CrossMark

ORIGINAL ARTICLE

Spectral Subtraction Denoising Preprocessing Block to Improve Slow Cortical Potential Based Brain–Computer Interface

Meena M. Makary[1,2] · Hani M. Bu-Omer[2] · Ramy S. Soliman[1] · Kyungmo Park[1] · Yasser M. Kadah[3]

Abstract The measured electroencephalographic (EEG) signals used in brain–computer interface (BCI) are usually contaminated with several additive random and physiological noise components. Even though several preprocessing strategies were proposed in the literature to extract useful EEG signal components for further processing and analysis, their performance is yet to meet practical needs to boost BCI applications performance. Of particular interest among those methods are the ones addressing random noise suppression that could help enhance the performance of low-cost BCI systems/headsets. A preprocessing block for signal denoising of slow cortical potential (SCP) that provides efficient noise removal and better classification accuracy is presented. This method is a variant of the spectral subtraction signal denoising whereby the noise power spectrum is estimated and removed adaptively. This method is applied to each EEG channel separately thus providing a preprocessing block that could be integrated with the existing spatial/temporal preprocessing methods. The impact of this method on classification accuracy is studied and compared to the conventional wavelet shrinkage method. The proposed method is verified using experimental data from the BCI competition II data set whereby a performance boost was obtained with the new preprocessing block. The improvement was quantitatively assessed using mean square error and mean absolute error measures and was also shown to be statistically significant. Moreover, spectral subtraction denoising is shown to have less computational complexity than wavelet shrinkage based methods. The new preprocessing block for SCP-based BCI signals based on spectral subtraction provides significant improvement in performance when added and offers an adaptive yet less computationally expensive alternative to existing methods such as wavelet shrinkage based denoising method. Also, given its independent channel processing, it shows potential for seamless integration into conventional processing chain for different BCI applications.

Keywords Brain–computer interface · Slow cortical potential · Signal denoising · Wavelet shrinkage · Spectral subtraction

1 Introduction

The brain–computer interface (BCI) technique offers alternative method of communication for patients with impaired motor abilities such as in amyotrophic lateral sclerosis (ALS) [1]. BCI collects and interprets multichannel brain activity data of the patient to directly map brain activity into specific tasks for assistive devices. Different techniques such as magnetoencephalography (MEG), electroencephalography (EEG), and functional magnetic resonance imaging (fMRI) can be utilized to capture the brain functional activity information at different spatial regions of the brain. These methods provide time course signals of brain activity from different locations at inherently different temporal and spatial resolutions. For example, EEG measures only a few electrical

✉ Meena M. Makary
mmakary@khu.ac.kr

1 Department of Biomedical Engineering, Kyung Hee University, Yongin, Gyeonggi 17104, South Korea

2 Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt

3 Electrical and Computer Engineering Department, King Abdulaziz University, Jeddah 21589, Saudi Arabia

🙋 Springer

signals from the surface of the scalp to represent the activity of the whole brain and hence offers poor spatial localization and high temporal resolution. On the other hand, fMRI offers much higher spatial resolution of the order of a few millimeters while suffering from a much poorer temporal resolution of its time course signals [2]. Even for BCI systems based on the same technology such as EEG, systems also range from simple devices with relatively inexpensive measuring electrodes connected to small processing unit that offer poor quality of signals while allowing mobility such as EEG headsets, to larger, expensive clinical EEG systems that provide higher quality of signals but would require a special setup that does not allow mobility. This demonstrates a clear trade-off between cost and mobility on one side and the acquired signal quality on the other.

EEG-based BCI systems offer a good choice with respect to cost and simplicity but the quality of collected signals is its major limitation. Presently, there are two broad types of BCI systems in terms of the type of EEG signals involved. The first is based on evoked EEG signals produced by external or internal neural stimulation of task-related brain areas such as steady-state visual evoked potential (SSVEP) and P300 [3–5]. The second type of systems is based on spontaneous EEG signals which are produced by human specific thoughts such as event-related desynchronization/event-related synchronization (ERD/ERS), EEG rhythmic waves (e.g., α, β, and γ waves), and slow cortical potentials (SCP) [6–8]. For example, SCP signals are voluntarily generated slow non-movement potential changes. They represent the cortical polarization changes lasting from 300 ms up to a few seconds of the subject's EEG signal [9]. SCP-based BCI usually targets to provide a binary decision that depends on the ability of the participant to voluntarily shift his/her SCP [10]. More recently, research studies indicate the possibility to differentiate several mental tasks using EEG signals [11]. So, the second type of BCI systems offers a simpler hardware implementation that does not include an integrated source of stimulation, its data processing is more challenging due to the difficulty to determine the start of the relevant brain activity given its spontaneous nature. So, any improvement in the signal quality for such type of systems would directly help improve their performance.

The problem of improving EEG signal quality has been addressed by many research studies aiming to suppress the additive random noise to improve the overall signal-to-noise ratio (SNR). This denoising procedure can be achieved using techniques in either spatial or temporal domains. Spatial domain methods include simple techniques such as local spatial averaging and extend to more elaborate forms of blind source separation (BSS) techniques as independent component analysis (ICA). In such methods, data from different channels are utilized to separate the true response signal from superimposed independent noise components [12–16]. For example, a study proposed an automatic technique based on BSS for EEG noise elimination and artifact removal (including ocular, high frequency muscle and ECG artifacts) [16]. In another study, a spectral signal space projection technique was proposed in which noise spatial patterns at particular frequencies are removed by applying time–frequency specific spatial projectors to the noisy time–frequency transformed data [12]. Alternatively, temporal domain methods utilize similarities within a single channel signal to identify then suppress the noise components in that signal. Such methods include simple averaging of consecutive epochs or linear filters such as band-pass filters [17, 18] and extend to more sophisticated time–frequency techniques based on wavelet shrinkage [19–23]. Hybrid techniques between temporal and spatial domain methods were also introduced to take advantage of both available channels and redundant signal epochs [24]. Even though present EEG denoising techniques offer promising results, there is still a need to further address their limitations. For instance, the need for large number of channels in spatial domain techniques that would increase the cost of the system. Furthermore, more expensive digital processing would be required for integrating computationally-demanding denoising methods into the BCI preprocessing chain. Such examples also lead to undesirable increase in power consumption that severely limit the practicality of mobile BCI system implementations. Therefore, it would be beneficial to propose a denoising method that would work with any number of electrodes and improve the BCI system performance beyond the existing methods at a reasonable computational complexity.

The aim of this study is to propose an EEG signal denoising strategy that improves BCI performance at a modest computational cost. The proposed approach is based on the spectral subtraction technique, which is an effective preprocessing framework originally developed for speech signals and later adopted to functional magnetic resonance imaging (fMRI) signal denoising [25], and P300-based BCI [2]. This method does not impose any constraints on the model for the true signal component while adaptively estimates the additive white Gaussian noise model parameters. This method is applied to experimental data from BCI competition II [26]. We follow the same steps of signal processing, feature vector formation and classification steps of the study in [27] since it offers the best achieved classification accuracy for this data set, and study the impact of adding the new method. The amount of removed noise, as well as the classification accuracy were quantitatively compared with conventional wavelet shrinkage denoising. The computational

complexity of the proposed spectral subtraction technique is also discussed and compared with the complexity of wavelet denoising method.

2 Methodology

Here, the mathematical basis of spectral subtraction denoising method and the applied modifications to meet the requirements of SCP BCI application are presented. Considering the conventional model of additive noise, temporal EEG signal can be given as the summation form of three components; namely: the true brain activity signal component, a baseline fluctuation component caused by physiological noise or instrumentation drift, and a random white Gaussian noise component [25]. The true response signal component from brain activity is the signal of interest in this mixture. On the other hand, the baseline fluctuation component will be assumed to be deterministic yet unknown component that is out of scope of this study (i.e., assumed to be effectively removed by existing preprocessing procedures [27, 28]). The random noise component consists of two portions: signals generated by neighboring brain areas that are not related to the true signal of the target BCI sought and thermal noise signal due to the data acquisition system electronics. While the latter is widely recognized as Gaussian white noise, the former can also be assumed to be so using central limit theorem applied to the summation of many signals known to have random patterns of activation. Taken together, we construct a model of additive noise in which the acquired EEG data is modeled as the summation of a deterministic component $d(t)$ (consisting of both true response EEG signal plus residual baseline/low frequency wander) and a random noise component $n(t)$ assumed to come from independent white Gaussian noise model. In a mathematical form, this model can be formulated as:

$$s(t) = d(t) + n(t). \tag{1}$$

Given the independence of both components of the model, the power spectrum of the signal can be obtained as:

$$P_{ss}(\omega) = P_{dd}(\omega) + P_{nn}(\omega). \tag{2}$$

Hence, given an estimate of the noise power spectrum, the deterministic component power spectrum can be calculated by subtracting the power spectrum of acquired signal and the power spectrum of an estimated random noise [25] such that:

$$P_{dd}(\omega) = P_{ss}(\omega) - P_{nn}(\omega). \tag{3}$$

Subsequently, the frequency domain magnitude of the deterministic signal can be given by taking its power spectrum square root. In order to fully estimate the time domain of the deterministic signal by inverse-transform of the frequency domain form, the phase part must be restored. Several methods can be followed to do that. One way is to calculate an estimate from the Fourier transform phase of the measure signal $S(\omega)$ such that the estimated deterministic signal spectrum $S_d(\omega)$ takes the form [25]:

$$S_d(\omega) = \sqrt{P_{dd}(\omega)} * e^{-(jPhase(S(\omega)))}. \tag{4}$$

By taking the real part of the inverse Fourier transform for $S_d(\omega)$, we can compute the denoised deterministic signal $s_d(t)$. Figure 1 illustrates a block diagram of the original spectral subtraction technique. Although, this method provided effective denoising for event-related fMRI time courses, there are two problems that exist for EEG signals in SCP-based BCI applications. The first problem is the method of estimating the phase component for the denoised signal from its original version. Because the phase component has important information just as the magnitude component, ignoring noise in this component is bound to limit the effectiveness of the denoising procedure. This issue was reported as a concern when applying this method to the fMRI time courses. The other problem is the noted discontinuity artifacts that come about when applying the discrete Fourier transform (DFT) that assumes periodic extensions of the limited record length used. Differences of EEG signal values between the initial and final time points induce a step in the time domain signal causing ringing at both ends of the signal. This is critical difference between EEG data and fMRI data since that artifact is much less severe in case of fMRI time courses due to their much higher baseline. Such jumps between the first and last time points in the processed EEG signals produce wrong high frequency components while computing the power spectrum. This results in artifacts that vary randomly depending on the amount of such jumps. To provide artifact-free denoising and to solve the above mentioned problem we proposed a modified formula of the previously developed spectral subtraction technique [2] in which we construct an even-symmetric signal from the original EEG signal (by obtaining its time inversed one and concatenating it with the original) before estimating the power spectrum by DFT. This offers two modifications to solve the above two problems in the previous denoising method. First, the even-symmetric EEG signal will have a real-valued spectrum with either zero phase (for the positive frequency amplitudes) or phase equal to $\prod$ (for negative frequency amplitudes) and hence eliminates the problematic process of phase estimation altogether. Second, it eliminates the edge artifacts by ensuring that the continuity is preserved at both boundaries of the EEG signal. Figure 2 shows the modified block diagram of the proposed spectral subtraction denoising.

Fig. 1 Block Diagram of the original spectral subtraction denoising technique. Block diagram of the original spectral subtraction denoising method that uses the original data to extract the noise model to calculate its power spectrum and eliminate it from the spectrum of the original data
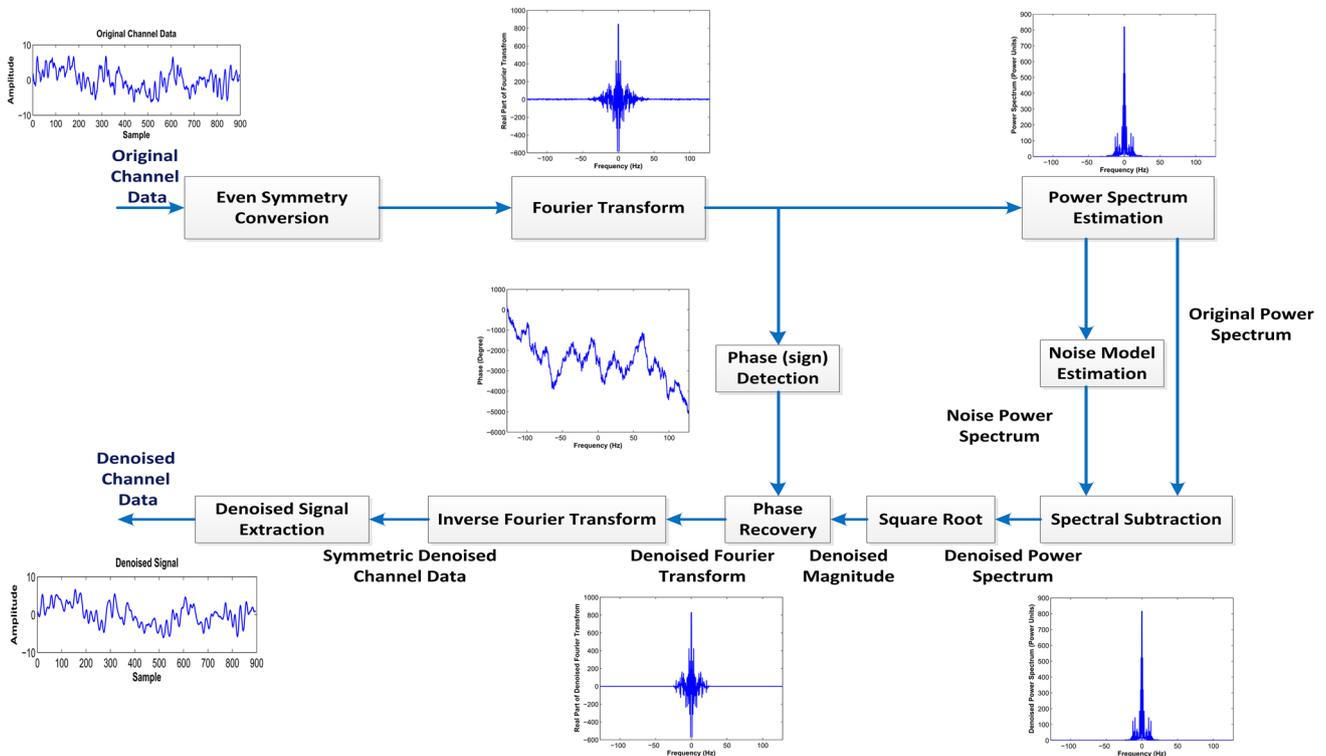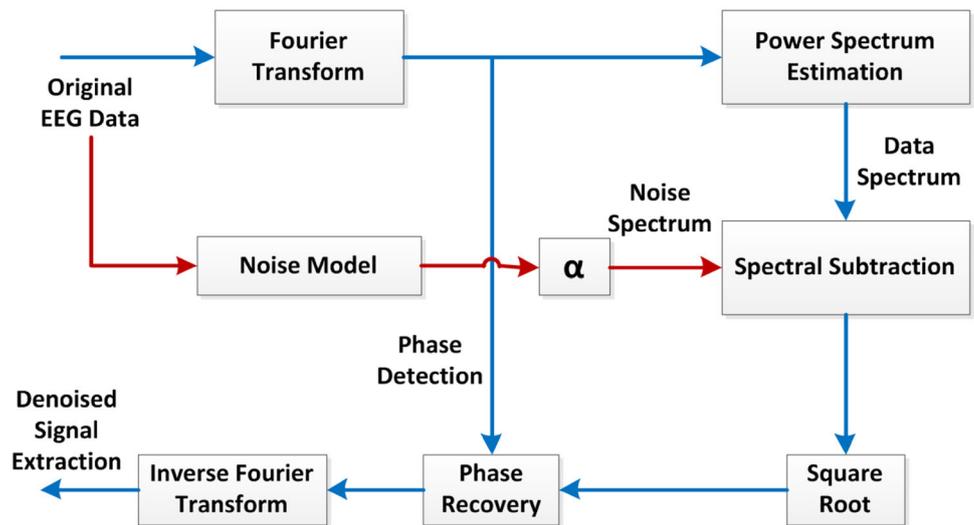
Fig. 2 Improved spectral subtraction denoising block diagram. Improved spectral subtraction denoising block diagram where a symmetric signal is constructed by concatenating the signal and its mirror (time-inversed version). This symmetric signal is the input instead of the original data in the conventional spectral subtraction technique. This helps to avoid artifacts coming from the discrepancy of signal levels at the ends by setting the phase of the signal to be zero

3 Power Spectrum Estimation of EEG Signal Noise

To implement the above denoising method as a preprocessing step for SCP signals, it is required to adaptively estimate the noise power spectrum. Given that the noise model was assumed as Gaussian white noise, it has a constant power spectrum over all frequency that is directly proportional to the variance of the noise. Therefore, the estimation of a single parameter would be sufficient to obtain the noise power spectrum. For adaptive denoising, the variance estimation of the noise is done from the raw EEG signals noting that such signals are usually acquired using a much higher sampling frequency than the known

frequency range of EEG signals. In particular, it can be assumed that the raw signal power spectra contain only noise components in their high frequency segments above the EEG frequency range. Consequently, taking the average of such parts of the raw signal power spectrum can provide a reliable estimate of the noise power spectrum as illustrated in Fig. 3. It should be noted that the use of average to estimate the noise power spectrum level from many points in the power spectrum is valid because at each point the power spectrum itself can be shown to be unbiased (i.e. with mean equals to the true value) and consistent (as the number of points goes to infinity, the variance decrease uniformly to zero) random variable. The average of these points is used to improve the estimation of the parameters of their underlying random processes because the magnitudes of these points are known to be independent and identically distributed. This noise level estimation and denoising procedure is performed for each EEG channel data separately to account for variability among different analog front-ends for different channels.

4 Implementation Steps

The detailed implementation steps of the modified spectral subtraction denoising are given as follows:

1. Reading the SCP EEG channel data $s(t)$ one at a time and constructing its symmetric form by simply concatenating the signal with its time-reflected version $s(-t)$.
2. Computing the discrete Fourier transform (DFT) of this symmetric raw SCP channel data.
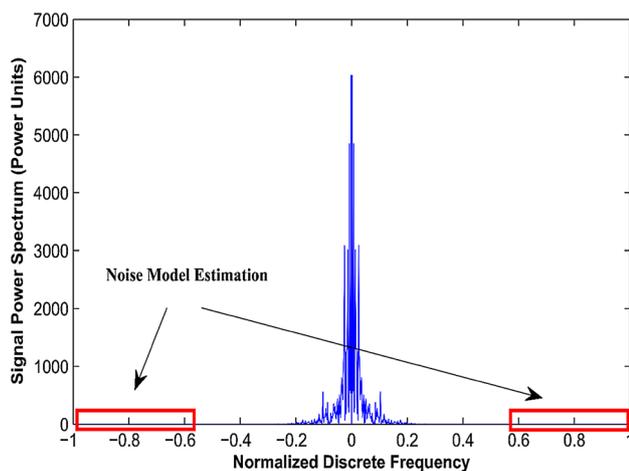
Fig. 3 Noise power spectrum estimation. Noise model estimation by calculating its power spectrum from the side parts of the original signal power spectrum on both frequency ends which is free of true components of the signal when averaging these areas

3. Obtaining the periodogram estimate of the power spectrum by squaring the magnitude of this obtained DFT.
4. Estimating the noise level by calculating the average of the power spectrum values in the 20% upper frequency segments of the frequency range that contain no true signal components as shown in Fig. 3.
5. Calculating the denoised signal power spectrum using Eq. (3) and clipping any negative value results from this subtraction to zero.
6. Estimating the denoised signal discrete Fourier spectrum as the square root of the estimated denoised signal power spectrum multiplied by the sign of the original discrete Fourier transform to recover the phase.
7. Estimating the denoised signal as the causal part of the inverse discrete Fourier transform of the estimated denoised signal discrete Fourier spectrum.

5 Experimental Verification

5.1 Description of the Data Set

The experimental verification of the proposed framework was performed using the data set from BCI competition II. This data set was collected from six healthy subjects (three males and three females; 22–35 years of age). Participants were instructed to perform two mental tasks to move a cursor up and down while their SCP signals were recorded. Each trial lasted for a total of 6 s divided into 1 s rest period, 1.5 s cue presentation period, and 3.5 s feedback period. The cue was a visual stimulus/target that appeared at the bottom or the top of a screen. Data were collected in the feedback period (3.5 s) at a sampling rate of 256 sample/s to give 896 samples for every trial per channel. The feedback was presented as a cursor and the SCP level (Cz-Mastoids) was indicated by its vertical position in the screen. Each participant completed 300 trials. Data were acquired using PsyLab EEG8 amplifier, analog to digital converter (PCIM-DAS1602/16 bit, Computer Boards), the range of amplitude was ± 1000 μV and at a sampling rate of 256 Samples/s. The positions of the electrodes followed the 10/20 system: CH01: A1-Cz (A1 at left mastoid), CH02: A2-Cz (A2 at right mastoid), CH03: 2 cm frontal of C3, CH04: 2 cm parietal of C3, CH05: 2 cm frontal of C4, CH06: 2 cm parietal of C4. The trials were divided into a training data set (consisting of 268 trials) and testing data set (consisting of 293 trials). Cue "0" and Cue "1" were used to label the two mental states in the training set for setting the classification algorithm parameters whose classification accuracy was evaluated on the testing data set.

5.2 Formation of the Feature Vector

In this work, we follow the same steps of processing, formation of feature vector, and classification as reported in [27], which presented the best achieved accuracy for this data set, in order to allow direct comparison and assessment of the impact of adding the spectral subtraction denoising block. The feature vector was computed based on the results of a wavelet package decomposition (WPD) step, which is a procedure derived from the conventional wavelet decomposition (WD) whereby samples of a discrete-time signal are passed through quadrature mirror filters that include multiple bases [29]. WD decomposes the input signal into two complementary and orthonormal subspaces, say V and W (containing low and high frequency information, respectively). Then, WPD continues splitting these two resultant subspaces into subsequent low and high frequency information forming a tree of wavelet packets. As illustrated in Fig. 4, $U_{j,n}$ is the nth subspace of wavelet decomposition at j^{th} level, and its orthonormal complementary basis is $U_{j,k}^n(t) = 2^{\frac{-j}{2}} u^n (2^{-j}t - k)$, where k is a shift. This could be satisfied by following equations:

$$u_{j,0}^n(t) = \sum_k h_0(k) u_{j-1,k}^j, \text{ when n is even.} \quad (5)$$

$$u_{j,0}^n(t) = \sum_k h_1(k) u_{j-1,k}^j, \text{ when n is odd.} \quad (6)$$

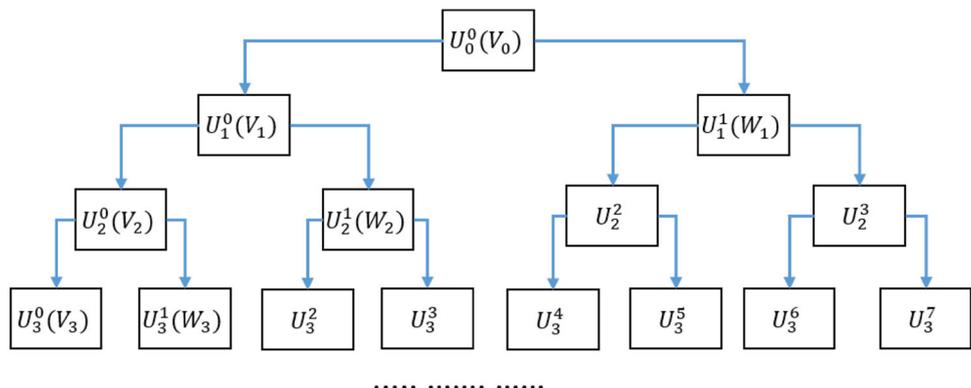where the quadruple mirror filters are $h_0(k), h_1(k)$ and related by:

$$h_1(k) = -1^{1-k} h_0(1-k). \quad (7)$$

Finally, the WPD coefficients at jth level and kth sample could be calculated as:

$$d_j^n = \sum_m h_0(m - 2k) d_{j-1}^{\frac{n}{2}}(m), \text{when n is even.} \quad (8)$$

$$d_j^n = \sum_m h_1(m - 2k) d_{j-1}^{\frac{n-1}{2}}(m), \text{when n is odd.} \quad (9)$$

The feature vector was formed from two parts: the sub-band average and sub-band energy coefficients which are described as follows:

A. Sub-band mean coefficients
    The first part of the feature vector is the sub-band means calculated from wavelet packet decomposition. The vector of sub-band means at the jth level can be calculated from [27]:

$$Mean_{j,n} = \mu_{j,n} = \frac{2^N}{2^j} \sum_k d_j^n(k). \quad (10)$$

The frequency ranges of WPD of the signal in the jth level can be obtained from:

$$\left\{ \left[0, \frac{f_s}{2^{(j+1)}}\right]; \left[\frac{f_s}{2^{(j+1)}}, \frac{2f_s}{2^{(j+1)}}\right]; \ldots ; \left[\frac{(2^j - 1)f_s}{2^{(j+1)}}, \frac{f_s}{2}\right] \right\}. \quad (11)$$

After performing WPD to the 6th level, as recommended by [27], and given that the sampling frequency for this data set is $f_s = 256$, then the frequency ranges of sub-bands at this level will be $\{[0 - 2]; [2 - 4], \ldots; [126 - 128]\}$. Given that the useful frequency range of EEG signals is between 0 and 50 Hz, which will be within the first 25 sub-bands, 25 features of the mean vector from each channel can be concatenated to form a total feature vector of 150 values (25 feature times 6 channels), $\mu = [\mu_1, \mu_2, \ldots, \mu_{150}]$.

B. Sub-band energy coefficients
    The second part of the final feature vector is the energy of sub-bands. WPD decomposes the energy of the signal into several time–frequency subdivisions. The integration of WPD sub-band squared amplitude is equivalent to the sub-band energy. That is,

Fig. 4 The WPD tree structures. The WPD decomposes the input signal into two complementary and orthonormal subspaces, say V and W (containing low and high frequency information, respectively). Then, it continues splitting these two resulting subspaces into subsequent low and high frequency information forming wavelet packet tree

$$Energy_{j,n} = E_{j,n} = \sum_k \left(d_j^n(k) \right)^2. \tag{12}$$

Similar to the mean vector, at the 6th level and within the useful frequency range 0–50 Hz we compute 25 values for each of the available six channel. Thus, the total sub-band energy vector for all channels is $E = [E_1, E_2, E_3, \ldots, E_{150}]$.

The significance of these features is because they provide information about the true response signal as extracted from different frequency sub-bands resulted from wavelet packet decomposition. The logic for choosing the mean and energy of wavelet decomposition sub-bands as recommended in [27] was to combine information from both frequency and time domains of the EEG signal in attempt to provide a more complete feature set.

Then, Fisher criterion was used for dimensionality reduction of the initial feature vector of sub-band means and energies. Among the energy and mean vectors, the most appropriate features were chosen based on fisher distance selection. $F(w)$, the Fisher criterion function [27, 30], satisfies the following equation:

$$F(w) = \frac{w^T S_B w}{w^T S_I w}, \tag{13}$$

where S_I is the within-class scatter matrix, S_B is between-class scatter matrix, and w is the fisher weight vector that maximize $F(w)$. One way to assess the separability of two classes using the value of $F(w)$ whereby the greater $F(w)$ value, the higher separability is achieved for the training data set. To assess the separability of the initial feature vector, we used Fisher distance criterion (J) [27, 30]:

$$J = (S_w^{-1} S_b)^T, \tag{14}$$

where S_w is the within-classes dispersion matrix, and S_b is the between-classes dispersion matrix. For each feature, the greater the value of separability J, the more promising this feature is in separating the data. After computing the separability distance J for every element in the vector of mean (μ) features, we chose the most relevant features with highest values of J (for example, the fist "d" features) for separating the data. Then, the final mean features vector is $\mu' = [\mu_1', \mu_2', \ldots, \mu_d']$. Similarly, for the energy vector E features, by choosing the first "l" elements with highest J values, the final energy features vector is $E' = [E_1', E_2', E_3', \ldots, E_l']$. The final features, which will be fed to the classifier, is $F = [\mu', E']$. The procedure of creating the final feature vector is shown in Fig. 5. As to the classifier, we used Probabilistic Neural Network (PNN)

with four layers, as suggested by [27], which can potentially fit the nonlinear characteristic of EEG signal and assuming the feature vectors to be non-linearly separable [31, 32].

We compared the proposed method results with wavelet-shrinkage denoising results when added as a preprocessing block to the above processing chain. The overall BCI performance and the amount of removed noise from the original signal in each case were quantitatively assessed. For wavelet denoising, we applied standard wavelet shrinkage denoising using MATLAB (MATLAB, The MathWorks, MA, USA) with choosing "Coiflet-3" as the basic wavelet as recommended by [33].
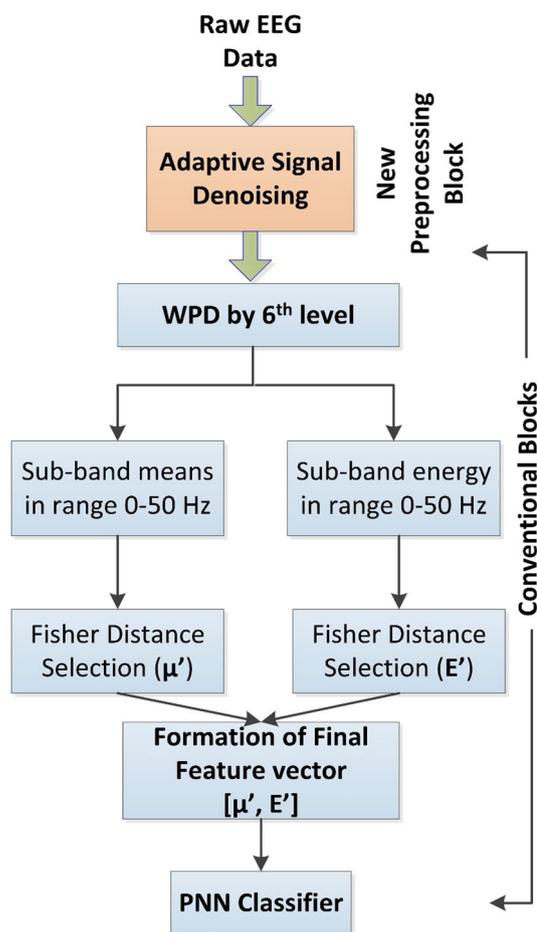
Fig. 5 Feature vector formation flow chart after adding new preprocessing block of denoising. Flow chart of feature vector creation after adding our proposed adaptive denoising block. The procedure starts with performing WPD to the 6th level and then sub-band means and energies were calculated in the range 0–50 Hz. Fisher distance was applied for dimensionality reduction and choosing the appropriate features. The final feature vector by concatenating the mean and energy final vectors, then fed to the PNN classifier
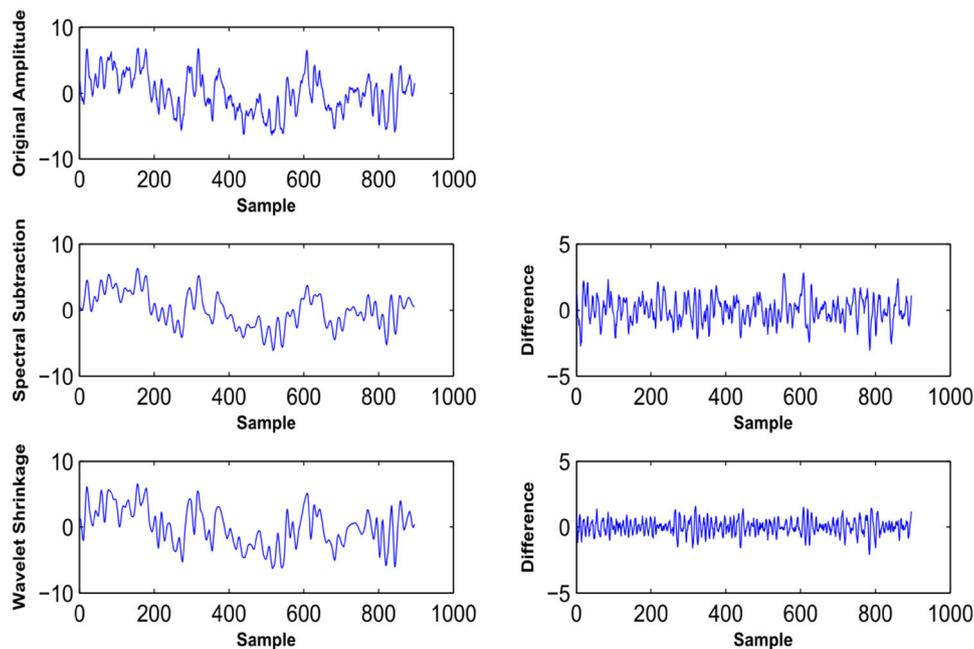
Fig. 6 Comparison between spectral subtraction denoising results and the widely used wavelet shrinkage denoising. Illustration of comparison results between spectral subtraction and wavelet shrinkage denoising as contrasted with the original signal. The *left column* shows the original signal in the top, the spectral subtraction denoised signal in the *middle* and the *wavelet* shrinkage denoised signal in the bottom. The *right column* shows the difference between the original signal and the denoised signal illustrating their random nature and showing the relative removed content of the signal between spectral subtraction (*top*) and wavelet shrinkage (*bottom*). The removed content looks higher in spectral subtraction as compared to wavelet shrinkage for the same EEG signal

5.3 Statistical Comparison between Spectral Subtraction and Wavelet Shrinkage Denoising

The performance of wavelet and spectral subtraction denoising was measured by comparing the original channel signal $s(t)$ and the denoised signal $d(t)$ from each method in time and frequency domains as recommended by [34]. In time domain, mean square error (MSE) is defined as [34]:

$$MSE = \sum_{t=1}^{T} [s(t) - d(t)]^2 / T, \qquad (15)$$

where T is the number of time samples in each channel. While in the frequency domain, mean absolute error (MAE) is defined as [34]:

$$MAE = \sum_{\omega=1}^{W} |P_{ss}(\omega) - P_{dd}(\omega)| / W, \qquad (16)$$

where P_{ss} and P_{dd} are the power spectrum of $s(t)$ and $d(t)$ respectively, and W is the number of frequency samples. In order to show that the difference between the two methods is statistically significant, we performed paired t test (SPSS v. 10.0.7, Chicago, IL, USA) to compare MSE and MAE from both methods against a significance level of p = 0.01.

6 Results and Discussion

Figure 6 shows a comparison of signal denoising using spectral subtraction and conventional wavelet shrinkage denoising and the raw SCP signal to qualitatively illustrate signal denoising and the amount of removed noise. The left column shows the original signal at the top, the denoised signal using spectral subtraction in the middle, while the denoised signal using wavelet shrinkage method at the bottom. The right column shows the difference between the original signal and the denoised signal illustrating their random nature and showing the relative eliminated noise content of the signal between spectral denoising (top) and wavelet shrinkage (bottom). The eliminated noise content looks higher in spectral subtraction as compared to wavelet shrinkage for the same EEG signal.
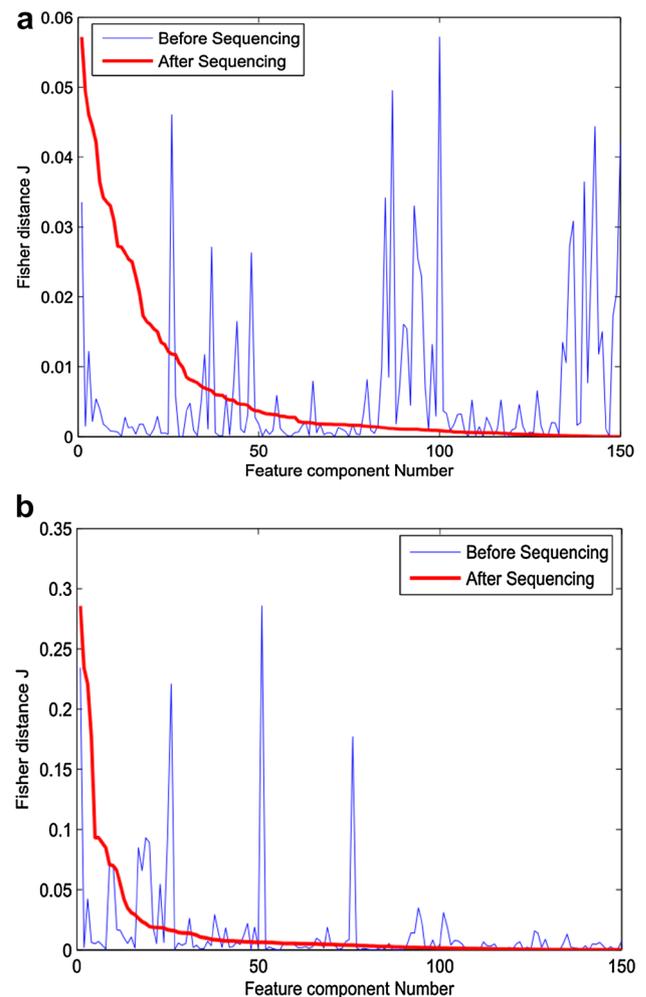
Also, the statistical comparison between spectral subtraction and wavelet shrinkage using paired t-test revealed statistically significant difference when comparing MSE (Spectral Subtraction: 4.6 ± 8.06, mean $\pm$ STD; Wavelet Shrinkage: 0.88 ± 0.37, p = 1.57e−36) and MAE (Spectral Subtraction: $7.86 \times 10^3 \pm 1.17 \times 10^3$, Wavelet Shrinkage: $1.68 \times 10^3 \pm 1.23 \times 10^3$, p = 1.36e−52). This shows that the difference in estimating the amount of noise
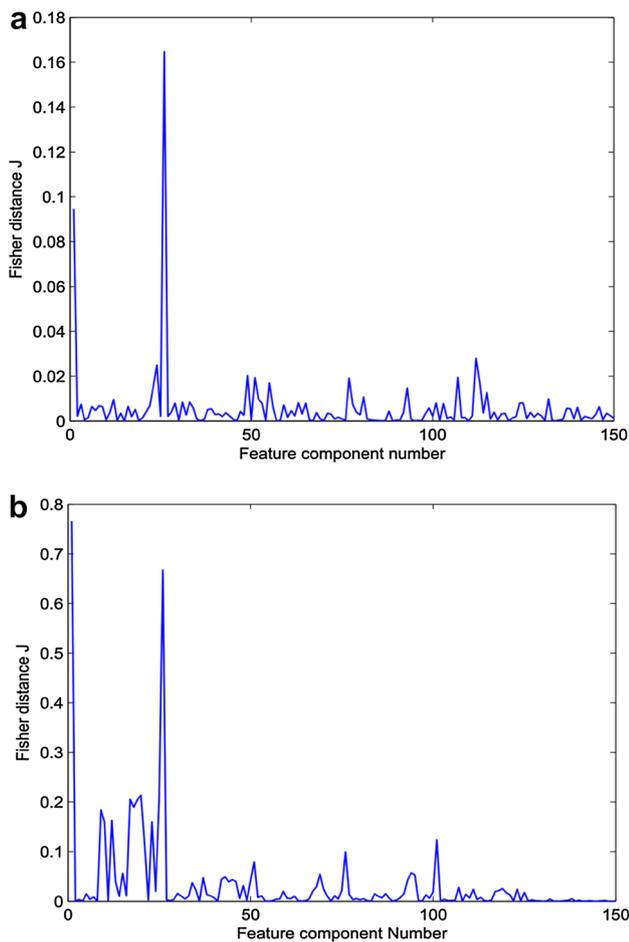
Fig. 7 Fisher distance separability for each element in the sub-band mean feature vector "μ" before and after adding spectral subtraction signal denoising block. **a** Each feature's Fisher distance separability in the sub-band mean vector without adding the spectral subtraction preprocessing block. **b** Each feature's Fisher distance separability in the sub-band mean vector after adding the spectral subtraction preprocessing block. With adding the spectral subtraction denoising block, the amplitude of fisher distance increases and more features appeared which could be new candidates for the final feature vector

Fig. 8 Fisher distance separability for each element in the sub-band energy feature vector "E" before and after adding spectral subtraction signal denoising block. **a** Each feature's Fisher distance separability in the sub-band energy vector without adding the spectral subtraction preprocessing block. **b** Each feature's Fisher distance separability in the sub-band energy vector after adding the spectral subtraction preprocessing block. With adding the spectral subtraction denoising block, the amplitude of fisher distance increases and more features appeared which could be new candidates for the final feature vector. The *red line* shows the features after sequencing by the higher J values

removed using each denoising method cannot be attributed to chance and reflects real increase in eliminating random noise component by the spectral subtraction method over the wavelet shrinkage method.

After forming the initial feature vector of sub-band means $\mu = [\mu_1, \mu_2, \ldots, \mu_{150}]$, and sub-band energies $E = [E_1, E_2, E_3, \ldots, E_{150}]$, we applied Fisher distance criterion for dimensionality reduction. The Fisher separability (J) for each element of the sub-band mean μ and sub-band energy E vectors is shown in Fig. 7(a) and Fig. 8(a) respectively, without adding spectral subtraction as a preprocessing block. As shown in Fig. 7(a), the mean vector have only two peak points which correspond to the most promising feature for separating the training data set, i.e. $\mu' = [\mu_1', \mu_2']$. Similarly, the energy vector as shown in

Fig. 8(a) contains several peak points indicating the ability of this feature to separate the training data set. To reproduce the exact results as [27], we chose the first 15 features (correspond to greater values of J), i.e. $E' = [E_1', E_2', E_3', \ldots, E_{15}']$. By concatenating the final mean and energy vectors, the final feature vector will be $F = [\mu', E']$ with 17 features. Note that the classifier accuracy will not be increased if more features are used in this case. Using the training data set of 268 trials to train the PNN with F, the resulted accuracy of the test data set was 90.8% which is higher than the best achieved results (88.7%) in BCI competition II by 2.1%. Interestingly, after

Table 1 Comparison of PNN classification accuracy with and without adding adaptive denoising preprocessing block

Method	Classification accuracy (%)	% Improvement
Best achieved result of BCI competition II	88.7	–
Tung et al. [27] results (without denoising block)	90.8	2.37
Results after signal denoising by wavelet shrinkage	91.1	2.71
Results after signal denoising by wavelet shrinkage and adding 2 more features	91.8	3.49
Results after signal denoising by spectral subtraction	91.4	3.04
Results after signal denoising by spectral subtraction block and adding 6 more features	94.1	6.09

adding the spectral subtraction denoising to all channels for both training and testing data sets and repeat the exact same steps as mentioned previously, the separability distance J of the mean and energy vectors as well as the classification accuracy changed. As shown in Figs. 7(b) and 8(b), after adding the denoising preprocessing block, the amplitude of Fisher distance separability J for both mean and energy vectors relatively increased in value and new peak points appeared which could be new candidates for the final feature vector. Using the same feature vector with 17 features, the classification accuracy increased to 91.4% by the effect of spectral subtraction denoising only. Moreover, with increasing the feature vector by adding 6 more features from the features that became significant because of the denoising step, the classification accuracy increased to 94.1% with insignificant increase in the processing time. Adding more features did not affect the classification results also in this case. In comparison, the conventional wavelet shrinkage denoising increased the classification accuracy to 91.1% with the effect of denoising. After adding two more features from the energy vector that became significant after wavelet shrinkage denoising, the classification accuracy increased to 91.8%. Comparison of all results is presented in the first column of Table 1. The second column of Table 1 shows the percentage of improvement over the best achieved results of this dataset as follows:

$$
\% \, Improvement \\
= \frac{Resulted \, Accuracy - Best \, Acheived \, \mathrm{A}ccuracy}{Best \, Acheived \, Accuracy} \times 100\%
$$

(17)

These results suggest the potential of denoising as useful preprocessing step for BCI applications and indicate the performance improvement with the proposed spectral subtraction denoising over the conventional wavelet shrinkage method.

As for the computational complexity of spectral subtraction, assuming a data set of L channels and each

channel have N time points, then the spectral subtraction denoising method will have a computational complexity of $O(LNlog_2N)$. On the other hand, signal denoising by wavelet shrinkage method has a computational complexity that varies depending on the implementation method with minimum complexity of $O(LN^2)$ which is significantly higher by a factor of $N/log_2(N)$ than denoising by spectral subtraction. Therefore, spectral subtraction is well-suited for real time processing required for BCI applications.

It should be noted that the existing frequency selective filters that amount to time domain convolution cause remaining noise in the output filtered EEG signal at different time points to be correlated. Therefore, one theoretical advantage of the proposed method is that it preserves the independence among the processed time points random components since no mixing of points via convolution is performed. Consequently, it is compatible with existing statistical analysis procedures where samples are required to be statistically independent. So, this shows the potential of seamless integration of spectral subtraction denoising as a standard step in the processing chain of BCI experiments to improve the SNR and hence increase the classification accuracy.

7 Conclusion

In this study, a denoising preprocessing block based on spectral subtraction for slow cortical potential signals was presented to improve the classification accuracy. Our denoising method was verified using BCI competition II experimental data and better performance was demonstrated. The performance and amount of removed noise were shown to be better than the wavelet shrinkage method. Moreover, the proposed method requires modest computational complexity that is suitable for real-time performance. The results of this study indicate the potential of adding the proposed method to the conventional preprocessing chain for different BCI applications.

Compliance with Ethical Standards

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

References

1. Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology*. doi:10.1016/S1388-2457(02)00057-3.

2. Alhaddad, M. J., Kamel, M. I., Makary, M. M., Hargas, H., & Kadah, Y. M. (2014). Spectral subtraction denoising preprocessing block to improve P300-based brain–computer interfacing. *Biomedical engineering online, 13*(1), 36. doi:10.1186/1475-925X-13-36.

3. Middendorf, M., McMillan, G., Calhoun, G., & Jones, K. S. (2000). Brain–computer interfaces based on the steady-state visual-evoked response. *IEEE Transactions on Rehabilitation Engineering, 8*(2), 211–214. doi:10.1109/86.847819.

4. Bayliss, J. D., & Ballard, D. H. (2000). Single trial P3 epoch recognition in a virtual environment. *Neurocomputing, 32–33,* 637–642. doi:10.1016/S0925-2312(00)00226-5.

5. Lin, K., Cinetto, A., Wang, Y., Chen, X., Gao, S., & Gao, X. (2016). An online hybrid BCI system based on SSVEP and EMG. *Journal of Neural Engineering, 13*(2), 26020. doi:10.1088/1741-2560/13/2/026020.

6. Avilés-Cruz, C., Villegas-Cortez, J., Ferreyra-Ramírez, A., & López, A. Z. (2016). *EEG pattern recognition: An efficient improvement combination of ERD/ERS/laterality features to create a self-paced BCI system* (pp. 231–240). New York: Springer. doi:10.1007/978-3-319-39393-3_23.

7. Dornhege, G., Blankertz, B., & Curio, G. (2003). Speeding up classification of multi-channel brain–computer interfaces: Common spatial patterns for slow cortical potentials. In *International IEEE/EMBS Conference on Neural Engineering, NER* (Vol. 2003-Janua, pp. 595–598). doi:10.1109/CNE.2003.1196898.

8. Mayer, K., Blume, F., Wyckoff, S. N., Brokmeier, L. L., & Strehl, U. (2016). Neurofeedback of slow cortical potentials as a treatment for adults with attention deficit-/hyperactivity disorder. *Clinical Neurophysiology, 127*(2), 1374–1386. doi:10.1016/j.clinph.2015.11.013.

9. N., Kübler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser, J., … Flor, H. (2000). The thought translation device (TTD) for completely paralyzed patients. *IEEE Transactions on Rehabilitation Engineering*, 8(2), 190–193. doi:10.1109/86.847812.

10. Hinterberger, T., Schmidt, S., Neumann, N., Mellinger, J., Blankertz, B., Curio, G., et al. (2004). Brain–computer communication and slow cortical potentials. *IEEE Transactions on Bio-Medical Engineering, 51*(6), 1011–1018. doi:10.1109/TBME.2004.827067.

11. Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors*. doi:10.3390/s120201211.

12. Ramírez, R. R., Kopell, B. H., Butson, C. R., Hiner, B. C., & Baillet, S. (2011). Spectral signal space projection algorithm for frequency domain MEG and EEG denoising, whitening, and source imaging. *NeuroImage, 56*(1), 78–92. doi:10.1016/j.neuroimage.2011.02.002.

13. de Cheveigné, A., & Simon, J. Z. (2008). Denoising based on spatial filtering. *Journal of Neuroscience Methods, 171*(2), 331–339. doi:10.1016/j.jneumeth.2008.03.015.

14. Vorobyov, S., & Cichocki, A. (2002). Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis. *Biological Cybernetics, 86*(4), 293–303. doi:10.1007/s00422-001-0298-6.

15. Geetha, G., & Geethalakshmi, S. N. (2012). Artifact removal from EEG using spatially constrained independent component analysis and wavelet denoising with Otsu's thresholding technique. In *Procedia Engineering* (Vol. 30, pp. 1064–1071). doi:10.1016/j.proeng.2012.01.964.

16. Romo Vázquez, R., Vélez-Pérez, H., Ranta, R., Louis Dorr, V., Maquin, D., & Maillard, L. (2012). Blind source separation, wavelet denoising and discriminant analysis for EEG artefacts and noise cancelling. *Biomedical Signal Processing and Control, 7*(4), 389–400. doi:10.1016/j.bspc.2011.06.005.

17. Mirghasemi, H., Shamsollahi, M. B., & Fazel-Rezai, R. (2006). Assessment of preprocessing on classifiers used in the P300 speller paradigm. In *Annual International Conference of the IEEE Engineering in Medicine and Biology—Proceedings* (pp. 1319–1322). doi:10.1109/IEMBS.2006.259520.

18. Hammon, P. S., & De Sa, V. R. (2007). Preprocessing and meta-classification for brain–computer interfaces. *IEEE Transactions on Biomedical Engineering, 54*(3), 518–525. doi:10.1109/TBME.2006.888833.

19. Quian Quiroga, R., & Garcia, H. (2003). Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology, 114*(2), 376–390. doi:10.1016/S1388-2457(02)00365-6.

20. Effern, A., Lehnertz, K., Grunwald, T., Fernández, G., David, P., & Elger, C. E. (2000). Time adaptive denoising of single trial event-related potentials in the wavelet domain. *Psychophysiology*, 37(6), 859–865. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11117466.

21. Sultan, H. (2010). Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison. *IEEE Signal Processing Letters, 17*(3), 237–240. doi:10.1109/LSP.2009.2037773.

22. Hammad, S., Corazzol, M., Kamavuako, E. N., & Jensen, W. (2012). Wavelet denoising and ANN/SVM decoding of a self-paced forelimb movement based on multi-unit intra-cortical signals in rats. In *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences* (pp. 990–994). IEEE. doi:10.1109/IECBES.2012.6498061.

23. Estrada, E., Nazeran, H., Sierra, G., Ebrahimi, F., & Setarehdan, S. K. (2011). Wavelet-based EEG denoising for automatic sleep stage classification. In *CONIELECOMP 2011—21st International Conference on Electronics Communications and Computers, Proceedings* (pp. 295–298). doi:10.1109/CONIELECOMP.2011.5749325.

24. Tu, Y., Huang, G., Hung, Y. S., Hu, L., Hu, Y., & Zhang, Z. (2013). Single-trial detection of visual evoked potentials by common spatial patterns and wavelet filtering for brain–computer interface. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2882–2885). IEEE. doi:10.1109/EMBC.2013.6610142.

25. Kadah, Y. M. (2004). Adaptive denoising of event-related functional magnetic resonance imaging data using spectral subtraction. *IEEE Transactions on Bio-Medical Engineering, 51*(11), 1944–1953. doi:10.1109/TBME.2004.831525.

26. BCI Competition II website. (n.d.). misc. Retrieved June 21, 2016, from http://www.bbci.de/competition/ii/.

27. Ting, W., Guo-zheng, Y., Bang-hua, Y., & Hong, S. (2008). EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement, 41*(6), 618–625. doi:10.1016/j.measurement.2007.07.007.

28. Hoffmann, U., Vesin, J.-M., Ebrahimi, T., & Diserens, K. (2008). An efficient P300-based brain–computer interface for disabled subjects. *Journal of Neuroscience Methods, 167*(1), 115–125. doi:10.1016/j.jneumeth.2007.03.005.

29. Xue, J. Z., Zhang, H., Zheng, C. X., & Yan, X. G. (2003). Wavelet packet transform for feature extraction of EEG during mental tasks. In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)* (pp. 360-363). IEEE. doi:10.1109/ICMLC.2003. 1264502.

30. Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification*. New York: John Wiley, Section. doi:10.1038/npp.2011.9.

31. Mao, K. Z., Tan, K. C., & Ser, W. (2000). Probabilistic neural-network structure determination for pattern classification. *IEEE Transactions on Neural Networks, 11*(4), 1009–1016. doi:10. 1109/72.857781.

32. Peters, B. O., Pfurtscheller, G., & Flyvbjerg, H. (1998). Mining multi-channel EEG for its information content: an ANN-based method for a brain–computer interface. *Neural networks*, *11*(7–8), 1429–1433. Retrieved from http://www.ncbi.nlm.nih. gov/pubmed/12662759.

33. Saavedra, C., & Bougrain, L. (2010). Wavelet denoising for p300 single-trial detection. In *Proceedings of the 5th French conference on computational neuroscience (Neurocomp'10)* (Vol. 10, pp. 227–231).

34. Peng, H., Hu, B., Shi, Q., Ratcliffe, M., Zhao, Q., Qi, Y., et al. (2013). Removal of ocular artifacts in EEG—An improved approach combining DWT and ANC for portable applications. *IEEE Journal of Biomedical and Health Informatics, 17*(3), 600–607. doi:10.1109/JBHI.2013.2253614.