statistical independence is the objective of this algorithm. JADE is another ICA algorithm which is based on joint diagonalization of cumulant matrices [16]. Results of the proposed method for real EEG data are compared with the results of well-known algorithms AAS [1], OBS [5] and FastICA.

### 4.1. Synthetic Data

A set of four sources including artifact and EEG rhythms with specific temporal structure are used to generate synthetic EEG signal (Figure 1.(a)). The sources are extracted from the real EEG data set which is used in the next section. In Figure 1.(a) the first two sources from the top are assumed to be the BCG artifacts, the third source is a sine wave and the last one is a uniform random noise. The sources are mixed by a full column rank random $4 \times 4$ matrix. Figure 1.(b) shows a typical mixture of the synthetic sources.

In order to determine the prediction distance for long-term prediction, the normalized autocorrelation of each simulated EEG signal is calculated. The distance between the peaks appeared in the result of estimation of autocorrelation function is approximately equal to the period of BCG artifact, and is selected as the prediction distance, $\tau$, for long-term prediction. For synthetic data utilized in this work, $\tau$ is obtained as 290 samples. An experiment including 100 trials is carried out to estimate the optimal values of prediction orders for short- and long-term prediction.

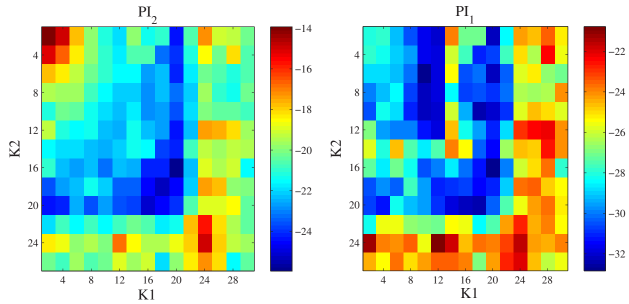



**Fig. 2**. Average performance indices of 100 trials for different values of prediction orders ($K_1$ and $K_2$). The image intensity represents the performance index value in dB.

In order to evaluate and compare the performance of different algorithms for synthetic EEG signal the following performance indices are used. The first index used in this work evaluates the success of a typical BSE algorithm in terms of extracting sources [11]. In all source extraction algorithms, it is desired to find a vector $\mathbf{w}$ such that the value of each source at time $t$ is estimated using $y_j(t) = \mathbf{w}_j^T \mathbf{x}(t)$. In the ideal case, $\mathbf{g}_j = \mathbf{w}_j^T \mathbf{A} = \pm c_j \mathbf{e}_j$ is a vector with only one non-zero element in the $j$th place, corresponding to the $j$th extracted source. $\mathbf{e}_j$ is a vector with only one unitary element, and $c_j$ is a positive scalar. Vector $\mathbf{g}_j$ is used to quantify the quality of source extraction algorithm as defined in the following equation:

$$PI_1 = \frac{1}{n}\sum_{i=1}^{n} 10\log_{10}\left(\frac{1}{n}\times\sum_{j=1}^{n}\frac{g_{ij}^2}{\max\{g_{i1}^2, g_{i2}^2, \ldots, g_{in}^2\}} - 1\right). \tag{18}$$

Smaller values for this index indicate better quality in the results of source extraction.

The next performance index presents the accuracy of the algorithm to extract the BCG sources. Assume that the number of sources which are extracted as BCG is denoted by $d_r$. The correlations between these sources and the actual artifact sources reflect the ability of the algorithm in terms of extracting artifacts:

$$PI_2 = \frac{1}{d_r}\sum_{i=1}^{d_r} 10\log_{10}\left|\frac{\langle \mathbf{y}_i(t), \mathbf{s}_i(t)\rangle}{\sqrt{\langle \mathbf{y}_i(t), \mathbf{s}_i(t)\rangle\langle \mathbf{y}_i(t), \mathbf{s}_i(t)\rangle}} - 1\right| \tag{19}$$

where $\mathbf{s}_i(t)$ is the actual source which has highest correlation with the extracted artifact $\mathbf{y}_i(t)$. Again, smaller value of this index shows better performance of the algorithm.

Figure 2 presents the performance index values, averaged over all 100 trials for $K_1, K_2 = 2, \ldots 30$. These results are obtained by calculated $PI_1$ and $PI_2$ for any combinations of $K_1$ and $K_2$ on a $30 \times 30$ grid. It is interestingly seen from Figure 2 that both $PI_1$ and $PI_2$ have minimum dB at $(K_1, K_2) = (20, 16)$. In fact, we observe co-occurrence of dark blue regions for both $PI_1$ and $PI_2$ within $K_1 =$ 16-20 and $K_2 =$ 16-20 in Figure 2.

For a comparison, we also present the average performance index values for different methods in Table 1. We used the optimal values of $K_1 = 20$ and $K_2 = 16$ for the proposed method in this table. It is seen that the performance of the proposed method is very close to that of FastICA and JADE for linearly mixed synthetic data. Although these three methods do not show significant difference in the value of $PI_1$, SLTP-BSE and FastICA provides smaller value for $PI_2$. The value reported in table 1 approved that the proposed BSE algorithm (coined SLTP-BSE) effectively extracts the sources of interest.

### 4.2. Real EEG-fMRI Data

The real EEG data set used in this work comprises of 64 EEG channels with sampling rate of 10KHz. The experiment is a simultaneous EEG-fMRI recording and was conducted at the King's College London. During the task, the subject is asked to move their hand toward the colored signs which are appeared on the screen. In order to decrease the computational time, the number of channels downsampled to 32. All the preprocessing steps used for this data set are: 1) removing gradient or imaging artifact using the method proposed by Niazy et al. [5]. 2) Down-sampling data to 250Hz. 3) Bandpass filtering using a Butterworth filter. The cutoff frequencies for low-cut and high-cut frequencies are selected as 0.5Hz and 45Hz respectively. A 10s segment of the recorded data includes BCG artifact is shown in Figure 3.

In order to remove the BCG artifact from EEG data, we first divided EEG signals into smaller segments. Then, the SLTP-BSE was applied to each segments. Five sources out of 32 were detected as BCG. We performed this step by computing the normalized autocorrelation of each extracted signal. These sources were then deflated and the remaining data backprojected into the sensor space. The algorithm parameters were selected as follows: prediction distance $\tau = 270$, the orders for short-term and long-term prediction errors $K_1 = 20$ and $K_2 = 16$ respectively. All step sizes were set to $\eta_w = \eta_b = \eta_d = 0.02$. Figure 4 shows the result of applying the proposed method to segments given in Figure 3, after deflating artifact sources. It is seen that, the proposed method removed a significant portion of BCG artifact.

The performance of the proposed method is compared with well established techniques using an appropriate performance index. This index measures the average correlation of the cleaned signals at time delays corresponding to the period of BCG ($\tau$) artifact [8]. The amount of BCG artifact which is still present in the data is shown by this index:

$$PI_3 = \frac{1}{n}\sum_{i=1}^{n} 10\log_{10} \frac{\sum_{\tau\pm\triangle\tau} \mid \sum_t(\hat{x}_i(t-\tau)-\mu_i)(\hat{x}_i(t)-\mu_i)) \mid}{\sum_t(\hat{x}_i(t)-\mu_i)^2} \tag{20}$$

where $\hat{\mathbf{x}}(t) = [\hat{x}_1(t), \hat{x}_2(t), \dots, \hat{x}_n(t)]^T$ is the remained data at time $t$ after removing BCG artifact and $\mu_i$ is the average of $\hat{x}_i(t)$. Smaller values for this index show better performance, meaning to have less amount of BCG artifact. We selected $\triangle\tau = 25$ samples for calculating $PI_3$. Figure 5 shows the value of $PI_3$ for AAS, OBS, FastICA and SLTP-BSE. We also note that the number of removed components using OBS and deflated BCG sources in FastICA were selected as 5. It can be seen that the proposed method achieves a smaller value compared to other methods. However, FastICA performance is close to that of the proposed method. Since this data set has been distorted by subject movement during the experiment, AAS and OBS show weaker performances compared to FastICA and the proposed method.

Figure 6 illustrates another segment of one EEG channel after applying different methods. Furthermore, the average power spectral density from all channels before and after artifact removal using the proposed method are illustrated in Figure 7. It is seen that the strong peaks existing in the original data at the main frequency and harmonies of the subject's heart beat frequency have been well removed.

**Table 1**. Different performance indices (in dB) for several BCG removal methods.

| | $PI_1$ | $PI_2$ |
|---|---|---|
| SLTP-BSE | -34.98 | -27.96 |
| Infomax | -19.49 | -14.14 |
| FastICA | -33.33 | -25.29 |
| JADE | -32.19 | -21.88 |

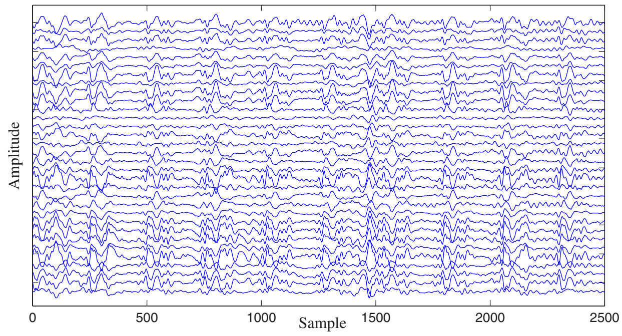



**Fig. 3**. 32 EEG channels recorded inside MR scanner after removing the gradient artifact.

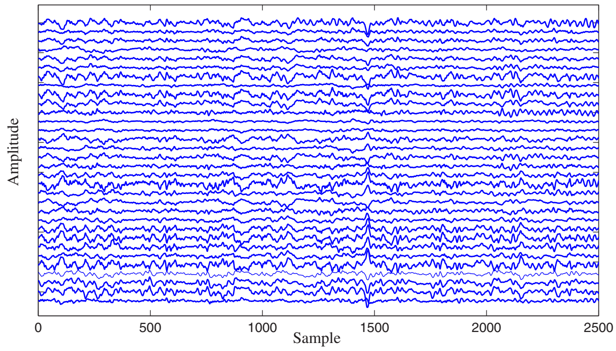



**Fig. 4**. 32 EEG channels after cleaning the BCG artifact using SLTP-BSE.

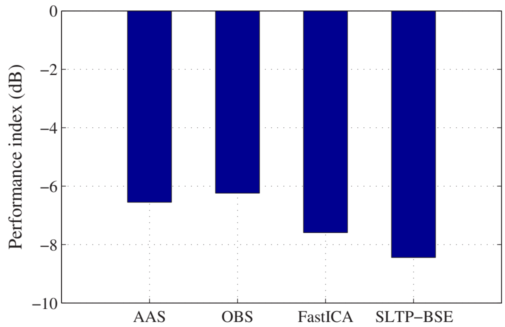



**Fig. 5**. Comparison of performance index $PI_3$ for different methods.

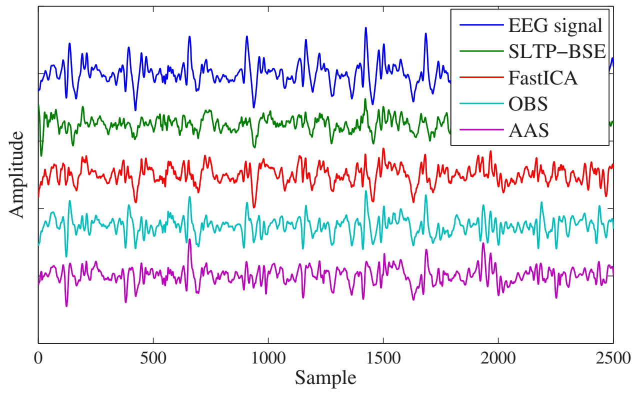



**Fig. 6**. The original EEG signal with corresponding cleaned signals using different techniques.

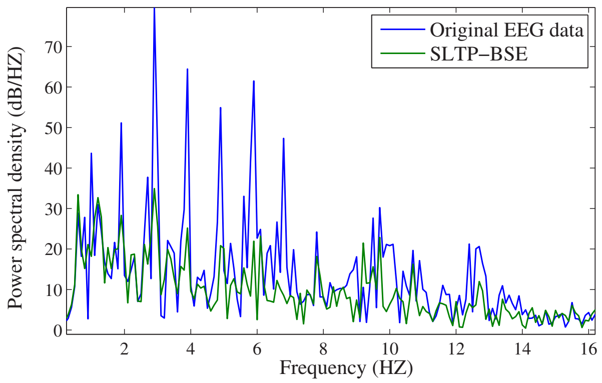



**Fig. 7**. Averaged power spectral density of the original EEG signal and the output of the proposed method.

## 5. CONCLUSION

We studied the problem of BCG artifact removal for both simulated signals and real EEG-fMRI recordings. In order to appropriately remove this kind of artifact, we proposed to use linear prediction approach to model the temporal structure of the BCG sources along with exploiting the prior knowledge about this source. Both short-term and long-term prediction were considered for this purpose. The results of our simulations on both synthetic and real data confirmed the effectiveness of the proposed approach for BCG artifact removal.

## 6. REFERENCES


[1] P. J. Allen, G. Polizzi, K. Krakow, D. R. Fish, and L. Lemieux, "Identification of EEG events in the MR scanner: The problem of pulse artifact and a method for its subtraction," *NeuroImage*, vol. 8, no. 3, pp. 229 – 239, 1998.

[2] P. J. Allen, O. Josephs, and R. Turner, "A Method for removing imaging artifact from continuous EEG recorded during functional MRI," *NeuroImage*, vol. 12, no. 2, pp. 230–239, Aug. 2000.

[3] G. Bonmassar, P. L. Purdon, I. P. Jaaskelainen, K. Chiappa, V. Solo, E. N. Brown, and J. W. Belliveau, "Motion and ballistocardiogram artifact removal for interleaved recording of EEG and EPs during MRI," *NeuroImage*, vol. 16, no. 4, pp. 1127 – 1141, 2002.

[4] F. Kruggel, C. J. Wiggins, C. S. Herrmann, and D. Y. von Cramon, "Recording of the event-related potentials during functional MRI at 3.0 Tesla field strength," *Mag. Res. in Med.*, vol. 44, no. 2, pp. 277–282, 2000.

[5] R. K. Niazy, C. F. Beckmann, G. D. Iannetti, J. M. Brady, and S. M. Smith, "Removal of fMRI environment artifacts from EEG data using optimal basis sets," *NeuroImage*, vol. 28, no. 3, pp. 720 – 737, 2005.

[6] G. Srivastava, S. Crottaz-Herbette, K. M. Lau, G. H. Glover, and V. Menon, "ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner," *NeuroImage*, vol. 24, no. 1, pp. 50 – 60, 2005.

[7] D. Mantini, M. G. Perrucci, S. Cugini, A. Ferretti, G. L. Romani, and C. D. Gratta, "Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis," *NeuroImage*, vol. 34, no. 2, pp. 598 – 607, 2007.

[8] W. Nakamura, K. Anami, T. Mori, O. Saitoh, A. Cichocki, and S. Amari, "Removal of ballistocardiogram artifacts from simultaneously recorded EEG and fMRI data using independent component analysis.," *IEEE Trans. on Biomed. Eng.*, vol. 53, no. 7, pp. 1294–1308, Jul. 2006.

[9] F. Ghaderi, K. Nazarpour, J. G. McWhirter, and S. Sanei, "Removal of Ballistocardiogram artifacts using the cyclostationary cource extraction method," *IEEE Trans. on Biomed. Eng.*, vol. 57, no. 11, pp. 2667 – 2676, Jul. 2010.

[10] Y. Leclercq, E. Balteau, T. Dang-Vu, M. Schabus, A. Luxen, P. Maquet, and C. Phillips, "Rejection of pulse related artefact (PRA) from continuous electroencephalographic (EEG) time series recorded during functional magnetic resonance imaging (fMRI) using constraint independent component analysis (cICA)," *NeuroImage*, vol. 44, no. 3, pp. 679 – 691, 2009.

[11] A. Cichocki and S. I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley & Sons, Inc., New York, NY, USA, 2002.

[12] A. Cichocki and R. Thawonmas, "On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics," *Neural Processing Letters*, vol. 12, no. 1, pp. 91–98, 2000.

[13] W. Liu, D. P. Mandic, and A. Cichocki, "Blind source extraction based on a linear predictor," *IET Signal Processing*, vol. 1, pp. 29 – 34, Mar. 2007.

[14] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution.," *Neural Comp.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[15] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley-Interscience, May 2001.

[16] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *IEE Proceedings-F*, vol. 140, pp. 362–370, 1993.