Enhancing Detection of Multi-Frequency-Modulated SSVEP Using Phase Difference Constrained Canonical Correlation Analysis

Chi Man Wong, Ze Wang, *Member, IEEE*, Boyu Wang, *Member, IEEE*, Agostinho Rosa, Tzyy-Ping Jung, *Fellow, IEEE*, and Feng Wan, *Senior Member, IEEE*

***Abstract*— *Objective*: Multi-frequency-modulated visual stimulation scheme has been shown effective for the steady-state visual evoked potential (SSVEP)-based brain-computer interfaces (BCIs) recently, especially in increasing the visual target number with less stimulus frequencies and mitigating the visual fatigue. However, the existing calibration-free recognition algorithms based on the traditional canonical correlation analysis (CCA) cannot provide the merited performance. *Approach*: To improve the recognition performance, this study proposes a phase difference constrained CCA (pdCCA), which assumes that the multi-frequency-modulated SSVEPs share a common spatial filter over different frequencies and have a specified phase difference. Specifically, during the CCA computation, the phase differences of the spatially filtered SSVEPs are constrained using the temporal concatenation of the sine-cosine reference signals with the pre-defined initial phases. *Main results*: We evaluate the performance of the proposed pdCCA-based method on three representative multi-frequency-modulated visual stimulation paradigms (i.e., based on the multi-frequency sequential coding, the dual-frequency, and the amplitude modulation). The evaluation results on four SSVEP datasets (Dataset Ia, Ib, II, and III) show that the pdCCA-based method can significantly outperform the current CCA method in terms of recognition accuracy. It improves the accuracy by 22.09% in Dataset Ia, 20.86% in Dataset Ib, 8.61% in Dataset II, and 25.85% in Dataset III. *Significance*: The pdCCA-based method, which actively controls the phase difference of the multi-frequency-modulated SSVEPs after spatial filtering, is a new calibration-free method for multi-frequency-modulated SSVEP-based BCIs.**

***Index Terms*— Brain-computer interface, multi-frequency-modulated visual stimulation, phase difference constrained canonical correlation analysis, steady-state visual evoked potential.**

Manuscript received 8 September 2022; revised 10 December 2022; accepted 21 December 2022. Date of publication 8 February 2023; date of current version 24 February 2023. This work is supported in part by The Science and Technology Development Fund, Macau SAR (File no. 0045/2019/AFJ and 0022/2021/APD), The University of Macau Research Committee (MYRG projects 2017-00207-FST and 2022-00197-FST), The Natural Science Foundation of Guangdong Province, China (Grant No. 2023A1515010844); the Laboratory for Robotics and Engineering Systems—Fundação para a Ciência e a Tecnologia (LARSyS—FCT) Plurianual funding 2020–2023, and the Natural Sciences and Engineering Research Council of Canada (NSERC), Discovery Grants program. This work was performed in part at the high performance computing cluster (HPCC) which is supported by information and communication technology office (ICTO) of the University of Macau. *(Corresponding author: Feng Wan.)*

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Chi Man Wong, Ze Wang, and Feng Wan are with the Centre for Cognitive and Brain Sciences, the Centre for Artificial Intelligence and Robotics, Institute of Collaborative Innovation, and the Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau, Taipa, Macau (e-mail: fwan@um.edu.mo).

Boyu Wang is with the Department of Computer Science and the Brain Mind Institute, University of Western Ontario, London, ON N6A 5B7, Canada.

Agostinho Rosa is with ISR and DBE-IST, Universidade de Lisboa, 1649-004 Lisbon, Portugal.

Tzyy-Ping Jung is with the Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla, CA 92093 USA.

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2023.3243290, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3243290

I. Introduction

THE non-invasive brain-computer interface (BCI) using the electroencephalography (EEG) has demonstrated its great potential in real-life applications [1], such as neural rehabilitation [2], [3], cognitive monitoring [4], speller [5], and environmental controller [6]. Recently, the steady-state visual evoked potential (SSVEP)-based BCI is recognized as one of the promising non-invasive BCI paradigms for practical use because of its outstanding information transfer rate (ITR) [7].

In general, the SSVEP is evoked by a periodic visual stimulation whose stimulation frequency ranges from 1Hz to 100Hz [8]. As the SSVEP is frequency-locked and phase-locked to the visual stimulation, the BCI based on the SSVEP usually adopts the frequency coding approach to encode subject's intentions, i.e., the frequency-division multiple access (FDMA) scheme in the telecommunication system [7], [9]. The SSVEP-based BCI exhibits a number of visual targets (flickering at different frequencies) for subjects to select. Different visual targets (or flickering frequencies) represent different control commands or messages. When the subject gazes at one specified visual target, the corresponding frequency components (e.g., the fundamental and harmonic

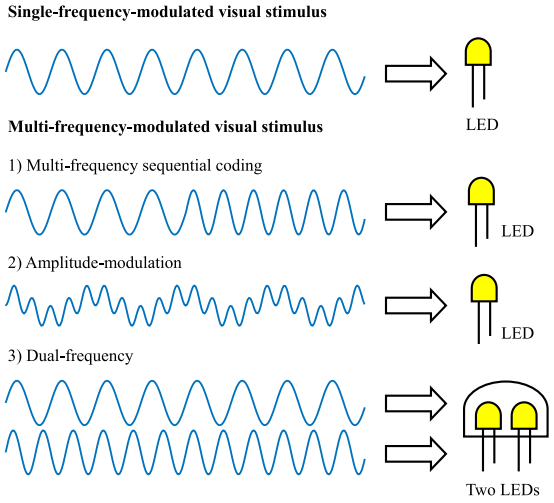

Fig. 1. The single-frequency-modulated and multi-frequency-modulated visual stimulation. The 'multi-frequency sequential coding', 'amplitude-modulation', and 'dual-frequency' are three examples.

frequencies of the flickering frequency) would be evoked in his/her measured SSVEP [10]. By checking the frequency information, it is possible to identify which visual target the subject wants to select. To date, the SSVEP-based BCIs can achieve the state-of-the-art performance among a variety of non-invasive BCIs [11], [12], [13].

In traditional SSVEP-based BCIs, a subject's single-trial SSVEP is encoded using only one stimulus frequency, which is called the single-frequency-modulated visual stimulation strategy. Apparently, N_{stim} visual targets occupy N_{stim} stimulus frequencies in this BCI. Empirically, the stimulus frequencies are selected from a specified frequency range (from 8 Hz to 16 Hz) and the frequency interval is not less than 0.2 Hz [9], [10], [11], [12]. Therefore, the number of visual targets is restricted in practice.

To address this issue, the multi-frequency-modulated visual stimulation has been introduced. In this case, a subject's single-trial SSVEP is encoded using more than one stimulus frequency. For example, the multiple frequencies are used to drive one flicker in different timeslots [14], [15], [16], [17], [18], drive one flicker at the same time [19], [20], and drive different flickers at the same time [21], [22], respectively, as illustrated in Fig. 1. The related studies reported that using the multi-frequency-modulated visual stimulation can bring the following advantages: 1) increase of the visual target number with the limited number of stimulus frequencies, 2) alleviation of the subject's fatigue, and 3) auto-calibration of the subject-specific phase lag. Specifically, in [14], Zhang et al. firstly introduced the multi-frequency sequential coding (MFSC) method. By adding the temporal coded information, the MFSC method could code a large number of targets with few frequencies. Recently, Chen et al. applied the MFSC method to implement a 160-target BCI speller with only eight frequencies [17]. Besides, two previous studies [19], [20] found that the flickers based on the frequency modulation (FM) or the amplitude modulation (AM) could induce lower subject's fatigue than using the traditional single-frequency-modulated visual stimulation. In fact, the subject's fatigue issue is a well-known problem in SSVEP-based BCIs [23], [24]. Moreover, Wu et al. showed that the multi-frequency-modulated visual stimulation can be used to calibrate the subject-specific phase lag automatically for the phase-coded SSVEP-BCIs in [25].

The canonical correlation analysis (CCA) is frequently used to identify the stimulus frequency from the single-frequency-modulated SSVEPs in the traditional SSVEP-based BCIs [26], [27], [28], [29]. The key idea in the CCA is to analyze the correlation coefficients between the SSVEP and the sine-cosine reference signals, which is ease-to-use and conveniently applied into many different scenarios. Importantly, the CCA is calibration-free. Therefore, the CCA has been extended from the single-frequency-modulated SSVEP recognition to the multi-frequency-modulated SSVEP recognition, in which the sine-cosine reference signals are modified from 'single-frequency' to 'multi-frequency' [14], [17], [19], [22], [30]. However, the phase features of the modified sine-cosine reference signals are always neglected, and thus the phase differences between different frequency components are ambiguous. It is speculated that the performance of the CCA using these modified reference signals might be non-optimal. The major reason is that the SSVEPs are phase-locked to the visual stimuli and thus their phase differences should be constrained.

For the traditional SSVEP-based BCIs using the single-frequency-modulated visual stimulation, Wong et al. proposed a new learning scheme to achieve an excellent recognition performance under two hypotheses: i) the phase difference between two SSVEPs evoked by two stimuli is dependent on the phase difference between these two stimuli, and ii) the SSVEPs evoked by different stimulus frequencies share a common spatial filter [31]. The key idea of the proposed learning scheme is to introduce a pre-defined phase difference among the sine-cosine reference signals and keep it constant during the CCA calculation, and eventually leading to a reliable recognition performance. Several following studies validated the superiority of the learning scheme in many different scenarios, such as in the supervised learning, the transfer learning, and the online learning scenarios [31], [32], [33], [34].

Here we assume that the multi-frequency-modulated SSVEPs also have the similar hypotheses (see Section II-B for details). Accordingly, the phase difference constrained CCA (pdCCA) is developed for the multi-frequency-modulated SSVEP-based BCIs. Unlike the method in [31], the pdCCA is calibration-free because the multi-frequency components can be elicited in each single-trial. To the best of our knowledge, this is the first study to utilize the phase difference constraint to enhance the recognition performance in the multi-frequency-modulated SSVEP-based BCIs. This study validates the superiority of the proposed pdCCA-based methods in SSVEP-based BCIs using three different multi-frequency-modulated visual stimulation approaches, namely 1) MFSC, 2) dual-frequency, and 3) AM, while compared to the traditional CCA method.

II. Materials and Methods

This study proposes a new CCA-based method for the multi-frequency-modulated SSVEP-based BCIs and also evaluates

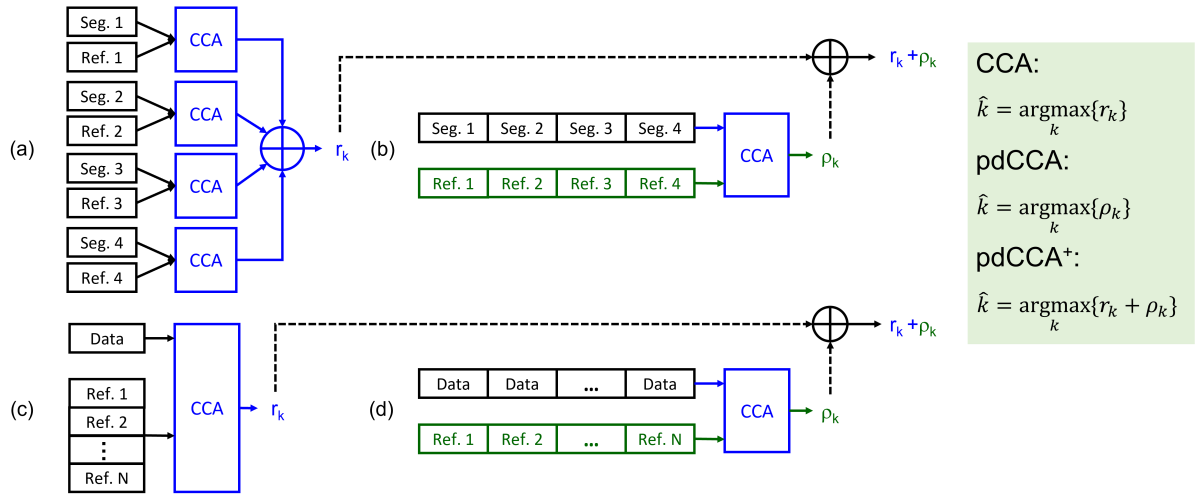

Fig. 2. The block diagrams of the CCA- and pdCCA-based frequency recognition methods for four datasets. (a) and (b) are the CCA-based and the pdCCA-based methods for Dataset Ia and Ib. (c) and (d) are the CCA-based and the pdCCA-based methods for Dataset II and III. The CCA, the pdCCA, and the pdCCA$^+$ utilize the features r_k, ρ_k, and $r_k + \rho_k$ in the final recognition, respectively (see (2), (4), (13) – (15)). Note that **Seg.** stands for 'segment' and **Ref.** stands for 'reference'.

TABLE I
TABLE OF NOTATIONS

Notation	Description
F_k, Φ_k	The stimulus frequency and phase for the k-th single-frequency-modulated visual stimulus ($k = 1, 2, \cdots, N_{\text{stim}}$)
$f_{k,j}, \phi_{k,j}$	The j-th stimulus frequency and phase for the k-th multi-frequency-modulated visual stimulus ($k = 1, 2, \cdots, N_{\text{stim}}$, $j = 1, 2, \cdots, N_f$)
N_{stim}	Number of visual stimuli
N_f	Number of frequencies for multi-frequency modulation
N_t	Number of trials
N_{fb}	Number of filter banks
N_{ch}	Number of channels
N_p	Number of sampling points
N_h	Number of harmonics
T_w	Data length, time-window length
$\mathbf{X}$	Single-trial data (Size: $N_p \times N_{ch}$)
$\mathbf{Y}_{F_k,\Phi_k}$ (or $\mathbf{Y}_k$)	Sine-cosine reference signal for the k-th stimulus (see (3). Size: $N_p \times 2N_h$)
$\mathbf{\Gamma}_{f,\phi}$	The component of $\mathbf{Y}_k$ (see (3). Size: $N_p \times 2$)

its performance on four multi-frequency-modulated SSVEP datasets. This section introduces these four SSVEP datasets. Then the idea of the pdCCA is presented (see Fig. 2). For convenience, Table I lists the notations in this study. Finally, the main content of the experimental study is presented.

A. Four SSVEP Datasets

1) Dataset Ia and Ib: Multi-Frequency Sequential Coding (MFSC) Visual Stimulation: In [17], Chen et al. used the MFSC approach to encode 160 visual targets in the SSVEP-based BCI experiment, in which each visual target is encoded with a unique sequential coding of four stimulus frequencies and phases. There were eight stimulus frequencies for the sequential coding, i.e. 8 Hz, 9 Hz, 10 Hz, 11 Hz, 12 Hz, 13 Hz, 14 Hz, and 15 Hz. The corresponding stimulus phases were 0, 0.5π, π, 1.5π, 0, 0.5π, π, and 1.5π. Due to the space limitation, the frequencies and phases of 160 visual targets based on the MFSC can be checked in Fig. S1 in the supplementary file. In the experimental study, they recruited eight (or twelve) subjects in the offline (or online) experiment. All subjects were instructed to gaze at 160 targets one by one in random orders and then repeated three times (or two times). In total, there were 480 trials in the offline experiment (or 320 trials in the online experiment). In each trial, the visual target flashed for 4 s, which could be divided into four 1-s segment and each segment was assigned one of eight frequencies. During the experiments, all subject data were recorded using a Neuroscan Synamps2 system through nine electrodes (Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, and O2). More details can be found in [17]. The corresponding datasets for offline and online experiments can be freely downloaded from the website of Tsinghua BCI group (http://bci.med.tsinghua.edu.cn/), which are termed as Dataset Ia and Dataset Ib in this study.

2) Dataset II: Dual-Frequency Visual Stimulation: Dataset II is the recorded data from the experiment 3 in [35], which is also downloadable from the website of Tsinghua BCI group. The dual-frequency approach was adopted to encode 40 targets, in which the stimulus frequencies were selected from 11 frequencies (i.e., 8.2 Hz, 9.0 Hz, 9.8 Hz, 10.6 Hz, 11.4 Hz, 12.2 Hz, 13.0 Hz, 13.8 Hz, 14.6 Hz, 15.4 Hz, and 16.2 Hz) and the stimulus phases were from four phases (i.e., 0, 0.5π, π, and 1.5π). The frequencies and phases of 40 visual targets based on the dual-frequency modulation can be checked in Fig. S2 in the supplementary file. In the experiment, twelve subjects were instructed to gaze at one of 40 targets for 2.25 s in random orders for eight rounds. Thereby, there were 320 trials in total. During the experiments, the EEG data over occipital area (i.e. Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, and O2) were recorded using a Neuroscan Synamps2 system. More details can be found in [35].

3) Dataset III: Amplitude-Modulated (AM) Visual Stimulation: Dataset III is the dataset from our experimental study. We designed an SSVEP-based BCI using the AM visual stimulation. According to [19], the AM signal can be imagined as either the product of two sinusoidal signals or the summation

TABLE II
THE FREQUENCIES AND PHASES OF THE SIX VISUAL TARGETS BASED ON AM (DATASET III)

	Target1	Target2	Target3	Target4	Target5	Target6
$f_1\|f_2$ (Hz)	10|9	13|12	16|15	10|9	13|12	16|15
$\phi_1\|\phi_2$ (rad)	0|0	0|0	0|0	π|0	π|0	π|0

of two sinusoidal signals, see (1). In our experiment, the visual stimulator consisted of six LEDs (or visual targets) driven by six AM signals using the mixed frequency and phase coding scheme. The frequencies and phases of these six AM signals are listed in Table II, in which three pairs of targets (i.e., Target1 vs. Target4, Target2 vs. Target5, Target3 vs. Target6) are driven by the same frequencies but different phases. We recruited 12 subjects (four females) to participate in this SSVEP-based BCI experiment. They were instructed to gaze at six visual targets one by one in random orders for seven times. In total, there were 42 trials. Each trial included 4 s visual stimulation time and 4 s rest time. During the experiments, their EEG signals (from O1, Oz, O2, POz, P3, Pz, and P4) were recorded using a g.USBamp amplifier (g.tec, Austria).

$$\begin{aligned} s(t) &= \sin(2\pi f_c t + \phi_c) \cdot \sin(2\pi f_m t + \phi_m) \\ &= -\frac{1}{2}\left[\cos(2\pi f_1 t + \phi_1) - \cos(2\pi f_2 t + \phi_2)\right], \end{aligned} \quad (1)$$

where $f_1 = f_c + f_m$, $f_2 = f_c - f_m$, $\phi_1 = \phi_c + \phi_m$, and $\phi_2 = \phi_c - \phi_m$.

B. Phase Difference Constrained CCA

According to [26] and [27], the CCA is utilized to compute the correlation coefficient between the single-frequency-modulated SSVEP data ($\mathbf{X}$) and different SSVEP reference signals ($\mathbf{Y}_k$ in (3)) after the projections. The correlation coefficient (r_k) can be obtained by solving the following optimization problem:

$$r_k = \underset{\mathbf{u},\mathbf{v}}{\text{argmax}} \frac{\mathbf{u}^\top \mathbf{X}^\top \mathbf{Y}_k \mathbf{v}}{\sqrt{\mathbf{u}^\top \mathbf{X}^\top \mathbf{X}\mathbf{u} \cdot \mathbf{v}^\top \mathbf{Y}_k^\top \mathbf{Y}_k \mathbf{v}}} = \text{CCA}(\mathbf{X}, \mathbf{Y}_k). \quad (2)$$

where $k = 1, 2, \cdots, N_{\text{stim}}$, $\mathbf{Y}_k$ is the SSVEP reference signal containing N_h pairs of the sine and cosine signals at N_h different frequencies and phases. More precisely, $\mathbf{Y}_k$ should be denoted as $\mathbf{Y}_{F_k,\Phi_k}$. For convenience, each pair of the sine and cosine signal can be denoted by $\Gamma_{h\cdot F_k, h\cdot\Phi_k}$ (see (3)), where $h = 1, 2, \cdots, N_h$.

$$\mathbf{Y}_{F_k,\Phi_k} = \begin{bmatrix} \sin(2\pi F_k t + \Phi_k) \\ \cos(2\pi F_k t + \Phi_k) \\ \vdots \\ \sin(2\pi N_h F_k t + N_h \Phi_k) \\ \cos(2\pi N_h F_k t + N_h \Phi_k) \end{bmatrix}^\top = \begin{bmatrix} \Gamma_{F_k,\Phi_k}^\top \\ \Gamma_{2\cdot F_k, 2\cdot\Phi_k}^\top \\ \vdots \\ \Gamma_{N_h F_k, N_h \Phi_k}^\top \end{bmatrix}^\top \quad (3)$$

Recently, Wong et al. proposed a learning across multi-stimulus scheme that learns from the SSVEP data corresponding to different frequencies to enhance the learning performance in the traditional SSVEP-based BCIs. Such an idea is under two assumptions: the SSVEP data corresponding to different frequencies i) share a common spatial filter, and ii) have the pre-defined phase difference [31]. Here we assume that the multi-frequency-modulated SSVEP also shares a common spatial filter over different modulation frequencies and different frequency components have the pre-defined phase difference. This motivates us to apply the learning across multi-stimulus scheme to analyze the multi-frequency-modulated SSVEPs, which can be formulated by:

$$\rho_k = \underset{\mathbf{u},\mathbf{v}}{\text{argmax}} \frac{\mathbf{u}^\top \mathcal{X}^\top \mathcal{Y}_k \mathbf{v}}{\sqrt{\mathbf{u}^\top \mathcal{X}^\top \mathcal{X}\mathbf{u} \cdot \mathbf{v}^\top \mathcal{Y}_k^\top \mathcal{Y}_k \mathbf{v}}} = \text{CCA}(\mathcal{X}, \mathcal{Y}_k), \quad (4)$$

where $\mathcal{X}$ is the temporal-concatenation of multi-frequency-modulated SSVEP data and $\mathcal{Y}_k$ is the temporal-concatenation of M sine-cosine reference signals for the k-th stimulus (M is the number of frequency components that the k-th multi-frequency-modulated visual stimulus evokes). Hereafter, (4) is termed as the phase difference constrained CCA (pdCCA).

1) Multi-Frequency Sequential Coding (MFSC) Visual Stimulation: According to [17], the input data is a 4-s data that can be divided into four 1-s segments: $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, and $\mathbf{X}_4$. The corresponding reference signals are $\mathbf{Y}_{f_{k,1},\phi_{k,1}}$, $\mathbf{Y}_{f_{k,2},\phi_{k,2}}$, $\mathbf{Y}_{f_{k,3},\phi_{k,3}}$, and $\mathbf{Y}_{f_{k,4},\phi_{k,4}}$. Then four correlation coefficients are computed by CCA($\mathbf{X}_j$,$\mathbf{Y}_{f_{k,j},\phi_{k,j}}$) and are synthesized for the final recognition, where $f_{k,j}$ and $\phi_{k,j}$ represent the stimulus frequency and phase of the j-th segment of the k-th MFSC stimulus, and $j = 1, 2, 3, 4$.

In this study, we analyze the 4-s data simultaneously, rather than analyze them separately. Specifically, the correlation coefficient is computed by CCA($\mathcal{X}$,$\mathcal{Y}_{\text{pdCCA}}$), where $\mathcal{X} = [\mathbf{X}_1^\top, \mathbf{X}_2^\top, \mathbf{X}_3^\top, \mathbf{X}_4^\top]^\top$. The key idea of the CCA and the pdCCA can be illustrated in Fig. 2 (a) and (b). Equations (5) and (6) describe the sine-cosine reference signal for the pdCCA and the pdCCA0, respectively. The pdCCA0 is the pdCCA with the zero phase setting, which is adopted in the comparison study in the offline data analysis (see Section II-C).

$$\mathcal{Y}_{\text{pdCCA}} = \left[\mathbf{Y}_{f_1,\phi_1}^\top, \mathbf{Y}_{f_2,\phi_2}^\top, \mathbf{Y}_{f_3,\phi_3}^\top, \mathbf{Y}_{f_4,\phi_4}^\top\right]^\top, \quad (5)$$

$$\mathcal{Y}_{\text{pdCCA0}} = \left[\mathbf{Y}_{f_1,0}^\top, \mathbf{Y}_{f_2,0}^\top, \mathbf{Y}_{f_3,0}^\top, \mathbf{Y}_{f_4,0}^\top\right]^\top. \quad (6)$$

For better understanding, the variables $\mathbf{Y}_{f_j,0}$ and $\mathbf{Y}_{f_j,\phi_j}$ can be replaced with $\mathbf{Y}_{f_{k,j},0}$ and $\mathbf{Y}_{f_{k,j},\phi_{k,j}}$ if they are dedicated for the k-th MFSC visual stimulus.

2) Dual-Frequency Visual Stimulation: Here, we assume that f_1 and f_2 are two stimulation frequencies used in the dual-frequency visual stimulation paradigm. According to the amplitude spectrum analysis in [35], the SSVEPs usually show strong responses at the fundamental and harmonic frequencies (nf_1 and nf_2), as well as the intermodulation frequency f_1+f_2. The other second-order and higher-order responses are relatively unstable and weak in many cases. Fig. 3 illustrates some representative examples. Hence we select the following frequency components to build up the sine-cosine reference signals: f_1, f_2, nf_1, nf_2, $f_1 + nf_2$, $nf_1 + f_2$, and $nf_1 + nf_2$, where $n = 1, 2, \cdots, N_h$.

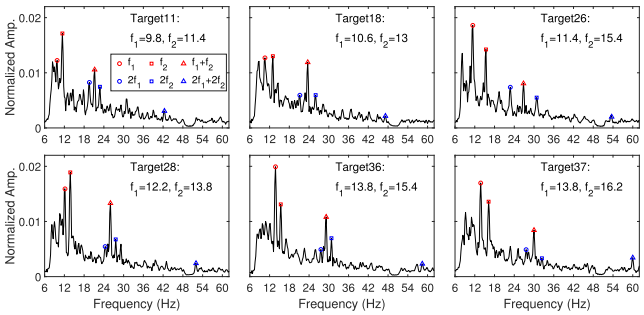

Fig. 3. The average amplitude spectrum of the EEG signals across 12 subjects in Dataset II. Note that only six amplitude spectrum corresponding to Target11, Target18, Target 26, Target28, Target36, and Target37 are presented. The frequency component f_1 is marked by the red circle, f_2 by the red square, and $(f_1 + f_2)$ by the red triangle. Their harmonics are depicted in blue colors.

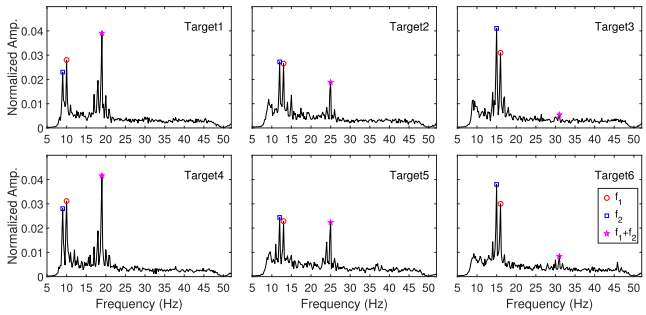

Fig. 4. The average amplitude spectrum of the EEG signals across 12 subjects in Dataset III. Note that the frequency component f_1 is marked by the red circle, f_2 by the blue square, and $(f_1 + f_2)$ by the magenta pentagram.

Given that the single-trial SSVEP data is $\mathbf{X}$, the correlation coefficients between $\mathbf{X}$ and the sine-cosine reference signals are computed by CCA($\mathbf{X}$,$\mathcal{Y}_{\text{CCA}}$) in the CCA approach and CCA($\mathcal{X}$,$\mathcal{Y}_{\text{pdCCA}}$) in the pdCCA approach, where $\mathcal{X} = [\mathbf{X}, \mathbf{X}, \cdots, \mathbf{X}]$, $\mathcal{Y}_{\text{CCA}}$ and $\mathcal{Y}_{\text{pdCCA}}$ are constructed using (7) and (8), as shown at the bottom of the page, respectively. Note that the phase differences between different sine-cosine components are kept constant in the pdCCA while the ones are not in the CCA [31]. $\mathcal{Y}_{\text{pdCCA0}}$ is for the pdCCA0, a special case of the pdCCA, which is included in the comparison study (see Section II-C for more information). For better understanding, the variables f_1, and f_2 can be replaced with $f_{k,1}$, and $f_{k,2}$ if they are dedicated for the k-th dual-frequency visual stimulus.

3) AM Visual Stimulation: According to Section II-A-III, f_1 and f_2 are two stimulation frequencies used in the AM visual stimulation paradigm. Fig. 4 shows that the frequency components at f_1, f_2, and $f_1 + f_2$ usually are predominant in frequency domain. Besides, the additional frequency components at the intermodulation frequencies, e.g., $f_1 \pm 1$, $f_2 \pm 1$, $f_1 \pm 2$ and $f_2 \pm 2$, are also existed, but they may cause interference in the recognition, e.g., the intermodulation frequency $2f_1 - f_2$ for Target1 is equal to $2f_2 - f_1$ for Target2. Therefore, in order to avoid the interference, we only select the frequency components of nf_1, nf_2, and $n(f_1 + f_2)$ in the sine-cosine reference signals, where $n = 1, 2, \cdots, N_h$.

Given that the single-trial SSVEP data is $\mathbf{X}$, the correlation coefficients between $\mathbf{X}$ and the sine-cosine reference signals are computed by CCA($\mathbf{X}$,$\mathcal{Y}_{\text{CCA}}$) in the CCA approach and CCA($\mathcal{X}$,$\mathcal{Y}_{\text{pdCCA}}$) in the pdCCA approach, where $\mathcal{X} = [\mathbf{X}, \mathbf{X}, \cdots, \mathbf{X}]$, $\mathcal{Y}_{\text{CCA}}$ and $\mathcal{Y}_{\text{pdCCA}}$ are generated using (10) and (11), respectively. $\mathcal{Y}_{\text{pdCCA0}}$ is for the pdCCA0, a special case of the pdCCA, which is included in the comparison study (see Section II-C for more information). For better understanding, f_1, and f_2 can be denoted as $f_{k,1}$, and $f_{k,2}$ if they are designed for the k-th AM visual stimulus.

$$\mathcal{Y}_{\text{CCA}} = \left[\mathbf{Y}_{f_1,0}, \mathbf{Y}_{f_2,0}, \mathbf{Y}_{f_1+f_2,0}\right], \tag{10}$$

$$\mathcal{Y}_{\text{pdCCA}} = \left[\mathbf{Y}^{\top}_{f_1,\phi_1}, \mathbf{Y}^{\top}_{f_2,\phi_2}, \mathbf{Y}^{\top}_{f_1+f_2,\phi_1+\phi_2}\right]^{\top}, \tag{11}$$

$$\mathcal{Y}_{\text{pdCCA0}} = \left[\mathbf{Y}^{\top}_{f_1,0}, \mathbf{Y}^{\top}_{f_2,0}, \mathbf{Y}^{\top}_{f_1+f_2,0}\right]^{\top}. \tag{12}$$

C. Offline Data Analysis

In the offline data analysis, we apply the same procedure to pre-process and analyze each trial of the EEG data from all datasets. The procedure can be described as below:

$$\mathcal{Y}_{\text{CCA}} = \left[\mathbf{Y}_{f_1,0}, \mathbf{Y}_{f_2,0}, \mathbf{Y}_{f_1+f_2,0}, \Gamma_{2\cdot f_1+f_2,0}, \cdots, \Gamma_{N_h\cdot f_1+f_2,0}, \Gamma_{f_1+2\cdot f_2,0}, \cdots, \Gamma_{f_1+N_h\cdot f_2,0}\right], \tag{7}$$

$$\mathcal{Y}_{\text{pdCCA}} = \begin{bmatrix} \Gamma^{\top}_{f_1,\phi_1} & \Gamma^{\top}_{f_2,\phi_2} & \Gamma^{\top}_{f_1+f_2,\phi_1+\phi_2} & \Gamma^{\top}_{f_1+f_2,\phi_1+\phi_2} & \Gamma^{\top}_{f_1+f_2,\phi_1+\phi_2} \\ \Gamma^{\top}_{2f_1,2\phi_1} & \Gamma^{\top}_{2f_2,2\phi_2} & \Gamma^{\top}_{2(f_1+f_2),2(\phi_1+\phi_2)} & \Gamma^{\top}_{2f_1+f_2,2\phi_1+\phi_2} & \Gamma^{\top}_{f_1+2f_2,\phi_1+2\phi_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Gamma^{\top}_{N_hf_1,N_h\phi_1} & \Gamma^{\top}_{N_hf_2,N_h\phi_2} & \Gamma^{\top}_{N_h(f_1+f_2),N_h(\phi_1+\phi_2)} & \Gamma^{\top}_{N_hf_1+f_2,N_h\phi_1+\phi_2} & \Gamma^{\top}_{f_1+N_hf_2,\phi_1+N_h\phi_2} \end{bmatrix}^{\top}, \tag{8}$$

$$\mathcal{Y}_{\text{pdCCA0}} = \begin{bmatrix} \Gamma^{\top}_{f_1,0} & \Gamma^{\top}_{f_2,0} & \Gamma^{\top}_{f_1+f_2,0} & \Gamma^{\top}_{f_1+f_2,0} & \Gamma^{\top}_{f_1+f_2,0} \\ \Gamma^{\top}_{2f_1,0} & \Gamma^{\top}_{2f_2,0} & \Gamma^{\top}_{2(f_1+f_2),0} & \Gamma^{\top}_{2f_1+f_2,0} & \Gamma^{\top}_{f_1+2f_2,0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Gamma^{\top}_{N_hf_1,0} & \Gamma^{\top}_{N_hf_2,0} & \Gamma^{\top}_{N_h(f_1+f_2),0} & \Gamma^{\top}_{N_hf_1+f_2,0} & \Gamma^{\top}_{f_1+N_hf_2,0} \end{bmatrix}^{\top}. \tag{9}$$

1) The data is filtered by N_{fb} bandpass filters to generate N_{fb} sub-band data for filter bank analysis [28]. The bandpass filter is the Chebyshev Type I infinite impulse response (IIR) filter and its lower cut-off and upper cut-off frequencies are denoted as f_L and f_U, respectively.
2) The filtered data is segmented from 0.14 s to 0.14+T_w s to exclude the latency delay (i.e., 0.14 s) caused by the human visual system [11], [36].
3) We calculate the correlation coefficients between the filtered data and the sine-cosine reference signals through the CCA (2) and the pdCCA (4). The resulting correlation coefficients are denoted as r_k and ρ_k (see Fig. 2), which are used for recognizing the stimulus frequency that the subject aims to select. Specifically, the recognition is performed by finding the maximum among N_{stim} coefficients or N_{stim} combined coefficients [27], [36], [11] based on (13)–(15), which may be termed as the CCA, the pdCCA, and the pdCCA$^+$ in this study. The pdCCA$^+$ is a combination of the CCA and the pdCCA.

$$\hat{k} = \underset{k}{\text{argmax}} \{r_k\} \tag{13}$$

$$\hat{k} = \underset{k}{\text{argmax}} \{\rho_k\} \tag{14}$$

$$\hat{k} = \underset{k}{\text{argmax}} \{r_k + \rho_k\} \tag{15}$$

4) When the filter-bank analysis approach is enabled, the resulting correlation coefficients for N_{fb} different sub-bands are weighted sum for frequency recognition. The weight value is designed as $w_{n_b} = n_b^{-1.25} + 0.25$ [28], and $n_b = 1, 2, \cdots, N_{fb}$. Thereby, (13)–(15) may be written as (16)–(18), in which r_{k,n_b} and ρ_{k,n_b} denote the correlation coefficients for the n_b-th sub-band data through the CCA and the pdCCA, respectively.

$$\hat{k} = \underset{k}{\text{argmax}} \{\sum_{n_b=1}^{N_{fb}} r_{k,n_b} \cdot w_{n_b}\} \tag{16}$$

$$\hat{k} = \underset{k}{\text{argmax}} \{\sum_{n_b=1}^{N_{fb}} \rho_{k,n_b} \cdot w_{n_b}\} \tag{17}$$

$$\hat{k} = \underset{k}{\text{argmax}} \{\sum_{n_b=1}^{N_{fb}} (r_{k,n_b} + \rho_{k,n_b}) \cdot w_{n_b}\} \tag{18}$$

To explore the importance of the phase difference constraints in the recognition performance, we compare the performance of the CCA, the pdCCA, and the pdCCA$^+$. In addition, the pdCCA with zero phase difference, i.e., the pdCCA0, is also explored. The main difference between the pdCCA and the pdCCA0 is the stimulus phases of the sine-cosine reference signals, see (5), (6), (8), (9), (11), (12). To sum up, there will be five methods in this comparison study, i.e., the CCA, the pdCCA0, the pdCCA, the pdCCA0$^+$, the pdCCA$^+$. Note that the pdCCA0$^+$ is the combination of the CCA and the pdCCA0.

To evaluate the recognition performance, we apply two commonly used performance indices: recognition accuracy (P) and ITR [1], [37]. They are computed by $P = N_c/N_t$ and $ITR = \frac{60}{T} \times [\log_2(N_{\text{stim}}) + P\log_2(P) + (1-P)\log_2(\frac{1-P}{N_{\text{stim}}-1})]$, where N_t is the number of trials for recognition, N_c is the number of correctly recognized trials, $T = T_w + T_s$ and $T_s = 0.5$ s for gaze shifting between two trials. Table III lists the key parameters in the offline data analysis. It can be noted that some of them may vary between datasets. In Dataset Ia and Ib, $N_h = 5$ and $N_{fb} = 6$ as we follow the setting in [17]. In Dataset II, $N_h = 3$ and $N_{fb} = 6$ because they can lead to the optimal accuracy of the CCA method as illustrated in Fig. 5. Similarly, in Dataset III, $N_h = 1$ and $N_{fb} = 3$ can achieve the highest CCA accuracy as exhibited in Fig. 5.

TABLE III
PARAMETER SETTINGS FOR DATASET IA, IB, II, AND III

Parameter	Dataset Ia and Ib	Dataset II	Dataset III
T_w (s)	0.4, 0.5, $\cdots$, 1.0	0.8, 1.0, $\cdots$, 2.0	0.5, 1.0, $\cdots$, 3.5
N_{ch}	9	9	7
N_{fb}	6	6	3
N_h	5	3	1
N_{stim}	160	40	6
N_f	4	2	2
N_t	480 (Ia), 320 (Ib)	320	42
$[f_L, f_U]$	[8·n_b, 90]	[8·n_b, 90]	[9·n_b, 90]

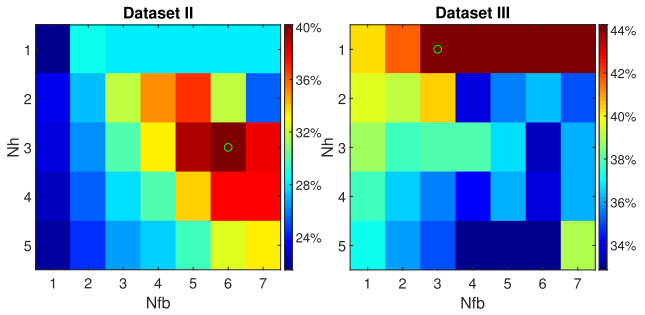

Fig. 5. The average accuracy of the CCA method under different N_h and N_{fb} for Dataset II (left panel) and Dataset III (right panel). The highest accuracy (in deep red) is marked by a green circle.

III. RESULTS

A. MFSC Visual Stimulation

Fig. 6(a) shows the average recognition accuracy of the CCA- and the pdCCA-based methods across all subjects on Dataset I. Besides, Table V shows the average recognition accuracy and ITR of the CCA- and the pdCCA-based methods for each subject across different T_w.

The accuracy comparison between any two methods is performed using the paired sample t-test. The corresponding statistical results after the Bonferroni correction are summarized in the second and third columns in Table IV. The pdCCA$^+$ accuracy is significantly higher than the CCA (Dataset Ia: $p < 0.001$; Dataset Ib: $p < 0.001$). In addition, when the phase difference constraints in the pdCCA$^+$ are disabled (e.g., all initial phases are set to zero), the resulting accuracies of the pdCCA0 and pdCCA0$^+$ suffer significantly. To sum up, the pdCCA and the pdCCA$^+$ outperform the CCA. Then, the pdCCA0 and pdCCA0$^+$ performs worse than the pdCCA and the pdCCA$^+$, even worse than the CCA.

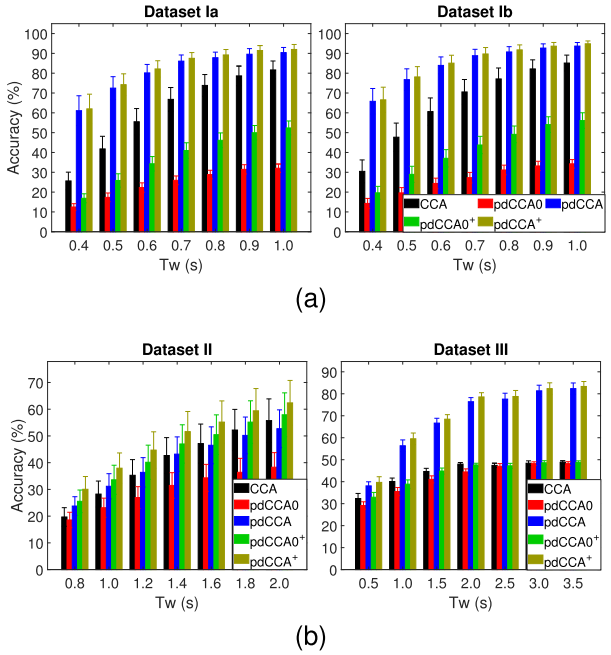

Fig. 6. The average recognition accuracy of the CCA- and the pdCCA-based methods on (a) Dataset Ia and Ib. (b) Dataset II and III. Note that the result is the average accuracy across all subjects under different T_w. The error bar indicates the standard error.

B. Dual-Frequency Visual Stimulation

First, the parameters N_h and N_{fb} are determined by a grid search method. The left panel of Fig. 5 shows the average CCA accuracy under different N_h ($1 \leq N_h \leq 5$) and N_{fb} ($1 \leq N_{fb} \leq 7$). Clearly, the CCA can achieve the highest accuracy when $N_h = 3$ and $N_{fb} = 6$. This setting would be used for both CCA and pdCCA-based methods.

Second, Fig. 6(b) shows the average recognition accuracy of the CCA- and the pdCCA-based methods across all subjects on Dataset II. It can be found that the pdCCA$^+$ and the pdCCA0$^+$ perform better than the CCA. In particular, when T_w is relatively small, the pdCCA accuracy is higher than the CCA accuracy. However, when T_w is relatively large, the pdCCA accuracy is no longer higher than the CCA accuracy. Table V shows the average performance of each subject across different T_w.

Third, a comparison study between any two methods is conducted using the paired sample t-test. The corresponding p-value after the Bonferroni correction is listed in the fourth column in Table IV. The pdCCA$^+$ accuracy is superior to the others. The pdCCA performs similarly to the CCA and the pdCCA0$^+$. Moreover, the pdCCA0 performs worse than the CCA.

C. AM Visual Stimulation

First, a grid search method is used to find the optimal parameters N_h and N_{fb} for the CCA method in the right panel of Fig. 5. It is evident that the parameters of $N_h = 1$ and $N_{fb} = 3$ can result in the highest CCA accuracy. This parameter setting would be used for the CCA method and the pdCCA-based methods in this experiment.

TABLE IV
ACCURACY COMPARISON RESULTS BASED ON THE PAIRED t-TEST WITH BONFERRONI CORRECTION

Method comparison (M1 vs. M2)	Dataset			
	Ia	Ib	II	III
pdCCA$^+$ vs. CCA	$>^{***}$	$>^{***}$	$>^{***}$	$>^{***}$
pdCCA$^+$ vs. pdCCA0	$>^{***}$	$>^{***}$	$>^{***}$	$>^{***}$
pdCCA$^+$ vs. pdCCA	$>^{*}$	$>^{*}$	$>^{*}$	$>^{**}$
pdCCA$^+$ vs. pdCCA0$^+$	$>^{***}$	$>^{***}$	$>^{***}$	$>^{***}$
pdCCA vs. CCA	$>^{***}$	$>^{***}$	N.S.	$>^{***}$
pdCCA vs. pdCCA0	$>^{***}$	$>^{***}$	$>^{***}$	$>^{***}$
pdCCA vs. pdCCA0$^+$	$>^{***}$	$>^{***}$	N.S.	$>^{***}$
pdCCA0$^+$ vs. CCA	$<^{***}$	$<^{***}$	$>^{*}$	N.S.
pdCCA0$^+$ vs. pdCCA0	$>^{***}$	$>^{***}$	$>^{**}$	$>^{***}$
pdCCA0 vs. CCA	$<^{***}$	$<^{***}$	$<^{*}$	$<^{***}$

1: $>^{***}$, $>^{**}$, and $>^{*}$ denote that M1 is better than M2 at different significant levels of 0.001, 0.01, and 0.05.
2: $<^{***}$, $<^{**}$, and $<^{*}$ denote that M1 is worse than M2 at different significant levels of 0.001, 0.01, and 0.05.
3: N.S. indicates no significant difference between M1 and M2.

Second, Fig. 6(b) shows the average recognition accuracy of the CCA- and the pdCCA-based methods across all subjects on Dataset III. Table V shows the individual performance of the CCA- and the pdCCA-based methods.

Third, the fifth column of Table IV shows the accuracy comparison results between any two methods based on the paired sample t-test. The pdCCA-based methods with the pre-defined phase difference constraints (i.e., pdCCA and pdCCA$^+$) perform better than the CCA. The pdCCA$^+$ outperforms the others. However, when the pdCCA-based methods are assigned the zero phase difference constraints (i.e., pdCCA0 and pdCCA0$^+$), their performance is apparently deteriorated, i.e., the pdCCA0 or the pdCCA0$^+$ cannot perform better than the CCA. As a summary, the pdCCA and the pdCCA$^+$ are superior to the CCA, pdCCA0, and pdCCA0$^+$. The CCA, pdCCA0, and pdCCA0$^+$ obtain the similar performance.

As expected, the CCA's accuracy in Table V and Fig. 6(b) is lower than 50% because the CCA is incapable of identifying the phase-coded feature without the offline calibration [38], [39]. In other words, the CCA cannot identify Target1/2/3 and Target4/5/6, and thus the maximum accuracy is 50%. On the contrary, the pdCCA can capture the phase difference features between the frequency components (e.g., f_1 and f_2), allowing for great accuracy. According to the experimental results, the average accuracy of the CCA is 44.25%, while the pdCCA is 68.40% and the pdCCA$^+$ is 70.10%.

IV. DISCUSSIONS

We introduce the pdCCA to take advantage of the multi-frequency-modulated SSVEP's phase difference feature for frequency recognition. The hypotheses are that the multi-frequency-modulated SSVEPs share a common spatial filter over different modulation frequencies and their phase differences are highly dependent on the phases of the multi-frequency-modulated visual stimulation. As stated in Section II, the optimization problem of the pdCCA is similar to the multi-stimulus CCA in [31], but there are substantial differences between them. At first, in the multi-stimulus CCA, the input is a temporal-concatenation of several single-frequency-modulated SSVEP data (e.g., M single-frequency-modulated SSVEP data). The multi-stimulus CCA uses the

TABLE V
AVERAGE RECOGNITION PERFORMANCE OF THE CCA AND THE PDCCA ON DATASET IA, IB, II, AND III (AVERAGE ACROSS TWS)

DS	Sub.	Accuracy (%)					ITR (bpm)				
		CCA	pdCCA0	pdCCA	pdCCA0$^+$	pdCCA$^+$	CCA	pdCCA0	pdCCA	pdCCA0$^+$	pdCCA$^+$
Ia	S1	74.14	28.90	94.58	45.71	94.91	83.49	22.45	126.31	42.02	126.99
	S2	81.04	33.42	92.29	53.33	92.80	96.91	28.03	122.56	52.80	123.74
	S3	69.11	22.41	89.24	35.04	89.96	75.83	16.06	115.10	29.39	116.58
	S4	67.59	29.14	82.68	45.83	85.60	72.80	22.78	100.76	42.19	106.10
	S5	44.85	17.02	65.06	27.59	66.70	40.41	10.52	70.83	20.64	73.21
	S6	51.90	25.15	75.63	37.53	78.66	49.42	18.22	86.69	31.68	92.09
	S7	58.18	21.01	81.93	33.48	83.57	58.16	14.18	98.64	27.05	101.63
	S8	37.59	17.80	67.26	26.07	68.96	31.40	11.11	72.49	18.97	75.21
	Mean	60.55	24.36	81.08	38.07	82.64	63.55	17.92	99.17	33.09	101.94
	SD	15.03	5.83	11.05	9.53	10.52	22.50	6.15	21.42	11.69	20.60
	p-val.	$***$	$***$	$*$	$***$	–	$***$	$***$	$*$	$***$	–
Ib	S1	77.19	31.56	90.09	50.18	91.83	89.32	25.55	116.38	48.07	120.04
	S2	84.60	37.14	94.29	59.51	95.09	103.37	32.40	125.91	61.41	127.62
	S3	32.19	14.91	57.90	23.53	60.18	25.18	8.54	58.18	16.32	61.48
	S4	51.56	20.00	89.38	30.00	88.35	48.77	13.24	113.75	23.07	111.48
	S5	70.00	29.64	92.54	45.54	93.17	76.54	23.24	121.10	41.55	122.58
	S6	43.08	19.11	67.72	28.13	70.04	38.03	12.24	73.70	21.05	77.37
	S7	49.60	19.42	73.39	31.25	75.54	46.36	12.59	82.81	24.37	86.47
	S8	73.84	24.73	95.67	39.91	95.76	82.96	17.92	128.58	34.60	128.81
	S9	47.23	17.32	78.13	24.91	78.57	43.37	10.92	91.85	18.11	92.73
	S10	84.15	37.72	93.35	59.29	93.88	103.01	33.50	124.49	61.55	125.54
	S11	92.32	38.57	98.44	59.55	98.44	120.30	34.94	136.97	62.38	136.95
	S12	72.10	25.67	84.69	43.17	87.28	80.27	18.90	104.11	38.77	109.35
	Mean	64.82	26.32	84.63	41.25	85.68	71.46	20.33	106.49	37.60	108.37
	SD	19.28	8.47	12.67	13.74	11.95	30.37	9.44	24.63	17.54	23.61
	p-val.	$***$	$***$	$*$	$***$	–	$***$	$***$	$*$	$***$	–
II	S1	55.40	42.86	49.51	63.04	65.31	61.57	41.22	51.74	76.20	80.72
	S2	53.26	46.61	68.75	62.90	70.67	57.69	46.82	87.47	75.42	91.38
	S3	52.14	26.16	34.64	50.00	55.71	55.76	18.26	28.91	52.17	62.06
	S4	5.09	2.77	2.46	3.88	3.71	0.49	0.08	0.07	0.19	0.16
	S5	67.01	41.12	50.36	69.60	74.06	83.97	38.79	53.40	89.63	99.34
	S6	56.56	35.40	55.13	58.35	65.36	63.60	30.50	61.52	67.03	80.90
	S7	64.60	53.84	67.77	73.26	76.07	78.95	59.50	85.67	97.54	104.04
	S8	21.34	27.95	38.17	29.38	32.10	12.86	20.58	34.01	22.13	25.66
	S9	7.10	8.93	11.74	9.64	10.85	1.36	2.45	4.20	2.78	3.57
	S10	29.96	29.60	44.15	38.62	47.14	22.78	22.49	42.70	34.48	47.45
	S11	25.76	16.47	27.10	26.92	32.86	17.74	8.12	19.55	19.14	26.80
	S12	43.48	26.92	37.01	45.63	51.21	41.72	19.20	32.61	45.12	54.33
	Mean	40.14	29.88	40.57	44.27	48.75	41.54	25.67	41.82	48.49	56.37
	SD	21.60	15.25	20.12	22.96	24.25	29.53	18.25	27.95	33.21	36.21
	p-val.	$***$	$***$	$*$	$***$	–	$**$	$**$	$*$	$***$	–
III	S1	44.56	42.18	73.47	43.54	74.83	7.75	6.41	28.01	7.02	29.42
	S2	47.28	43.20	73.81	47.62	75.51	9.56	6.82	29.20	9.88	31.36
	S3	47.96	44.56	71.09	47.62	71.09	10.37	7.86	25.88	9.99	26.00
	S4	47.62	45.92	75.17	47.62	76.19	9.85	8.78	30.08	9.85	31.31
	S5	46.94	47.28	68.71	47.96	70.07	9.29	9.56	24.61	10.16	25.87
	S6	40.14	35.03	62.93	39.12	67.35	5.44	3.50	18.64	4.93	22.28
	S7	46.60	45.92	71.77	46.60	74.49	9.32	8.50	27.79	9.32	30.83
	S8	39.46	37.07	62.24	40.48	63.27	5.62	4.33	18.37	6.47	19.32
	S9	40.82	41.50	68.37	40.82	68.37	5.79	5.94	23.72	5.79	23.99
	S10	42.18	40.82	55.10	42.18	57.48	7.30	6.09	14.89	7.22	16.97
	S11	44.90	40.82	69.73	43.20	71.43	8.20	5.77	25.14	6.95	26.94
	S12	42.52	39.12	68.37	42.86	71.09	6.52	4.97	23.44	6.70	26.34
	Mean	44.25	41.95	68.40	44.13	70.10	7.92	6.54	24.15	7.86	25.89
	SD	3.11	3.70	5.77	3.21	5.46	1.78	1.85	4.70	1.86	4.64
	p-val.	$***$	$***$	$**$	$***$	–	$***$	$***$	$***$	$***$	–

1: In this table, the paired sample t-test is applied to compare the CCA and the pdCCA$^+$, the pdCCA0 and the pdCCA$^+$, the pdCCA and the pdCCA$^+$, the pdCCA0$^+$ and the pdCCA$^+$. Note that the p values are corrected with the Bonferroni correction ($*$: $p < 0.05$, $**$: $p < 0.01$, $***$: $p < 0.001$).

shared knowledge among M SSVEP data in the calculation. On the other hand, in the pdCCA, the input is a single-trial multi-frequency-modulated data that includes M frequency-modulated components. The pdCCA uses the shared knowledge among M components in the calculation. Second, the pdCCA is a calibration-free approach, so that the CCA calculation is performed every trial, which is similar to the traditional CCA in [27]. In contrast, the multi-stimulus CCA is a calibration-based approach and the main calculation of the multi-stimulus CCA is performed only once in offline calibration. As a result, to avoid the confusion, we call (4) the pdCCA, rather than the multi-stimulus CCA. In addition, the name 'pdCCA' can also emphasize the importance of the phase difference constraint.

Table IV, Table V and Fig. 6 show that the pdCCA with a well-defined phase constraint is superior to the CCA (except Dataset II) while the pdCCA with a zero phase constraint (i.e., pdCCA0) is inferior to the CCA. These results suggest that the well-designed phase difference constraints play an important role in improving the recognition performance of

multi-frequency-modulated SSVEPs. Furthermore, the combination of the CCA and the pdCCA (i.e., pdCCA$^+$) can provide the highest performance. To sum up, the proposed phase difference constraint can boost the average CCA accuracy: i) from 60.55% to 82.64% in Dataset Ia, ii) from 64.82% to 85.68% in Dataset Ib, iii) from 40.14% to 48.75% in Dataset II, and vi) from 44.25% to 70.10% in Dataset III. Accordingly, the average ITR can be increased: i) from 63.55 bpm to 101.94 bpm in Dataset Ia, ii) from 71.46 bpm to 108.37 bpm in Dataset Ib, iii) from 41.54 bpm to 56.37 bpm in Dataset II, and vi) from 7.92 bpm to 25.89 bpm in Dataset III.

Clearly, the performance improvement in Dataset Ia, Ib, and III is greater than that in Dataset II, e.g., the improvement of the average accuracy is 22.09% in Dataset Ia, 20.86% in Dataset Ib, 8.61% in Dataset II, and 25.85% in Dataset III. A possible reason is that the dual-frequency-modulated SSVEP has relatively large inter-subject variability, especially for the intermodulation frequency components, which means that the SSVEP power is not fixed at the specified frequencies [35]. Thereby, using the pre-designed sine-cosine reference signals with a fixed combination of frequencies and phases in the pdCCA is not optimal for all individuals. On the other hand, Fig. 4 shows that the SSVEP power is predominant at three specified frequencies, i.e. f_1, f_2, and $f_1 + f_2$, in Dataset III. In Dataset I, there is no intermodulation frequency because the modulation frequencies are well separated in temporal domain.

In Section III-A, we tried to repeat the CCA performance as mentioned in [17]. As a matter of fact, the presented performance in this study is slightly different from theirs (81.65±12.75% vs. 82.72±10.80% for Dataset Ia and 85.13±13.86% vs. 87.16±11.46% for Dataset Ib when $T_w = 1$), maybe due to some unknown parameters, e.g., the parameters of the filter design. In order to further validate the superiority of the proposed method, we compare the CCA performance in [17] and the pdCCA$^+$ performance (when $T_w = 1$). The paired t-test results indicate that the pdCCA$^+$ performance is significantly higher than the CCA performance in [17], i.e., Dataset Ia: $p < 0.01$ (accuracy) and $p < 0.01$ (ITR); Dataset Ib: $p < 0.05$ (accuracy) and $p < 0.05$ (ITR).

The future work will consider several limitations of this study. First, in the current pdCCA, the initial phases and the frequencies of the sine-cosine reference signals may not be optimal for some frequency components, subjects, visual stimulation paradigms, etc. For example, the initial phase of the intermodulation frequency component is the linear combination of the modulated phases, which may be too simple to deal with the non-linearity in the intermodulation. Besides, the SSVEP phase is known to be subject-specific [38], [39], such that the phase difference between frequency components might also vary by individual. It is worthwhile to investigate how to select the optimal frequency component and how to design the optimal sine-cosine reference signal for different scenarios. Second, as mentioned before, the proposed pdCCA could not achieve an excellent improvement for Dataset II because the predominant frequency components of the dual-frequency-modulated SSVEPs have large individual difference [22], [40], [41], [42]. Recently, a new binocular vision scheme as introduced in [43] can be used to suppress the individual difference. Thereby, the following study might investigate the performance of the proposed pdCCA on this new visual stimulation paradigm. Third, in the either CCA or pdCCA, it is possible to mine much more discriminative information under a tensor-based framework, like in [44]. Last but not least, the existing study is only dedicated to the synchronous BCIs, which cannot achieve the autonomous control of the external equipment. The following work could extend it to the asynchronous mode, like in [45] and [46], which can greatly increase the user's autonomy in the real-life applications [47].

V. Conclusion

This study provides new insights into the improvement of the calibration-free recognition algorithms in the multi-frequency-modulated SSVEP-based BCIs. The experimental results validate that the phase difference constraints can help improve the CCA's recognition accuracy in the multi-frequency-modulated SSVEP-BCIs without calibration. The proposed pdCCA is a promising recognition approach for the multi-frequency-modulated SSVEP-based BCIs.

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] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nature Rev. Neurol.*, vol. 12, no. 9, pp. 513–525, 2016.

[3] M. Zhuang, Q. Wu, F. Wan, and Y. Hu, "State-of-the-art non-invasive brain–computer interface for neural rehabilitation: A review," *J. Neurorestoratology*, vol. 8, no. 1, pp. 12–25, Mar. 2020.

[4] T. O. Zander and C. Kothe, "Towards passive brain–computer interfaces: Applying brain–computer interface technology to human–machine systems in general," *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025005.

[5] A. Rezeika, M. Benda, P. Stawicki, F. Gembler, A. Saboor, and I. Volosyak, "Brain–computer interface spellers: A review," *Brain Sci.*, vol. 8, no. 4, p. 57, Apr. 2018.

[6] X. Gao, D. Xu, M. Cheng, and S. Gao, "A BCI-based environmental controller for the motion-disabled," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 137–140, Feb. 2003.

[7] 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.

[8] C. S. Herrmann, "Human EEG responses to 1–100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena," *Experim. Brain Res.*, vol. 137, nos. 3–4, pp. 346–353, 2001.

[9] G. Bin, X. Gao, Y. Wang, B. Hong, and S. Gao, "VEP-based brain-computer interfaces: Time, frequency, and code modulations [research frontier]," *IEEE Comput. Intell. Mag.*, vol. 4, no. 4, pp. 22–26, Nov. 2009.

[10] Y. Wang, R. Wang, X. Gao, B. Hong, and S. Gao, "A practical VEP-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 234–240, Jun. 2006.

[11] 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, Nov. 2015.

[12] M. Nakanishi, Y. Wang, X. Chen, Y. 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.

[13] Y. S. Zhang et al., "Two-stage frequency recognition method based on correlated component analysis for SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1314–1323, Jul. 2018.

[14] 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, Mar. 2012, Art. no. e29519.

[15] Y. Kimura, T. Tanaka, H. Higashi, and N. Morikawa, "SSVEP-based brain–computer interfaces using FSK-modulated visual stimuli," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2831–2838, Oct. 2013.

[16] S. Ge et al., "SSVEP-based brain–computer interface with a limited number of frequencies based on dual-frequency biased coding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 760–769, 2021.

[17] Y. Chen, C. Yang, X. Ye, X. Chen, Y. Wang, and X. Gao, "Implementing a calibration-free SSVEP-based BCI system with 160 targets," *J. Neural Eng.*, vol. 18, Jun. 2021, Art. no. 046094.

[18] X. Ye, C. Yang, Y. Chen, Y. Wang, X. Gao, and H. Zhang, "Multisymbol time division coding for high-frequency steady-state visual evoked potential-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1693–1704, 2022.

[19] M. H. Chang, H. J. Baek, S. M. Lee, and K. S. Park, "An amplitude-modulated visual stimulation for reducing eye fatigue in SSVEP-based brain–computer interfaces," *Clin. Neurophysiol.*, vol. 125, no. 7, pp. 1380–1391, 2014.

[20] A. M. Dreyer and C. S. Herrmann, "Frequency-modulated steady-state visual evoked potentials: A new stimulation method for brain–computer interfaces," *J. Neurosci. Meth.*, vol. 241, pp. 1–9, Feb. 2015.

[21] K.-K. Shyu, P.-L. Lee, Y.-J. Liu, and J.-J. Sie, "Dual-frequency steady-state visual evoked potential for brain computer interface," *Neurosci. Lett.*, vol. 483, no. 1, pp. 28–31, Oct. 2010.

[22] X. Chen, Z. Chen, S. Gao, and X. Gao, "Brain–computer interface based on intermodulation frequency," *J. Neural Eng.*, vol. 10, no. 6, Dec. 2013, Art. no. 066009.

[23] T. Cao, F. Wan, C. M. Wong, J. N. da Cruz, and Y. Hu, "Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces," *Biomed. Eng. Online*, vol. 13, no. 1, p. 28, 2014.

[24] Y. Peng et al., "Fatigue evaluation using multi-scale entropy of EEG in SSVEP-based BCI," *IEEE Access*, vol. 7, pp. 108200–108210, 2019.

[25] H.-Y. Wu, P.-L. Lee, H.-C. Chang, and J.-C. Hsieh, "Accounting for phase drifts in SSVEP-based BCIs by means of biphasic stimulation," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1394–1402, May 2011.

[26] 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. 53, no. 12, pp. 2610–2614, Dec. 2006.

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

[28] 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, Aug. 2015, Art. no. 046008.

[29] C. M. Wong, B. Wang, Z. Wang, K. F. Lao, A. Rosa, and F. Wan, "Spatial filtering in SSVEP-based BCIs: Unified framework and new improvements," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 11, pp. 3057–3072, Nov. 2020.

[30] M. Hye Chang and K. Suk Park, "Frequency recognition methods for dual-frequency SSVEP based brain–computer interface," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 2220–2223.

[31] C. M. Wong et al., "Learning across multi-stimulus enhances target recognition methods in SSVEP-based BCIs," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016026.

[32] C. M. Wong et al., "Inter- and intra-subject transfer reduces calibration effort for high-speed SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2123–2135, Oct. 2020.

[33] C. M. Wong et al., "Transferring subject-specific knowledge across stimulus frequencies in SSVEP-based BCIs," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 552–563, Apr. 2021.

[34] C. M. Wong et al., "Online adaptation boosts SSVEP-based BCI performance," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 6, pp. 2018–2028, Jun. 2021.

[35] L. Liang et al., "Optimizing a dual-frequency and phase modulation method for SSVEP-based BCIs," *J. Neural Eng.*, vol. 17, no. 4, Aug. 2020, Art. no. 046026.

[36] 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.

[37] D. E. Thompson et al., "Performance measurement for brain–computer or brain–machine interfaces: A tutorial," *J. Neural Eng.*, vol. 11, no. 3, Jun. 2014, Art. no. 035001.

[38] Y. Li, G. Bin, X. Gao, B. Hong, and S. Gao, "Analysis of phase coding SSVEP based on canonical correlation analysis (CCA)," in *Proc. 5th Int. IEEE/EMBS Conf. Neural Eng.*, Apr. 2011, pp. 368–371.

[39] J. Pan, X. Gao, F. Duan, Z. Yan, and S. Gao, "Enhancing the classification accuracy of steady-state visual evoked potential-based brain–computer interfaces using phase constrained canonical correlation analysis," *J. Neural Eng.*, vol. 8, no. 3, Jun. 2011, Art. no. 036027.

[40] M. H. Chang, J. S. Lee, J. Heo, and K. S. Park, "Eliciting dual-frequency SSVEP using a hybrid SSVEP-P300 BCI," *J. Neurosci. Methods*, vol. 258, pp. 104–113, Jan. 2016.

[41] X. Chen, Y. Wang, S. Zhang, S. Gao, Y. Hu, and X. Gao, "A novel stimulation method for multi-class SSVEP-BCI using intermodulation frequencies," *J. Neural Eng.*, vol. 14, no. 2, Apr. 2017, Art. no. 026013.

[42] J. Mu, Y. Tan, D. B. Grayden, and D. Oetomo, "Multi-frequency canonical correlation analysis (MFCCA): A generalised decoding algorithm for multi-frequency SSVEP," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 6151–6154.

[43] Y. Sun et al., "A binocular vision SSVEP brain-computer interface paradigm for dual-frequency modulation," *IEEE Trans. Biomed. Eng.*, early access, Oct. 5, 2022, doi: 10.1109/TBME.2022.3212192.

[44] Y. Pei et al., "A tensor-based frequency features combination method for brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 465–475, 2022.

[45] C. Yang, X. Yan, Y. Wang, Y. Chen, H. Zhang, and X. Gao, "Spatio-temporal equalization multi-window algorithm for asynchronous SSVEP-based BCI," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 0460b7.

[46] F. W. Gembler, M. Benda, A. Rezeika, P. R. Stawicki, and I. Volosyak, "Asynchronous c-VEP communication tools—Efficiency comparison of low-target, multi-target and dictionary-assisted BCI spellers," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, Oct. 2020.

[47] L. Chen et al., "Adaptive asynchronous control system of robotic arm based on augmented reality-assisted brain–computer interface," *J. Neural Eng.*, vol. 18, no. 6, Dec. 2021, Art. no. 066005.