Biomedical Signal Processing and Control 18 (2015) 186–194

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Technical Note

EEG–fMRI: Dictionary learning for removal of ballistocardiogram artifact from EEG

Vahid Abolghasemi [a,*], Saideh Ferdowsi [b]

[a] Faculty of Electrical Engineering, University of Shahrood, Shahrood 3619995161, Iran
[b] Faculty of Electrical Engineering, University of Shahrood, Shahrood 3619995161, Iran

ARTICLE INFO

Article history:
Received 21 August 2014
Received in revised form
22 December 2014
Accepted 5 January 2015
Available online 24 January 2015

Keywords:
Artifact removal
Ballistocardiogram (BCG)
Dictionary learning
Electroencephalography (EEG)
Sparse representation

ABSTRACT

In this paper, artifact removal from biomedical signals is addressed. We particularly focus on removing ballistocardiogram (BCG) artifact from EEG. BCG mainly appears in EEG signals during simultaneous EEG–fMRI recordings. Different from most existing artifact removal techniques, we propose a method based on dictionary learning framework. Due to strength of sparsifying dictionaries in applications such as image denoising, it is expected to succeed in BCG removal task as well. This is investigated in the proposed approach where a dictionary is learned from original EEG recording. The dictionary is designed to locally model BCG characteristics. After achieving the dictionary, BCG can be simply subtracted from the original signal and the clean EEG is obtained. Our experimental results on both synthetic and real data confirm the effectiveness of the proposed method. The results reveal the flexibility of learned dictionary for modeling the fluctuations in artifact, and removing it from original EEG signals.

1. Introduction

Electroencephalography (EEG) acquired from the brain scalp, reflects the synaptic activity inside the brain. This type of biomedical signal has shown advantages in diagnosis of different brain illness and also brain functions. EEG, which is normally recorded in a number of channels (at different locations of the brain), is known to convey significant *temporal* information about the brain. However, it cannot reveal detailed spatial information about the brain and therefore it is not a perfect brain representative in cases where exact locations in the brain is to be found. In contrast to EEG, functional magnetic resonance imaging (fMRI) reveals useful *spatial* information and represents the local hemodynamic changes in the brain known as blood oxygenation level dependent (BOLD). Although EEG and fMRI have been separately used in many applications, simultaneous EEG–fMRI has been recently received too much interests due to its spatial–temporal characteristics [1–6]. This specific multimodal property of EEG–fMRI opens a new window to study unknown human brain functions which is an ongoing field of study [7–9].

In spite of advantages of simultaneous EEG–fMRI, it has major limitations. The most obvious one is the effects of artifacts appeared in EEG. Several dominant artifacts which contaminate EEG signals recorded inside the MRI scanner are vibration-related artifacts, gradient artifact, and ballistocardiogram (BCG). Vibration-related artifacts are caused due to the helium pump of the MR scanner [10]. Several techniques such as template subtraction or spatial filtering were used in this regard. However, these have not shown to be very effective mainly because of irregular temporal shape of the artifact. Another artifact in the EEG signal is induced due to the scanner's internal ventilation system (for fresh air supply). The properties of this artifact have been recently studied by Nierhaus et. al [10]. They show that the ventilation system causes a peak in the power spectrum of the EEG, mainly around the gamma frequency range of the EEG, which can be problematic. Gradient artifact results from variations of magnetic field in the MRI scanner during image acquisition, and severely distorts the EEG signals. Fortunately, gradient artifact appears in high frequencies and in very high amplitudes up to 100 times larger than the EEG amplitudes [11]. Therefore, it can be removed using averaging and subtraction approaches [1]. BCG artifact, which is more difficult to be removed, is the result of EEG electrode movements over the subject scalp laid inside the scanner. Small electrode movements, resulted from cardiac pulsation in the magnetic field, induce a voltage into each electrode. Primary EEG recordings inside the MR scanner appeared to have heart-beat-related artifact. The artifact was supposed to be due to whole body movement as a result of acceleration and abrupt reversal in blood flow in the aortic arch [12]. After this observation, further studies

* Corresponding author.
E-mail address: vabolghasemi@ieee.org (V. Abolghasemi).

http://dx.doi.org/10.1016/j.bspc.2015.01.001
1746-8094/

have been achieved to characterize and interpret BCG behavior with the aim of performing a successful removal technique. It is obvious that strength of electromagnetic field has a direct effect on the BCG and this is the fundamental reason of causing BCG. In other words, movement of any electrical conductor (including body movements) inside a static magnetic field causes electromagnetic induction [12]. In other studies, the pulsatile movement of the scalp is interpreted as a result of expansion and contraction of adjacent blood vessels [12]. Following this conclusion, a method was proposed for BCG removal by using a motion sensor, combined with an adaptive filtering approach. It is also believed that blood flow by itself, as a conductive fluid, could cause BCG. In fact, when blood flow is exposed in a static magnetic field, its movement induces electrical potentials [13]. These studies show that electrodes adjacent to blood vessels, such as frontal–temporal and occipital channels, would severely be affected and thus leading to major contribution of BCG-related artifacts in EEG signals.

BCG artifact obscures the EEG signal mainly at alpha frequencies (8–13 Hz) and below [14]. The amplitude of BCG is approximately 150 μV inside a magnetic field with the strength of 1.5 T [14]. It is very important to remove BCG artifact in an EEG–fMRI regime. This should be carried out very carefully as losing EEG information during BCG removal may lead the output signal to be meaningless. There are essentially two approaches in this regard. One family is devoted to designing different data recording techniques, and the second category aims at developing signal processing algorithms to remove BCG artifacts.

Investigating of BCG-related artifact properties would be very important for the researchers to develop novel removal techniques either at the acquisition stage or through signal processing methods. One primary approach to decrease the electrode movements, and consequently reduce BCG artifact, is firmly bandaging the wires and electrodes to the subject scalp [11]. Recently, a specific data recording called DRPE (direct recording prior encoding) is proposed for better separation of BCG and EEG [15]. This recording technique leads to representative bases of BCG- and EEG-only signals, alleviating the risk of model mismatches. In another study, an integrated learning and inference approach is proposed to estimate the BCG artifacts in a noisy recording [16]. This is achieved by utilizing a high-density EEG-cap and setting a special-designed experiment. Yan et al. [17] studied the magnitude and spatial variation of the artifact voltages. They expect that these effects are due to cardiac-pulse driven head rotation. Moreover, the Hall effect [17] is expected to be caused by blood flow in the main arteries of the brain, which conclusively lead to BCG artifact. Deeper study has been recently demonstrated by Mullinger et al. [18] for identifying the sources of the pulse artifact in a combined EEG–fMRI recording. They design extensive sets of experiments to measure the amplitude and variability of three sources; cardiac-pulse-driven head rotation, the Hall effect due to pulsatile blood flow and pulse-driven expansion of the scalp. They could significantly reduce the effects of pulse artifact by using head restraint and utilizing an insulating layer during EEG acquisition.

The second and more popular category of BCG removal approaches relies mainly on signal processing techniques. One of the pioneering simple BCG removal methods is average artifact subtraction (AAS) [11]. In this method, the ballistocardiogram is removed from EEG signal by subsequent template subtraction from each channel. The main challenge in using AAS is that the template cannot well follow the artifact changes over time. This is due to the fact that shape of BCG artifact is affected by both variation of heart beat and subject movements and thus does not occur similarly in all trials and time segments. Improved versions of AAS using different techniques, such as a dynamic template [19], or clustering of pulse artifacts [2], to address the variability of the artifact over time was later proposed. Recently, a statistical modeling based on singular value decomposition (SVD) is used to remove both gradient and BCG artifacts in EEG–fMRI recordings [20]. Although these techniques could mitigate the lacks of original AAS, they still have a major limitation and that is the need for a reference electrocardiogram (ECG) signal which may not always be available. Other previous BCG removal studies are mainly based on blind source separation (BSS) techniques. Principal component analysis (PCA) is used by Niazy et al. [21] to develop an algorithm known as optimal basis set (OBS). This method, first recognizes BCG components in the contaminated signals, then adaptively subtracts a fixed number of components. OBS performs well when BCG-only artifact exists in the recordings. In cases where additional artifacts exist in EEG data, e.g. due to subject movement, OBS fails. There are several attempts to address this issue by applying Infomax to the results of OBS [12,22]. Independent component analysis (ICA) is another well-known and powerful BSS technique which has been used to remove the BCG from EEG [23]. ICA-based methods assume that the EEG recordings can be represented as a linear mixture of several occurrences, namely, neural activities inside the brain, BCG, the artifacts caused by muscles, and noise. ICA separates these components by decomposing the EEG recording into a set of independent components (ICs). After identifying the BCG component and removing it, the clean EEG can be obtained by backprojection of the remaining components [24,25]. The most important issue in BCG removal by ICA, is choosing the correct number of extracted components that should be deflated. Numerous studies discuss and propose different strategies to select most convenient number of component to be deflated [24–26]. There is always a trade-off in removing the number of ICs; small number of sources may leave some artifacts in the deflated signals, whereas large number of sources may eliminate vital EEG information. Incorporating prior knowledge (if exist) into BSS algorithms is also another stream of works interested to some researchers [27,28]. These techniques are conventionally called blind source extraction (BSE) as these attempt to extract a specific source(s) with known behaviors. Another such method is the work of Ferdowsi et al. [29] which use linear predictors (LP) to model temporal characteristics of BCG for gaining an acceptable removal performance.

In this paper, we address the BCG artifact removal problem using a different technique. Our proposed method is based on the idea of sparse representation and dictionary learning which has been recently shown to be very effective in variety of applications [30,31], particularly denoising [32]. Although advantages of sparse representation has been studied in many biomedical applications [33,34], it rarely studied for artifact removal purposes. For instance, in [35] and [36], the authors apply predefined basis to sparsely represent EEG recordings and remove the existing artifacts. These works neither consider EEG–fMRI data nor take advantage of learned dictionaries. Another related work that addresses BCG removal is the work of [3] where combination of a set of predefined basis called discrete hermite transform (DHT) is used to model BCG artifact. However, it does not address the advantages of learned dictionaries and is mainly successful for removing BCG-only artifacts. It is evident that due to strong separability property of learned dictionaries [37], much work needs to be done in this field. In this work, we learn a sparsifying dictionary from the EEG recordings, contaminated with BCG, acquired in an EEG–fMRI experiment. The learned dictionary mainly captures characteristics of BCG due to relatively higher amplitude compared to EEG. In order words, the learned dictionary can well describe main features of BCG. To remove the artifact, we simply backproject the signal using the learned dictionary and sparse coefficients, and then subtract it from the original EEG recordings. We have conducted extensive experiments to examine the effectiveness of the proposed method. The obtained results confirm accurate BCG removal performance.

In addition, we observed that the learned dictionary is capable of removing not only BCG but other types of existing artifacts.

The rest of the paper is organized as follows. In the next section dictionary learning is described in detail. In Section 3 we mathematically present the proposed technique for the purpose of BCG artifact removal. The simulation results for both synthetic and real data are presented in Section 4. Finally, the conclusion is drawn in Section 5.

2. Dictionary learning

The objective in dictionary learning is to find a dictionary that can *sparsely* represent a signal or image. Dictionary is an *overcomplete* matrix in which its columns (also called atoms) describe the features of the corresponding signals/images. The term *sparse* refers to signals with small number of non-zero samples compared to signal length. Assume an overcomplete dictionary $\mathbf{D} \in \mathbb{R}^{n\times K}$, $n<K$, is available. A signal $\mathbf{x}_1 \in \mathbb{R}^n$ can be represented as a linear combination of the columns of dictionary, i.e. $\{\mathbf{d}_i\}_{i=1}^K$, such that $\mathbf{x}_1 = \mathbf{D}\mathbf{s}_1$. The key point is that if $\mathbf{x}_1$ belongs to the class of dictionary atoms, then $\mathbf{s}_1 \in \mathbb{R}^K$ is said sparse. Consequently, $\mathbf{x}_1$ will be the result of linear combination of *few* columns of $\mathbf{D}$ with a good approximation. The question now is how to learn such dictionary? In order to mathematically define the dictionary learning problem assume m training signal of length n: $\{\mathbf{x}_i\}_{i=1}^m$. By collecting all training vectors into one matrix $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^m$ of size $n\times m$, and constituting $\mathbf{S} = \{\mathbf{s}_i\}_{i=1}^m$ of size $K\times m$, the dictionary learning problem can be expressed as:

$$\min_{\mathbf{D},\mathbf{S}} \left\{ \left\|\mathbf{X}-\mathbf{D}\mathbf{S}\right\|_F^2 = \sum_{i=1}^m \left\|\mathbf{x}_i - \mathbf{D}\mathbf{s}_i\right\|_2^2 \right\} \quad s.t. \quad \|\mathbf{s}_i\|_0 \le \tau. \tag{1}$$

where $\|.\|_F$ denotes the Frobenius norm. The term $\|.\|_0$ is called ℓ_0-norm which counts the number of non-zeros and is a measure of sparsity. τ is the upper bound for acceptable number of non-zero in each $\mathbf{x}_i$ and should satisfy $\tau \ll K$. Many solutions to (1) or its variations have been reported in the corresponding literature.[1] Most existing techniques tackle (1) alternatingly, i.e. obtaining $\mathbf{S}$ assuming $\mathbf{D}$ is fixed, then, obtaining $\mathbf{D}$ assuming $\mathbf{S}$ is fixed. This process is iteratively executed until a local minimum is met. One of the popular and powerful methods that follows the above alternating minimization process is K-SVD [30]. As its name implies, it applies combination of singular value decomposition (SVD) and K-means clustering to find $\mathbf{D}$. For finding $\mathbf{S}$, any of many sparse recovery techniques can be used [38,39]. As theoretical details of K-SVD is out of the scope of this paper we refer the interested readers to [30] for more details. In this paper, we use K-SVD as the dictionary learning algorithm. Some of other existing proposed dictionary learning methods are method of optimal directions (MOD) [31] and maximum likelihood (ML)-based method [40].

3. Proposed BCG removal technique

In the previous section, the mathematical tools to be used in this section were introduced. This section is devoted to discussing how we can apply these tools to corrupted EEG data for removing its BCG contribution. In fact, we can simply say that our problem here is similar to the classical denoising problem [32] but with a different interpretation. If we define the clean EEG signal by $\mathbf{v}$, the corrupted acquired EEG by $\mathbf{y}$, and the BCG contribution by $\mathbf{x}$, the following additive model can be addressed:

$$\mathbf{y} = \mathbf{x} + \mathbf{v}. \tag{2}$$

The desire is to model BCG part, i.e. $\mathbf{x}$, to make it separable from $\mathbf{y}$. This can be accurately achieved using dictionary learning techniques, particularly due to relatively higher magnitude of $\mathbf{x}$, compared to $\mathbf{v}$. Applying this model to $\mathbf{y}$ will leave portions that cannot be modeled by the dictionary, and that would be the clean EEG, i.e. $\mathbf{v}$.

It is worthy to note here that the proposed method is applied individually to each EEG recording channel. Hence, we denote $\mathbf{y}$ of length N as one of the EEG recording signal channels, acquired in an EEG–fMRI scheme. As seen in the previous section, we require m training signals as input of any dictionary learning method. A simple and common approach to provide these training samples is to segment the original signal into m smaller signal of length n. Note that these m segments can have any percentage overlap with each other. In this paper, full overlap between the adjacent segments is considered. It means that each segment is different from its adjacent segment only by one sample. The advantage of this strategy, i.e. learning the dictionary from small segments, is that the learned dictionary is *local*, meaning that it can perfectly capture local behavior of the input signals. In order words, the learned dictionary is expected to contain local existing features in the given input signal, which could be valuable for later analysis. In fact, the nature of using small signal segments (regardless of the amount of overlaps), for learning the dictionary, makes it local. This is in contrast to *global* analytical transforms such as discrete cosine transform (DCT) which are instantaneously applied to the entire signal. In addition, if non-overlapping segments are used for dictionary learning, the existence dependence between the adjacent samples (in the segments borders) is lost. Therefore, we use overlapping segments to learn informative dictionary atoms. However, our main purpose in this paper is not just learning the dictionary from EEG recordings; more importantly we aim at removing its BCG contribution. Therefore, we need to modify the original dictionary learning problem (1), taking model (2) into account:

$$\min_{\mathbf{D},\{\mathbf{s}_i\},\mathbf{x}} \lambda\|\mathbf{y}-\mathbf{x}\|_2^2 + \sum_i \rho_i \left\|\mathbf{s}_i\right\|_0 + \sum_i \left\|\mathbf{D}\mathbf{s}_i - \mathcal{R}_i\mathbf{x}\right\|_2^2. \tag{3}$$

The first term in the above expression is the Euclidean norm between the acquired EEG, $\mathbf{y}$, and BCG contribution, $\mathbf{x}$. Parameter λ is a positive scalar that controls the amount of energy that should be devoted to $\mathbf{x}$ or $\mathbf{v}$ (note that based on (2): $\|\mathbf{y}-\mathbf{x}\|_2^2 = \|\mathbf{v}\|_2^2$). The second and third terms in (3), is equivalent to the dictionary learning problem given in (1) with slight differences in writing. Here, $\mathbf{s}_i$ is the sparse coefficient vector for i-th signal segment $\mathbf{x}_i$ such that $\mathbf{x}_i = \mathbf{D}\mathbf{s}_i$. Moreover, $\mathbf{x}_i = \mathcal{R}_i\mathbf{x}$, where $\mathcal{R}_i$ is an $n\times N$ binary (0,1) matrix that extracts the i-th segment from $\mathbf{x}$. Parameter ρ_i is a positive scalar to control the level of sparsity of $\mathbf{s}_i$, and is adjusted within the dictionary learning algorithm.

The result of solving (3) is the dictionary, $\mathbf{D}$, sparse coefficients, $\{\mathbf{s}_i\}$, and BCG signal, $\mathbf{x}$. It is obvious that this problem is not simultaneously convex in all these three variables. The most straightforward solution to these kind of problems is alteraning minimization, similar to what used to solve original dictionary learning problem (1). To apply an alternating minimization process, we first initialize $\mathbf{x}$ and $\mathbf{D}$ respectively with $\mathbf{y}$ and overcomplete DCT dictionary.[2] Then, as explained above, all the segments of $\mathbf{x}$ are extracted and re-arranged. The segments are then processed by K-SVD [30] algorithm which updates $\mathbf{D}$ and estimates the sparse coefficients $\{\mathbf{s}_i\}$ in an iterative procedure. Up to this point, $\mathbf{x}$ was assumed fixed and dictionary learning was updating the dictionary and sparse coefficients. Afterward, the situation is reversed, i.e. $\mathbf{D}$ and $\{\mathbf{s}_i\}$ are assumed fixed and only $\mathbf{x}$ varies. So, $\mathbf{x}$ can be estimated

[1] One popular variation of (1) is a relaxed version, by replacing ℓ_0-norm with ℓ_1-norm defined as $\|\mathbf{s}\|_1 = \sum_i |\mathbf{s}(i)|$.

[2] Note that in (3), only $\mathbf{y}$ and $\mathcal{R}$ are known at first.

by computing the derivative of (3) with respect to $\mathbf{x}$ and setting it to zero:

$$\lambda(\mathbf{y}-\mathbf{x})+\sum_i \mathcal{R}_i^T(\mathbf{D}\mathbf{s}_i-\mathcal{R}_i\mathbf{x})=0, \tag{4}$$

leading to:

$$\hat{\mathbf{x}}=\left(\lambda\mathbf{I}+\sum_i \mathcal{R}_i^T\mathcal{R}_i\right)^{-1}\left(\lambda\mathbf{y}+\sum_i \mathcal{R}_i^T\mathbf{D}\mathbf{s}_i\right). \tag{5}$$

In this equation, $\mathbf{I}$ is the identity matrix and $\hat{\mathbf{x}}$ is the extracted BCG portion of original EEG recording. Also, $(.)^T$ stands for matrix transpose. Finally, the clean EEG is simply obtained via: $\hat{\mathbf{v}}=\mathbf{y}-\hat{\mathbf{x}}$. It is important to note that, calculation of (5) is achieved much simpler than what appears. Since, $\sum_i \mathcal{R}_i^T\mathcal{R}_i$ is diagonal, $\hat{\mathbf{x}}$ can be calculated sample-by-sample, and matrix inverse and huge matrix computations can be avoided. The steps of proposed method are summarized in Algorithm 1.

Algorithm 1: The proposed algorithm.

Input: EEG recording $\mathbf{y}$, dictionary dimension $n \times K$, and segment size m.
Output: Estimated clean EEG $\hat{\mathbf{v}}$.
begin
1. Set $\mathbf{D}$ to overcomplete DCT;
2. Set $\mathbf{x}=\mathbf{y}$;
3. Extract segments from $\mathbf{x}$: $\{\mathbf{x}_i\}_{i=1}^m$;
4. Apply K-SVD to find $\{\mathbf{s}_i\}_{i=1}^m$ and $\mathbf{D}$;
5. Compute $\hat{\mathbf{x}}$ using (5);
6. Compute clean EEG: $\hat{\mathbf{v}}=\mathbf{y}-\hat{\mathbf{x}}$
end

4. Experiments

We divide the experimental results into two parts; synthetic and real data. Extensive sets of experiments are conducted for both data types, and the performance of the proposed method are evaluated. Since our proposed method is regarded as a single channel technique, other relative methods such as AAS [11], OBS [21], and discrete time hermit (DHT) [3] are used in this paper for comparison purposes, both in synthetic and real data experiments.

4.1. Synthetic data

The synthetic signal is composed of two parts: a $1/f$ (sometimes called Pink) noise, and a BCG-like signal. Due to nonuniform characteristics of EEG signal we choose to use $1/f$ noise to resemble the EEG. The BCG-like signal is created by combining several sinusoidal with different frequencies and amplitudes. These signals and their summation are shown in Fig. 1. The input SNR of contaminated signal, i.e. $\mathbf{Y}$ with respect to $\mathbf{S}1$, in Fig. 1 is 10 dB. As seen from the figure, the $1/f$ noise which resembles EEG has a significantly lower amplitude than BCG-like part; the fact that is observed in the real scenario as well. In order to apply the proposed method, segments of length 50 with full overlap is extracted from the noisy EEG signal. Then, Algorithm 1 is applied to learn a 50×100 dictionary, followed by other steps to separate EEG and BCG parts. We also applied AAS, OBS and DHT methods to the synthetic signal. Fig. 2 gives the results of separation using different methods. It is visible from this figure that the separated BCG using the proposed method matches well with the original signal in Fig. 1. DHT method can well recognize the high-amplitude parts of the BCG, while it fails to fully reconstruct the low-amplitude parts (Fig. 2(c)). Nonetheless, the results of AAS and OBS are less accurate. To provide quantitative measures for performance evaluation, the calculated root mean square error (RMSE) and output signal to noise ratio (SNR) for the extracted BCG are shown in Table 1. These results are obtained at different $1/f$ noise levels (equivalent to three different input SNRs), averaged over 500 trials. The values in the table reveals the higher performance of the proposed method compared to other techniques.

Table 1
Averaged SNRs and RMSEs of extracted BCG for different techniques.

Method	Input SNR (dB)	Output SNR (dB)	RMSE
AAS	5	13.4038	0.0685
	10	20.0167	0.0314
	15	24.7173	0.0182
OBS	5	13.1220	0.0707
	10	19.9904	0.0315
	15	24.6277	0.0184
DHT	5	14.1812	0.0611
	10	21.1946	0.0274
	15	26.1920	0.0154
Proposed	5	14.7821	0.0556
	10	22.8452	0.0228
	15	27.1542	0.0137

Table 2
Performance indices for several BCG removal methods.

	AAS	OBS	DHT	Proposed method
PI_1 (dB)	−5.27	−5.16	−7.595	−9.31
PI_2 (dB)	−22.38	−22.37	−28.73	−31.35

Furthermore, in order to evaluate and compare the performance of different algorithms for synthetic EEG signals the following performance indices are used. The first performance index presents the performance of the artifact removal algorithms by measuring the improvement in normalized power spectrum ratio (INPS) [41,42]:

$$PI_1=10\log_{10}\frac{\sum_f \phi_{\hat{\mathbf{v}}}(f)}{\sum_f \phi_{\mathbf{y}}(f)} \tag{6}$$

where $\phi_{\mathbf{y}}(f)$ and $\phi_{\hat{\mathbf{v}}}(f)$ are the power spectral density of recorded EEG and cleaned EEG (i.e. after removing BCG), respectively. f is a set of intervals around the basic frequency of BCG artifact and its harmonics. The length of this interval is selected 0.5 Hz in our simulations. Smaller value of this index show better performance of the algorithm in terms of removing the frequency components of BCG artifact.

The next performance index is utilized to measure the amount of distortion after BCG removal. The core meaning of the following performance index is to measure the amount of correlation between the cleaned EEG, i.e. $\hat{\mathbf{v}}$, and the original EEG, i.e. $\mathbf{v}$, available in the synthetic experiment. In other words, this index gives a measure of the amount of BCG which is still remained in the cleaned EEG. This can be formulated using PI_2 as follows:

$$PI_2=10\log_{10}\left|\frac{\langle\hat{\mathbf{v}},\mathbf{v}\rangle}{\|\mathbf{v}\|\cdot\left\|\hat{\mathbf{v}}\right\|}-1\right| \tag{7}$$

where $\langle .,. \rangle$ denotes the vector inner product operation. Again, smaller value of this index shows higher ability of the algorithm to remove artifact without distorting the data.

It is important to note that both above performance indices are applied separately to each EEG channel. Hence, in case of multichannel recordings, the obtained performance indices at each channel should be averaged over all channels. Table 2 shows the calculated performance indices after applying different methods to the synthetic data. It is observed that proposed approach outperforms other methods since it achieves smaller values for PI_1 and PI_2 compared to other techniques.

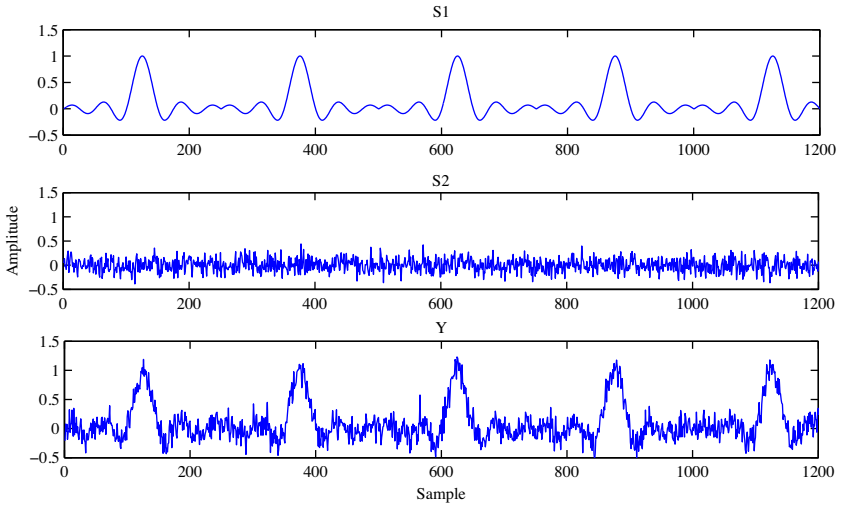

Fig. 1. Simulated EEG data; S1: BCG (sinusoidals), S2: EEG ($1/f$ noise), and Y: summation of S1 and S2. The SNR of Y is 10 dB.

4.2. Real EEG–fMRI data

The same real EEG–fMRI data from five subjects as detailed in the work of [29] is used in this section. The 64-channel EEG is acquired by a Neuroscan Maglink RT system with a cap having 10–20 kΩ resistor. The sampling rate of raw EEG data was set at 10 kHz. However, after applying the preprocessing steps, i.e. (a) gradient artifact removal, (b) downsampling, and (c) bandpass filtering, the data is converted to a 32-channel EEG with sampling frequency of 250 Hz. More details about the data and experimental settings can be found in [29].

Similar to synthetic data experiments, we first divide all 32 recorded EEG signals into segments of size 50 samples. One primary approach is to apply Algorithm 1 separately to each signal to obtain 32 individual dictionaries and then remove the BCG contribution. However, in order to gain a faster learning process requiring less memory space we follow a different approach. Since all 32 channels pose similar nature, Algorithm 1 is applied simultaneously

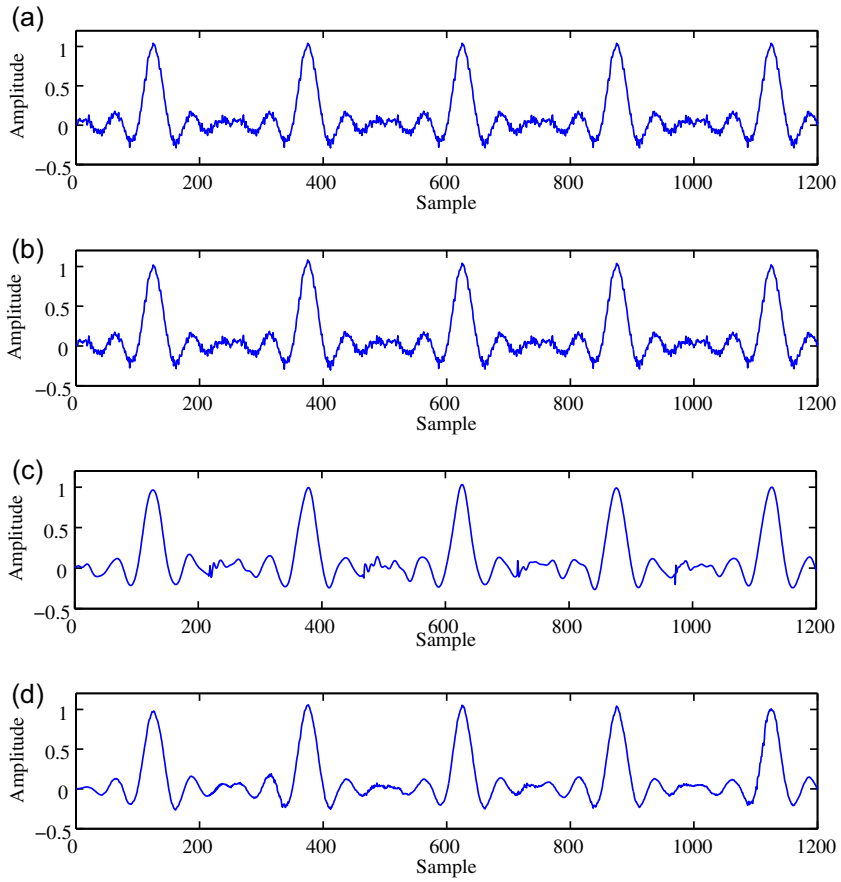

Fig. 2. Extracted BCG from synthetic data using (a) AAS, (b) OBS, (c) DHT, and (d) proposed method.

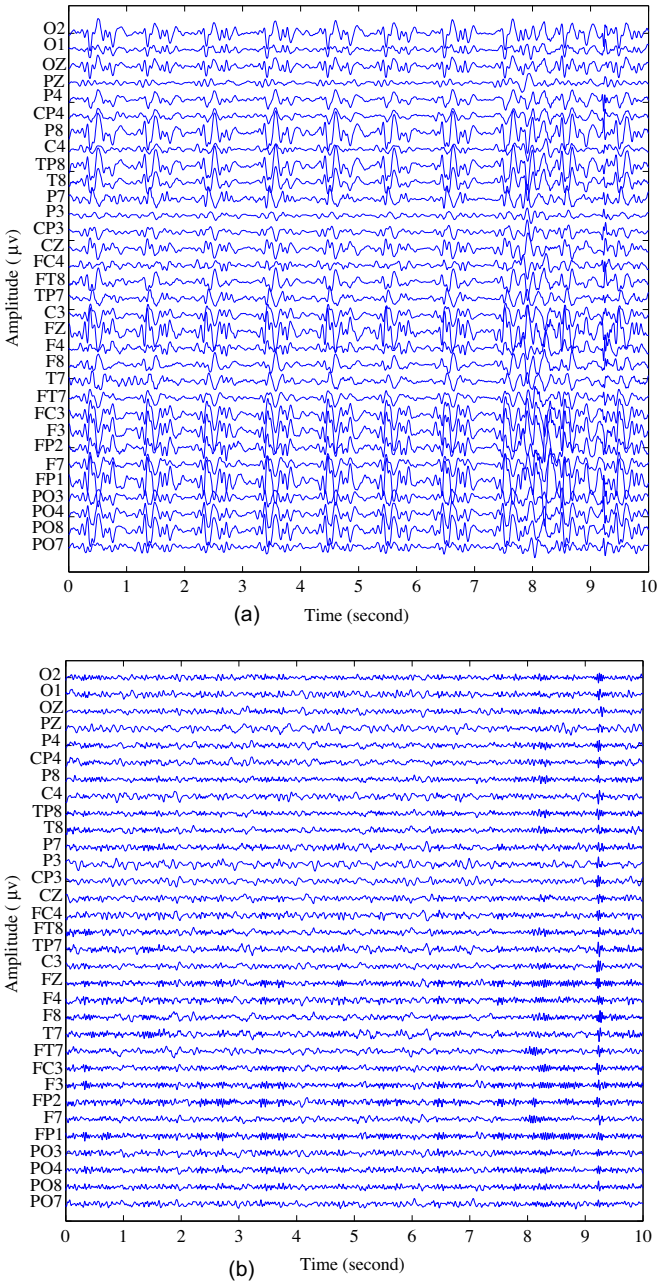

Fig. 3. 32 real EEG channels recorded inside the MR scanner; (a) pre-processed to remove the gradient artifact. (b) Processed using the proposed method to remove BCG artifact.

to all 32 channels to learn *one* single dictionary. To achieve this, the extracted segments from all channels are stacked into a single matrix which is then given to Algorithm 1. The learned dictionary based on this procedure contains atoms which describe existing features in all channels. Finally, Eq. (5) is executed for the corresponding segments to produce BCG contribution of each channel.

Fig. 3(b) illustrates the result of applying the proposed method to the distorted real EEG segments given in Fig. 3(a). It is observed that the proposed method has successfully removed significant portions of BCG artifacts from all channels. Furthermore, we demonstrate some of the atoms of the learned dictionary in Fig. 4. The curves in this figure can be considered as basis functions to model BCG behavior which is then used to remove it from original signals.

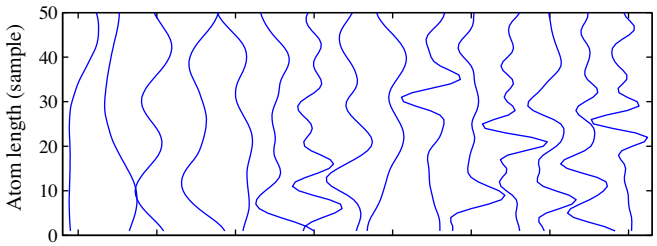

Fig. 4. Selected dictionary columns (atoms). Note that dictionary size is 50×100.

Table 3
Averaged values of PI_1 and RMSE over five subjects calculated for different artifact removal methods.

	AAS	OBS	DHT	Proposed method
PI_1 (dB)	−6.155	−6.706	−6.882	−7.469
RMSE	18.306	21.108	31.204	38.241

In the next part of the experiments we investigate the effects of varying the penalty terms, i.e. λ and ρ. Instead of directly dealing with λ and ρ, it is more feasible in practice to constraint the term $\sum_i \left\|\mathbf{D}\mathbf{s}_i - \mathcal{R}_i\mathbf{x}\right\|_2^2 \leq T$ with a scalar T for solving (3). The main intuition for this replacement is that λ and ρ depend on each other. In other words, one cannot simultaneously have a nearly zero reconstruction error (i.e. large λ) and perfectly sparse coefficients (i.e. large ρ). Thus, choosing a small T, makes $\mathbf{s}_i$s not to be too sparse, leading to a *fine* fitting to the original data and therefore a small error $\|\mathbf{y}-\mathbf{x}\|_2^2$. In contrast, a large T, makes $\mathbf{s}_i$s to be too sparse, leading to a *coarse* fitting to the original data and therefore a large error $\|\mathbf{y}-\mathbf{x}\|_2^2$. To illustrate this effect, the results of using the proposed method for different values of T are studied. These results are given in Fig. 5 for a small segment of severely distorted EEG. Fig. 5 demonstrates the strength and flexibility of the proposed method to following the input signal behavior. It further indicates that the proposed method is not only able to model the BCG artifact but other strong existing distortions in the signal.

4.2.1. Parameter selection

In order to determine proper value for T a five-fold cross-validation was used. First, a subset of contaminated EEG data was randomly selected for validation. Then, complementary subsets were selected as training set. The proposed method was first applied to the training set with a T, leading to the best removal performance. The criteria for assessing the best BCG removal were INPS, RMSE, and VEP (will be shortly explained). After obtaining the optimal $T = 0.15$, this process was repeated where the testing and training subsets were replaced in all five subsets. The obtained cross-validation results were consistent and confirmed the stability of the proposed method for the selected T.

4.2.2. PSD evaluation

In the next part of experiments we compute the average power spectral density (PSD) of all channels before and after artifact removal using the proposed method for one of the subjects. These results are illustrated in Fig. 6. It is visible from this figure that the existing strong harmonics in the original data that fall within the main frequency of the subjects heart beat have been entirely removed.

4.2.3. Performance index calculation

Table 3 presents the averaged PI_1, improvement in normalized power spectrum ratio (INPS) (6), over all 32 channels using different artifact removal methods. Also, the averaged root mean square error (RMSE) of the recorded EEG signal and the clean EEG after

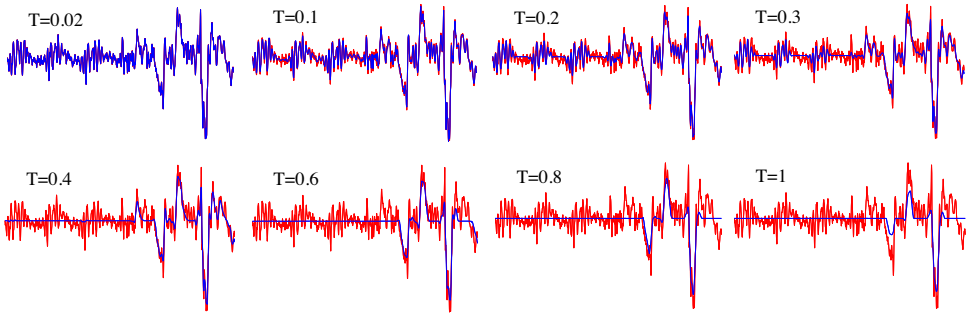

Fig. 5. Segments of original recorded EEG (red), and the results of applying the proposed method (blue) with different value of *T*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

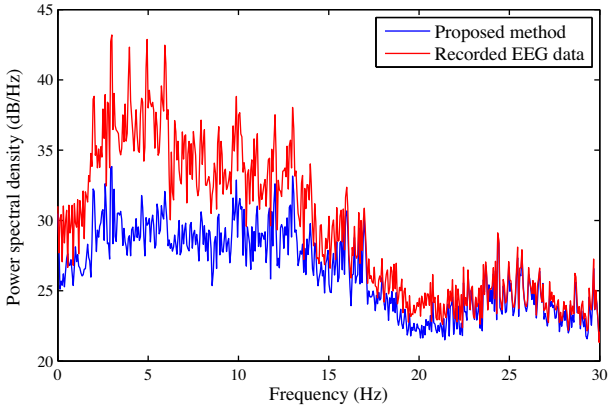

Fig. 6. Averaged power spectral density of the original EEG signal and the output of the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

applying different methods are presented in this table. The obtained values for PI_1 and RMSE show that the proposed method outperforms other conventional techniques. We note that smaller values of PI_1 and larger values of RMSE means higher performance of the corresponding algorithm. One limitation of AAS and OBS is that these methods require an extra ECG channel to be applicable to real data. Nevertheless, the proposed method is data-driven and not dependent on any extra signal. Moreover, since AAS and OBS assume periodicity in BCG, they are very sensitive to its variations. Relying on this assumption may degrade the results particularly if the BCG is not exactly periodic which is normally the case. In contrast, the proposed method does not rely on a periodic behavior for BCG that makes it robust, leading to promising results. Furthermore, compared to DHT method that uses predefined basis functions, the proposed method takes advantage of local data-driven functions (learned dictionary atoms) which is the main reason of achieving improved results.

4.2.4. ERD evaluation

In order to observe and measure the amount of remained meaningful information in the cleaned EEG, the following experiment is carried out. The main task performed by the subjects in our dataset is hand movement. One of the important effects of hand movement is decreasing the power of signal in the μ frequency band (8–13 Hz) [43]. In other words, while the subject is preparing to move, μ rhythm is suppressed and the signal power is decreased over sensory motor cortex areas [44]. This phenomenon is called event-related desynchronization (ERD) and is an appropriate measure to evaluate the success of BCG removal stage [29,45]. Unfortunately, BCG artifact co-occurs with ERD in μ frequency band. Therefore,

Table 4
SNRs (dB) of ERPs for all subjects after removing BCG.

Method	Subject				
	1	2	3	4	5
AAS	2.041	2.517	3.307	1.985	2.749
OBS	3.184	3.162	3.684	2.316	3.957
DHT	3.663	3.738	4.131	3.210	3.941
Proposed	4.056	4.487	4.964	3.904	4.142

information retrieval at this frequency band becomes a crucial and challenging task which highly depends on the performance of BCG removal stage. Similar to the work of [29], we use Morlet wavelet transform to estimate ERD in μ band. For this purpose, we first define the complex Morlet wavelets $w(t, f_0)$ [46]:

$$w(t, f_0) = A \exp\left(\frac{-t^2}{2\sigma_t^2}\right) exp(2i\pi f_0 t) \tag{8}$$

where $\sigma_f = 1/2\pi\sigma_t$ and $A = (\sigma_t\sqrt{\pi})^{1/2}$. The value of (f_0/σ_f) known as trade-off ratio is commonly set to 7 for generating the wavelet family. Since the subjects are right-handed, C3 channel is selected for these calculations. Using the Morlet wavelets, we obtain the time-frequency map for each segment of clean data and original recording of C3 channel to detect the power changes of μ rhythm (frequencies between 8 and 13 Hz). The obtained time–frequency distributions are then averaged over all the segments. These results for both original EEG and cleaned EEG are given in Fig. 7. It can be clearly observed from Fig. 7(b) that the power reduction starts at zero second (movement preparation). The dark blue region after zero second and within 8–13 Hz frequencies indicates the ERD at the lowest power level. These results cannot be deduced from original EEG, Fig. 7(a), implying a successful BCG removal using the proposed method.

4.2.5. ERP evaluation

In the next part of the experiments we aim at measuring the quality improvement for event related potentials (ERP). We observe the amount of ERPs, preserved during the BCG removal, when comparing the results of the proposed approach with those of other methods. The appropriate assessment used for this purpose is SNRs for peaks of averaged ERPs after BCG removal. SNR is calculated by dividing the amplitude of ERP peak of interest to the standard deviation of the pre-stimulus interval. For this purpose, ERP at electrode C3 was used as the electrode of interest after averaging over the trials. Table 4 demonstrates the calculated SNRs for different techniques. It is observed from Table 4 that the proposed method gives the highest SNR among other techniques, confirming that ERP information is preserved.

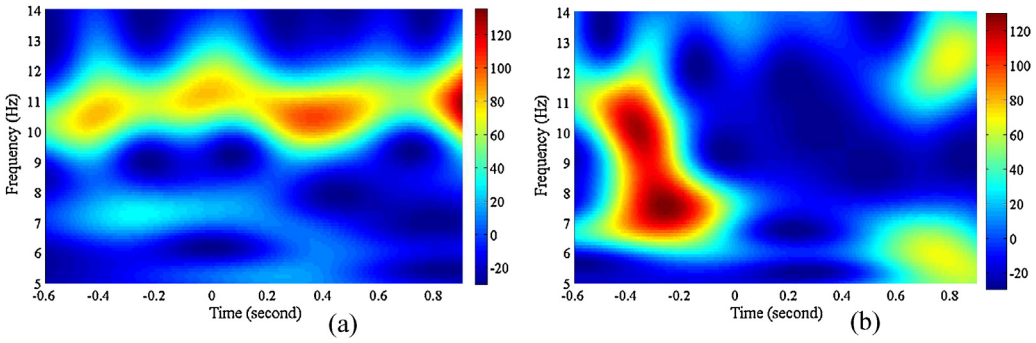

Fig. 7. Averaged time–frequency distributions of C3 channel (in dB) for ERD analysis; (a) original EEG, and (b) cleaned EEG as a result of applying the proposed method.

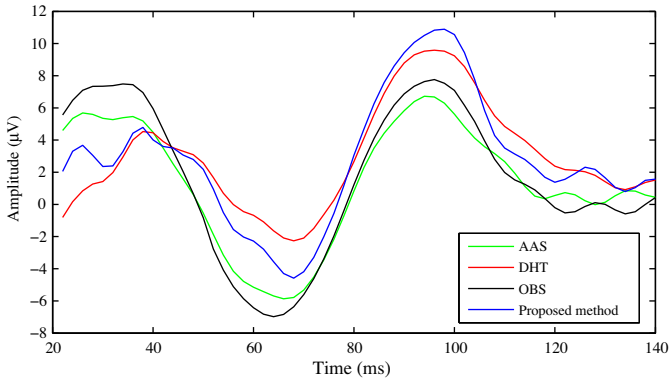

Fig. 8. VEP on PO4 electrode obtained from EEG recorded inside the MR scanner.

4.2.6. VEP evaluation

In order to investigate more about the quality of extracted information from clean EEG, the visual evoked potential (ERP) is calculated. The waveforms of the obtained VEP from PO4 electrode for one subject are illustrated in Fig. 8. The results are given for several methods. The P100 peak (100 ms after the stimuli) is visible in this figure which indicates preserving the useful visual information.

5. Conclusions

In this paper a new method for removal of BCG artifact from EEG signals was proposed. The method first learns a sparsifying dictionary from the given data. Then, the dictionary is used to model the existing BCG contribution in the original signals. The obtained BCG model is then subtracted from the original EEG to achieve clean EEG. To the best of our knowledge, this is the first reported study about applicability of dictionary learning for BCG removal. The key-point here is that the learned dictionary, which is obtained from input data, can adaptively and locally represent the BCG characteristics. This is in contrast to methods based on predefined global basis such as Fourier, wavelet, etc. In other words, the proposed method specifically designs a dictionary for the given data, while well-known fixed transforms are publicly designed for a broader class of signals and may not be able to follow all aspects of a given signal. The experiments on both synthetic and real data confirmed that promising results can be achieved using the proposed method. Moreover, unlike other existing methods which rely on periodicity of BCG (which is not always met), or ECG channel information, no strong presumption is utilized in the proposed algorithm. This leads to higher robustness and stability especially in badly acquired EEG signals contaminated with BCG artifacts.

Further evaluation of the proposed method for other EEG–fMRI recordings is still an open issue. The performance of the proposed method could differ for other EEG–fMRI experiments. In fact, depending on the tasks performed by subject, different types of brain activities are expected to occur. For instance, we were not able to verify steady state evoked potentials (SSEP) for the dataset used in this work due to lack of a repeated flashing. However, it might be measurable for other EEG–fMRI experiments. It would be valuable to know if such brain activity is preserved after applying the proposed removal algorithm. These are open streams of work which should be addressed in the future.

References

[1] P.J. Allen, O. Josephs, R. Turner, A method for removing imaging artifact from continuous EEG recorded during functional MRI, NeuroImage 12 (2) (2000) 230–239.
[2] S. Gonçalves, P. Pouwels, J. Kuijer, R. Heethaar, J. de Munck, Artifact removal in coregistered EEG/fMRI by selective average subtraction, Clin. Neurophysiol. 118 (11) (2007) 2437–2450.
[3] A. Mahadevan, A. Acharya, S. Sheffer, D.H. Mugler, Ballistocardiogram artifact removal in EEG–fMRI signals using discrete hermite transforms, Sel. Top. Signal Process. 2 (6) (2008) 839–853.
[4] S. Sanei, J. Chambers, EEG Signal Processing, John Wiley & Sons, Inc., New York, NY, USA, 2007.
[5] S. Sanei, Adaptive Processing of Brain Signals, John Wiley & Sons, Inc., New York, NY, USA, 2014.
[6] H. Xia, D. Ruan, M.S. Cohen, Separation and reconstruction of BCG and EEG signals during continuous EEG and fMRI recordings, Front. Neurosci. 8 (2014) 163.
[7] J. Jorge, W. van der Zwaag, P. Figueiredo, EEG–fMRI integration for the study of human brain function, NeuroImage 102P1 (2014) 24–34, Multimodal Data Fusion.
[8] I. Neuner, J. Arrubla, J. Felder, N.J. Shah, Simultaneous EEG–fMRI acquisition at low, high and ultra-high magnetic fields up to 9.4 T: perspectives and challenges, NeuroImage 102P1 (2014) 71–79, Multimodal Data Fusion.
[9] P. LeVan, J. Maclaren, M. Herbst, R. Sostheim, M. Zaitsev, J. Hennig, Ballisto-cardiographic artifact removal from simultaneous EEG–fMRI using an optical motion-tracking system, NeuroImage 75 (0) (2013) 1–11.
[10] T. Nierhaus, C. Gundlach, D. Goltz, S.D. Thiel, B. Pleger, A. Villringer, Internal ventilation system of MR scanners induces specific EEG artifact during simultaneous EEG–fMRI, NeuroImage 74 (2013) 70–76.
[11] P.J. Allen, G. Polizzi, K. Krakow, D.R. Fish, L. Lemieux, Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction, NeuroImage 8 (3) (1998) 229–239.
[12] S. Debener, K.J. Mullinger, R.K. Niazy, R.W. Bowtell, Properties of the ballistocardiogram artefact as revealed by EEG recordings at 1.5, 3 and 7 T static magnetic field strength, Int. J. Psychophysiol. 67 (3) (2008) 189–199.
[13] T.S. Tenforde, C.T. Gaffey, B.R. Moyer, T.F. Budlnger, Cardiovascular alterations in Macaca monkeys exposed to stationary magnetic fields: experimental observations and theoretical analysis, Bioelectromagnetics 4 (1) (1983) 1–9.
[14] G. Bonmassar, P.L. Purdon, I.P. Jääskeläinen, K. Chiappa, V. Solo, E.N. Brown, J.W. Belliveau, Motion and ballistocardiogram artifact removal for interleaved recording of EEG and EPs during MRI, NeuroImage 16 (4) (2002) 1127–1141.
[15] H. Xia, D. Ruan, M.S. Cohen, Separation and reconstruction of BCG and EEG signals during continuous EEG and fMRI recordings, Front. Neurosci. 8 (2014) 163.
[16] H. Xia, D. Ruan, M. Cohen, BCG artifact removal for reconstructing full-scalp EEG inside the MR scanner, in: 2013 International Workshop on Pattern Recognition in Neuroimaging (PRNI), 2013, pp. 178–181.
[17] W.X. Yan, K.J. Mullinger, G.B. Geirsdottir, R. Bowtell, Physical modeling of pulse artefact sources in simultaneous EEG/fMRI, Hum. Brain Mapp. 31 (4) (2010) 604–620.
[18] K.J. Mullinger, J. Havenhand, R. Bowtell, Identifying the sources of the pulse artefact in EEG recordings made inside an MR scanner, NeuroImage 71 (2013) 75–83.

[19] F. Kruggel, C.J. Wiggins, C.S. Herrmann, D.Y. von Cramon, Recording of the event-related potentials during functional MRI at 3.0 T field strength, Magn. Reson. Med. 44 (2) (2000) 277–282.
[20] Z. Liu, J.A. de Zwart, P. van Gelderen, L.-W. Kuo, J.H. Duyn, Statistical feature extraction for artifact removal from concurrent fMRI-EEG recordings, NeuroImage 59 (3) (2012) 2073–2087.
[21] R.K. Niazy, C.F. Beckmann, G.D. Iannetti, J.M. Brady, S.M. Smith, Removal of fMRI environment artifacts from EEG data using optimal basis sets, NeuroImage 28 (3) (2005) 720–737.
[22] S. Debener, A. Strobel, B. Sorger, J. Peters, C. Kranczioch, A.K. Engel, R. Goebel, Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI: removal of the ballistocardiogram artefact, NeuroImage 34 (2) (2007) 587–597.
[23] G. Srivastava, S. Crottaz-Herbette, K. Lau, G. Glover, V. Menon, ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner, NeuroImage 24 (1) (2005) 50–60.
[24] D. Mantini, M.G. Perrucci, S. Cugini, A. Ferretti, G.L. Romani, G. Del Gratta, Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis, NeuroImage 34 (2) (2007) 598–607.
[25] W. Nakamura, K. Anami, T. Mori, O. Saitoh, A. Cichocki, S. Amari, Removal of ballistocardiogram artifacts from simultaneously recorded EEG and fMRI data using independent component analysis, IEEE Trans. Biomed. Eng. 53 (7) (2006) 1294–1308.
[26] M. Dyrholm, R. Goldman, P. Sajda, T.R. Brown, Removal of BCG artifacts using a non-kirchhoffian overcomplete representation, IEEE Trans. Biomed. Eng. 56 (2) (2009) 200–204.
[27] F. Ghaderi, K. Nazarpour, J.G. McWhirter, S. Sanei, Removal of ballistocardiogram artifacts using the cyclostationary source extraction method, IEEE Trans. Biomed. Eng. 57 (11) (2010) 2667–2676.
[28] Y. Leclercq, E. Balteau, T. Dang-Vu, M. Schabus, A. Luxen, P. Maquet, 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 44 (3) (2009) 679–691.
[29] S. Ferdowsi, S. Sanei, V. Abolghasemi, J. Nottage, O. O'Daly, Removing ballistocardiogram artifact from EEG using short- and long-term linear predictor, IEEE Trans. Biomed. Eng. 60 (7) (2013) 1900–1911.
[30] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[31] K. Engan, S.O. Aase, J. Hakon Husoy, Method of optimal directions for frame design, in: IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP'99, Washington, DC, USA, 1999, pp. 2443–2446.
[32] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Trans. Image Process. 15 (12) (2006) 3736–3745.
[33] I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'Ardenne, W. Richter, J.D. Cohen, J. Haxby, Independent component analysis for brain fMRI does not select for independence, Proc. Natl. Acad. Sci. 106 (26) (2009) 10415–10422.
[34] M. Lustig, D. Donoho, J.M. Pauly, Sparse MRI: the application of compressed sensing for rapid MR imaging, Magn. Reson. Med. 58 (6) (2007) 1182–1195.
[35] X. Yong, R. Ward, G. Birch, Artifact removal in EEG using morphological component analysis, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009, 2009, pp. 345–348.
[36] X. Yong, R. Ward, G. Birch, Generalized morphological component analysis for EEG source separation and artifact removal, in: 4th International IEEE/EMBS Conference on Neural Engineering, 2009. NER'09, 2009, pp. 343–346.
[37] V. Abolghasemi, S. Ferdowsi, S. Sanei, Blind separation of image sources via adaptive dictionary learning, IEEE Trans. Image Process. 1 (6) (2012) 2921–2930.
[38] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666.
[39] S.S. Chen, D. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1999) 33–61.
[40] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Res. 37 (23) (1997) 3311–3325.
[41] G. Srivastava, S. Crottaz-Herbette, K. Lau, G. Glover, V. Menon, ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner, NeuroImage 24 (1) (2005) 50–60.
[42] S. Tong, A. Bezerianos, J. Paul, Y. Zhu, N. Thakor, Removal of ECG interference from the EEG recordings in small animals using independent component analysis, J. Neurosci. Methods 108 (1) (2001) 11–17.
[43] G. Pfurtscheller, F. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, Clin. Neurophysiol. 110 (11) (1999) 1842–1857.
[44] E. Formaggio, S. Storti, M. Avesani, R. Cerini, F. Milanese, A. Gasparini, M. Acler, R. Pozzi Mucelli, A. Fiaschi, P. Manganotti, EEG and fMRI coregistration to investigate the cortical oscillatory activities during finger movement, Brain Topogr. 21 (2) (2008) 100–111.
[45] H. Jasper, W. Penfield, Electrocorticograms in man: effect of voluntary movement upon the electrical activity of the precentral gyrus, Eur. Arch. Psychiatry Clin. Neurosci. 183 (1949) 163–174.
[46] A. Teolis, Computational Signal Processing with Wavelets, Applied and Numerical Harmonic Analysis, Birkhauser, 1998.