Research Article

Real-Time EEG Signal Enhancement Using Canonical Correlation Analysis and Gaussian Mixture Clustering

Chin-Teng Lin [ID],[1,2] **Chih-Sheng Huang**,[2] **Wen-Yu Yang**,[2] **Avinash Kumar Singh** [ID],[1,2,3] **Chun-Hsiang Chuang**,[1,2] **and Yu-Kai Wang**[1]

[1]*School of Software, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia*
[2]*Brain Research Center, National Chiao Tung University, Hsinchu, Taiwan*
[3]*College of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan*

Correspondence should be addressed to Chin-Teng Lin; chin-teng.lin@uts.edu.au

Electroencephalogram (EEG) signals are usually contaminated with various artifacts, such as signal associated with muscle activity, eye movement, and body motion, which have a noncerebral origin. The amplitude of such artifacts is larger than that of the electrical activity of the brain, so they mask the cortical signals of interest, resulting in biased analysis and interpretation. Several blind source separation methods have been developed to remove artifacts from the EEG recordings. However, the iterative process for measuring separation within multichannel recordings is computationally intractable. Moreover, manually excluding the artifact components requires a time-consuming offline process. This work proposes a real-time artifact removal algorithm that is based on canonical correlation analysis (CCA), feature extraction, and the Gaussian mixture model (GMM) to improve the quality of EEG signals. The CCA was used to decompose EEG signals into components followed by feature extraction to extract representative features and GMM to cluster these features into groups to recognize and remove artifacts. The feasibility of the proposed algorithm was demonstrated by effectively removing artifacts caused by blinks, head/body movement, and chewing from EEG recordings while preserving the temporal and spectral characteristics of the signals that are important to cognitive research.

1. Introduction

Electroencephalography (EEG), which is the most convenient brain imaging tool that reveals electrical activity in the brain, has the most near-term potential for real-time applications in everyday environments [1]. However, EEG signals are contaminated with various artifacts, such as signal associated with muscle activity, eye movement, and body motion, which are not of cerebral origin. These artifacts appear as high-amplitude or high-frequency bursts of activity and may be confused with abnormalities, leading to misinterpretation [2]. Numerous signal-preprocessing methods have been proposed for eliminating artifacts that are generated by eye movements or muscle activity [3–10]. Low- and high-pass filters are commonly utilized to remove muscle artifacts and drift artifacts, respectively.

However, the spectral patterns of artifacts usually overlap those of brain signals of interest. Frequency filters not only remove artifacts but also suppress informative brain signatures [10]. Recently, advanced machine learning methods [3–10] were developed to deal with artifacts. Most of these methods have three main parts which are source separation, feature extraction, and classification. Concerning source separation, independent component analysis (ICA) [11–13], which yields maximally temporally independent signals from the EEG recording, is a powerful tool for separating brain activity from artifacts. However, the iterative process for measuring independence from EEG signals within multichannel recordings is computationally intractable [14]. Additionally, manually excluding the ICA components that are associated with artifacts is challenging and time-consuming.

Canonical correlation analysis (CCA) has been demonstrated to outperform ICA and frequency filters in eliminating muscle artifacts (electromyography, EMG) [8, 10, 15]. The CCA-based method exploits the fact that the autocorrelation of muscle activity is weaker than that of brain activity. The use of a correlation threshold enables the CCA to identify muscle activity automatically. The objective function of CCA has a closed-form solution that facilitates the real-time implementation of CCA for removing artifacts.

An efficient method for identifying artifacts is informative features, the frequency, spatial, and temporal domains of EEG signals [6]. These features are used jointly rather than independently. For example, spectral powers are not efficient for identifying ocular artifacts, because EEG signals and ocular activity exhibit similar spectral properties around the frequency band of brain activity of interest but temporal features, such as fractal dimension [16] and higher-order statistics [17], are more useful for identifying ocular artifacts.

Feature extraction is often followed by classification. Most classification systems for artifact removal use a supervised learning method [18, 19] that requires labeled data as input for training model. The process of distinguishing artifacts from nonartifacts by visual inspection can be time-consuming, and the result is based on previous experience of the inspector. Therefore, an effective method for labeling artifacts and nonartifacts automatically should be developed. This work proposes a real-time artifact removal algorithm that performs three main tasks—source separation using CCA, artifact feature extraction using spectral and temporal features, and classification based on Gaussian mixture model. To evaluate the performance of the proposed method, two standard visual-evoked tasks, a visual-evoked potential (VEP) task and a steady-state visually evoked potential (SSVEP) task, are performed.

2. Materials and Methods

2.1. Participants and Experiment. Eleven right-handed adults (seven males and four females, aged from 18 to 27 years) were recruited to participate in the study. None of the participants had a history of psychological disorders. Following a detailed explanation of the experimental procedure, all participants completed a consent form before participating. All subjects were required to wear a wired EEG cap with 62 Ag/AgCl electrodes, including 60 EEG electrodes and two reference electrodes (opposite lateral mastoids) (Figure 1(b)). The placement of the EEG electrodes was consistent with the modified international 10–20 system. The contact impedance between all electrodes and cortex was maintained below 5 kΩ. The EEG recordings were collected using a Scan SynAmps2 Express system (Compumedics Ltd., VIC, Australia). The EEG recordings were digitally sampled at 1 kHz with a 16-bit resolution. EEG signals were referenced to the opposite lateral mastoids.

All subjects were instructed to look at the center of a screen and follow the audible instructions to generate artifacts of various types purposefully. Subjects participated in four experimental runs; in one, no motion was performed,

and in the other three, common artifacts (blinking, chewing, and head rotation) were generated. Each run involved two sessions. Each session comprised three parts, which were instruction, stimulation, and resting (Figure 1(a)). In the instruction period, the speaker instructed the subjects to perform a particular motion when the flicker stimulus was presented. To prevent habituation, in the stimulus period, the flicker stimulus was presented for 10 s, and the subjects rested for 10 s after that (Figure 1(a)). A 60 s resting break after each session enabled the eye muscles to relax. The screen monitor (DELL U2311Hb, 23″, 60 Hz, 1920 × 1080) was used to present the visual stimuli. The distance between the subject's head and the screen was 30 cm (Figure 1(b)).

The flicker stimulus experiment was conducted in a shielded room to prevent any unwanted artifact from appearing in the EEG data. The two flickering stimuli had frequencies of 1 Hz and 15 Hz to induce visual-evoked potential (VEP) and steady-state visual-evoked potential (SSVEP), respectively.

2.2. Proposed Artifact Removal Algorithm. The proposed algorithm comprises three main parts, which are CCA, artifact feature extraction, and the Gaussian mixture model after preprocessing, which involves downsampling to 256 Hz and passband filtering of 0.1–60 Hz. Figure 2 shows the flowchart of proposed artifact removal algorithm.

2.2.1. Canonical Correlation Analysis and EEG Reconstruction. Let the observed EEG signals be $\mathbf{X}(t) = [\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_M(t)]^T$, $t = 1, 2, \ldots, N$, where N is the number of samples and M represents the number of EEG electrodes used to make the observations. The purpose of blind source separation (BSS) is performed to recover the sources $S(t) = [s_1(t), s_2(t), \ldots, s_M(t)]^T$ from only sensor observations $\mathbf{X}(t)$. In BSS, $\mathbf{X}(t) = [\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_M(t)]^T$ is the mixture of a set of unknown source signals $S(t)$, which is regarded as a linear combination,

$$\mathbf{X}(t) = \mathbf{A} \cdot S(t), \tag{1}$$

where $\mathbf{A}$ is the unknown mixing matrix. The unknown source signals $S(t)$ are derived by introducing the demixing matrix $\mathbf{W}$,

$$\widehat{S}(t) = \mathbf{W}\mathbf{X}(t), \tag{2}$$

where $\widehat{S}(t)$ approximates the unknown source signals in $S(t)$, and ideally, $\mathbf{W}$ is the inverse of the unknown mixing matrix $\mathbf{A}$.

Canonical correlation analysis (CCA) is one of the BSSs, which solves the problem by forcing the sources to be maximally autocorrelated and mutually uncorrelated; it has been extensively used to separate muscle artifacts from other EEG activity [8, 10]. In CCA, $\mathbf{Y}(t)$ is the instantly delayed version of the observed EEG signals $\mathbf{X}(t)$ such that $\mathbf{Y}(t) = \mathbf{X}(t-1)$, and the mean of each row of the matrix $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ is removed to derive two sets of basis vectors for $\mathbf{X}$ and $\mathbf{Y}$. Suppose two canonical
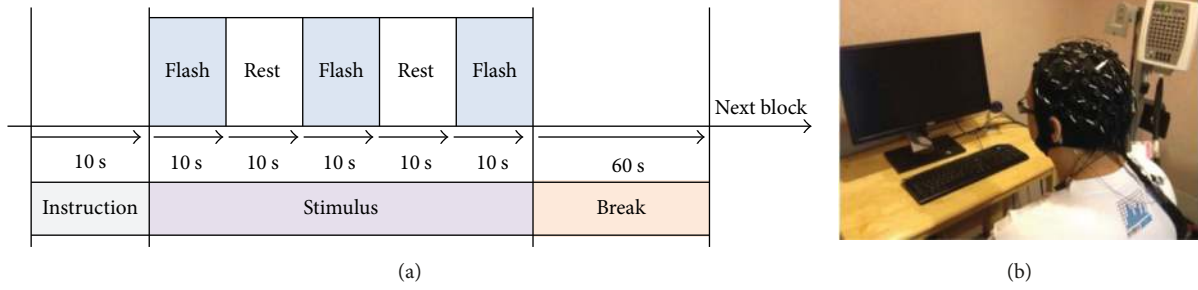
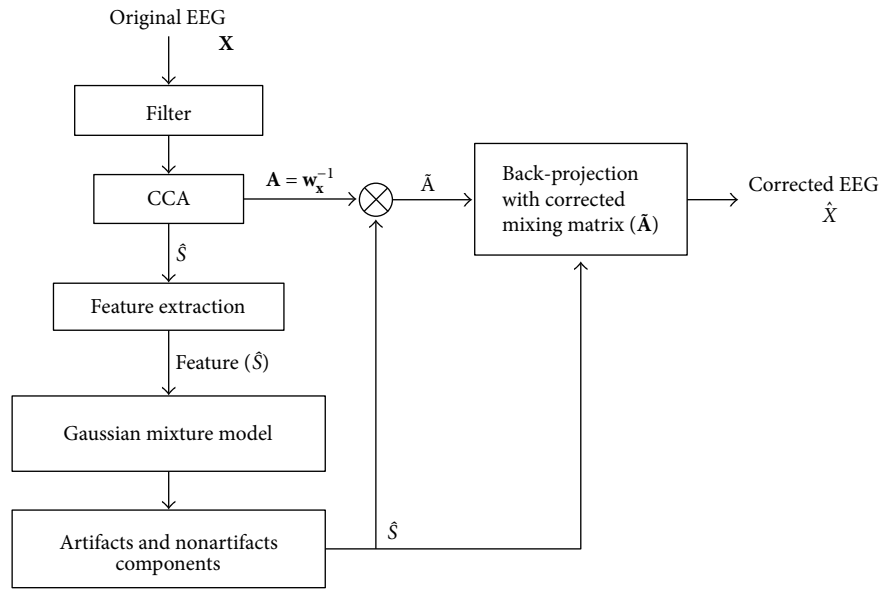Figure 1: (a) Experimental paradigm. (b) Experimental setup.

Figure 2: Flowchart of proposed artifact removal algorithm.

variables, U and V, are linear combinations of the components in $\mathbf{X}$ and $\mathbf{Y}$,

$$
\begin{aligned}
U(t) &= \mathbf{w}_{\mathbf{x}}^T \mathbf{X}(t), \\
V(t) &= \mathbf{w}_{\mathbf{y}}^T \mathbf{Y}(t).
\end{aligned}
\tag{3}
$$

CCA is used to find the matrices $\mathbf{w}_{\mathbf{x}} = [w_{x_1} \cdots w_{x_M}]$ and $\mathbf{w}_{\mathbf{y}}[w_{y_1} \cdots w_{y_M}]$ that maximize the correlation ρ between U and V; the following problem has to be solved:

$$
\max_{\mathbf{W_X}, \mathbf{W_Y}} \rho(U, V) = \frac{\mathbf{w}_{\mathbf{x}}^T \mathbf{C_{xy}} \mathbf{w}_{\mathbf{y}}}{\sqrt{\left(\mathbf{w}_{\mathbf{x}}^T \mathbf{C_{xx}} \mathbf{w}_{\mathbf{x}}\right)\left(\mathbf{w}_{\mathbf{y}}^T \mathbf{C_{yy}} \mathbf{w}_{\mathbf{y}}\right)}},
\tag{4}
$$

where $\mathbf{C_{xx}}$ and $\mathbf{C_{yy}}$ are the autocovariance matrices of $\mathbf{X}$ and $\mathbf{Y}$, and $\mathbf{C_{xy}} = \mathbf{C_{yx}^T}$ are the cross-covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$. This optimization problem can be solved by solving the partial derivative with respect to $\mathbf{w}_{\mathbf{x}}$ and $\mathbf{w}_{\mathbf{y}}$, respectively, and transforming it into the following eigenvalue problem:

$$
\begin{aligned}
\mathbf{C_{xx}^{-1}} \mathbf{C_{xy}} \mathbf{C_{yy}^{-1}} \mathbf{C_{yx}} \mathbf{w_x} &= \rho^2 \mathbf{w_x}, \\
\mathbf{C_{yy}^{-1}} \mathbf{C_{yx}} \mathbf{C_{xx}^{-1}} \mathbf{C_{xy}} \mathbf{w_y} &= \rho^2 \mathbf{w_y},
\end{aligned}
\tag{5}
$$

where $\mathbf{w_x}$ and $\mathbf{w_y}$ are eigenvectors, and the canonical autocorrelation coefficient $\rho^2 \in [0\ 1]$ is the eigenvalue. $(u_i(t), v_i(t))$ represents the ith pair of canonical variates, and ρ_i is the correlation between $u_i(t)$ and $v_i(t)$. In CCA, $U(t)$ is related to the derived components $\widehat{S}(t)$ as follows:

$$
\widehat{S}(t) = \mathbf{W}\mathbf{X}(t) = U(t) = \mathbf{w}_{\mathbf{x}}^T \mathbf{X}(t).
\tag{6}
$$

The unknown mixing matrix $\mathbf{A}$ can be obtained from the inverse of the demixing matrix $\mathbf{W} = \mathbf{w_x}$: $\mathbf{A} = \mathbf{w_x^{-1}}$.

The corrected EEG signals $\widehat{\mathbf{X}}(t)$ can be derived from the back-projection of derived components,

$$
\widehat{\mathbf{X}}(t) = \tilde{\mathbf{A}} \cdot \widehat{S}(t),
\tag{7}
$$

where $\tilde{\mathbf{A}}$ is the corrected mixing matrix, whose columns, which represent artifact components, contain elements that are set to zero.
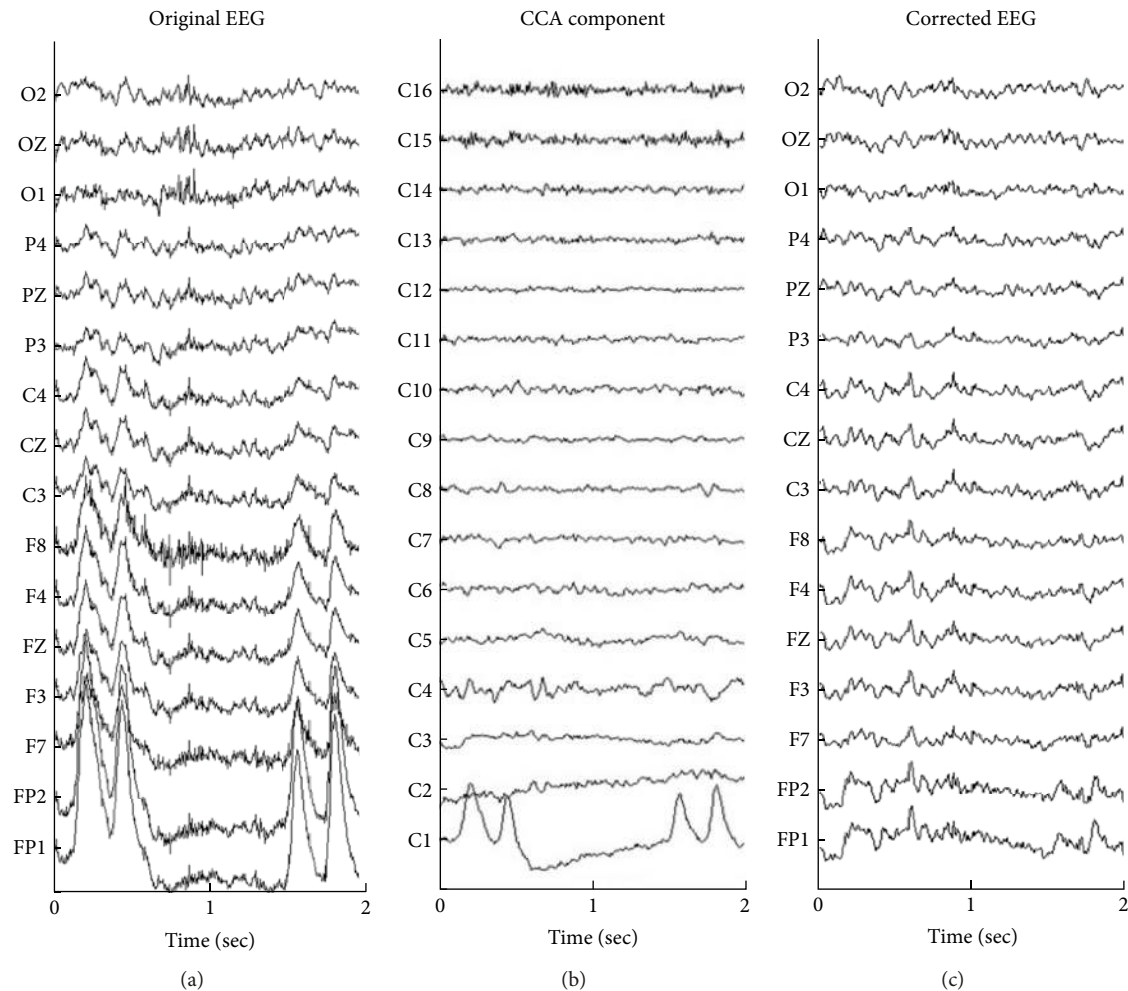
FIGURE 3: Demonstration of removal of EEG artifacts by BSS-CCA. (a) A 2 s portion of EEG time series that contains blinking. (b) Corresponding CCA component activations. (c) EEG corrected by removing C1, C2, C15, and C16 from (b).

Figure 3(a) shows a 2 s portion of the recorded EEG time series selected from 16 EEG electrodes on scalps, and Figure 3(b) presents the derived CCA component activations, with the order of CCA components sorted (from largest to smallest) by the autocorrelation coefficient. The blink artifact was isolated to as C1; the slow artifact was isolated as C2, and the muscle artifacts were isolated as C15 and C16. Figure 3(c) shows the artifact-free corrected EEG signals that were obtained by removing four selected components (C1, C2, C15, and C16). Although the CCA was initially proposed to eliminate muscle artifacts, it also successfully separates out slow artifacts.

2.2.2. Feature Extraction. Different types of artifact in EEG typically have different characteristics. For example, the amplitudes of ocular or body movement artifacts are usually much higher than those of the EEG activities of interest. High-frequency and low-amplitude activities accompany muscle artifacts. Therefore, this work proposes ten features—six spectral and four temporal—to reflect the variability of CCA components.

Six spectral features are extracted from the power spectral density (PSD) by fast Fourier transform (FFT) into six specific frequency bands (lower δ (0.5~2 Hz), higher δ (0.5~4 Hz), θ (4~8 Hz), α (8~12 Hz), β (13~30 Hz), and low γ (30~60 Hz)). Four temporal features are extracted from the canonical autocorrelation coefficient, kurtosis, skewness, and fractal dimension (FD), respectively. The canonical autocorrelation coefficient was obtained by CCA. The separated muscle artifact components have the lowest autocorrelation coefficient, and the slow artifact components have the highest autocorrelation coefficient. Kurtosis here is the degree of the peaked distributions of EEG signal and is used to compare degrees of non-Gaussianity of random variables [6, 20]. If the component waveform is highly gathered around the central distribution value, then, the kurtosis is positive. Skewness measures the degree of asymmetry of distribution [6, 20]. If the distribution of the EEG signal of the component waveform is symmetrical, then, the skewness is zero. Ocular or body movement-induced components may exhibit an asymmetrical distribution. The fractal dimension (FD) can quantify the complexity of a waveform, and it is extensively

used in the detection of ocular artifacts [21]. In this work, the FD is obtained using Sevcik's algorithm, which is quite robust and allows for fast computation. Sevcik's algorithm firstly maps n-point waveform, which has coordinates (x_i, y_i), $i = 1, \ldots, n$, into a unit square. Let the normalized abscissa of the square be x_i^*, and the normalized ordinate of the square be y_i^*, which can be defined as

$$
\begin{aligned}
x_i^* &= \frac{x_i}{x_{\max}}, \\
y_i^* &= \frac{y_i - y_{\min}}{y_{\max} - y_{\min}},
\end{aligned}
\tag{8}
$$

where $x_{\max}$ is the maximum of x, and $y_{\max}$ and $y_{\min}$ are the maximum and minimum of y, respectively. The FD of the waveform is approximated as

$$
\mathrm{FD} = 1 + \frac{\ln(l)}{\ln(2(n-1))},
\tag{9}
$$

where l denotes the total length of the waveform.

2.2.3. Gaussian Mixture Model. The Gaussian mixture model (GMM) is used for unsupervised learning. It is a probabilistic model that assumes that all data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters [22]. The GMM consists of K Gaussian components as

$$
P(x|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{K} g(x|\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}),
\tag{10}
$$

with mean vector $\boldsymbol{\mu_i}$ and covariance matrix $\boldsymbol{\Sigma_i}$. Each probability density of GMM component has the form

$$
g(x|\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma_i}|}} \exp\left\{ -\frac{1}{2}(x - \boldsymbol{\mu_i})^T \boldsymbol{\Sigma_i}^{-1}(x - \boldsymbol{\mu_i}) \right\}.
\tag{11}
$$

Consider n samples, with the following likelihood function of GMM:

$$
P(X|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} P(x_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}).
\tag{12}
$$

The parameters are estimated from the maximum likelihood function. The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are obtained by maximizing the likelihood function of GMM ($P(X|\boldsymbol{\mu}\boldsymbol{\Sigma})$). The GMM function implements the expectation-maximization (EM) algorithm for fitting mixed Gaussian models.

2.3. Data Processing. In this work, the continuous EEG signals are processed using a fixed length window (2 s) with an overlapping window (1.5 s). Therefore, the 50 s of raw EEG data in each stimulus block were segmented into 117 2 s epochs, yielding a total of 936 epochs for each subject. The EEG signals for each epoch were decomposed into components by CCA. Traditionally, a neurophysiologist manually labels all data as ocular artifacts, EMG components, or EEG components of interest by inspecting the time series,

the power spectral density, or topography of the components. The EEG dataset contains tens of thousands of individual EEG patterns from and across individual participants. In this work, $11 \times 936 \times 50 = 514{,}800$ (number of subjects × number of epochs × number of EEG channels) CCA components are involved. Manually labeling various types of artifact among these decomposed CCA components is difficult and time-consuming. Therefore, GMM clusters them automatically to identify various extracted features. Typically, researchers/experts manually score and label many common artifacts, such as those associated with blinking, muscle activity, and motion, and the EEG signals of interest by checking the temporal and spectral properties of decomposed components in each GMM cluster. In this work, the Fisher criterion $\mathrm{tr}(S_w^{-1} S_b)$ [23] was used to assess the number of GMM clusters; S_b and S_w are the class between and within scatter matrices, and $\mathrm{tr}(\mathbf{A})$ represents the trace of the square matrix $\mathbf{A}$. A larger $\mathrm{tr}(S_w^{-1} S_b)$ implies a larger separability in feature space, so the number of GMM clusters can be obtained by finding the largest value of the Fisher criterion variable.

2.4. Performance Evaluation. Performance of algorithm has been compared with the most popular method of artifact removal called artifact subspace reconstruction (ASR) [24, 25]. ASR is the most common and widespread online method of removing transient, high-amplitude-related artifacts from different sources like eye blinks, muscle bursts, and movement [26] while recovering essential EEG background activities that lie in the subspace spanned. ASR method is available with EEGLAB [27], and comparison has been performed with nondefault parameters of a sliding window (500 ms), a threshold of three standard deviations, and without any channel rejection.

3. Experimental Results

3.1. Number of Clusters of GMM and Extracted Features. In this work, the number of the cluster was obtained by finding the maximum value of the Fisher criterion variable over a set (2, 3,…,20). Figure 4 shows the estimated values of the Fisher criterion as the number of clusters was varied from two to 20. The number of GMM clusters was set to 12, based on a grid search.

In the experiment design, various typical artifacts, which were blinking, chewing, and head rotation, were generated following the instructions during the experiment. GMM clusters were grouped into four classes, which corresponded to muscle artifacts, ocular artifacts, body movement artifacts, and nonartifacts, by visually inspecting the temporal waveforms and PSDs of EEG signals. Figure 5 presents the feature histograms for four classes, and most of the classes differed significantly ($p < 0.001$, Wilcoxon signed rank test) between any two classes, with the following exceptional pairs. The first pair was α-power features between the muscle activity class and the ocular artifacts class and between the body movement class and the nonartifact activity class. The second pair was the low γ-power features between the body movement class and the nonartifact activity class, and the third pair was
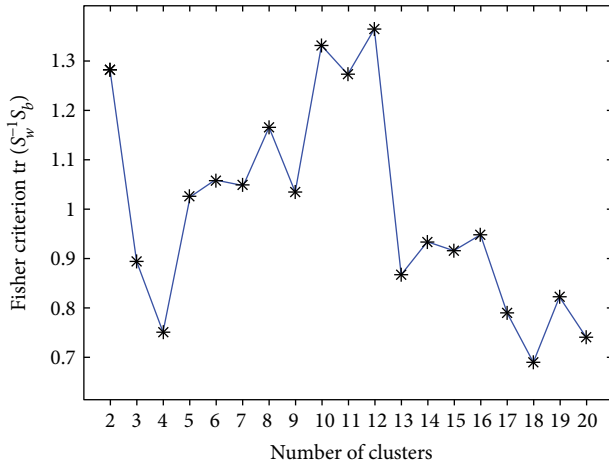
Figure 4: Fluctuation of Fisher criterion value from two to 20 clusters.

the FD feature between the ocular artifacts and the body movement, and the kurtosis feature was between the muscle activity class and the nonartifact activity class.

3.2. EEG Data. Three 5 s long EEG data were selected to illustrate the effectiveness of the proposed artifact removal algorithm (Figure 6). Under all conditions, the amplitude of EEG signals before artifact removal was larger than that after artifact removal. For chewing and head rotation, the EEG signals were contaminated with not only muscle artifacts (Figures 6(b) and 6(c)) but also slow fluctuations. Figure 6(a) shows EEG signals recorded over the prefrontal and frontal area, and Figures 6(b) and 6(c) show EEG signal recorded over all channels. Exclusion of the CCA components made of slow fluctuations and muscle activity was visible in EEG signals using the proposed artifact removal algorithm.

Temporal (ERPs) and spectral (PSDs) responses of EEG in VEP and SSVEP tasks were used to evaluate the effectiveness of the proposed artifact removal algorithm. In the VEP task, EEG epochs were extracted from 500 ms before to 1000 ms after the visual stimulus onset. In the SSVEP task, EEG epochs were extracted from 0 ms to 1000 ms after the beginning of the visual stimulus onset. The power spectrum activities of the EEG signals were calculated by fast Fourier transformation (FFT) and converted to decibels by taking their log power. The EEG waveforms and the power spectrum activities were vertically stacked by epochs, yielding 2D images (Figure 7), and the averaged ERPs and power spectrum activities from all epochs are shown in Figures 7 and 8.

Visual stimuli accompany visual-evoked potentials (VEPs) at most recording scalp sites, including the occipital, parietal, central, and frontal electrode sites [28]. Therefore, only results from EEG channels (Fz, Cz, Pz, and Oz) are shown. The proposed artifact removal algorithm reduced the EEG spectral power at high frequency (>20 Hz) during chewing and head rotation (Figure 7). Under all artifact-generating conditions, before artifact removal, the amplitudes and waveforms of ERPs did not appear similar to

those under motionless conditions; following artifact removal, the ERPs were similar to those under motionless conditions (Figure 8). Comparison with ASR has been made which shows (Figure 9) that the proposed artifact removal algorithm not only eliminates the influence of artifacts on EEG signals but also maintains the specific frequency response and the temporal profiles in the defined task (Figure 9(a)).

4. Discussion

The performance of the proposed artifact removal algorithm was evaluated by classical visual-evoked tasks with common artifacts. Results thus obtained demonstrated that common artifacts in EEG were successfully suppressed by the proposed artifact removal algorithm. EEG signals appeared much cleaner after artifact removal than before it (Figure 6). Moreover, temporal and spectral properties of EEG signals in visual-evoked tasks were maintained after the proposed algorithm was applied and EEG signals had similar profiles to those under motionless condition. The proposed algorithm effectively eliminated artifacts from EEG signals while retaining the featured related properties of EEG activity.

One widely used BSS method in EEG studies is independent component analysis (ICA), which decomposes EEG signals into a set of statically independent components by maximizing its statistical independence. Several methods of ICA-based artifact removal have been proposed for removing artifacts from contaminated EEG signals [6, 9, 20, 29, 30]. Although ICA method can clearly capture the artifacts from EEG signals and separate them from the components of brain activity, it may not be ideal for detecting muscle-related components [31]. ICA requires an iterative process to solve the optimization problem of maximizing the statistical independence among components, which requires extensive computational resources and time. Therefore, the ICA-based artifact removal algorithm cannot be used in real-time applications. Recent studies [8, 10, 15] have demonstrated that CCA outperforms ICA and the traditional filter approach to solve the muscle artifact problem. A CCA-based online approach for artifact removal [8] has also been proposed for removing muscle artifacts from EEG signals. However, these CCA-based methods do not recognize or remove slow artifacts, such as eye movements including blinking. Based on our observations of CCA components derived from CCA-method, CCA can completely separate slow artifacts from brain signals (C1 and C2 in Figure 3(b)); in particular, slow artifacts have higher autocorrelation coefficients and comparable amplitudes than brain signals. Therefore, this work also uses CCA for the real-time separation of brain and artifact activities.

Standard methods involve specific feature extraction and learning classification. Feature extraction is a highly efficient means of achieving satisfactory artifact classification performance. If extracted features can achieve high separability, then, it is easier to distinguish between different classes. Several studies [6, 10, 20, 21, 32] have demonstrated that spectral activities, the canonical autocorrelation coefficient,
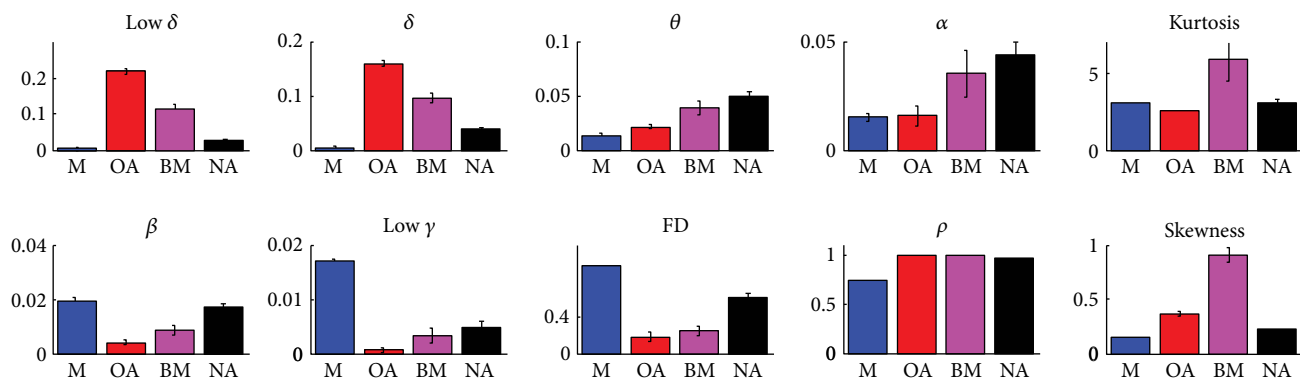
FIGURE 5: Histogram of artifact and nonartifact significance for extracted features. M, OA, BM, and NA represent muscle artifact, ocular artifact, body movement artifact, and nonartifact activity.
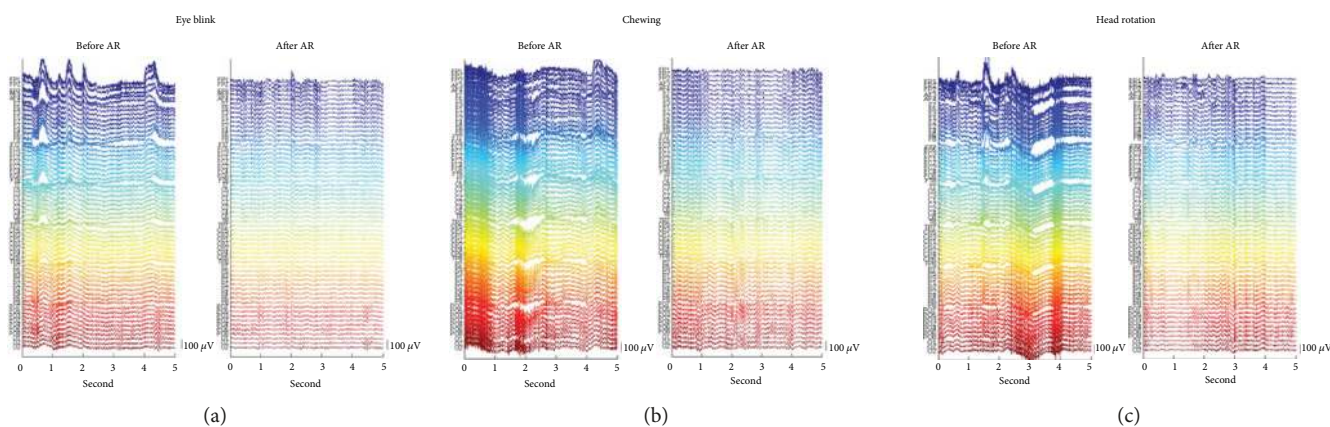

(a)

(b)

(c)

FIGURE 6: Five seconds of EEG data from a representative subject with (a) blinking, (b) chewing, and (c) head rotation. In each subfigure, left and right plots are obtained before and after the proposed artifact removal algorithm is applied.

kurtosis, skewness, and fractal dimension are useful indices for identifying visual, cardiac, or muscle artifacts. This work uses these features for identifying artifact and presents evidence that the extracted features are effective in distinguishing between artifacts or between nonartifact components. Most artifact learning classification requires supervised learning, which requires labeled data to be input to a training model [6, 18, 19]. Although manual labeling can identify artifact and nonartifact component activities, the selection of artifactual components is subjective and needs to be carried out by an expert in the field [33]. In fact, the EEG dataset contains tens of thousands of individual EEG patterns of every participant. This work involves recorded data with 514,800 CCA components; for such a large EEG dataset, manual scoring is impractical, time-consuming, and inconvenient. Unsupervised learning methods are considered in artifact component classification. This work considers GMM because it utilizes well-studied statistical inference techniques and provides flexibility in the mixing of distributions and overall density for each cluster by the mixture model [34]. The number of clusters was obtained by finding the maximum value of the Fisher criterion variable, which provides the highest separability in

distinguishing among clusters. Based on the results of the GMM, typical artifacts and EEG activities of interest for all clusters are examined. Typical artifacts (such as those associated with eye movement, muscle activity, and other motion) should be scored for assessment by checking the overall properties of CCA component in GMM cluster.

This work evaluated the efficacy of the proposed artifact removal algorithm for EEG signals for the activities that are likely to produce artifacts during the experiment. Despite its contributions, this work has certain limitations. The collected data are limited to typical artifacts (ocular, muscle, and body movement artifacts). Unlike in laboratory settings, many unknown and typical application-specific artifacts may be generated in real-world situations. For example, an EEG experiment involving car driving on a real road, a speed breaker, or emergency braking may produce atypical artifacts in the EEG recording. If extracerebral artifacts make informative cortical-generated signals very noisy, then, the presented visual-evoked phenomenon in EEG signals does not exist; therefore, no such components were separated and recovered. Finally, the proposed algorithm was applied to EEG signals in two typical visual-evoked tasks. Other tasks should be carried out to evaluate the efficacy of the proposed
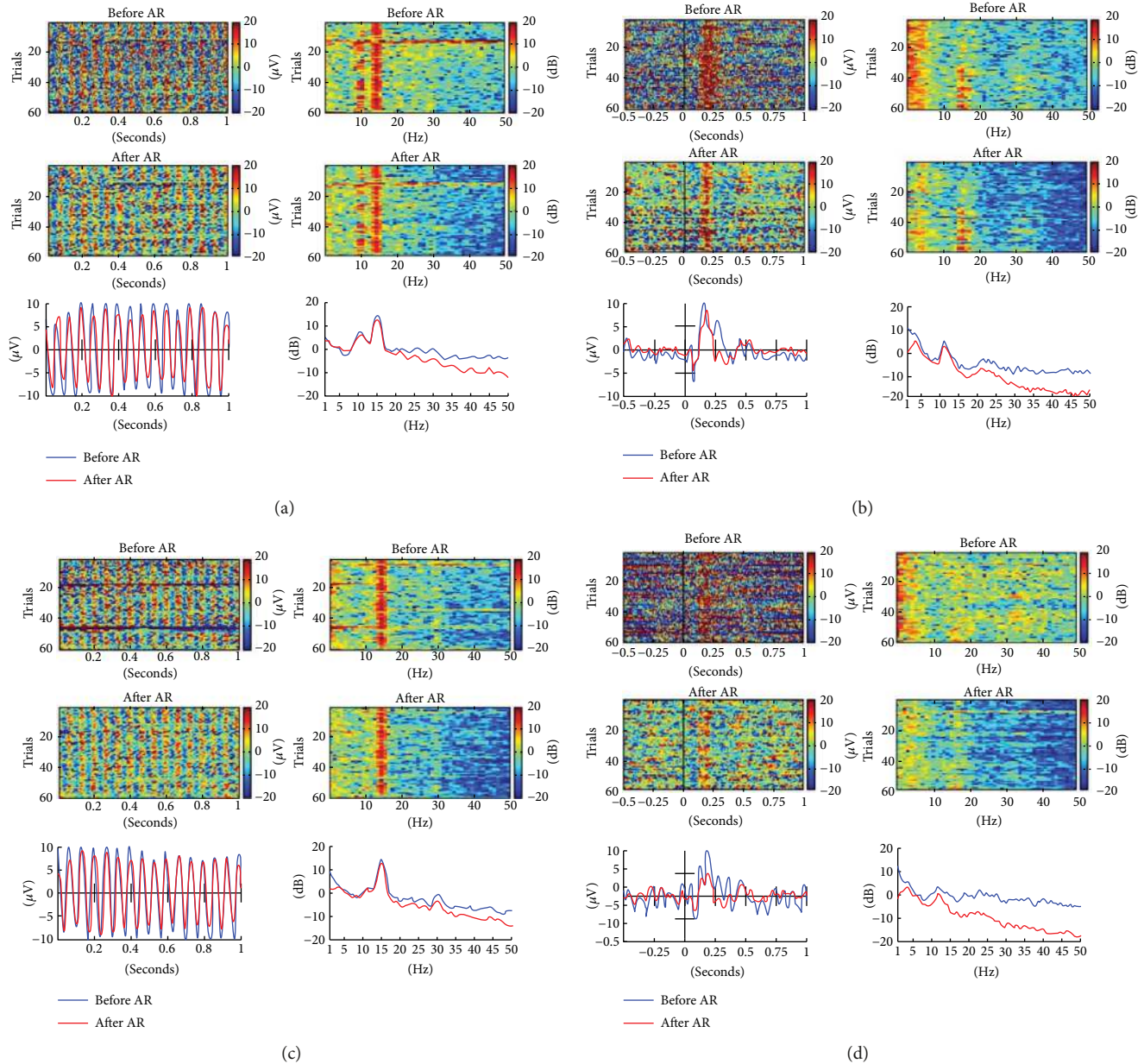
FIGURE 7: Temporal waveform and power spectrum activities of Oz-EEG channel from a representative subject. (a and c) SSVEP task. (b and d) VEP task. (a and b) Chewing. (c and d) Head rotation. Left and right-hand images in subfigures show EEG waveforms and corresponding log power activity, respectively. Top and middle 2D images show EEG results before and after artifact removal, respectively, and bottom images present averaged results. Each horizontal line in 2D images represents a single trial. The time zero on the x-axis represents the time of onset of the visual stimulus. Blue and red lines in bottom images show averaged results obtained before and after artifact removal, respectively. Note: AR indicates artifact removal.

artifact removal algorithm. Efforts are underway to examine these problems.

5. Conclusion

This work proposes a real-time EEG artifact removal algorithm, involving CCA, artifact feature extraction, and the GMM. The efficacy of the proposed method was demonstrated using two classical visual-evoked tasks, which were regular screen flashes with frequencies of 1 Hz and 15 Hz and user-generated artifacts. This work

proposes the GMM-based methodology to cluster all components automatically and intelligently to reduce the inconvenience and complexity of labeling the training dataset manually. Experimental results show the feasibility of using the proposed approach to remove artifacts from EEG signals while retaining the properties of *visual-evoked potential information*. In this work, the GMM is projecting component in the labeled dataset which may cause artifact components to be wrongly classified in CCA. Efforts are underway to use transfer learning to classify artifact components more accurately.
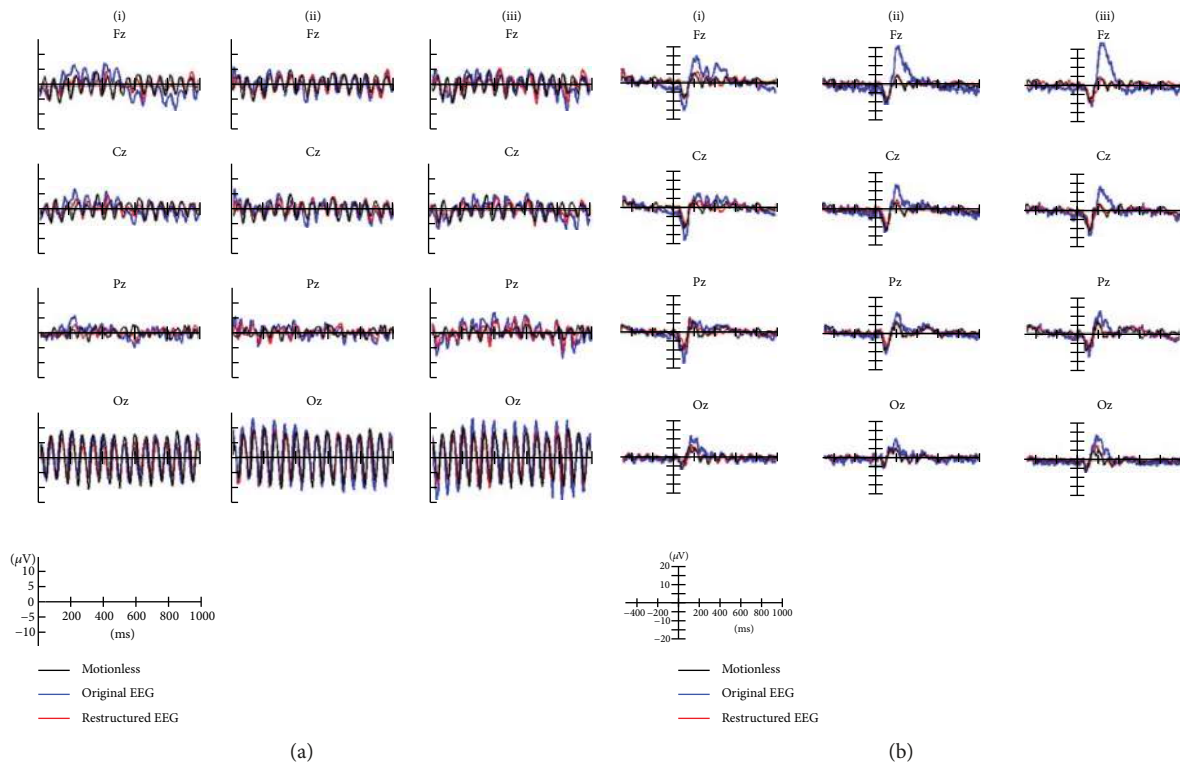
FIGURE 8: Averaged evoked potentials in Fz, Cz, Pz, and Oz channels during (a) SSVEP and (b) VEP tasks from a representative subject. Each subplot figure shows the averaged waveforms for three artifacts: (i) blinking artifacts, (ii) head rotation artifacts, and (iii) chewing artifacts. In each panel, black line presents averaged waveform without motion, whereas blue and red lines present averaged waveforms before and after artifact removal, respectively.
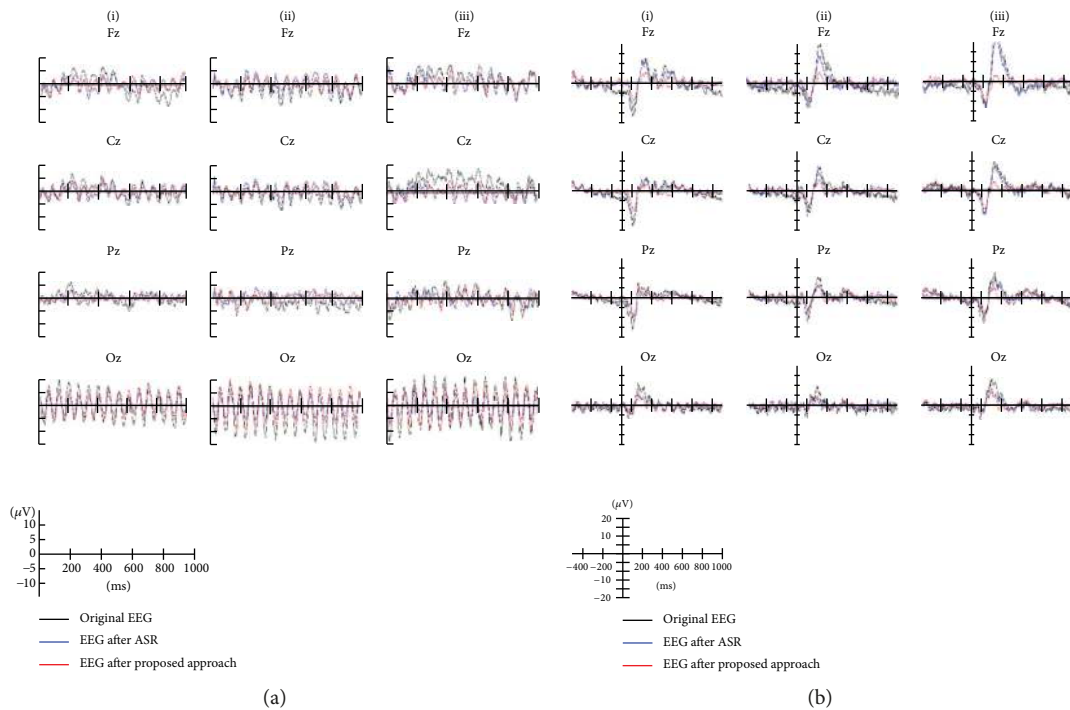

FIGURE 9: Averaged evoked potentials in Fz, Cz, Pz, and Oz channels during (a) SSVEP and (b) VEP tasks from a representative subject. Each subplot figure shows the averaged waveforms for three artifacts: (i) blinking artifacts, (ii) head rotation artifacts, and (iii) chewing artifacts. In each panel, black line presents averaged waveform without motion, whereas blue and red lines present averaged waveforms after ASR and after proposed artifact removal method, respectively.

Disclosure

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L.-d. Liao, C.-T. Lin, K. McDowell et al., "Biosensor technologies for augmented brain–computer interfaces in the next decades," *Proceedings of the IEEE*, vol. 100, Special Centennial Issue, pp. 1553–1566, 2012.

[2] W. O. Tatum, B. A. Dworetzky, and D. L. Schomer, "Artifact and recording concepts in EEG," *Journal of Clinical Neurophysiology*, vol. 28, no. 3, pp. 252–263, 2011.

[3] G. Barbati, C. Porcaro, F. Zappasodi, P. M. Rossini, and F. Tecchio, "Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals," *Clinical Neurophysiology*, vol. 115, no. 5, pp. 1220–1232, 2004.

[4] M. Crespo-Garcia, M. Atienza, and J. L. Cantero, "Muscle artifact removal from human sleep EEG by using independent component analysis," *Annals of Biomedical Engineering*, vol. 36, no. 3, pp. 467–475, 2008.

[5] V. Lawhern, D. Slayback, D. Wu, and B. J. Lance, "Efficient labeling of EEG signal artifacts using active learning," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Kowloon, China, October 2015.

[6] I. Winkler, S. Haufe, and M. Tangermann, "Automatic classification of artifactual ICA-components for artifact removal in EEG signals," *Behavioral and Brain Functions*, vol. 7, no. 1, pp. 1–15, 2011.

[7] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2010.

[8] J. Gao, C. Zheng, and P. Wang, "Online removal of muscle artifact from electroencephalogram signals based on canonical correlation analysis," *Clinical EEG and Neuroscience*, vol. 41, no. 1, pp. 53–59, 2010.

[9] W. Zhou and J. Gotman, "Automatic removal of eye movement artifacts from the EEG using ICA and the dipole model," *Progress in Natural Science*, vol. 19, no. 9, pp. 1165–1170, 2009.

[10] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2583–2587, 2006.

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

[12] S. Makeig, T. P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 20, pp. 10979–10984, 1997.

[13] T.-P. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T. W. Lee, and T. J. Sejnowski, "Imaging brain dynamics using independent component analysis," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1107–1122, 2001.

[14] R. Martín-Clemente, S. Hornillo-Mellado, and J. L. Camargo-Olivares, "Independent component analysis: a low-complexity technique," in *New Challenges on Bioinspired Applications: 4th International Work-Conference on the Interplay between Natural and Artificial Computation, IWINAC 2011*, J. M. Ferrández, J. R. Álvarez Sánchez, F. Paz and F. J. Toledo, Eds., pp. 324–332, Springer, Berlin, Heidelberg, 2011.

[15] A. Vergult, W. de Clercq, A. Palmini et al., "Improving the interpretation of ictal scalp EEG: BSS–CCA algorithm for muscle artifact removal," *Epilepsia*, vol. 48, no. 5, pp. 950–958, 2007.

[16] G. Gomez-Herrero, W. Clercq, H. Anwar et al., "Automatic removal of ocular artifacts in the EEG without an EOG reference channel," in *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, Rejkjavik, Iceland, June 2006.

[17] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *NeuroImage*, vol. 34, no. 4, pp. 1443–1449, 2007.

[18] S.-Y. Shao, K.-Q. Shen, C. J. Ong, E. Wilder-Smith, and X.-P. Li, "Automatic EEG artifact removal: a weighted support vector machine approach with error correction," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 336–344, 2009.

[19] W.-Y. Hsu, C. H. Lin, H. J. Hsu, P. H. Chen, and I. R. Chen, "Wavelet-based envelope features with automatic EOG artifact removal: application to single-trial EEG data," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2743–2749, 2012.

[20] J. Escudero, R. Hornero, D. Abasolo, A. Fernandez, and M. Lopez-Coronado, "Artifact removal in magnetoencephalogram background activity with independent component analysis," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 11, pp. 1965–1973, 2007.

[21] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 2, pp. 177–183, 2001.

[22] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 286–298, 2007.

[23] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, vol. 592, Elsevier Science Publishing Co Inc., San Diego, 1990.

[24] T. Mullen, C. Kothe, Y. M. Chi et al., "Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2184–2187, Osaka, Japan, July 2013.

[25] C. A. E. Kothe and T. P. Jung, *Artifact removal techniques with signal reconstruction*, 2016, Google Patents.

[26] T. C. Bulea, S. Prasad, A. Kilicarslan, and J. L. Contreras-Vidal, "Sitting and standing intention can be decoded from scalp EEG recorded prior to movement execution," *Frontiers in Neuroscience*, vol. 8, p. 376, 2014.

[27] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[28] G. R. Mangun and S. A. Hillyard, "Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 17, no. 4, pp. 1057–1074, 1991.

[29] T. P. Jung, S. Makeig, C. Humphries et al., "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[30] S. Choi, A. Cichocki, H. M. Park, and S. Y. Lee, "Blind source separation and independent component analysis: a review," *Neural Information Processing - Letters and Reviews*, vol. 6, pp. 1–57, 2005.

[31] C. A. Joyce, I. F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," *Psychophysiology*, vol. 41, no. 2, pp. 313–325, 2004.

[32] R. Romo Vázquez, H. Vélez-Pérez, R. Ranta, V. Louis Dorr, D. Maquin, and L. Maillard, "Blind source separation, wavelet denoising and discriminant analysis for EEG artefacts and noise cancelling," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 389–400, 2012.

[33] T.-P. Jung, C. Humphries, T. W. Lee et al., "Extended ICA removes artifacts from electroencephalographic recordings," in *Proceedings of the 1997 conference on Advances in neural information processing systems 10*, pp. 894–900, MIT Press, Denver, 1998.

[34] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.