TSINGHUA SCIENCE AND TECHNOLOGY
ISSNll1007-0214 02/10 pp262–270
DOI: 10.26599/TST.2018.9010010
Volume 24, Number 3, June 2019

Boosting the Information Transfer Rate of an SSVEP-BCI System Using Maximal-Phase-Locking Value and Minimal-Distance Spatial Filter Banks

Ke Lin, Shangkai Gao, and Xiaorong Gao*

Abstract: For Brain-Computer Interface (BCI) systems, improving the Information Transfer Rate (ITR) is a very critical issue. This study focuses on a Steady-State Visually Evoked Potential (SSVEP)-based BCI because of its advantage of high ITR. Unsupervised Canonical Correlation Analysis (CCA)-based method has been widely employed because of its high efficiency and easy implementation. In a recent study, an ensemble-CCA method based on individual training data was proposed and achieved an excellent performance with ITR of 267 bit/min. A 40-target SSVEP-BCI speller was investigated in this study, using an integration of Minimal-Distance (MD) and Maximal-Phase-locking value (MP) approaches. In the MD approach, a spatial filter is developed to minimize the distance between the training data and the reference sine signal, and in this study, two different types of distance were compared. In the MP approach, a spatial filter is developed to maximize the Phase-Locking Value (PLV) between the training calibration data and the reference sine signal. In addition to the fundamental frequency of stimulation, the harmonics were used to train MD and MP spatial filters, which formed spatial filter banks. The test data epoch was multiplied by the MP and MD spatial filter banks, and the distances and PLVs were extracted as features for recognition. Across 12 subjects with a 0.4 s-data length, the proposed method realized an average classification accuracy and ITR of 93% and 307 bit/min, respectively, which is significantly higher than the current state-of-the-art method. To the best of our knowledge, these results suggest that the proposed method has achieved the highest ITR in SSVEP-BCI studies.

Key words: SSVEP-BCI; Information Transfer Rate (ITR); spatial filter; distance; Phase-Locking Value (PLV)

1 Introduction

Brain-Computer Interfaces (BCIs) are pathways by which the brain communicates directly with external devices that extract brain signals; these interfaces provide a new way to communicate and have received increased attention in recent years[1–3]. Steady-State Visually Evoked Potential (SSVEP) is a widely used BCI technology which has a high classification accuracy and high Information Transfer Rate (ITR)[4–6]; however, further improvement of the ITR in SSVEP-BCI systems remains a critical concern. The unsupervised learning methods based on Canonical Correlation Analysis (CCA) achieve good performances in BCIs[7] especially SSVEP-BCI systems[5,8–10]. The CCA method is used to construct spatial filters that maximize the correlation coefficient between the multichannel electroencephalogram (EEG) signal and the reference signal. The reference signal is usually a combination of sinusoidal and cosine signals with the flickering stimulus frequency and its harmonics, and it is based on an unsupervised learning

• Ke Lin, Shangkai Gao, and Xiaorong Gao are with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China. E-mail: lincoln1128@gmail.com; gsk-dea@tsinghua.edu.cn; gxr-dea@tsinghua.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2017-06-27; revised: 2017-08-12; accepted: 2017-08-24

method. This CCA-based unsupervised method can be quickly and easily used because no training is required. Moreover, superior ITRs can be obtained by such methods. For example, the ITR of a speller proposed by Chen et al.[5] is up to 105 bit/min.

Recently, supervised learning methods based on training data in SSVEP-BCI systems have received much attention[6, 11–15]. In a recent study, both sine-cosine and training data were used as reference signal in CCA algorithm, and with the information provided by the training data, an ITR of 267 bit/min was achieved[6]; to the best of our knowledge, this is the highest ITR reported in BCI studies. Canonical correlation coefficients obtained by training data can also be used to train classifiers for testing[14]. A multi-way CCA proposed by Zhang et al.[11] applied tensor decomposition on training data to find the appropriate reference signals. An L1-regularized multi-way CCA was developed based on a multi-way CCA by imposing L1 regularization on the trial-way array optimization to improve system performance[12]. Zhang et al.[13] also proposed an Mset-CCA method which learns multiple linear transforms to maximize the overall correlation among canonical variates from multiple EEGs sets. A maximum evoked response spatial filter obtained from training data which extract validated SSVEP signals while suppressing the background EEG signals and noises was introduced to detect idle states in an asynchronous SSVEP-BCI system[15].

For the Visual Evoked Potential (VEP) signal and other evoked signals, the signal recorded by a single electrode in the scalp EEG is considered to be a weighted summation of each individual source signal on the cerebral cortex through the volume conductor effect[16, 17]. Therefore, it is a common method to obtain the task-related source signal by the space filtering method to solve the EEG inverse problem[18–24].

For the SSVEP signal, since it is a steady-state sinusoidal signal, a good spatial filter should make the spatial filtered source signal more similar to the sinusoidal signal with the same frequency. In addition, phase-based methods have also been introduced in SSVEP signal processing[25], in which the Phase-Locking Value (PLV) was used to measure the phase synchronization of two signals by their instantaneous phase difference. The PLV between an SSVEP signal and a reference complex signal with the estimated stimulus frequency was extracted as features to train a support vector machine classifier. The results showed that the PLV-based method outperformed the CCA method.

The aim of this study is to increase the ITR of a multi-frequency SSVEP-BCI system (40 frequencies in this study). Minimal-distance and maximal-PLV spatial filters with information from the amplitude and phase dimensions, respectively, were designed to extract the SSVEP. For the multi-channel EEG signal corresponding to a particular frequency, one spatial filter was developed to minimize the distance between the single channel data after weighted averaging and the reference sine signal, where the reference sine signal was the averaged data across trials. Two types of distances, Pearson's correlation distance and Euclidean distance, were used in this study. Another spatial filter was developed to maximize the PLV between the weighted spatial averaged single channel data and the reference sine signal. In addition to the fundamental frequency, spatial filters were also trained using harmonics. All the spatial filters formed Maximal-Phase-locking value (MP) and Minimal-Distance (MD) spatial filter banks. The classification accuracy and ITR were compared with a state-of-the-art method[6] to demonstrate the effectiveness of the proposed method.

2 Materials and Methods

2.1 Data acquisition

In this study, the online data from the state-of-the-art method[6] was used. Twelve healthy subjects (five females of 23–29 years old) participated in the offline experiment. The visual flickers were presented using the sampled sinusoidal stimulation method by the joint frequency-phase modulation method[6]. The frequency and initial phase of each stimulus are illustrated in Fig. 1. Each stimulus was rendered in a 140×140 pixel square. Electroencephalogram data were recorded at a sampling rate of 1000 Hz. Nine EEG electrode sites over parietal and occipital areas (Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, and O2) were selected. Electrode impedances were kept below 10 kΩ.

The experiment was separated into a training session and a test session. The training session consisted of 12 blocks and each block contained 40 trials, corresponding to each frequency. The training blocks were used to generate the spatial filters and derive SSVEP templates. The test session included five blocks and each block also contained 40 trials. Spatial filters and SSVEP templates obtained in the training session were used to classify the data recorded in each trial.

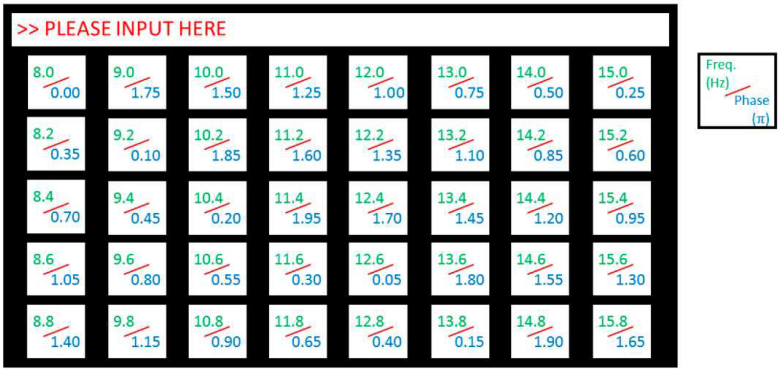

Fig. 1 Frequencies and initial phases for all stimuli[6].

Each trial lasted 1 s, which includes 0.5 s for visual stimulation and 0.5 s for gaze shifting and rest.

2.2 Data processing

Before analyzing the data, all data epochs were down-sampled to 250 Hz. The data epochs were also manually delayed for 0.14 s because of the latency delay in the visual pathway[26].

2.2.1 MP spatial filter

PLV was defined using Eq. (1)[25–28].

$$\mathrm{PLV}(\boldsymbol{x}_1,\boldsymbol{x}_2)=|E(\mathrm{e}^{\mathrm{j}\Delta\varphi_t})|=\left|\frac{1}{T}\sum_t(\mathrm{e}^{\mathrm{j}\Delta\varphi_t})\right| \tag{1}$$

Here, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two signals with T samples; $\Delta\varphi_t$ is the phase difference between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ at temporal sample t, and $E[\cdot]$ denotes the expected value. If $\boldsymbol{x}_2$ is a sine signal, the PLV is between 0 and 1, which is positively correlated with the instantaneous phase stability of $\boldsymbol{x}_1$. When PLV = 1, it indicates that the instantaneous phase of $\boldsymbol{x}_1$ is perfectly stable, and when PLV = 0, it indicates a totally chaotic signal.

For the multi-trial EEG signal $\widetilde{\boldsymbol{X}}=(\boldsymbol{X}_{(1)},\boldsymbol{X}_{(2)},\ldots,\boldsymbol{X}_{(L)})$ and reference sine signal $\boldsymbol{y}_2=\sin(\omega t)=\sin(2\pi ft)$, we aim to find a spatial filter $\boldsymbol{w}_p$ that maximizes the PLV between the spatial filtered data and reference, where L represents the number of trials. Therefore, the optimization criterion of the maximum PLV of the spatially filtered single channel data can be expressed by Formula (2):

$$\arg\max_{\boldsymbol{w}_p}\sum_l\mathrm{PLV}(\boldsymbol{w}_p\boldsymbol{X}_{(l)},\boldsymbol{y}_2) \tag{2}$$

where l is the index of trial.

The Hilbert transform was used to obtain the instantaneous phase, and a gradient descent algorithm can be used to optimize $\boldsymbol{w}_p$. Denoting $\boldsymbol{y}_{(l)}=\boldsymbol{w}_p\boldsymbol{X}_{(l)}$ and taking the derivative of Formula (2), after a series of simplifications, we obtain

$$\begin{aligned}\frac{\partial\sum_l\mathrm{PLV}(\boldsymbol{w}_p\boldsymbol{X}_{(l)},\boldsymbol{y}_2)}{\partial\boldsymbol{w}_p}=&\frac{1}{L\cdot T}\sum_l\Bigg[\Bigg(\Bigg(\sum_l\frac{A}{\sqrt{A^2+B^2}}\Bigg)^2+\\&\Bigg(\sum_l\frac{B}{\sqrt{A^2+B^2}}\Bigg)^2\Bigg)^{-\frac{1}{2}}\times\Bigg(\sum_l\frac{A}{\sqrt{A^2+B^2}}\cdot\\&\sum_l\frac{A'\sqrt{A^2+B^2}-A\frac{1}{\sqrt{A^2+B^2}}(A\cdot A'+B\cdot B')}{A^2+B^2}+\\&\sum_l\frac{B}{\sqrt{A^2+B^2}}\cdot\\&\sum_l\cdot\frac{B'\sqrt{A^2+B^2}-B\frac{1}{\sqrt{A^2+B^2}}(A\cdot A'+B\cdot B')}{A^2+B^2}\Bigg)\Bigg]\end{aligned} \tag{3}$$

Here, $A=\boldsymbol{y}_{(l)}(t)\sin(\omega t)-\hat{\boldsymbol{y}}_{(l)}(t)\cos(\omega t)$ and $B=\boldsymbol{y}_{(l)}(t)\cos(\omega t)+\hat{\boldsymbol{y}}_{(l)}(t)\sin(\omega t)$. $\hat{\boldsymbol{y}}_{(l)}(t)=\mathrm{HT}(\boldsymbol{y}_{(l)}(t))$. $\mathrm{HT}(\cdot)$ is the Hilbert transform.

$$A'=\frac{\partial A}{\partial\boldsymbol{w}_p}=\boldsymbol{X}_{(l)}(t)\sin(\omega t)-\hat{\boldsymbol{X}}_{(l)}(t)\cos(\omega t) \tag{4}$$

$$B'=\frac{\partial B}{\partial\boldsymbol{w}_p}=\boldsymbol{X}_{(l)}(t)\cos(\omega t)+\hat{\boldsymbol{X}}_{(l)}(t)\sin(\omega t) \tag{5}$$

The iterative refinement formula is given as

$$\boldsymbol{w}_{p,(i+1)}=\boldsymbol{w}_{p,(i)}+\alpha\frac{\partial\sum_l\mathrm{PLV}(\boldsymbol{w}_{p,(i)}\cdot\boldsymbol{X}_{(l)},\boldsymbol{y}_2)}{\partial\boldsymbol{w}_{p,(i)}} \tag{6}$$

Here, i denotes the iteration index, and α denotes the iteration step. A line search algorithm was applied with $\alpha=10^q, q=[-5,5]$, and the α with the largest increment in one step was used. When the variation between two iterations was lower than the bottom threshold 10^{-6} or the iteration time was greater than the upper threshold (1000), the iteration was stopped, and the current $\boldsymbol{w}_p$ was the optimal MP spatial filter.

2.2.2 MD spatial filter

To make the spatial-filtered single channel signal more similar to the sine template of the same initial phase,

we aim to obtain the minimum distance between the two signals. We use two common distances: Pearson correlation distance and Euclidean distance as MD_1 and MD_2, respectively.

(1) Pearson correlation distance (MD_1)

For multiple channel EEG data $\hat{\boldsymbol{X}} = (\boldsymbol{X}_{(1)}, \boldsymbol{X}_{(2)}, \ldots, \boldsymbol{X}_{(L)})$, where L represents the number of trials, we aim to find a spatial filter $\boldsymbol{w}_{d_1}$ that minimizes the Pearson correlation distance between the spatial filtered single channel data $\boldsymbol{y} = (\boldsymbol{y}_{(1)}, \boldsymbol{y}_{(2)}, \ldots, \boldsymbol{y}_{(L)})$ and the reference sine signal $\boldsymbol{y}_2 = \sin(\omega t) = \sin(2\pi f t)$. The optimization goal is shown as

$$\arg\max_{\boldsymbol{w}_{d_1}} \sum_l (\mathrm{Corr}(\boldsymbol{y}_{(l)}, \boldsymbol{y}_2)) = \arg\max_{\boldsymbol{w}_{d_1}} \frac{\sum_i (\boldsymbol{y}_{(l)}(i) - \hat{\boldsymbol{y}}_{(l)})(\boldsymbol{y}_2(i) - \hat{\boldsymbol{y}}_2)}{\sqrt{\sum_i (\boldsymbol{y}_{(l)}(i) - \hat{\boldsymbol{y}}_{(l)})^2} \sqrt{\sum_i (\boldsymbol{y}_2(i) - \hat{\boldsymbol{y}}_2)^2}} \quad (7)$$

Here, $\mathrm{Corr}(\cdot)$ denotes Pearson correlation coefficient and $\hat{\boldsymbol{y}}$ denotes the mean of $\boldsymbol{y}$. Equation (7) is the CCA between $X_{(l)}$ and $\boldsymbol{y}_2$. $\boldsymbol{w}_{d_1,(l)}$ can be obtained by solving the generalized eigenvalue. For multi-channel EEG data, the average value of $\boldsymbol{w}_{d_1,(l)}$ was the minimal Pearson correlation distance spatial filter.

$$\boldsymbol{w}_{d_1} = \frac{1}{L} \sum_l \boldsymbol{w}_{d_1,(l)} \quad (8)$$

(2) Euclidean distance (MD_2)

Similar as MD_1, the optimization goal of MD_2 was shown as

$$\arg\min_{\boldsymbol{w}_{d_2}} \sum_l \mathrm{Euc}(\boldsymbol{w}_{d_2} \boldsymbol{X}_{(l)}, \boldsymbol{y}_2) = \arg\min_{\boldsymbol{w}_{d_2}} \sqrt{\sum_l (\boldsymbol{w}_{d_2} \boldsymbol{X}_{(l)}(i) - \boldsymbol{y}_2(i))^2} \quad (9)$$

where $\mathrm{Euc}(\cdot)$ denotes Euclidean distance. Equation (9) can be solved by least square method. For multi-channel EEG data, the average value of $\boldsymbol{w}_{d,(l)}$ was the minimal Euclidean distance spatial filter.

$$\boldsymbol{w}_{d_2} = \frac{1}{L} \sum_l \boldsymbol{w}_{d_2,(l)} \quad (10)$$

Figure 2 shows the schematic diagram of MP, and MD_1 and MD_2.

2.2.3 MP&MD spatial filter bank

Each data epoch was z-scored before training MD and MP spatial filters. The initial iterative value for each MP spatial filter is the spatial filter obtained by

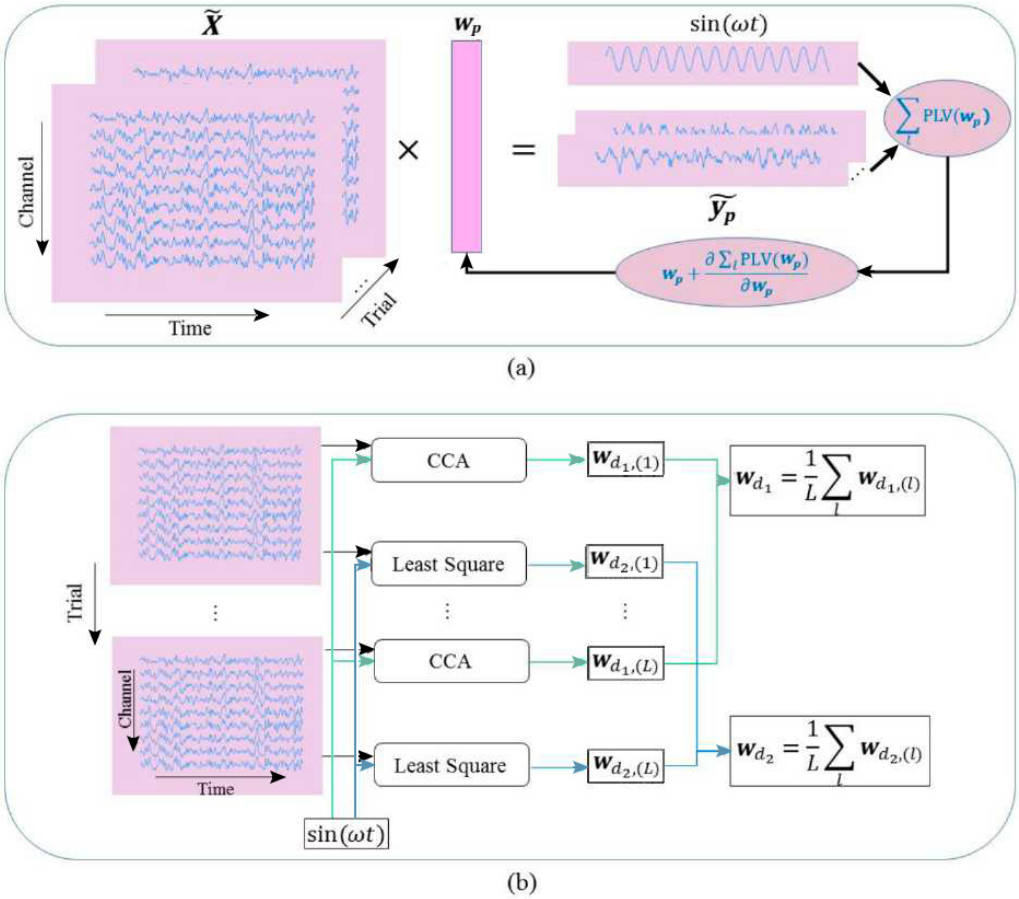

Fig. 2 Schematic diagram of the generation of (a) MP spatial filter and (b) two MD spatial filters.

cannonical correlation analysis between the training epochs and traditional sine-cosine references. The data used to train MP spatial filter was narrow band-pass filtered to the frequency range of interest. Once the iterative initial value was set, the MP spatial filters were obtained using the algorithm described in Section 2.2.1. Likewise, MD spatial filters can be obtained using the method described in Section 2.2.2. In addition to the fundamental frequency, the associated harmonics frequencies were also useful for classification[29–31]. Training data epochs $\tilde{\boldsymbol{X}}^{(f)}$ of frequency f is first band-pass filtered into N banks $(\tilde{\boldsymbol{X}}_1^{(f)},\ldots,\tilde{\boldsymbol{X}}_N^{(f)})$ using the similar method proposed in a previous study[30]. The passband frequency was $[8k-2, 90], k=1,\ldots,N$. N was 4 in this study based on the result obtained in Section 3.1. Using $(\tilde{\boldsymbol{X}}_1^{(f)},\ldots,\tilde{\boldsymbol{X}}_N^{(f)})$, we can obtain a series of MP filters $\boldsymbol{w}_{p,1}^{(f)},\ldots,\boldsymbol{w}_{p,N}^{(f)}$ and MD filters $\boldsymbol{w}_{d,1}^{(f)},\ldots,\boldsymbol{w}_{d,N}^{(f)}$. These spatial filters formed spatial filter banks. Averaging $(\tilde{\boldsymbol{X}}_1^{(f)},\ldots,\tilde{\boldsymbol{X}}_N^{(f)})$ by trial, we can obtain the template signal $\boldsymbol{Y}_1^{(f)},\ldots,\boldsymbol{Y}_N^{(f)}$. In the test session, test data epoch $\boldsymbol{X}$ was also filtered into different banks $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N$.

By multiplying $\boldsymbol{X}_k$ with $\boldsymbol{w}_{p,k}^i, i=1,2,\ldots,40$, we can obtain 40 single channel data. Likewise, by multiplying $\boldsymbol{Y}_k^{(i)}$ with $\boldsymbol{w}_{p,k}^i, i=1,2,\ldots,40$, we can also obtain 40 single channel data. By calculating the PLV between these two single channel data series, we can obtain $\gamma_k^{(i)}, i=1,2,\ldots,40$. A similar operation was performed with $\boldsymbol{w}_{d,k}^i, i=1,2,\ldots,40$, and the Pearson correlation distance (or Euclidean distance) between two single channel data series was obtained, denote as $\rho_k^{(i)}, i=1,2,\ldots,40$. By a weighted summation of $\gamma_k^{(i)}$ and $\rho_k^{(i)}$, we obtain

$$\tilde{\gamma}^{(i)}=\sum_{k=1}^{N}a(k)\cdot(\gamma_k^{(i)})^2 \tag{11}$$

$$\tilde{\rho}^{(i)}=\sum_{k=1}^{N}a(k)\cdot(\rho_k^{(i)})^2 \tag{12}$$

Here, $a(k)=k^{-1/2}$ based on the principle of power law. By adding $\tilde{\gamma}^{(i)}$ and $\tilde{\rho}^{(i)}$, we can obtain $\tilde{\sigma}^{(i)}=\tilde{\gamma}^{(i)}+\tilde{\rho}^{(i)}$. The frequency with the maximal value among $\tilde{\sigma}^{(i)}, i=1,2,\ldots,40$, was the final output. The classification procedure was illustrated as Fig. 3.

2.2.4 Evaluation of the proposed method

To evaluate the proposed MP&MD method, it was compared with the state-of-the-art method[6]. In addition to the standard CCA method, the state-of-the-art method combined correlation analysis between test data and SSVEP template signals. The detailed description of the ensemble-CCA method is omitted here and can be obtained in that study[6].

Classification accuracy and ITR are the two criterions used in this study. ITR was defined as follows:

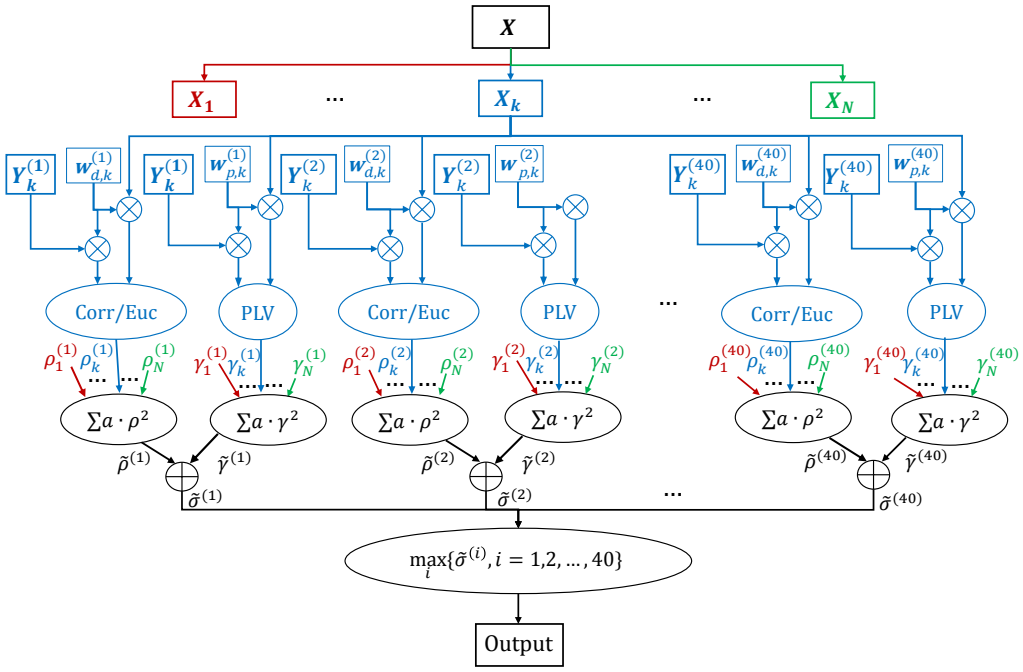

Fig. 3 Flowchart of the proposed MP&MD filter bank method in test session.

$$\mathrm{ITR}=\frac{60}{T}\left[\log_2 n+p\log_2 p+(1-p)\log_2\left(\frac{1-p}{n-1}\right)\right] \quad (13)$$

Here, n is the number of targets, p is the mean accurate classification rate, and T (seconds/target) is the time for a selection. In this study, $n = 40$, $T = 1$ s.

3 Results

3.1 Parameter optimization

Figure 4 shows the result of parameter optimization. Figure 4a shows the averaged classification accuracy using different spatial filter combinations (MD_1 + MP, MD_2 + MP) and different numbers of sub-bands with 0.4 s-data length. It can be seen that classification accuracy increased with N. When N was larger than 4, the classification accuracy became stable. The two filter combinations had no significant difference. The maximum accuracy was obtained with $N = 4$ and MD_2 + MP. In the following section, we compare with the state-of-the-art method, using MD_2 + MP and $N = 4$. Figure 4b shows the accuracies using MP, MD_1, MD_2, MD_1 + MP, and MD_2 + MP with $N = 4$. It can be seen that the accuracy using MD_i + MP ($i = 1, 2$) was significantly higher than that of MP or MD_i ($i = 1, 2$) using paired t-test (MD_i + MP vs. MP, $^{*}p < 0.05$; MD_i + MP vs. MD, $^{**}p < 0.01$). It can be concluded that both MP and MD_i are helpful in classification.

3.2 Recognition performance

Figure 5 shows the averaged classification accuracy and ITR of all subjects with different data lengths from 0.1 s to 0.6 s with an interval of 0.1 s. From Fig. 5a, it can be seen that the MD_2 + MP method achieved higher accuracies than the state-of-the-art method for

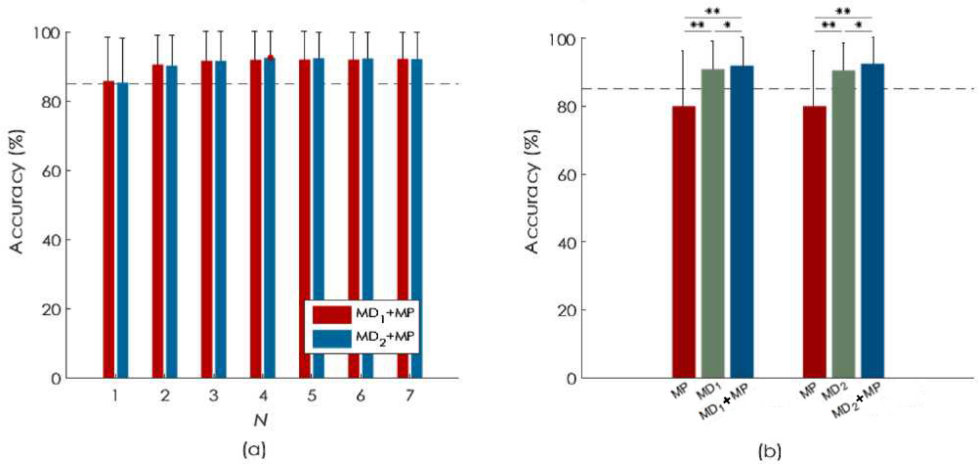

Fig. 4 (a) Classification accuracy using different spatial filter combinations (MD_1 + MP, MD_2 + MP) and different numbers of sub-bands. The red circle indicates the highest value; (b) Classification accuracy using different spatial filter combinations. ($^{*}p < 0.05$, $^{**}p < 0.01$). The date length was 0.4 s for both (a) and (b). The dashes represent the ensemble-CCA method accuracy.

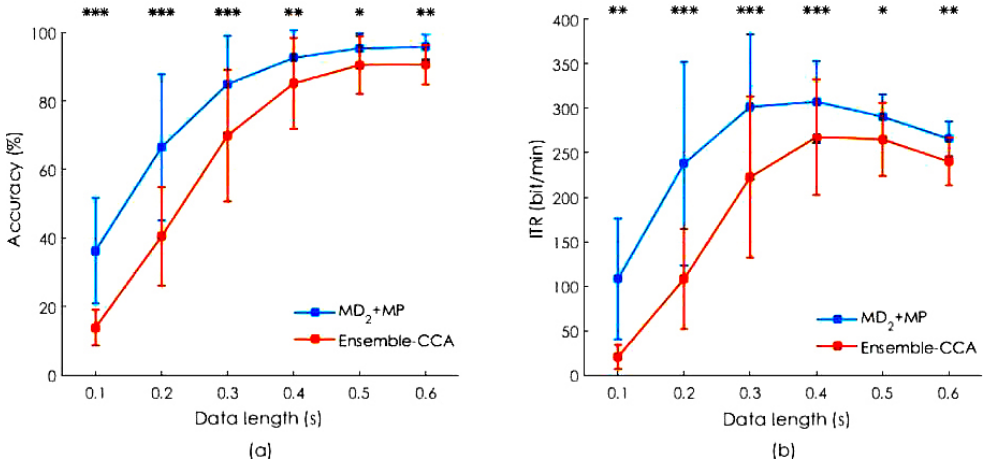

Fig. 5 Comparison of the classification accuracy (a) and ITR (b) of the proposed method and the state-of-the-art method. The asterisks indicate significant difference between the two methods by paired t-test ($^{*}p < 0.01$, $^{**}p < 10^{-3}$, and $^{***}p < 10^{-4}$).

all data lengths. The statistical significance between the two methods were $^{***}p < 10^{-4}$ for 0.1 s, 0.2 s, and 0.3 s; $^{**}p < 10^{-3}$ for 0.4 s and 0.6 s; and $^{*}p < 0.01$ for 0.5 s using paired t-test. From Fig. 4b, it can be seen that MD_2 + MP method achieved a significant higher ITR than the ensemble-CCA method for all data lengths. The statistical significances are $^{***}p < 10^{-4}$ for 0.2 s, 0.3 s, and 0.4 s; $^{**}p < 10^{-3}$ for 0.1 s and 0.6 s; $^{*}p < 0.01$ for 0.5 s. The above results demonstrate that the proposed method outperforms the state-of-the-art method.

The highest ITR achieved by the proposed method was 307.4 bit/min at data length of 0.4 s, which is much better than the state-of-the-art method with 267.7 bit/min.

The classification accuracy and ITRs of all subjects with 0.4 s-data length is given Table 1. The classification accuracy and ITRs of MD_2 + MP are all higher than those of ensemble-CCA for all subjects. The average classification accuracy was 92.6% for MD_2 + MP and 85.1% for ensemble-CCA. The average ITR was 307.4 bit/min for MD_2 + MP and 267.7 bit/min for ensemble-CCA.

4 Discussion

For BCI systems, increasing the ITR is a very important issue. In this study, we designed minimal-distance and maximal-PLV spatial filters to optimize the algorithm from the amplitude and phase dimensions, respectively. The results show that through the integration of MD and MP, the system performance was significantly improved than when only MD or MP system is used. In addition, by using the frequency harmonics information of SSVEP, we designed the MD and MP spatial filters of each frequency harmonic band to form spatial filter banks. The results show that the frequency harmonics information also contributes to the correct identification. The algorithm of this study achieved the highest ITR of SSVEP-BCI study at 307.4 bit/min, which is significantly higher than the 267.7 bit/min of the state-of-the-art method[6].

Table 1 Classification accuracy and ITR of all subjects with 0.4 s-data length.

Subject ID	Accuracy (%)		ITR (bit/min)	
	MD_2+MP	Ensemble-CCA	MD_2+MP	Ensemble-CCA
S1	98.5	97.5	342.0	334.7
S2	74.5	53.0	210.3	122.7
S3	100.0	96.5	354.8	327.9
S4	99.5	93.5	350.0	308.8
S5	87.0	79.5	271.8	233.8
S6	82.5	71.5	248.5	196.9
S7	93.5	85.5	308.8	263.9
S8	99.5	95.0	350.0	318.1
S9	93.0	87.0	305.7	271.8
S10	89.0	76.0	282.7	217.2
S11	99.0	97.0	345.9	331.3
S12	95.0	89.5	318.1	285.5
Mean	**92.6**	**85.1**	**307.4**	**267.7**
Std	8.0	13.3	46.1	65.0

This supervised learning-based approach used in this study is advantageous in the following areas compared to previous methods based on supervised learning or unsupervised learning. First, compared with CCA and other unsupervised learning methods[5,8,9], the algorithm was optimized specifically for each subject because of the introduction of training data. Second, compared with other supervised learning methods[14,15], the algorithm also introduced the phase information, and the spatial filter designed based on phase information makes the algorithm more comprehensive. The proposed fusion of amplitude and phase outperformed previous amplitude-only systems. Third, the proposed gradient descent-based approach can be used to design the optimization goal more specifically for SSVEP signal with stability, sinusoidal similarity, and phase-locking.

Since the data used in this study were from a previous study[6], we could not optimize some experimental parameters such as the number of channels, the frequency of stimulation, and the number of stimulus frequencies. The channel used in this study is the nine channel around the occipital area. Whether other electrodes will be beneficial may be discussed in the future. The stimulation frequencies used in this study were from 8 Hz to 15.8 Hz with an interval of 0.2 Hz. The use of higher stimulation frequencies to reduce the visual fatigue of the subjects may be explored in a future study[32]. To mitigate the misjudgment caused by the small frequency interval in this study, the interval may be increased in later studies. In addition, the number of stimulus frequencies may also be increased to further improve the ITR.

In this study, 12 samples were used for training. In future studies, the impact of the number of training samples on system performance may be investigated. For the online system, there could be a compromise between the training sample number and classification accuracy to meet practical application

needs. In addition, to reduce the training time, the data collected from this subject a few days ago or data from other subjects could also be used to train spatial filters. The idea of transfer learning[33, 34] or reinforcement learning[35, 36] could be applied to realize further optimization. Moreover, it can be found from the result of the proposed study that MD significantly outperforms MP, which inspired us to further optimize the MP filter to improve the overall performance. Furthermore, the two MD filters used in the study have similar performances. Other distances might be better than the proposed two distances, such as cosine similarity, Mahalanobis distance, K-L divergence, and Riemannian distance.

5 Conclusion

The MP&MD method, which is an integration of MD and MP approaches, was proposed in this study to boost the ITR of an SSVEP-BCI. In the MP approach, a spatial filter maximizes the PLV between the training data and a reference sine signal, while in the MD approach, a spatial filter minimizes the distance between the training data and a reference sine signal. In addition to the fundamental frequency of stimulation, the harmonics were used to train MD and MP spatial filters, which formed spatial filter banks. During the test session, the test data epoch was multiplied by the MP and MD spatial filter banks. The results show that the proposed method achieved a significant higher classification accuracy and ITR than the state-of-the-art method. The ITR obtained by this method was about 307.4 bit/min, which to the best of our knowledge, is the highest in SSVEP-BCI studies.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 61431007 and 91320202).

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, Brain-computer interfaces for communication and control, *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] S. K. Gao, Y. J. Wang, X. R. Gao, and B. Hong, Visual and auditory brain-computer interfaces, *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1436–1447, 2014.

[3] M. A. Lebedev and M. A. L. Nicolelis, Brain-machine interfaces: Past, present and future, *Trends Neurosci.*, vol. 29, no. 9, pp. 536–546, 2006.

[4] Y. J. Wang, X. R. Gao, B. Hong, C. Jia, and S. K. Gao, Brain-computer interfaces based on visual evoked potentials, *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 5, pp. 64–71, 2008.

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

[6] X. G. Chen, Y. J. Wang, M. Nakanishi, X. R. Gao, T. P. Jung, and S. K. Gao, High-speed spelling with a noninvasive brain-computer interface, *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 44, pp. 6058–6067, 2015.

[7] X. J. Wu, L. L. Zeng, H. Shen, M. Li, Y. A. Hu, and D. W. Hu, Blind source separation of functional MRI scans of the human brain based on canonical correlation analysis, *Neurocomputing*, vol. 269, pp. 220–225, 2017.

[8] G. Y. Bin, X. R. Gao, Z. Yan, B. Hong, and S. K. Gao, An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method, *J. Neural Eng.*, vol. 6, no. 4, p. 046002, 2009.

[9] K. Lin, A. Cinetto, Y. J. Wang, X. G. Chen, S. K. Gao, and X. R. Gao, An online hybrid BCI system based on SSVEP and EMG, *J. Neural Eng.*, vol. 13, no. 2, p. 026020, 2016.

[10] E. W. Yin, Z. T. Zhou, J. Jiang, Y. Yu, and D. W. Hu, A dynamically optimized SSVEP Brain-Computer Interface (BCI) speller, *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1447–1456, 2015.

[11] Y. Zhang, G. X. Zhou, Q. B. Zhao, A. Onishi, J. Jin, X. Y. Wang, and A. Cichocki, Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCIs, in *Neural Information Processing*, B. L. Lu, L. Q. Zhang, and J. Kwok, eds. Berlin, Germany: Springer, 2011, pp. 287–295.

[12] Y. Zhang, G. X. Zhou, J. Jin, M. J. Wang, X. Y. 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, 2013.

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

[14] P. Poryzala and A. Materka, Cluster analysis of CCA coefficients for robust detection of the asynchronous SSVEPs in brain-computer interfaces, *Biomed. Signal Process. Control.*, vol. 10, pp. 201–208, 2014.

[15] D. Zhang, B. S. Huang, W. Wu, and S. L. Li, An idle-state detection algorithm for SSVEP-based brain-computer interfaces using a maximum evoked response spatial filter, *Int. J. Neural Syst.*, vol. 25, no. 7, p. 1550030, 2015.

[16] J. Onton, M. Westerfield, J. Townsend, and S. Makeig, Imaging human EEG dynamics using independent component analysis, *Neurosci. Biobehav. Rev.*, vol. 30, no. 6, pp. 808–822, 2006.

[17] P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch, EEG coherency I: Statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales, *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 5, pp. 499–515, 1997.

[18] S. Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski, Independent component analysis of electroencephalographic data, in *Proc. 8th Int. Conf.*

Neural Information Processing Systems, Denver, CO, USA, 1996, pp. 145–151.
[19] A. Delorme, T. Sejnowski, and S. Makeig, Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis, *NeuroImage*, vol. 34, no. 4, pp. 1443–1449, 2007.
[20] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task, *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, 1999.
[21] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, Optimizing spatial filters for robust EEG single-trial analysis, *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.
[22] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, 2000.
[23] N. Xu, X. R. Gao, B. Hong, X. B. Miao, S. K. Gao, and F. S. Yang, BCI competition 2003-data set IIb: Enhancing P300 wave detection using ICA-based subspace projections for BCI applications, *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1067–1072, 2004.
[24] 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, 2013.
[25] K. Suefusa and T. Tanaka, Phase-based detection of intentional state for asynchronous brain-computer interface, in *Proc. 2015 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Brisbane, Australia, 2015, pp. 808–812.
[26] F. Di Russo and D. Spinelli, Electrophysiological evidence for an early attentional mechanism in visual processing in humans, *Vision Res.*, vol. 39, no. 18, pp. 2975–2985, 1999.
[27] S. Aydore, D. Pantazis, and R. M. Leahy, A note on the phase locking value and its properties, *NeuroImage*, vol. 74, pp. 231–244, 2013.
[28] J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, Measuring phase synchrony in brain signals, *Hum. Brain Mapp.*, vol. 8, no. 4, pp. 194–208, 1999.
[29] J. R. Hughes, Human brain electrophysiology, Evoked potentials and evoked magnetic fields in science and medicine, *Electroencephalogr. Clin. Neurophysiol.*, vol. 73, no. 1, p. 255, 1989.
[30] X. G. Chen, Y. J. Wang, S. K. Gao, T. P. Jung, and X. R. 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.
[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, pp. 123–130, 2005.
[32] Y. J. Wang, R. P. Wang, X. R. Gao, B. Hong, and S. K. Gao, A practical VEP-based brain-computer interface, *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 234–240, 2006.
[33] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, Self-taught learning: Transfer learning from unlabeled data, in *Proc. 24th Int. Conf. Machine Learning*, Corvalis, OR, USA, pp. 759–766, 2007.
[34] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.
[35] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction, *IEEE Trans. Neural Networks*, vol. 9, no. 5, p. 1054, 1998.
[36] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

Ke Lin received the BE degree from Nanjing University in 2012, and the PhD degree from Tsinghua University in 2017. His research interests focus on brain-computer interface, biomedical signal processing, and machine learning.

Xiaorong Gao received the BS degree from Zhejiang University in 1986, MS degree from Peking Union Medical College in 1989, and PhD degree from Tsinghua University in 1992. He is currently a professor of the Department of Biomedical Engineering, Tsinghua University. His current research interests are biomedical signal processing and medical instrumentation, especially the study of brain-computer interface.

Shangkai Gao graduated from Tsinghua University, Beijing, China, in 1970, and received the MEng degree in 1982 from Tsinghua University. She is now a professor of the Department of Biomedical Engineering in Tsinghua University. Her research interests include neural engineering and medical imaging, especially the study of brain-computer interface. She is now on the editorial board member of *Journal of Neural Engineering* and *Physiological Measurement*. She is a fellow of American Institute for Medical and Biological Engineering (AIMBE). She is now on the editorial board member of *IEEE Transactions on Biomedical Engineering*, as well as the senior editor of *IEEE Transactions on Neural System and Rehabilitation Engineering*.