Correlated Component Analysis for Enhancing the Performance of SSVEP-Based Brain-Computer Interface

Yangsong Zhang, Daqing Guo, Fali Li, Erwei Yin, Yu Zhang, Peiyang Li, Qibin Zhao, Toshihisa Tanaka, Dezhong Yao, and Peng Xu

***Abstract*— A new method for steady-state visual evoked potentials (SSVEPs) frequency recognition is proposed to enhance the performance of SSVEP-based brain-computer interface (BCI). Correlated component analysis (CORCA) is introduced, which originally was designed to find linear combinations of electrodes that are consistent across subjects and maximally correlated between them. We propose a CORCA algorithm to learn spatial filters with multiple blocks of individual training data for SSVEP-based BCI scenario. The spatial filters are used to remove background noises by combining the multichannel electroencephalogram signals. We conduct a comparison between the proposed CORCA-based and the task-related component analysis (TRCA) based methods using a 40-class SSVEP benchmark data set recorded from 35 subjects. Our experimental study validates the efficiency of the CORCA-based method, and the extensive comparison results indicate that the CORCA-based method significantly outperforms the TRCA-based method. Superior performance demonstrates that the proposed method holds the promising potential to achieve satisfactory performance for SSVEP-based BCI with a large number of targets.**

Manuscript received December 15, 2017; revised March 26, 2018; accepted April 8, 2018. Date of publication April 13, 2018; date of current version May 8, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 81571770, Grant 31771149, Grant 81401484, Grant 61522105, Grant 61527815, Grant 61703407, and Grant 61773129, in part by the Longshan academic talent research supporting program of SWUST under Grant 17LZX692, and in part by JSPS KAKENHI under Grant 17K00326. *(Corresponding authors: Daqing Guo; Dezhong Yao; Peng Xu.)*

Y. Zhang is with the MOE Key Laboratory for Neuroinformation, Clinical Hospital of Chengdu Brain Science Institute, University of Electronic Science and Technology of China, Chengdu 610054, China, and also with the School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China (e-mail: zhangysacademy@gmail.com).

D. Guo, F. Li, P. Li, D. Yao, and P. Xu are with the MOE Key Laboratory for Neuroinformation, Clinical Hospital of Chengdu Brain Science Institute, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: dqguo@uestc.edu.cn; xupeng@uestc.edu.cn; dyao@uestc.edu.cn).

E. Yin is with the National Institute of Defense Technology Innovation, Academy of Military Sciences China, Beijing 100081, China.

Y. Zhang is with the Department of Psychiatry and Behavior Sciences, Stanford University, CA 94305 USA.

Q. Zhao is with the Tensor Learning Unit, RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan, and also with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China.

T. Tanaka is with the Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan.

Digital Object Identifier 10.1109/TNSRE.2018.2826541

***Index Terms*— Brain-computer interface (BCI), electroencephalogram (EEG), steady-state visual evoked potentials (SSVEP), correlated component analysis (CORCA), task-related component analysis (TRCA).**

I. Introduction

BRAIN-COMPUTER interface (BCI) provides an alternative communication channel to humans by using the brain activities, which has achieved unprecedented progress in the past decades [1]–[3]. Among various BCI modalities, the steady-state visual evoked potential (SSVEP)-based BCI has received increasing attention due to its high information transfer rate (ITR) and low requirement of training [4]–[7]. Till now, the SSVEP-based BCI can provide the highest ITR compared to other typical BCIs [8], such as motor imagery [9]–[13], P300 [14]–[16], etc.

Based on the metric ITR, two factors that can significantly affect the high performance of an SSVEP-based BCI are the number of coded targets and an efficient frequency recognition algorithm. For the number of targets, plenty of paradigms have been proposed, e.g., multi-frequency coding method [17], frequency-phase mixed coding method [18], joint frequency-phase modulation (JFPM) method [4], hybrid SSVEP and with other modalities [19]–[22], etc. Among them, the research with the JFPM method achieved ITR above 300 bits/min [8], [23]. For the frequency recognition algorithm, the communities have mainly focused on the multiple channel methods which provides more robust performance than single channel methods. For example, canonical correlation analysis (CCA) [24], minimum energy combination (MEC) [25], multivariate synchronization index (MSI) methods [26], have been developed to implement robust frequency recognition. In fact, between the two factors, the latter one serves as to guarantee the success of realizing high performance of BCI system. In order to yield robust results, the researchers investigated the reference signals optimization, explored the temporal and frequency band information, and introduced advanced machine learning and advanced signal processing techniques to recognize the SSVEP components and optimize some existed algorithms.

Although there exist some state-of-the-art algorithms, how to accurately recognize the SSVEP frequency with a short time window is still a challenging issue for the development of high performance BCI. Most of the multiple

channel methods required template signals corresponding to the stimulus frequencies. In the early stage, the template signals are the artificial reference signals that constructed with sine-cosine functions, which lack subject-specific information. Zhang *et al.* [27] proposed to learn more effective reference signals from the cross session data for each subject, such as multiway CCA, multiset extension of CCA (MsetCCA) [28], multilayer correlation maximization model [29], etc. Yuan *et al.* [30] proposed to exploit inter-subject information by grand averaging EEG data across source subjects as the template signals.

Under each stimulation frequency, the SSVEPs consist of brain responses at fundamental and harmonic frequency components, some studies present to combine harmonic components in frequency detection [31], among which the filter bank analysis may have yielded the best performance [32]. The filter bank analysis can decompose the SSVEPs into multiple sub-band components, then fusion the results from all sub-bands to recognize the frequency [32], [33]. Besides the frequency characteristics, the temporal structure in EEG data could also help to improve the performance of the algorithm [34]. Zhang *et al.* proposed to explore the temporal information using covariance matrix modelling, and integrate time lag version of the EEG into its original EEG. The experimental results show both two strategies could enhance the performance of MSI [35], [36].

In previous studies, many advanced signal processing and machine learning methods have been introduced to recognize SSVEPs, such as the MEC, CCA and MSI [24]–[26], etc. Recently, the convolutional neural network has been used for SSVEP recognition with only five stimulus frequencies [37], its performance on the system with large number of targets has not been investigated. For the large number of visual flickers, two studies are worth to mention. In a 40-target BCI system, Chen *et al*. [8] proposed a method that combine the filter bank analysis and CCA with individual calibration data, the system reaches ITR up to 5.32 bits per second. In another study, a method based on the task-related component analysis (TRCA) was presented, which achieved an averaged ITR of 325.33 bits per minute in the cue-guided task [23]. The TRCA is a method that extracts task-related components efficiently by maximizing the reproducibility during the task period [38], TRCA was used to learn the spatial filters that could eliminate background unrelated activities from EEG.

Supervised learning with calibration dataset is an important direction to design new algorithms for SSVEP frequency detection, we could use the transfer learning to learn effective spatial filters from available data to improve the signal-to-noise ratio (SNR) of the signals. In the current study, we proposed an novel algorithm based on correlated component analysis (CORCA) for SSVEP detection. The CORCA was employed to design spatial filters that make the linear combinations of multiple channels consistent across trials and simultaneous maximally correlated between them. The proposed algorithm was evaluated on an open benchmark dataset for SSVEP-based BCI [39]. This dataset was collected with an 40-target BCI speller using JFPM coding paradigm. For comparison, we chose the TRCA-based method [23] as a baseline method, which has achieved promising results with the same JFPM coding system as this dataset.

The rest of the paper is organized as follows. Section II presents the materials and methods. Section III gives experimental results. Discussion and conclusion are given in the last two sections.

II. Methods and Materials

A. Benchmark Dataset

As introduced in [39], the dataset was collected from thirty-five healthy subjects (seventeen females, eighteen males, mean age: twenty-two years) in an offline 40-target BCI speller BCI experiment. The user interface contains 40 characters flickering at different frequencies (8-15.8 Hz with an interval of 0.2 Hz). For each subject, the experiment consisted of six blocks. Each block consisted of 40 trials corresponding to all 40 characters indicated in a random order. Each trial started with a visual cue (a red square) indicating a target stimulus. In each block, the cue appeared for 0.5 s on the screen. Subjects were asked to shift their gaze to the target as soon as possible within the cue duration. Following the cue offset, all stimuli started to flicker on the screen concurrently and lasted 5 s. After stimulus offset, the screen was blank for 0.5 s before the next trial began. Each trial lasted a total of 6 s. To facilitate visual fixation, a red triangle appeared below the flickering target during the stimulation period. In each block, subjects were asked to avoid eye blinks during the stimulation period. To avoid visual fatigue, there was a rest for several minutes between two consecutive blocks.

EEG data were acquired using a Synamps2 system (Neuroscan, Inc.) with a sampling rate of 1000 Hz from sixty-four channels that were placed on the standard positions according to the international 10-20 system. The ground electrode (GND) was placed on midway between Fz and FPz. The reference electrode was located on the vertex. The amplifier frequency passband ranged from 0.15 Hz to 200 Hz. Electrode impedances were kept below 10 $k\Omega$. To remove the common power-line noise, a notch filter at 50 Hz was applied in data recording. Event triggers generated by the computer to the amplifier and recorded on an event channel synchronized to the EEG data.

The continuous EEG data have been segmented into 6 s epochs (500 ms pre-stimulus, 5.5 s post-stimulus onset). The epochs were subsequently downsampled to 250 Hz. More detailed information about this dataset can be found in [39].

B. Data Preprocessing

All data epochs were band-pass filtered from 7 Hz to 90 Hz with an IIR filter for the subsequent analysis. Considering a latency delay in the visual system [39], the used data were extracted in [0.64 s $0.64+d$ s] from each epoch, where d was the time window length for frequency recognition.

C. TRCA-Based Method

This SSVEP detection method utilized the TRCA-based spatial filters. TRCA is a method can learn spatial filters

to extract task-related components by maximizing the reproducibility during the task period [38], [40]. Denote the EEG signals of the h-th trial as $\boldsymbol{X}^{(h)} \in \mathbb{R}^{Nc\times Ns}$, $h = 1, 2, \cdots, N_t$. Hereinafter, Nc, Ns and Nt denote the number of channels, the number of sampling points and the number of trials, respectively. A linear combination of the h-th trial is computed as:

$$\boldsymbol{y}^{(h)} = \boldsymbol{w}^T \boldsymbol{X}^{(h)} \tag{1}$$

The goal of TRCA is to optimize the weight vector $\boldsymbol{w} \in \mathbb{R}^{Nc\times 1}$ in such a way that temporal profiles of $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \cdots, \boldsymbol{y}^{(N_h)}$ exhibits a maximal temporal similarity among task trials. The weight vector can be obtained by maximizing the sum of covariance between all possible combinations of N_t signals obtained by formula (1) [38]. These all possible combinations of covariance are summed as:

$$\begin{aligned}
&\sum_{\substack{h_1,h_2=1\\ h_1\neq h_2}}^{N_t} Cov(\boldsymbol{y}^{(h_1)}, \boldsymbol{y}^{(h_2)}) \\
&= \sum_{\substack{h_1,h_2=1\\ h_1\neq h_2}}^{N_t} \sum_{j_1,j_2=1}^{N_c} w_{j_1} w_{j_2} Cov(X_{j_1,.}{}^{(h_1)}, X_{j_2,.}{}^{(h_2)}) \\
&= \boldsymbol{w}^T \boldsymbol{S} \boldsymbol{w}
\end{aligned} \tag{2}$$

Here, the $X_{j_1,.}{}^{(h_1)}$ and $, X_{j_2,.}{}^{(h_2)}$ denote j_1-th and j_2-th channel signals of $\mathbf{X}^{(h_1)}$ and $\mathbf{X}^{(h_2)}$, respectively. $Cov(\cdot,\cdot)$ indicates computing the covariance. Denote $\bar{\boldsymbol{X}} = [\boldsymbol{X}^{(1)}\ \boldsymbol{X}^{(2)}\ \cdots\ \boldsymbol{X}^{(N_t)}]$ be a concatenated matrix of all the training trials. In order to obtain a finite solution of $\mathbf{w}$, the constraint condition is that the variance of $\boldsymbol{y} = \boldsymbol{w}^T \bar{\boldsymbol{X}}$ is normalized to one as:

$$\begin{aligned}
Var(\boldsymbol{y}) &= \sum_{j_1,j_2=1}^{N_c} w_{j_1} w_{j_2} Cov(\bar{X}_{j_1,.}, \bar{X}_{j_2,.}) \\
&= \boldsymbol{w}^T \boldsymbol{Q} \boldsymbol{w} \\
&= 1.
\end{aligned} \tag{3}$$

The $\bar{X}_{j_1,.}$ and $\bar{X}_{j_1,.}$ denote j_1-th and j_2-th channel signals of the trial-aggregated data $\bar{\boldsymbol{X}}$, respectively. Then, the constrained optimization becomes as following:

$$\bar{\boldsymbol{w}} = arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^T \boldsymbol{S} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{Q} \boldsymbol{w}} \tag{4}$$

The optimal weight vector $\bar{\boldsymbol{w}}$ is the eigenvector of matrix $\boldsymbol{Q}^{-1}\boldsymbol{S}$ corresponding to the largest eigenvalue. For SSVEP-based BCI, the above method was used to obtain spatial filters to remove background activities from EEG data.

In the TRCA-based frequency recognition method, the spatial filters can be calculated with all the training trials with formula (4). For individual training data $\bar{\boldsymbol{X}}_n \in \mathbb{R}^{Nc\times Ns\times Nt}$, we can obtain N_f spatial filters, $\boldsymbol{w}_n \in \mathbb{R}^{N_c\times 1}$, $n = 1, 2, \cdots, N_f$. N_f is the number of stimulus frequencies. For a new test sample $\mathbf{X} \in \mathbb{R}^{Nc\times Ns}$, and the template signals $\boldsymbol{Y}_n \in \mathbb{R}^{Nc\times Ns}(n = 1, \cdots, N_f)$, we can calculate the correlation coefficient between the $\boldsymbol{X}$ and each template signal as:

$$\lambda_n = \rho(\boldsymbol{w}_n^T \boldsymbol{X}, \boldsymbol{w}_n^T \boldsymbol{Y}_n), \quad n = 1, 2, \cdots, N_f \tag{5}$$

where $\rho(\cdot,\cdot)$ indicates the one-dimensional correlation analysis computation between two signals.

Then, the frequency of the test sample $\mathbf{X}$ could be recognized by the following formula:

$$f_t = \max_n \lambda_n, \quad n = 1, 2, \cdots, N_f \tag{6}$$

Owing to the N_f spatial filters were similar to each other, the ensemble TRCA-based method was also proposed [23]. The authors integrated all spatial filters into a matrix $\mathbf{W} \in \mathbb{R}^{N_c\times N_f}$, whose columns correspond to the N_f spatial filters were obtained. Then, the correlation coefficients between a test data and a template signal for n-th visual stimulus are modified to be as:

$$\bar{\lambda}_n = \phi(\boldsymbol{W}^T \boldsymbol{X}, \boldsymbol{W}^T \boldsymbol{Y}_n), \quad n = 1, 2, \cdots, N_f \tag{7}$$

where $\phi(\cdot,\cdot)$ indicates the two-dimensional correlation analysis computation between two signals, which was implemented using the function 'corrcoef.m' in the MATLAB. Then, the frequency of the test sample $\boldsymbol{X}$ could be recognized by the same rule defined by formula (6).

D. CORCA-Based Method

In this subsection, we present a promising frequency detection method based on the correlated component analysis (CORCA) [41]–[44]. The CORCA can be employed to extract maximally correlated signal components of multiple EEG records from different subjects. In other words, it seeks to project the data of all subjects such that the resulting projections exhibit maximal intersubject correlation (ISC) across the subject pool. Denote $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots,$ and $\boldsymbol{X}_N$ be the data matrices of size $Nc \times Ns$ from N subjects. Here, N is the number of subjects. Let $I_i = \{I_{i1}, I_{i2}\} = \{(1, 2), (1, 3), \cdots, (N-1, N)\}$ denote the set of all $P = N \times (N-1)/2$ all possible combinations of subjects pairs. Then, we can define two subject-aggregated data matrices as:

$$\bar{\boldsymbol{X}}_1 = [\boldsymbol{X}_{I_{11}}\ \boldsymbol{X}_{I_{21}}\ \cdots\ \boldsymbol{X}_{I_{P1}}]; \tag{8}$$

$$\bar{\boldsymbol{X}}_2 = [\boldsymbol{X}_{I_{12}}\ \boldsymbol{X}_{I_{22}}\ \cdots\ \boldsymbol{X}_{I_{P2}}]; \tag{9}$$

With CORCA, we can find a weight vector $\mathbf{w} \in \mathbb{R}^{N_c\times 1}$ such that the projections $\boldsymbol{y}_1 = \bar{\boldsymbol{X}}_1^T \boldsymbol{w}$ and $\boldsymbol{y}_2 = \bar{\boldsymbol{X}}_2^T \boldsymbol{w}$ exhibit maximal correlation. Then, the ISC between the subject-aggregated data is defined as [42] and [43]:

$$\frac{\boldsymbol{y}_1{}^T \boldsymbol{y}_2}{\|\boldsymbol{y}_1\|\ \|\boldsymbol{y}_2\|} = \frac{\boldsymbol{w}^T \boldsymbol{R}_{12} \boldsymbol{w}}{\sqrt{\boldsymbol{w}^T \boldsymbol{R}_{11} \boldsymbol{w}}\sqrt{\boldsymbol{w}^T \boldsymbol{R}_{22} \boldsymbol{w}}} \tag{10}$$

Here, $\boldsymbol{R}_{12}$ is the pooled intersubject cross-covariance, $\boldsymbol{R}_{11}$ and $\boldsymbol{R}_{22}$ are the pooled intrasubject covariance matrices, respectively. Assuming $\boldsymbol{w}^T \boldsymbol{R}_{11} \boldsymbol{w} = \boldsymbol{w}^T \boldsymbol{R}_{22} \boldsymbol{w}$ [41], [42], the solution to maximize the ISC of equation (10) can be estimated through a generalized eigenvalue problem as following:

$$(\boldsymbol{R}_{12} + \boldsymbol{R}_{21})\boldsymbol{w} = \lambda(\boldsymbol{R}_{11} + \boldsymbol{R}_{22})\boldsymbol{w} \tag{11}$$

Where, $\boldsymbol{R}_{21}$ is the transposition of $\boldsymbol{R}_{12}$. These matrices are calculated as following:

$$\boldsymbol{R}_{12} + \boldsymbol{R}_{21} = \frac{1}{N(N-1)} \sum_j \sum_{i,i\neq j,} \boldsymbol{C}_{ji} \tag{12}$$

$$\boldsymbol{R}_{11} + \boldsymbol{R}_{22} = \frac{1}{N} \sum_j \boldsymbol{C}_{jj} \tag{13}$$

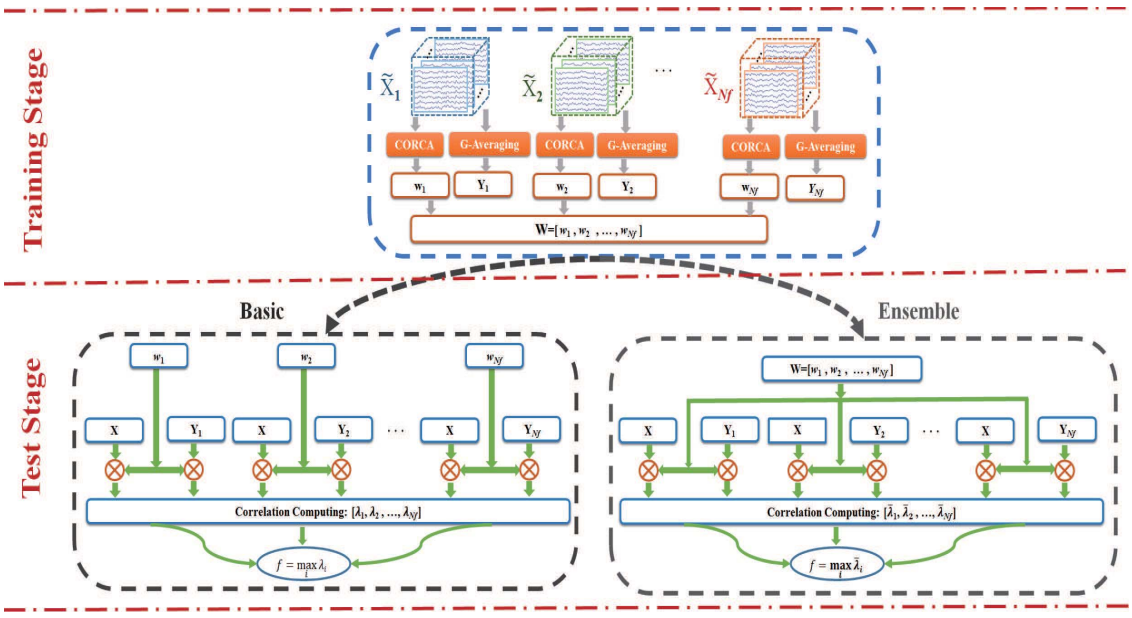

Fig. 1. Flowchart of the proposed CORCA-based method. For each subject, N_f training data corresponding to all the stimulus frequencies were available, $\widetilde{X}_i \in \mathbb{R}^{Nc\times Ns\times Nt}$, $i = 1, 2, \cdots, N_f$. In the training stage, we can use the CORCA to calculate the spatial filters for each frequency, i.e., $\boldsymbol{w_1}, \boldsymbol{w_2}, \cdots, \boldsymbol{w}_{N_f}$, and obtain the reference signals by group averaging across multiple training blocks, $Y_1, Y_2, \cdots, Y_{N_f}$. In the test stage, when the basic CORCA-based method was used, each single spatial filter w_i was used in calculating the correlation coefficient between the test signal and the reference Y_i ($i = 1, 2, \cdots, N_f$) by formula (15). When the ensemble CORCA-based method was used, all the N_f spatial filters were integrated together to a matrix $\boldsymbol{W}$, which was then used for the test signal and all reference signals to compute the correlation coefficients by formula (17). The SSVEP frequency is recognized by formula (16). G-Averaging denotes the group averaging across multiple training blocks.

The $\boldsymbol{C}_{ji}$ measures the cross-covariance of all electrodes in subject j with all electrodes in subject i, it is calculated as following:

$$\boldsymbol{C}_{ji} = \frac{1}{N_s}\boldsymbol{X}_j\boldsymbol{X}_i{}^T \tag{14}$$

$\boldsymbol{X}_j$ and $\boldsymbol{X}_i$ are the data matrices of subjects j and i, respectively.

The optimal component projection that captures the largest correlation between subjects is the eigenvector $\boldsymbol{w}$ of matrix $(\boldsymbol{R}_{11}+\boldsymbol{R}_{22})^{-1}(\boldsymbol{R}_{12}+\boldsymbol{R}_{21})$ with the maximal eigenvalues. More details about CORCA can be found in references [41]–[43].

Inspired by the principle of the CORCA for across subjects analysis, in this study, we generalized its application for SSVEP-based BCI to obtain the spatial filters which render the linear combinations of multiple channels SSVEP data consistent across trials and maximally correlated between them, i.e, maximal inter-trial correlation (ITC). Specifically, we used the CORCA to learn spatial filters for each stimulus frequency with the training data for each subject. With the spatial filters, we could obtain the projection signals of a single-trial test data and a template signal, and then calculate the correlation coefficient between the two signals as the classification feature. To the best of our knowledge, CORCA has never been introduced in any SSVEP-based BCI study before.

With individual training data $\widetilde{\boldsymbol{X}}_i \in \mathbb{R}^{Nc\times Ns\times Nt}$ of the i-th stimulus frequency of one subject, we can obtain spatial filters $\boldsymbol{w}_i \in \mathbb{R}^{Nc\times 1}$ with CORCA, $i = 1, 2, \cdots, N_f$. Denote $\boldsymbol{Y}_i \in \mathbb{R}^{Nc\times Ns}$ as the template signals which obtained by averaging across multiple training trials $\widetilde{\boldsymbol{X}}_i$, $i = 1, 2, \cdots, N_f$. When a test sample $\boldsymbol{X} \in \mathbb{R}^{Nc\times Ns}$ is available, a list of correlation coefficients can be computed between the test data and all the template signals as following:

$$\lambda_i = \rho(\boldsymbol{w}_i^T\boldsymbol{X}, \boldsymbol{w}_i^T\boldsymbol{Y}_i), \quad i = 1, 2, \cdots, N_f \tag{15}$$

The frequency of the test data is then recognized as the frequency of the template signals with the maximal correlation coefficients.

$$f_t = \max_i \lambda_i, \quad i = 1, 2, \cdots, N_f \tag{16}$$

In fact, the theoretical and computational analysis demonstrate that the obtained N_f spatial filters are similar to each other [23], [45]. The experimental results with TRCA-based method confirmed that integrating all the spatial filters could further improve the performance [23]. Therefore, a ensemble CORCA method was also presented here. With the N_f spatial filters $(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_{N_f})$, we concatenated them to obtain a matrix $\boldsymbol{W} \in \mathbb{R}^{Nc\times N_f}$, then computation of the correlation coefficient between a test data and a template signal for i-th visual stimulus is modified to be as:

$$\bar{\lambda}_i = \phi(\boldsymbol{W}^T\boldsymbol{X}, \boldsymbol{W}^T\boldsymbol{Y}_i), \quad i = 1, 2, \cdots, N_f \tag{17}$$

where $\phi(\cdot,\cdot)$ indicates the two-dimensional correlation analysis computation between two signals. Then, the frequency of the test trial $\boldsymbol{X}$ could be recognized by the formula (16). The diagram of the proposed methods is shown in Fig.1. Hereinafter, we termed this method using formula (15) as the basic CORCA-based method, and using formula (17) as the ensemble CORCA-based method.

E. Frequency Recognition With Filter Bank

In signal processing, filter bank methods have been widely used to analyze signals that exhibit multiple sub-band frequency components. A filter bank indicates an array of band-pass filters that separates an input signal into multiple sub-band components. The filter bank method has been applied in the development of BCI algorithms. For example, it was integrated with common spatial pattern to improve the motor imagery classification [46]. And, this technology was also introduced to enhance the SSVEP frequency recognition, such as the filter bank CCA which show significantly outperformed the standard CCA method [32]. Accordingly, we further explore the performance of CORCA under the filter bank framework. We will see that the standard CORCA based method is a special case of this framework.

To implement filter bank analysis, the band-pass filters are used to decompose SSVEP data into sub-band components so that independent information embedded in the harmonic components can be extracted more efficiently [30]. According to the study in reference [30], and the same paradigm was used for collecting benchmark dataset, the lower and upper cut-off frequencies of the m-th sub-band were set to $m \times 8$ Hz and 90 Hz, respectively. The zero-phase Chebyshev Type I Infinite impulse response (IIR) was used to extract each sub-band signals. Optimization of the filter banks may enhance the performance, but this is beyond the focus of this study.

With the individual training data, we run the same routes of CORCA-based method or TRCA-based method for each sub-band signals, independently. With all sub-bands, we can obtain a set of correlation vectors: $\beta_n = (\lambda_n^1, \lambda_n^2, \cdots, \lambda_n^{N_m})^T$, $n = 1, 2, \cdots, N_f$. Here, N_m is the total number of sub-bands. Then, β_n was merged to a single metric as the feature of the test data at the n-th frequency. The linear combination method was adopted in this study, i.e.,

$$\rho_n = \Phi^T \beta_n \tag{18}$$

According to the fact that the response of SSVEP harmonics decreases as considered frequency increases, the weights vector W was defined as follows:

$$\Phi = \begin{pmatrix} 1^{(-a)} \\ 2^{(-a)} \\ \vdots \\ N_m{}^{(-a)} \end{pmatrix} + b, \tag{19}$$

where a and b are two constant parameters that maximize the classification performance. Their values can be determined using a grid search method. Here, we empirically set these two parameters as: $a = 1.25$ and $b = 0.25$ respectively. Surely, the optimization of these two values can benefit the classification performance, which will explore in our future studies. In the end, the intended target can be identified by the rule provided in formula (16).

F. Performance Evaluation

Classification accuracy and ITR were used to evaluated the performance of all methods. For the estimation of ITR in current offline dataset, the 0.5 s gaze-shifting time was added when calculating the ITRs. This study calculated these two measurements with different data lengths (0.2 s to 1 s, with a step of 0.1 s). For comparison, the TRCA-based method was used to compare the performance with the proposed CORCA-based method. A leave-one-out cross validation was used to estimate classification accuracy and ITR. Specifically, the EEG samples from five blocks were used for learning the spatial filters while the samples from the left-out one block for validation. The procedure was repeated for six times for each subject to ensure each block serves as the validation set once.

Paired-sample t-test was implemented to investigate the statistical difference of accuracies obtained by CORCA-based and TRCA-based methods.

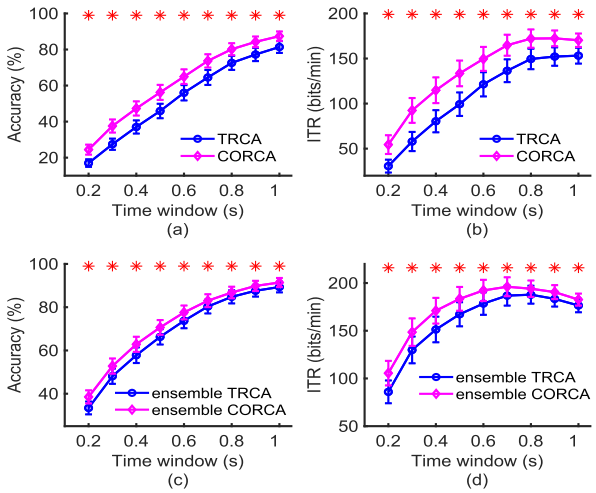

Fig. 2. The average accuracies and ITRs across all subjects obtained by the CORCA-based and TRCA-based methods (a-b), and the ensemble CORCA-based and TRCA-based methods (c-d), at various time windows, respectively. * indicates significant difference between two methods by paired t-tests ($p < 0.001$). The error bars indicate standard errors.

III. Results

Firstly, we investigated the results of the two methods without the filter bank. Fig. 2 shows the averaged accuracies and simulated ITRs across all the subjects with the basic and ensemble CORCA-based and TRCA-based methods at various time windows, respectively. The number of training blocks and the number of electrodes were 5 and 9, respectively. The ensemble methods yielded much better performance when combining the spatial filters of all stimulus frequencies, especially at the short time windows. The classification accuracy and the ITRs show consistent tendency. Furthermore, we investigated how the performance of the two methods varied with different numbers of training blocks and electrodes. During the calculation, the time

window was set to 0.8 s. As shown in Fig. 3, the performance was improved for both versions of CORCA-based and TRCA-based methods, i.e., basic and ensemble ones. By the statistical analysis with paired t-tests, we could find that the proposed CORCA-based method shows superiority compared to the TRCA-based method at all conditions, especially with

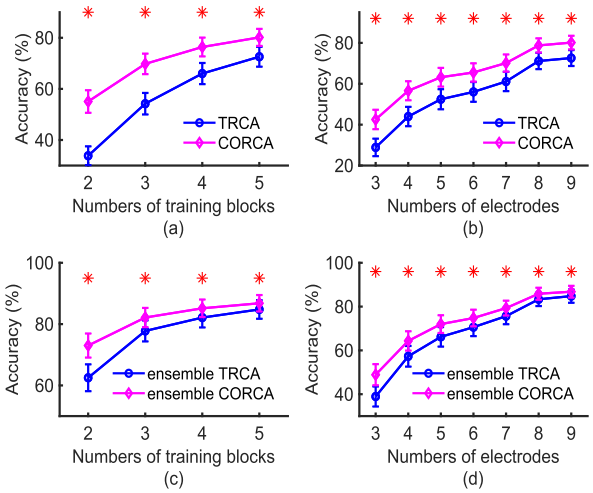

Fig. 3. The averaged accuracies with the different numbers of training blocks (a) and (c), different numbers of electrodes (b) and (d). The first and second rows present the results of basic and ensemble (CORCA-based and TRCA-based) methods, respectively. * indicates significant difference between two methods by paired *t*-tests ($p < 0.001$). The error bars indicate standard errors.

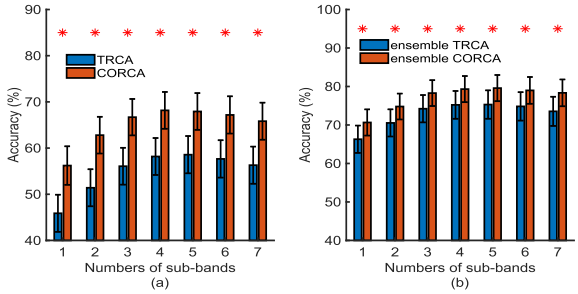

Fig. 4. The averaged accuracies with the different numbers of sub-bands. (a) basic CORCA-based and TRCA-based methods; (b) ensemble CORCA-based and TRCA-based methods. * indicates significant difference between two methods by paired *t*-tests ($p < 0.001$). The error bars indicate standard errors.

small numbers of electrodes and training blocks ($p < 0.001$). With the ensemble spatial filters, the difference between the two method become smaller than with the single spatial filter. In this study, the paradigm adopted 40 frequencies. Sometimes, the system may have much smaller number of stimulus frequencies, such as four frequencies [28], it seems that the basic CORCA-based method may be more effective in such system.

Secondly, the performance of the two methods using the filter bank was explored. The results show that the number of training blocks and the number of electrodes have an effect on the classification accuracy. Thus, we only investigated the number of sub-bands, and set these two numbers to 5 and 9. The time window was set as 0.5 s. In Fig. 4, the highest accuracy was obtained when this numbers of sub-bands were 4 and 5 for the two basic versions, and both 5 for two ensemble versions of CORCA-based and TRCA-based methods. Then, we set the numbers of sub-bands to be 5, and investigated how the accuracies of these methods vary with the various time windows. The results are presented in Figs. 5 and 6. As we expected, the filter bank can further enhance the performance

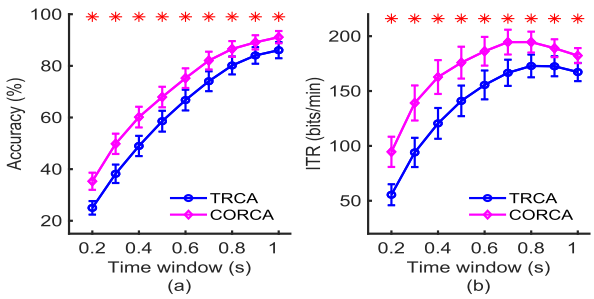

Fig. 5. The average accuracies (a) and ITRs (b) across all subjects obtained by the basic CORCA-based and TRCA-based methods with filter bank at various time windows, respectively. * indicates significant difference between two methods by paired *t*-tests ($p < 0.001$). The error bars indicate standard errors.

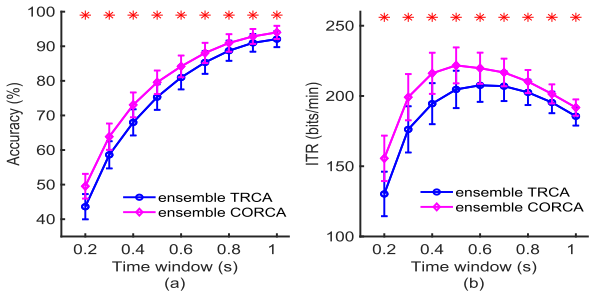

Fig. 6. The average accuracies (a) and ITRs (b) across all subjects obtained by the ensemble CORCA-based and TRCA-based methods with filter bank at various time windows, respectively. * indicates significant difference between two methods by paired *t*-tests ($p < 0.001$). The error bars indicate standard errors.

of the two methods. The proposed CORCA-based method also outperformed the TRCA-based method.

Overall, the extensive experimental evaluation demonstrates the superiority of the CORCA-based method compared to TRCA-based method at different settings.

IV. Discussions

Enhancing the performance of BCIs is a critical issue for various possible applications. In this study, we proposed the CORCA-based frequency recognition method for SSVEP-based BCIs. Its better performance may attribute to the effective spatial filters learned during the training stage. With the spatial filters, the EEG background artifacts were effectively removed during computing the weighted sum of the signal across all electrodes, then enhancing SNR and discriminability of the signals [29]. The CORCA and TRCA both learn spatial filters with training data by solving the generalized eigenvalue problem and the constrained optimization problem in formulas (11) and (4), respectively. Their spatial filters can be employed to improve the SNR of SSVEP signals. For the TRCA, it seeks to find spatial filters by maximizing the sum of the covariance of all possible combinations of trials, but each trial was used separately as shown in formula (1). For the CORCA, it seeks to find spatial filters by maximizing ITC, all trials are aggregated together as shown in formulas (8-10). This technique can make the linear combinations of EEG signals consistent across trials and simultaneous maximally correlated between them [41], [42]. The theoretical

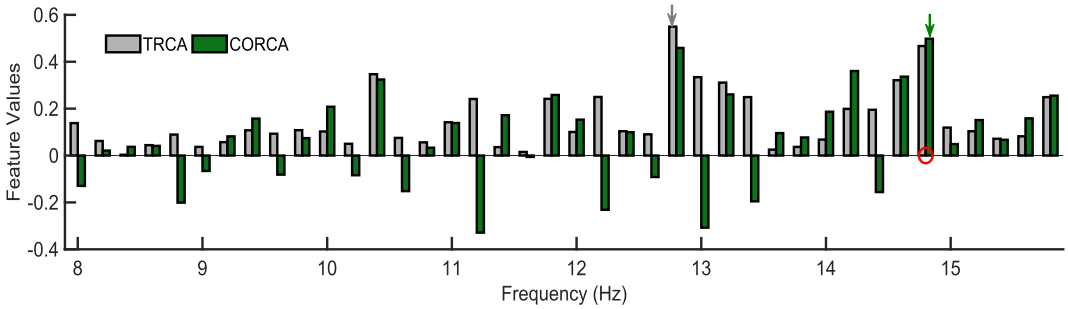

Fig. 7. The feature values of a test sample obtained with basic CORCA-based and TRCA-based method without filter bank. The two colored arrows indicate the frequency locations with the two methods. The red cirlce indicates the true frequency location of this sample. The time window was 1 s.

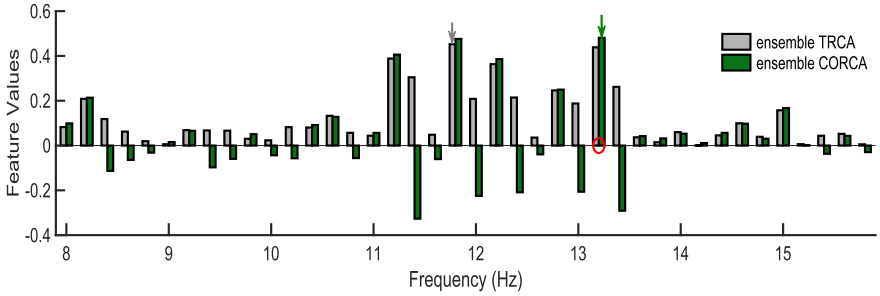

Fig. 8. The feature values of a test sample obtained with ensemble CORCA-based and TRCA-based method without filter bank. The two colored arrows indicate the frequency locations with the two methods. The red cirlce indicates the true frequency location of this sample. The time window was 1 s.

design of CORCA could guarantee this method to learn robust spatial filters, then improve the SNR of the projected signals.

As a representative example, Figs. 7 and 8 further present feature values calculated with basic and ensemble CORCA-based and TRCA-based methods without filter bank. The time window was 1 s, and the number of training blocks and the number of electrodes were 5 and 9, respectively. The largest feature value of CORCA-based method was obtained just right on the location of the true frequency, whereas that of the TRCA-based method was shifted to another location. Furthermore, we adopted the r-square to investigate the discriminability of features obtained by the basic and ensemble CORCA-based and TRCA-based methods. The r-square was computed with the feature values corresponding to the target stimulus and the maximal feature values corresponding to non-target stimuli [23]. An example of r-square values for SSVEPs at 8.2 Hz for each method (data length 0.8 s) is presented in Fig. 9. Paired t-tests showed the r-squares of the CORCA-based methods are significantly larger than those of the TRCA-based methods. Compared with the basic methods, the ensemble methods could enhance the discriminability of the features. It demonstrates that the proposed method could provide effectively discriminative features, and may explain why it yields better performance than TRCA-based method. In the current study, the performance of CORCA-based method was test on an offline datasets, online validation will be investigated in future studies.

Although there exist many other methods, the TRCA-based method was used as the only baseline method for the comparison, which achieved the best performance using the same paradigm. Besides, some parameters were not optimal, such

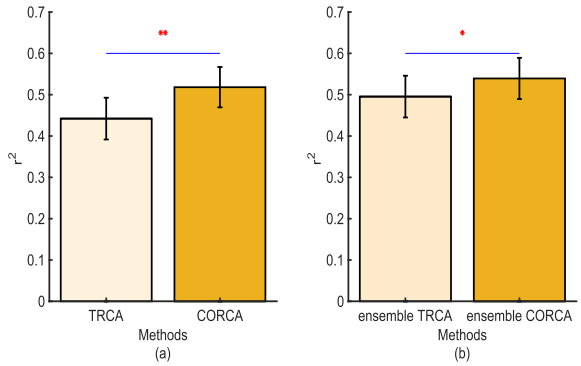

Fig. 9. An example of r-square values for SSVEPs at 8.2 Hz for each method (data length 0.8 s). (a) basic CORCA-based and TRCA-based methods; (b) ensemble CORCA-based and TRCA-based methods. * ($p < 0.05$) and ** ($p < 0.01$) indicates significant difference between two methods by paired t-tests. The error bars indicate standard errors.

as the phase interval value, a and b in formula (19), the performance could be further improved by carefully optimizing these parameters [32]. The optimization could be realized by grid-search which could be considered in the future studies.

Both CORCA-based and TRCA-based methods require individual training data, and more data could enhance the performance as shown in Fig.3. To some extent, the training stage for collecting efficient data is cumbersome, time-consuming, and may depress the subject owing to the visual fatigue caused by the flickering stimuli. Reducing the training time and enhance the performance is critical for both methods. One possible solution might resort to the transfer learning method to obtain the spatial filters with existing datasets

from other subjects. Considering the differences between subjects, the adaptively updating technology also needs to utilize in online operation. For the reference signals, sine-cosine reference signals are without subject-specific information, the methods can hardly yield the optimal performance, especially in short time windows. Some sophisticated methods, such as MsetCCA [28], multilayer correlation maximization model [29], could be used to generate the reference signals with available datasets, except for the group averaging across blocks. Besides, in current study, the resolution of the stimulus frequencies is 0.2 Hz in the experiment, and the proposed method could achieve satisfactory performance. In future, it needs to further investigate how the resolution of stimulus frequencies affects the performance of the method in subsequent studies.

V. Conclusion

In this study, a novel frequency recognition method with CORCA was proposed to enhance the performance of SSVEP-based BCI, which significantly outperformed the state-of-the-art method. The CORCA method effectively learned spatial filters from individual training data to remove EEG background artifacts of the multichannel EEG signals for later feature extraction. Experimental results based on a benchmark dataset including thirty-five subjects suggest that the proposed method yields significantly higher detection accuracies than the competing method, it holds the promise for achieving accurate frequency recognition in SSVEP-based BCI application.

Acknowledgement

The authors would sincerely like to thank Dr. Masaki Nakanishi for providing the TRCA codes and discussing the CORCA method in detail. They would also acknowledge Prof. Rui Zhang for valuable comments on an early version of this manuscript.

References

[1] S. Gao, Y. Wang, X. Gao, and B. Hong, "Visual and auditory brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1436–1447, May 2014.

[2] B. He, B. Baxter, B. J. Edelman, C. C. Cline, and W. W. Ye, "Noninvasive brain-computer interfaces based on sensorimotor rhythms," *Proc. IEEE*, vol. 103, no. 6, pp. 907–925, Jun. 2015.

[3] Y. Li *et al.*, "Multimodal BCIs: Target detection, multidimensional control, and awareness evaluation in patients with disorder of consciousness," *Proc. IEEE*, vol. 104, no. 2, pp. 332–352, Feb. 2016.

[4] X. Chen, Z. Chen, S. Gao, and X. Gao, "A high-ITR SSVEP-based BCI speller," *Brain-Comput. Interfaces*, vol. 1, nos. 3–4, pp. 181–191, 2014.

[5] E. Yin, Z. Zhou, J. Jiang, Y. Yu, and D. Hu, "A dynamically optimized SSVEP brain–computer interface (BCI) speller," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1447–1456, Jun. 2015.

[6] M. Nakanishi, Y. Wang, Y. Wang, Y. Mitsukura, and T. Jung, "A high-speed brain speller using steady-state visual evoked potentials," *Int. J. Neural Syst.*, vol. 24, no. 6, p. 1450019, 2014.

[7] H. Wang *et al.*, "Discriminative feature extraction via multivariate linear regression for SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 5, pp. 532–541, May 2016.

[8] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 44, pp. E6058–E6067, 2015.

[9] F. Qi, Y. Li, and W. Wu, "RSTFC: A novel algorithm for spatio-temporal filtering and classification of single-trial EEG," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3070–3082, Dec. 2015.

[10] Z. Qiu *et al.*, "Optimized motor imagery paradigm based on imagining Chinese characters writing movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 7, pp. 1009–1017, Jul. 2017.

[11] T. Uehara, M. Sartori, T. Tanaka, and S. Fiori, "Robust averaging of covariances for EEG recordings classification in motor imagery brain-computer interfaces," *Neural Comput.*, vol. 29, no. 6, pp. 1631–1666, 2017.

[12] J. Feng *et al.*, "Towards correlation-based time window selection method for motor imagery BCIs," *Neural Netw.*, vol. 102, pp. 87–95, Jun. 2018.

[13] X. Xie, Z. L. Yu, Z. Gu, J. Zhang, L. Cen, and Y. Li, "Bilinear regularized locality preserving learning on Riemannian graph for motor imagery BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 698–708, Mar. 2018.

[14] J. Jin *et al.*, "An adaptive P300-based control system," *J. Neural Eng.*, vol. 8, no. 3, p. 036006, Apr. 2011.

[15] J. Wang, Z. Gu, Z. Yu, and Y. Li, "An online semi-supervised P300 speller based on extreme learning machine," *Neurocomputing*, vol. 269, pp. 148–151, Dec. 2017.

[16] J. Jin, H. Zhang, I. Daly, X. Wang, and A. Cichocki, "An improved P300 pattern in BCI to catch user's attention," *J. Neural Eng.*, vol. 14, no. 3, p. 036001, 2017.

[17] Y. Zhang, P. Xu, T. Liu, J. Hu, R. Zhang, and D. Yao, "Multiple frequencies sequential coding for SSVEP-based brain-computer interface," *PLoS ONE*, vol. 7, no. 3, p. e29519, 2012.

[18] C. Jia, X. Gao, B. Hong, and S. Gao, "Frequency and phase mixed coding in SSVEP-based brain–computer interface," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 200–206, Jan. 2011.

[19] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu, "A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm," *J. Neural Eng.*, vol. 10, no. 2, p. 026012, 2013.

[20] Y. Li, J. Pan, F. Wang, and Z. Yu, "A hybrid BCI system combining P300 and SSVEP and its application to wheelchair control," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3156–3166, Nov. 2013.

[21] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu, "A speedy hybrid BCI spelling approach combining P300 and SSVEP," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 473–483, Feb. 2014.

[22] E. Yin, T. Zeyl, R. Saab, T. Chau, D. Hu, and Z. Zhou, "A hybrid brain–computer interface based on the fusion of P300 and SSVEP scores," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 693–701, Jul. 2015.

[23] M. Nakanishi, Y. Wang, X. Chen, Y.-T. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, Jan. 2018.

[24] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIS," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 6, pp. 1172–1176, Jun. 2007.

[25] O. Friman, I. Volosyak, and A. Graser, "Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 742–750, Apr. 2007.

[26] Y. Zhang, P. Xu, K. Cheng, and D. Yao, "Multivariate synchronization index for frequency recognition of SSVEP-based brain–computer interface," *J. Neurosci. Methods*, vol. 221, pp. 32–40, Jan. 2014.

[27] Y. Zhang, G. Zhou, J. Jin, M. Wang, X. Wang, and A. Cichocki, "L1-regularized multiway canonical correlation analysis for SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 6, pp. 887–896, Nov. 2013.

[28] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis," *Int. J. Neural Syst.*, vol. 24, no. 4, pp. 1450013-1–1450013-14, 2014.

[29] Y. Jiao, Y. Zhang, Y. Wang, B. Wang, J. Jin, and X. Wang, "A novel multilayer correlation maximization model for improving CCA-based frequency recognition in SSVEP brain-computer interface," *Int. J. Neural Syst.*, vol. 28, no. 4, p. 1750039, 2018.

[30] P. Yuan, X. Chen, Y. Wang, X. Gao, and S. Gao, "Enhancing performances of SSVEP-based brain–computer interfaces via exploiting inter-subject information," *J. Neural Eng.*, vol. 12, no. 4, p. 046006, 2015.

[31] G. R. Müller-Putz, R. Scherer, C. Brauneis, and G. Pfurtscheller, "Steady-state visual evoked potential (SSVEP)-based communication: Impact of harmonic frequency components," *J. Neural Eng.*, vol. 2, no. 4, p. 123, 2005.

[32] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface," *J. Neural Eng.*, vol. 12, no. 4, p. 046008, 2015.
[33] M. R. Islam, M. K. I. Molla, M. Nakanishi, and T. Tanaka, "Unsupervised frequency-recognition method of SSVEPs using a filter bank implementation of binary subband CCA," *J. Neural Eng.*, vol. 14, no. 2, p. 026007, 2017.
[34] H. Wang, "Temporally local maximum signal fraction analysis for artifact removal from biomedical signals," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4919–4925, Sep. 2010.
[35] Y. Zhang, D. Guo, P. Xu, Y. Zhang, and D. Yao, "Robust frequency recognition for SSVEP-based BCI with temporally local multivariate synchronization index," *Cognit. Neurodynamics*, vol. 10, no. 6, pp. 505–511, Dec. 2016.
[36] Y. Zhang, D. Guo, D. Yao, and P. Xu, "The extension of multivariate synchronization index method for SSVEP-based BCI," *Neurocomputing*, vol. 269, pp. 226–231, Dec. 2017.
[37] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLOS ONE*, vol. 12, no. 2, p. e0172578, 2017.
[38] H. Tanaka, T. Katura, and H. Sato, "Task-related component analysis for functional neuroimaging and application to near-infrared spectroscopy data," *NeuroImage*, vol. 64, pp. 308–327, Jan. 2013.
[39] Y. Wang, X. Chen, X. Gao, and S. Gao, "A benchmark dataset for SSVEP-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1746–1752, Oct. 2017.
[40] H. Tanaka, T. Katura, and H. Sato, "Task-related oxygenation and cerebral blood volume changes estimated from NIRS signals in motor and cognitive tasks," *NeuroImage*, vol. 94, pp. 107–119, Jul. 2014.
[41] J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra, "Correlated components of ongoing EEG point to emotionally laden attention—A possible marker of engagement," *Frontiers Hum. Neurosci.*, vol. 6, no. 6, p. 112, 2012.
[42] J. P. Dmochowski, M. A. Bezdek, B. P. Abelson, J. S. Johnson, E. H. Schumacher, and L. C. Parra, "Audience preferences are predicted by temporal reliability of neural processing," *Nature Commun.*, vol. 5, p. 4567, Jul. 2014.
[43] S. S. Cohen and L. C. Parra, "Memorable audiovisual narratives synchronize sensory and supramodal neural responses," *eNeuro*, vol. 3, no. 6, pp. e0203-16.2016-1–e0203-16.2016-11, 2016.
[44] J. J. Ki, S. P. Kelly, and L. C. Parra, "Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli," *J. Neurosci.*, vol. 36, no. 10, pp. 3092–3101, 2016.
[45] R. Srinivasan, F. A. Bibi, and P. L. Nunez, "Steady-state visual evoked potentials: Distributed local sources and wave-like dynamics are sensitive to flicker frequency," *Brain Topogr.*, vol. 18, no. 3, pp. 167–187, Mar. 2006.
[46] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *Int. J. Neural Syst.*, vol. 27, no. 2, p. 1650032, 2017.