Discriminative Ocular Artifact Correction for Feature Learning in EEG Analysis

Xinyang Li*, Cuntai Guan, *Senior Member, IEEE*, Haihong Zhang, *Member, IEEE*, and Kai Keng Ang, *Senior Member, IEEE*

***Abstract*—Electrooculogram (EOG) artifact contamination is a common critical issue in general electroencephalogram (EEG) studies as well as in brain–computer interface (BCI) research. It is especially challenging when dedicated EOG channels are unavailable or when there are very few EEG channels available for independent component analysis based ocular artifact removal. It is even more challenging to avoid loss of the signal of interest during the artifact correction process, where the signal of interest can be multiple magnitudes weaker than the artifact. To address these issues, we propose a novel discriminative ocular artifact correction approach for feature learning in EEG analysis. Without extra ocular movement measurements, the artifact is extracted from raw EEG data, which is totally automatic and requires no visual inspection of artifacts. Then, artifact correction is optimized jointly with feature extraction by maximizing oscillatory correlations between trials from the same class and minimizing them between trials from different classes. We evaluate this approach on a real-world EEG dataset comprising 68 subjects performing cognitive tasks. The results showed that the approach is capable of not only suppressing the artifact components but also improving the discriminative power of a classifier with statistical significance. We also demonstrate that the proposed method addresses the confounding issues induced by ocular movements in cognitive EEG study.**

***Index Terms*—Brain–computer interface (BCI), electroencephalogram (EEG), feature learning, ocular artifacts.**

I. Introduction

THE technical field of electroencephalogram (EEG)-based brain–computer interfaces (BCIs) has seen rapid growth in recent years, with a wide range of promising applications, such as motor rehabilitation and cognitive training [1], [2]. Machine learning or signal processing techniques for detecting mental conditions from EEG signals receive much attention [3]–[6].

EEG is highly susceptible to artifact contamination, especially for artifacts induced by ocular movements such as blinks. Therefore, algorithms aiming at recovering artifact-free signal have been investigated intensively [7]–[11]. For example, eye movement correction procedure (EMCP) uses electrooculogram (EOG) recorded along with EEG, and subsequently, subtracts the EOG components from EEG after the scaling based on regression [12], [13]. However, as EOG may also contain components from brain activities, such subtraction based on regression would cause the loss of relevant EEG signals.

For high-dimensional EEG data, independent component analysis (ICA) proves to be a more favorable method in eliminating the ocular artifact components in EEG [13]–[16]. With sources estimated by ICA, the source components corresponding to ocular movements are identified and removed either manually or automatically using prior knowledge about the spatial pattern of the ocular artifacts [17]–[19]. In ICA-based analysis, EOG is not necessary, while it is desirable to record sufficient EEG channels to capture as many sources as possible [20]. Usually, ICA is applied to datasets recorded from at least ten EEG channels [21]. And it is found that as few as 35 channels are needed for source estimation in the study of concurrent locomotor and cognitive tasks [20]. Although the requirement of the minimum number of channels may vary in different experiment tasks, the source separation would be less suitable when only a few EEG channels are available.

Those existing methods require either extra measurements of ocular movements or multiple EEG channels. However, in practical BCI systems, the number of available channels could be limited for comfort and convenience of subjects, and there is only one channel of EEG in certain BCI systems [2], [22]. Thus, methods recovering artifact-free signal for single-channel signal have been proposed. In [23], multichannel signal is obtained using time-delayed coordinates of the single-channel signal, followed by standard ICA-based artifact correction. In the ensemble empirical-mode decomposition (EMD) ICA in [24], intrinsic-mode functions (IMFs) of single-channel signal are obtained by EMD, and used as multichannel data. Similarly, there are methods decomposing a single-channel signal into multiple components using wavelet decomposition, followed by standard source separation methods, such as ICA [25].

For ocular artifact removal in BCI, the most challenging issue is to remove the artifacts with the minimal loss of the cerebral

Manuscript received June 22, 2016; revised September 16, 2016; accepted October 31, 2016. Date of publication November 16, 2016; date of current version July 15, 2017. *Asterisk indicates corresponding author.*

*X. Li is with the University of Essex, Colchester CO4 3SQ, U.K. (e-mail: x.li@essex.ac.uk).

C. Guan is with the Institute for Infocomm Research, Agency for Science, Technology and Research, with Nanyang Technological University, and also with Curtin University.

H. Zhang is with the Institute for Infocomm Research, Agency for Science, Technology and Research.

K. K. Ang is with the Institute for Infocomm Research, Agency for Science, Technology and Researchand also with Nanyang Technological University.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2016.2628958

information [26]. However, most of the artifact removal algorithms are designed to be preprocessing procedures, which are independent of the following classification or detection in BCI. It has not been addressed sufficiently to avoid accuracy drop caused by the loss of discriminative information in EEG signals.

To solve the problem, in this paper, we propose a novel discriminative ocular artifact correction (OAC) approach for feature learning in EEG analysis by adopting oscillatory correlations across EEG trials as the objective function [27], [28]. Joint OAC and feature learning are achieved by integrating the interclass dissimilarity and within-class similarity in a regularized model, which is learnt in a supervised manner. Components related to ocular movements are extracted from the raw data as pseudoartifact channels so that it is applicable to single-channel EEG data without any dedicated EOG or eye-tracker. The proposed method is evaluated on a binary EEG dataset containing 68 subjects performing cognitive tasks, and the confounding issues brought by the ocular artifacts in cognitive EEG study is also discussed. Based on the above discussion, we highlight the contributions of this paper as follows.

1) automatic artifact extraction for single-channel EEG without visual inspection is introduced;
2) a discriminative model for joint OAC and feature learning is proposed; and
3) the confounding issues brought by ocular artifacts in cognitive EEG study are investigated.

This paper is organized as follows. In Section II, the framework of discriminative artifact correction has been introduced. In Section III, the validity of the proposed method is verified by an experimental study on attention detection based on EEG, followed by detailed analysis of the relationship between the ocular artifacts and attentive state. Concluding remarks are given in Section IV.

II. Discriminative OAC

A. Ocular Artifact Detection

When there is no direct measure of ocular movement for OAC, the artifacts need to be extracted from the raw EEG data $X_0(t) \in \mathbb{R}^{n_c \times n_t}$, where n_c is the number of channels and n_t is the number of time samples. Given the analysis of the morphology characteristic of the eye movements related potentials in [12] and [29], a moving average filter is applied to the raw EEG data to obtain the smoothen signal $x_s(t)$ for further artifact extraction as

$$x_s(t) = \frac{1}{m} \sum_{j=-\frac{m}{2}}^{j=\frac{m}{2}} x_0(t+j) \tag{1}$$

where m is the number of the neighboring points used in the moving average filter, and $x_0(t) \in \mathbb{R}^{n_t}$ is the EEG signal from one arbitrary channel, or in other words, one arbitrary row of $X_0(t)$.

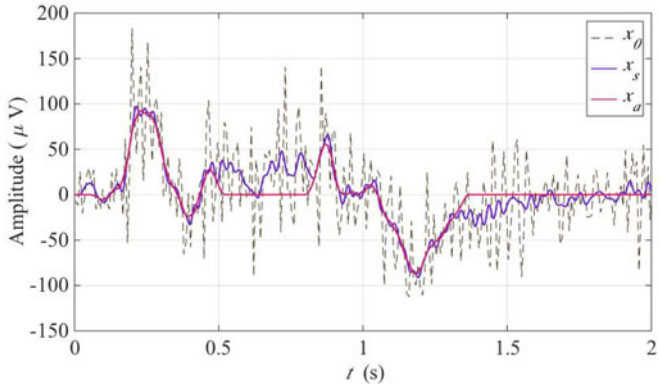

Fig. 1. Examples of the construction of artifact signal. x_0 is the raw data, x_s is the smoothen data, and x_a is the constructed artifact, which are all zeros except for the segments containing the peaks within the certain peak amplitude range.

The relative amplitude of the peaks is calculated as

$$\begin{aligned} h(t) &= \max\{|x_s(t) - x_s(t+\tau_i)|\} \\ \tau_i &= -\tau, -\tau+1, \ldots, \tau-1, \tau \end{aligned} \tag{2}$$

Remark 1: One ocular artifact could include both positive and negative peaks. In other words, a peak could consist of an ocular artifact together with the peak either before it or after. To construct the artifact as complete as possible, in (2), the maximum of the relative amplitude is used as the measurement of the peak.

Define the peak amplitude range parameter h_r as

$$h_r = [h_b, h_u]. \tag{3}$$

Then, find the set $\mathcal{P}_t$ containing time indexes of those peaks with amplitude in the range h_r as

$$\mathcal{P}_t = \{t_i : \frac{m}{2} < t_i < n_t - \frac{m}{2} \text{ and } h_b < h(t_i) < h_u\}. \tag{4}$$

For each element $t_i \in \mathcal{P}_t, i = 1, 2, \ldots, |\mathcal{P}_t|$, let t_i^{zb} and t_i^{za} be the nearest zero points before and after t_i, i.e.,

$$t_i^{zb} = \arg\max_t t \text{ s.t. } t < t_i \text{ and } x_s(t) = 0 \tag{5}$$

$$t_i^{za} = \arg\min_t \ t \text{ s.t. } t > t_i \text{ and } x_s(t) = 0. \tag{6}$$

Remark 2: It is impossible to find zero points for real discrete signals. Thus, in practical implementation, we set a small threshold, and signal points with absolute values below the threshold are regarded as zero points.

With the time period $[t_i^{zb}, t_i^{za}]$ obtained for each peak point $t_i \in \mathcal{P}_t$, the artifact signal $x_a(t)$ is constructed as

$$x_a(t) = \begin{cases} x_s(t), & t \in [t_i^{zb}, t_i^{za}] \text{ with } i = 1, 2, \ldots, |\mathcal{P}_t| \\ 0, & \text{else.} \end{cases} \tag{7}$$

An example of constructing $x_a(t)$ from $x_s(t)$ is illustrated in Fig. 1. As shown by the figure, $x_a(t)$ is zero except those points belonging to peaks whose amplitudes are within a certain range. In this way, EEG data that are not contaminated with the ocular artifacts could be kept as intact as possible after artifact correction.

Moreover, with different amplitude ranges $h_r^j = [h_b^j, h_u^j], j = 1, 2, \ldots, n_h$, $x_a^j(t)$ can be extracted correspondingly, where n_h is the number of peak amplitude ranges. Define $X_a(t)$ as the matrix containing all artifact signals $x_a^j(t)$ as follows:

$$X_a(t) = \begin{bmatrix} x_a^1(t) \\ \vdots \\ x_a^{n_h}(t) \end{bmatrix} \tag{8}$$

$X_a(t)$ in (8) can be regarded as the pseudoartifact signal, which is even more advantageous than the real EOG signal in artifact correction. As it is zero at most of the time points, it would cause less information loss with the artifact removal. Besides, by separating the artifacts by amplitudes of the peaks, artifacts corresponding to different ocular movements could be processed with different filtering parameters assigned. It is more flexible to maintain the discriminative information in EEG signals than the conventional EMCP, where one propagation factor is estimated for one EEG-EOG pair. In the following section, we will introduce the discriminative learning for artifact correction.

B. Regularization Framework Based on Oscillatory Correlation

Let i be the trial index, and we define the signal after correction as $x_{c,i}(t)$, i.e.,

$$x_{c,i}(t) = x_{0,i}(t) - \theta_a^T X_{a,i}(t) \tag{9}$$

where $\theta_a \in \mathbb{R}^{n_h}$ is the artifact correction coefficient or filtering coefficient, scaling the artifacts in EEG to be removed and similar to the propagation factor in the conventional EMCP. In [27] and [28], the oscillatory correlation has been proved to be effective for source separation. In this paper, we propose to optimize the correction coefficient θ_a using the oscillatory correlations between EEG trials, as ocular artifacts should be more sporadic and irregular compared to the oscillatory modulation caused by mental activities.

Rewrite (9) as

$$x_{c,i}(t) = \theta^T X_i(t) \tag{10}$$

where

$$X_i(t) = \begin{bmatrix} x_{0,i}(t) \\ X_{a,i}(t) \end{bmatrix} \tag{11}$$

$$\theta = \begin{bmatrix} 1 \\ -\theta_a \end{bmatrix}. \tag{12}$$

Define the instantaneous power of $x_{c,i}(t)$ as $\phi_i(t)$, i.e.,

$$\phi_i(t) = \sqrt{(\theta^T X_i(t))^2 + (\theta^T H_i(t))^2} \tag{13}$$

where $H_i(t)$ is the Hilbert transform of $X_i(t)$. To obtain an average oscillatory correlation between multiple trials, for each trial i, the average instantaneous power for all trials except i is defined as $\psi_i(t)$, i.e.,

$$\psi_i(t) \equiv \frac{1}{n-1} \sum_{j \neq i} \sqrt{(\theta^T X_j(t))^2 + (\theta^T H_j(t))^2}. \tag{14}$$

Thus, the objective function maximizing the cross-trial oscillatory correlation is

$$\hat{\theta} = \max_{\theta} \ \frac{1}{n} \sum_i \rho_{\phi_i(t),\psi_i(t)} \tag{15}$$

where

$$\rho_{\phi_i(t),\psi_i(t)} = \frac{\int \bar{\phi}_i(t)\bar{\psi}_i(t)\mathrm{d}t}{\sqrt{\int \bar{\phi}_i^2(t)\mathrm{d}t \cdot \int \bar{\psi}_i^2(t)\mathrm{d}t}} \tag{16}$$

with

$$\bar{\phi}_i(t) = \ \phi_i(t) - \frac{1}{n_t} \int \phi_i(t)\mathrm{d}t \tag{17}$$

$$\bar{\psi}_i(t) = \ \psi_i(t) - \frac{1}{n_t} \int \psi_i(t)\mathrm{d}t. \tag{18}$$

Optimizing θ using (15) could maximize the average cross-trial oscillatory correlation so that sporadic ocular artifacts could be subdued, but it is not enough to maintain the discriminative information. To ensure that the artifact correction could benefit the classification in BCI, θ should be learnt in a discriminative manner, which is different from the regressive coefficient estimation or source separation. Thus, the interclass oscillatory correlation r_i is taken into consideration, which could be calculated as

$$r_i = \frac{1}{|\mathcal{Q}^+||\mathcal{Q}^-|} \sum_{i \in \mathcal{Q}^+} \sum_{j \in \mathcal{Q}^-} \rho_{\phi_i(t),\phi_j(t)} \tag{19}$$

where $\mathcal{Q}^c$ is the set of trial index belonging to class c, and $|\mathcal{Q}^c|$ is the number of the elements in $\mathcal{Q}^c$ with the class label $c \in \{+, -\}$. Similarly, the within-class oscillatory correlation for class c, r_w^c, could be calculated as

$$r_w^c = \frac{1}{|\mathcal{Q}^c|} \sum_{i \in \mathcal{Q}^c} \rho_{\phi_i(t),\psi_i(t)}. \tag{20}$$

For joint artifact correction and discriminative feature learning, we propose a regularized oscillatory correlation objective function as

$$\hat{\theta} = \arg\max_{\theta} \ (1-\lambda) \sum_c \lambda^c r_w^c - \lambda r_i, \ \text{with} \ \sum_c \lambda^c = 1 \tag{21}$$

where λ^c is the weight given to the within-class oscillatory correlation for class c, and λ controls the weights of within-class and interclass oscillatory correlation. The regularization has been widely used in the computational model development in BCI to address the within-class similarity and interclass dissimilarity at the same time [30]. With (21), θ is optimized so that the within-class oscillatory correlation could be maximized while the interclass oscillatory correlation is minimized.

C. Joint Optimization of Features and Artifact Correction Coefficients

Given the discriminative oscillatory correlation, we propose to extract two kinds of features, the correlation feature and the

power feature. The oscillatory correlation feature $\mathbf{f}_{r,i}^c$ can be obtained as

$$\mathbf{f}_{r,i}^c = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \rho_{\phi_i(t), \bar{\phi}^c(t)} \mathrm{d}t \tag{22}$$

where $\bar{\phi}^c(t)$ is the average instantaneous power of class c, i.e.,

$$\bar{\phi}^c(t) = \frac{1}{|\mathcal{Q}^c|} \sum_{j \in \mathcal{Q}^c} \sqrt{(\theta^T X_j(t))^2 + (\theta^T H_j(t))^2}. \tag{23}$$

Thus, for each trial for each time window, a pair of correlation features is extracted. The power feature $\mathbf{f}_{p,i}$ for trial i could be extracted as

$$\mathbf{f}_{p,i} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} ||x_{c,i}(t)||^2 \mathrm{d}t \tag{24}$$

where $[t_1, t_2]$ is the time window for the power calculation. With (21), for a certain time window, if the power of the signals from one class is high, than that from the other class would be low, and vice versa. Therefore, the band power feature $\mathbf{f}_p$ is consistent with the objective function.

Regarding selecting regularization parameter in (21), in this paper, we propose to use mutual information between the feature $\mathbf{f}$ and class label c, i.e., $I(\mathbf{f}, c)$, instead of cross validation to reduce the computational complexity. The mutual information has been widely used for feature optimization in BCI, and details of the calculation can be found in [31]–[33].

Let $\Lambda_k = [\lambda, \lambda^+, \lambda^-], k \in \{1, 2, \ldots, n_k\}$, which contains all n_k combinations of regularization parameters, e.g., $\Lambda_1 = [0, 0.5, 0.5]$, $\Lambda_2 = [0.1, 0.5, 0.5]$, etc. With different Λ_k, we could optimize θ using (21), followed by the calculation of feature $\mathbf{f}$ and the mutual information $I(\mathbf{f}, c)$. Given $I(\mathbf{f}, c)$ calculated based on different Λ_k, we choose θ that yields the highest mutual information $I(\mathbf{f}, c)$. By introducing $I(\mathbf{f}, c)$, on the one hand, we could select the best combination of the regularization term Λ_k. On the other hand, it is guaranteed that the feature discrimination improves during the optimization.

To ensure that the artifacts will not be enhanced, we add one more constraint for the optimization of θ, which is

$$|\theta_1 + \theta_j| < \ 1, j = 2, \ldots, n_h \tag{25}$$

$$\theta_1 \equiv \ 1 \tag{26}$$

where θ_j is the jth element of θ. θ_1 is the weight corresponding to the raw EEG signal $x_{0,i}(t)$, which is constrained to be 1. With (25), θ_j cannot be positive, and subsequently, the detected artifact will not be enhanced.

By maximizing the interclass oscillatory differences, the artifact correction parameter could also be driven toward increasing the amplitude of the ocular artifacts if the artifacts contribute to the discrimination between two classes. Although it could be addressed by adding extra constraint terms for θ, the optimization would be more complicated. Thus, in this paper, we propose to avoid enhancing artifacts by only accepting the solutions that suppress the ocular artifacts. Details of the optimization process are described in Algorithm 1.

Algorithm 1: Optimization of the OAC coefficients.

Input: Training set $\mathcal{Q}_{tr}$;
Output: Artifact correction parameter $\hat{\theta}$.
begin
  Initiate $\hat{\theta} = [1, 0, \cdots, 0]^T$;
  Calculate $I_0(\mathbf{f}, c)$;
  Initiate $k = 1$;
  while $k < n_k$ **do**
    Optimize θ using (21) with Λ_k;
    Calculate $I(\mathbf{f}, c)$;
    Normalize θ by θ_1 so that $\theta_1 = 1$.
    if $|\theta_1 + \theta_j| < 1, j = 2, \cdots, n_h$ **then**
      if $I(\mathbf{f}, c) > I_0(\mathbf{f}, c)$ **then**
        update $\hat{\theta} = \theta$;
        update I_0;
    k=k+1.
end

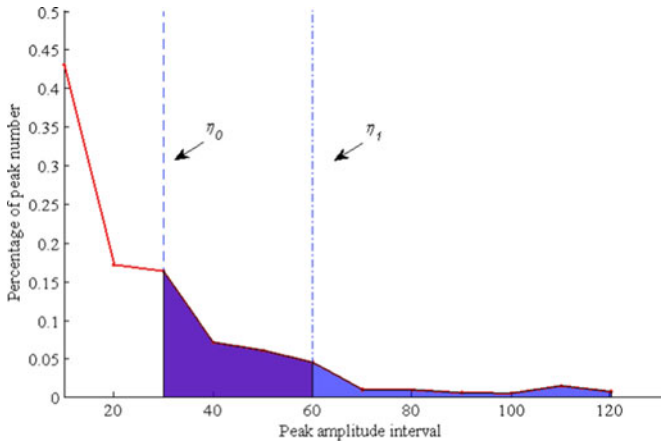

Fig. 2. Selection of the peak amplitude range according to the peak number distribution.

Remark 3: During the optimization, we do not constrain θ_1 to be 1. Instead, we normalize it by θ_1 upon the completion of the optimization as $\theta = \theta/\theta_1$.

Remark 4: Algorithm 1 selects the best result given all different Λ_k, and subsequently, it is not necessarily to be recursive and irrelevant to the sequence of Λ_k. The same solution will be obtained upon the completion of the algorithm with the initialization of $\hat{\theta} = [1, 0, ..., 0]$.

D. Selection of Peak Amplitude Range

Ocular movements suffer from significant cross-subject variations so it is very difficult to find a reasonable h_r^j for all subjects. We propose to determine h_r^j according to the peak distribution with regard to peak amplitude. Fig. 2 shows an example of the percentage of the number of peaks from different amplitude bins in the total number of trials. In this example, totally 12 peak amplitude bins ranging from 10 to 120 are investigated with the width of each bin as 10. In other words, the x-axis represents h with the amplitude bin defined as $h_r = [h, h + 10]$. Let $p_c(h)$

be the peak number percentage for each bin, and it can be found that $p_c(h)$ decreases as h increases.

With the histogram of the given peak amplitude bins, we propose to calculate h_r^j $(j = 0, \ldots, n_h)$ as

$$h_b^0 \equiv 0 \tag{27}$$

$$h_b^j = \arg\min_{h_b} \frac{\sum_{h=h_b}^{h=h_b^{j-1}} p_c(h)}{\sum_{h=h_b^{j-1}}^{h=\infty} p_c(h)} > \eta_{j-1}, \quad j > 1 \tag{28}$$

$$h_u^j = h_b^{j+1} \tag{29}$$

where $\eta_j (j = 0, \ldots, n_h)$ is a predefined ratio threshold for the accumulated peak number percentage. For a better understanding, (27) and (28) are illustrated in Fig. 2, where the accumulated peak number percentage is shown in a shadowed area. With (28), the boundary of each peak range is determined by whether accumulated peak number percentage exceeds a certain threshold η_j. η_0 defines the starting amplitude range of the artifacts, while $\eta_j (j = 1, \ldots, n_h)$ divides signals with the artifacts in different peak ranges into different pseudochannels. The advantage of using the ratio threshold η_j is that ineffective optimization caused by insufficient data within a certain peak range could be avoided. Although η_j is the parameter that needs to be predefined, the robustness of the model against different settings of η_j is guaranteed and will be shown in the next section.

III. Experimental Study

A. Experiment Setup

The method proposed in Section II could be used for artifact correction in any BCI tasks, while in this paper, we focus on the ocular artifacts in attention detection. In total, 68 subjects participated in a cognitive experiment. For each subject, three sessions of Color Stroop test were conducted [34]. In each session, there were 40 Stroop trials, during which the subject was assumed to be concentrating on the test. Each Stroop trial was followed by an idle period when the subject could relax. The Stroop trial lasted around 10 s, while the idle period between 2 Stroop trials was around 15 s.

EEG data were recorded using a dry EEG headband with one bipolar channel, which was positioned at the frontal site, and the sampling rate is 256 Hz. To increase the number of trials, a 4 s window with a window shift of 2 s was applied to segment EEG data recorded during Stroop trials, which yielded data of the attention class. The same segmentation is also applied to EEG recorded during idle periods, which yields data of the idle class. We used the segments only at the beginning of the idle periods so that the final dataset was balanced between the two classes, i.e., the attention class and idle class. Moreover, the first and second halves of the original Stroop and idle trials were truncated into training trials and test trials, respectively. In this way, the test set was totally independent of the training set. For each subject, the number of total truncated trials was around 240.

B. Ocular Artifact in Attentive State Detection

Ocular movements are closely related to attentive states. In other words, whether a subject is attentive or concentrating could be reflected by his or her ocular movements to a certain extend. For a quantitative study, the number of peaks in different peak amplitude ranges h_r are compared between attentive and idle states. In total, 17 peak amplitude bins ranging from 10 to 170 are investigated with the width of each bin as 10. We found that for both training and test sets, the number of peaks of idle state is consistently larger than that of the attentive state in the amplitude range of around 30–80.

To further investigate the role ocular movements play in attentive states, we conduct correlation analysis between the classification accuracies and the number of peaks in different amplitude ranges, the results of which are shown in Fig. 3. The correlation coefficients for different amplitude ranges for sessions 1, 2, and 3 are shown in Fig. 3(a)–(c), respectively, with the corresponding $\boldsymbol{p}$-values of Pearson correlation tests shown in (d)–(f), respectively. Moreover, we plot $-\log(p)$ with $p = 0.05$ in dotted lines, and thus, the values above these lines indicate the significance of the correlation. It can be found that the classification accuracies are positively correlated with the number of peaks in the amplitude range of around 30–80, with higher correlation coefficients and $\boldsymbol{p}$-values lower than 0.05. To obtain the proper neurofeedback for attention training, a BCI should capture the differences in mental state rather than that in ocular activities. Thus, OAC is a significant issue for the BCI differentiating attentive states from idle states [2], [22].

One of the common approaches to address the artifact issue is to reject all the contaminated trials so that only the "clean" data are used. By calculating the numbers of trials containing the peaks in different amplitude ranges, we find that almost half of the trials need to be discarded due to rejecting trials with artifact contamination. To make up the loss of the trials, the treatment session will be inevitably prolonged for extra repeated recordings, and will become more tedious and tiring for subjects. Thus, to address the issue of the ocular artifacts in real applications of BCI, we propose the ocular artifacts method, with which the collected data could be sufficiently exploited.

C. Feature Extraction and Classification

In this study, raw EEG data are smoothen with $m = 10$ in (1), and $\tau = 1$ in (2), which is around 4 ms with the sampling rate at 256 Hz, for the measurement of the peak. The number of the amplitude ranges is 2, i.e., $n_h = 2$, yielding $X_a(t) \in \mathbb{R}^{2\times n_t}$ in (8). For each trial, the threshold to find the zero points t^{za} and t^{zb} is two times of the minimal absolute value of $x_s(t)$ for the trial. The regularization parameters λ and λ^c are preset to be in the range $[0, 0.1, \ldots, 0.5]$, yielding totally 36 combinations contained in Λ with $n_k = 36$. h_r^1 and h_r^2 are obtained as introduced in Section II-D with $\eta_0 = \eta_1 = \eta$, and η is set differently to investigate the sensitivity of the proposed method with regard to η. Equation (21) is optimized by limited-memory Broyden-Fletcher-Goldfarb-Shanno (implemented using MATLAB's minFunc) [28]. After the OAC, a filter bank containing nine frequency bands (2–6 Hz, 6–10 Hz, ..., 34–38 Hz) is applied on $x_c(t)$. For each frequency band, we calculate the features for every 2 s window with 1 s window overlapping as introduced in Section II-C. Mutual

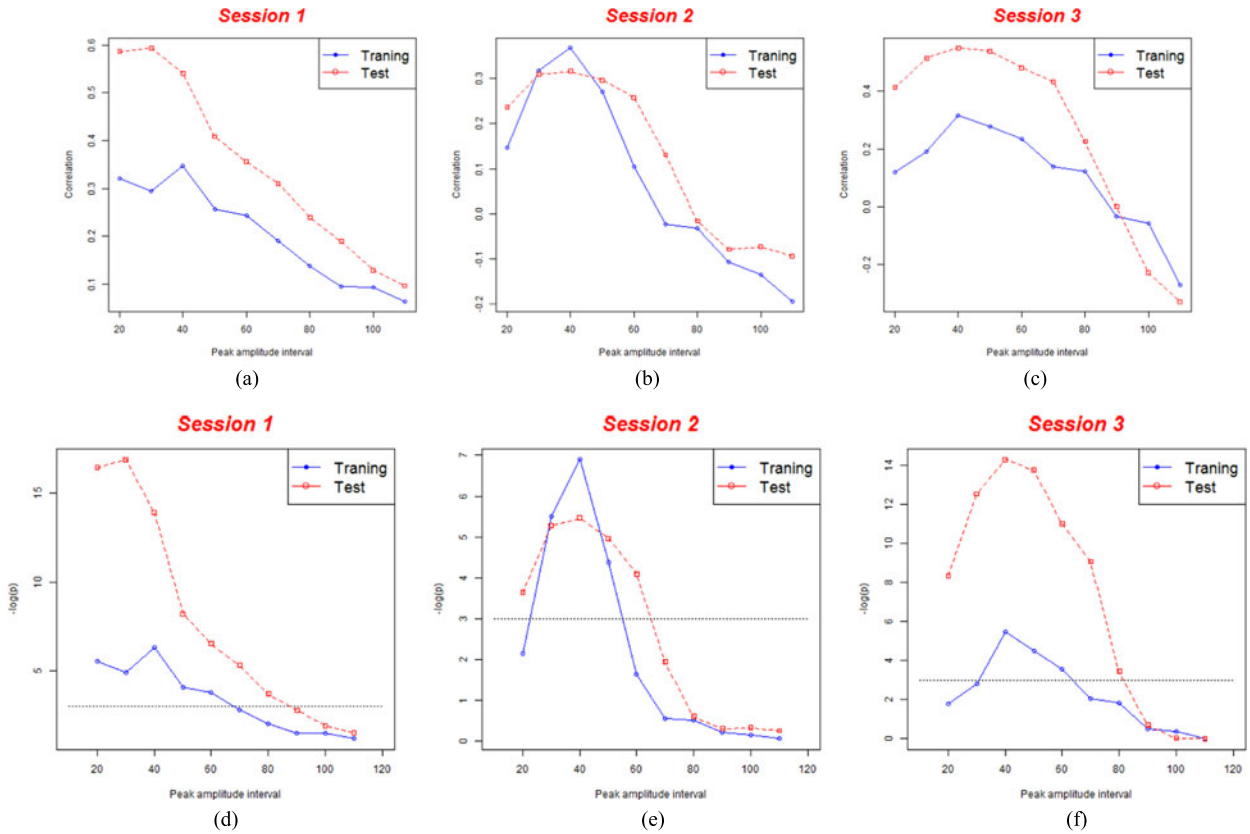

Fig. 3. Correlation between classification accuracy and number of peaks in different amplitude ranges. The classification accuracies are positively correlated with the number of peaks in the amplitude range of around 30–80, with higher correlation coefficients and ***p***-values lower than 0.05. (a) Correlation coefficient (session 1). (b) Correlation coefficient (session 2). (c) Correlation coefficient (session 3). (d) *p*-value (session 1). (e) *p*-value (session 2). (f) *p*-value (session 3).

TABLE I
TEST CLASSIFICATION RESULTS (%)

		BL η	0.60	0.65	0.70	0.75	0.80
SS1	Mean	77.75	80.24	80.79	80.41	80.21	80.04
	Median	79.22	82.28	83.32	80.82	81.02	84.15
	p-value	-	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.05**
SS2	Mean	78.94	82.46	81.71	82.19	82.10	81.67
	Median	79.96	84.45	84.01	85.40	85.54	85.11
	p-value	-	**<0.001**	**<0.01**	**<0.001**	**<0.001**	**<0.001**
SS3	Mean	74.47	76.85	76.05	77.33	77.95	77.35
	Median	73.00	75.64	76.37	77.19	78.21	76.75
	p-value	-	**<0.05**	>0.05	**<0.001**	**<0.001**	**<0.001**
Total	Mean	77.05	79.85	79.51	79.98	80.08	79.69
	Median	77.45	81.25	81.67	81.82	81.67	82.14
	p-value	-	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**

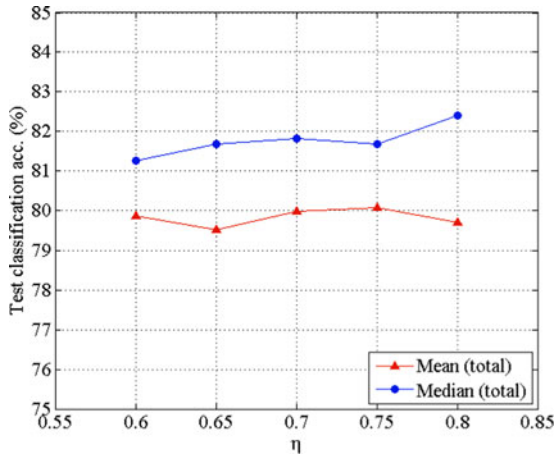

Fig. 4. Change in test classification accuracy with regard to η.

information is applied to select the best four features, and, subsequently, the selected features are classified into the attention class or the idle class by a linear discriminant analysis (LDA) classifier [32].

D. Classification Results

Table I summarizes the classification results of the proposed OAC method compared with the baseline (BL) method for which no artifact correction is applied. Results of Session 1, Session 2, and Session 3 are indicated by "SS1," "SS2," and "SS3," respectively. As shown in Table I, for all three sessions the proposed method improves both the median and average classification accuracies, the significance of which is validated by paired t-test with almost all **p**-values below 0.05. More importantly, by comparing the mean and median accuracies under different η, it can be found that the results are quite consistent. In Fig. 4, the mean and median of the test classification accuracies of all three sessions are illustrated, as shown by which the results of

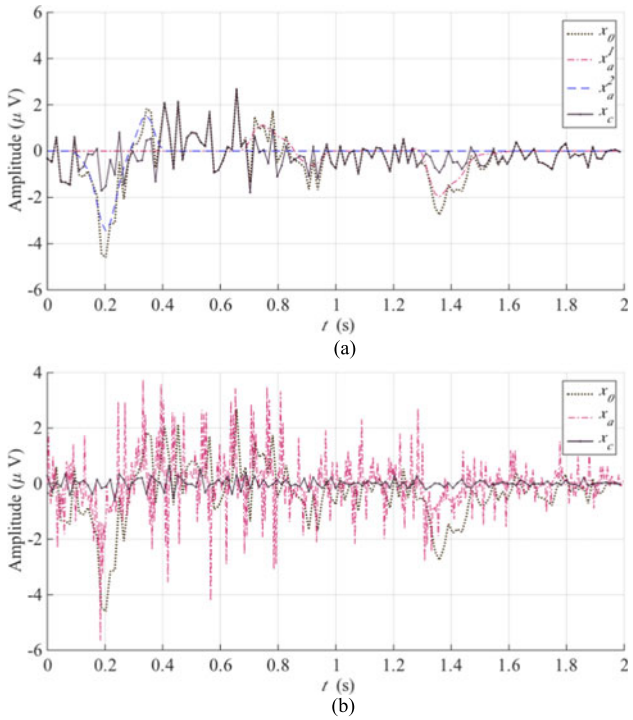

Fig. 5. Comparison of artifact correction. (a) OAC. (b) SCICA.

the proposed artifact correction method is robust against change in the parameter η within this range.

E. Comparing With Other Single-Channel Artifact Correction Methods

For further validation, OAC is compared with single-channel ICA (SCICA), details of which could be found in the supplementary material and [23] and [24]. An example of the comparison between SCICA and OAC is illustrated in Fig. 5, where the raw signals x_0, the extracted artifacts x_a, and the corrected signals x_c in OAC and SCICA are presented in (a) and (b), respectively. It can be found that both SCICA and OAC could remove the ocular artifacts in the form of spikes. Artifacts x_a extracted by SCICA contain more high-frequency components, while x_c obtained by OAC are more similar to the raw signals x_0 for the time segments where there are no artifacts.

For the comparison of classification accuracy, given that the dataset in this paper contains 68 subjects who attended three sessions of experiments with around 240 trials per session, it is very time consuming and difficult to perform visual inspection for all the subjects. Thus, we performed SCICA to EEG recordings of three sessions from ten randomly selected subjects to obtain the classification accuracy. For this subset of dataset, the average classification accuracies for BL, SCICA, and OAC are 72.95%, 74.35%, and 74.85%, respectively, and the median classification accuracies for BL, SCICA, and OAC are 71.25%, 72.36%, and 76.25%, respectively. Both SCICA and OAC could improve the test classification accuracies, while OAC achieves more improvement.

Most of the existing single-channel artifact correction methods require manual identification of artifact components and are usually implemented trial by trial, which means that visual inspection needs to be preformed for each trial. Requiring neither manual selection nor visual inspection, the proposed method is automatic for both artifact detection and correction, which makes it more feasible and efficient in processing practical BCI data. Moreover, for EEMD-ICA, it is impossible to obtain a generalized model using the training data due to the varying numbers of IMFs across trials. In contrast, in the proposed method, the model trained by the training data is effective for the test data, as validated by the test accuracy. Most importantly, the proposed method is a discriminative model and jointly optimizes features, while the conventional artifact correction methods usually serve as data-driven preprocessing procedures.

F. Discussion

The proposed method achieves improvement for most of the subjects, while there are still some subjects, for whom the classification accuracy drops with the artifact correction method. For a better understanding of the OAC, we have investigated the data of subject 19, session 1, the classification accuracies with LDA classifier of which are 86.84% and 59.21% under the BL method, and the proposed artifact correction method, respectively. We calculate the number of training trials containing peaks in different amplitude ranges, speculating that the accuracy drop could be due to insufficient training trials contaminated with artifacts. We found that 56 and 10 out of 74 trials are contaminated with peaks with amplitudes in ranges h_r^1 and h_r^2, below the 30% and 10% percentile ranks among all subjects, respectively. Thus, it is possible that less trials with artifacts result in ineffective optimization. For further investigation, we conduct statistical comparison of the number of trials contaminated with artifacts between subjects with and without accuracy improvements. However, the results show that there is no significant difference.

Regarding the parameter setting, although η_j could be set differently, we set all η_j to be the same in this paper. With η_j representing the percentile of the number of peaks, η_0 should be no less than 0.50 as the ocular artifacts should not constitute a large portion of the data. And, generally, η_j should not be larger than 0.80, or there will be insufficient data with artifacts for certain pseudoartifacts channels. Thus, $0.6 < \eta < 0.8$ is a reasonable range, within which we have shown that the classification results are quite consistent.

In this paper, powers from multiple subbands are used as features, while the optimization is performed on a broad band. The framework could be more consistent if the optimization is conducted in a band-wise manner. However, the higher the frequency band, the more difficult for reliable artifact detection. Moreover, the band-wise optimization will inevitably increase the computational complexity, which makes the artifact correction algorithm prone to overfitting. Thus, in this paper, the optimization is based on a broad band, followed by selecting the best subbands using mutual information. Further improvement could be made to find a proper band-wise optimization approach in our future works.

Despite the offline analysis presented in this paper, the proposed artifact correction method could be used for online processing as long as training trials are available. The feasibility

of the online implementation depends on artifact detection rather than optimization. In this paper, given the moving average applied with five neighboring points, there would be a delay of around 20 ms for the test data artifact correction, which is acceptable for online processing. Moreover, the application of the proposed method is not confined to ocular artifacts. By constructing different artifact templates, the optimization in Section II-B could also be used to remove other artifact types, such as teeth-grinding or swallowing. Extension for multiclass classification problems is also possible by adopting a one-versus-all strategy.

IV. Conclusion

This study investigates the OAC with minimal cerebral information loss for discriminative EEG feature learning. In the proposed method, multiple artifact signals are constructed from the raw EEG data, and the filtering parameters are optimized using oscillatory correlation objective function. In particular, the optimization is performed in a supervised manner to ensure that artifact correction benefits the classification. Applicable to single-channel EEG, artifact signal extraction in the proposed method requires no EOG or eye-tracker measurement, and more importantly, it is automatic without any manual selection or visual inspection of artifact. The proposed method is evaluated using a single-channel EEG dataset containing 68 subjects, and there are two classes, subjects performing cognitive tasks and staying in idle states. The results show that artifact correction is of special importance in addressing the related confounding issues. The effectiveness of the proposed method is validated by the significance of the improvement in classification accuracy in statistical tests. The example of the comparison of EEG signals before and after the correction shows the merit of the proposed method in keeping the signal as intact as possible with alteration applied only to segments with artifacts.

References

[1] K. K. Ang and C. Guan, "Braincomputer interface for neuro-rehabilitation of upper limb after stroke," *Proc. IEEE*, vol. 103, no. 6, pp. 944–953, Jun. 2015.

[2] C. G. Lim *et al.*, "A brain-computer interface based attention training program for treating attention deficit hyperactivity disorder," *PLoS ONE*, vol. 7, no. 10, 2012, Art. no. e46692.

[3] A. S. Aghaei *et al.*, "Separable common spatio-spectral patterns for motor imagery BCI systems," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 15–29, Jan. 2016.

[4] B. J. Edelman *et al.*, "EEG source imaging enhances the decoding of complex right hand motor imagery tasks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 4–14, Jan. 2016.

[5] X. Li *et al.*, "Discriminative learning of propagation and spatial pattern for motor imagery EEG analysis," *Neural Comput.*, vol. 25, no. 10, pp. 2709–2733, Oct. 2013.

[6] X. Li *et al.*, "Adaptation of motor imagery EEG classification model based on tensor decomposition," *J. Neural Eng.*, vol. 11, 2014, Art. no. 056020.

[7] P. P. Acharjee *et al.*, "Independent vector analysis for gradient artifact removal in concurrent EEG-fMRI data," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1750–1758, Jul. 2015.

[8] J. W. Kelly *et al.*, "Fully automated reduction of ocular artifacts in high-dimensional neural data," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 598–606, Mar. 2011.

[9] I. W. Selesnick *et al.*, "Transient artifact reduction algorithm (TARA) based on sparse optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6596–6611, Dec. 2014.

[10] H. Zeng *et al.*, "EOG artifact correction from EEG recording using stationary subspace analysis and empirical mode decomposition," *Sensors*, vol. 13, no. 11, pp. 14 839–14 859, 2013.

[11] D. Looney *et al.*, "Subspace denoising of EEG artefacts via multivariate EMD," in *Proc. 2014 IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2014, pp. 4688–4692.

[12] G. Gratton, M. G. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," *Electroencephalogr. Clin. Neurophysiol.*, vol. 55, pp. 468–484, 1983.

[13] A. Hoffmann and M. Falkenstein, "The correction of eye blink artefacts in the EEG: A comparison of two prominent methods," *PLoS ONE*, vol. 3, no. 8, 2008, Art. no. e3004.

[14] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0165168494900299

[15] T.-P. Jung *et al.*, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects," *Clin. Neurophysiol.*, vol. 111, pp. 1745–1758, 2000.

[16] N. Bigdely-Shamlo, "Eyecatch: Data-mining over half a million EEG independent components to construct a fully-automated eye-component detector," in *Proc. 2013 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2013, pp. 5845–5848.

[17] C. Guerrero-Mosquera and A. Navia-Vázquez, "Automatic removal of ocular artefacts using adaptive filtering and independent component analysis for electroencephalogram data," *IET Signal Process.*, vol. 6, no. 2, pp. 99–106, 2012.

[18] Q. Zhao *et al.*, "Automatic identification and removal of ocular artifacts in EEG-improved adaptive predictor filtering for portable applications," *IEEE Trans. Nanobiosci.*, vol. 13, no. 2, pp. 109–117, Jun. 2014.

[19] A. Mognon *et al.*, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011.

[20] T. M. Lau *et al.*, "How many electrodes are really needed forEEG-based mobile brain imaging?" *J. Behavioral Brain Sci.*, vol. 2, no. 3, pp. 387–393, 2012.

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

[22] C. G. Lim *et al.*, "Effectiveness of a brain-computer interface based programme for the treatment of ADHD: A pilot study," *Psychological Bull.*, vol. 43, no. 1, pp. 73–82, 2010.

[23] M. E. Davies and C. J. James, "Source separation using single channel ICA," *Signal Process.*, vol. 87, no. 8, pp. 1819–1832, 2007.

[24] B. Mijovic *et al.*, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2188–2196, Sep. 2010.

[25] J. Lin and A. Zhang, "Fault feature separation using wavelet-ICA filter," *NDT E Int.*, vol. 38, no. 6, pp. 421–427, 2005.

[26] C. Zhao and T. Qiu, "An automatic ocular artifacts removal method based on wavelet-enhanced canonical correlation analysis," in *Proc. 2011 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 4191–4194.

[27] S. Dähne *et al.*, "SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters," *NeuroImage*, vol. 86, pp. 111–122, 2014.

[28] S. Dähne *et al.*, "Finding brain oscillations with power dependencies in neuroimaging data," *NeuroImage*, vol. 96, pp. 334–348, 2014.

[29] A. S. Keren, S. Yuval-Greenberg, and L. Y. Deouell, "Saccadic spike potentials in gamma-band EEG: Characterization, detection and suppression," *NeuroImage*, vol. 49, no. 3, pp. 2248–2263, 2010.

[30] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[31] K. K. Ang *et al.*, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Front. Neurosci.*, vol. 6, no. 39, 2012.

[32] K. K. Ang *et al.*, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recognit.*, vol. 45, no. 6, pp. 2137–2144, 2012.

[33] J. Meng *et al.*, "Simultaneously optimizing spatial spectral features based on mutual information for eeg classification," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 227–240, Jan. 2015.

[34] C. M. MacLeod, "Half a century of research on the stroop effect: An integrative review," *Psychological Bull.*, vol. 109, pp. 163–203, 1991.

Author's photographs and biographies not available at the time of publication.