

# Cross-Subject Transfer Method Based on Domain Generalization for Facilitating Calibration of SSVEP-Based BCIs

Jiayang Huang<sup>ID</sup>, Zhi-Qiang Zhang<sup>ID</sup>, *Member, IEEE*, Bang Xiong<sup>ID</sup>, Quan Wang<sup>ID</sup>,  
Bo Wan<sup>ID</sup>, Fengqi Li<sup>ID</sup>, and Pengfei Yang<sup>ID</sup>, *Member, IEEE*

**Abstract**—In steady-state visual evoked potential (SSVEP)-based brain-computer interfaces (BCIs), various spatial filtering methods based on individual calibration data have been proposed to alleviate the interference of spontaneous activities in SSVEP signals for enhancing the SSVEP detection performance. However, the time-consuming calibration session would increase the visual fatigue of subjects and reduce the usability of the BCI system. The key idea of this study is to propose a cross-subject transfer method based on domain generalization, which transfers the domain-invariant spatial filters and templates learned from source subjects to the target subject with no access to the EEG data from the target subject. The transferred spatial filters and templates are obtained by maximizing the intra- and inter-subject correlations using the SSVEP data corresponding to the target and its neighboring stimuli. For SSVEP detection of the target subject, four types of correlation coefficients are calculated to construct the feature vector. Experimental results estimated with three SSVEP datasets show that the proposed cross-subject transfer method improves the SSVEP detection performance compared to state-of-art methods. The satisfactory results demonstrate that the proposed method provides an effective transfer learning strategy requiring no tedious data collection process for new users, holding the potential of promoting practical applications of SSVEP-based BCI.

**Index Terms**—Brain-computer interfaces (BCIs), cross-subject, domain generalization, steady-state visual evoked potential (SSVEP), transfer learning.

Manuscript received 5 June 2023; revised 15 July 2023; accepted 10 August 2023. Date of publication 14 August 2023; date of current version 22 August 2023. This work was supported in part by the Shaanxi Key Technology Research and Development Program under Grant 2021ZDLGY07-01, and in part by the National Natural Science Foundation of China under Grant 61972302 and Grant 61962019. (*Corresponding authors: Zhi-Qiang Zhang; Pengfei Yang.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of Xidian University.

Jiayang Huang, Bang Xiong, Quan Wang, Bo Wan, Fengqi Li, and Pengfei Yang are with the Key Laboratory of Smart Human-Computer Interaction and Wearable Technology of Shaanxi Province, School of Computer Science and Technology, Xidian University, Xi'an 710071, China (e-mail: jyhuang1@stu.xidian.edu.cn; bxiong@stu.xidian.edu.cn; qwang@xidian.edu.cn; wanbo@xidian.edu.cn; 22031212345@stu.xidian.edu.cn; pfyang@xidian.edu.cn).

Zhi-Qiang Zhang is with the School of Electronic and Electrical Engineering, Institute of Robotics, Autonomous Systems and Sensing, University of Leeds, LS2 9JT Leeds, U.K. (e-mail: z.zhang3@leeds.ac.uk).

Digital Object Identifier 10.1109/TNSRE.2023.3305202

## I. INTRODUCTION

**B**RAIN-COMPUTER interface (BCI) is a new connection pathway for information transfer and control between the brain and a device with computing capabilities [1], [2]. Among different types of BCIs, electroencephalography (EEG)-based systems with the characteristics of non-invasiveness and portability have been widely used in real-world applications [3], [4]. Steady-state visual evoked potential (SSVEP)-based BCIs, as one of the popular paradigms in EEG-based BCIs, have been extensively applied in varieties of scenarios due to their high efficiency, ease of use, and high reliability [5], [6], [7], [8].

For SSVEP-based BCIs, developing SSVEP decoding methods is the main task to guarantee the high efficiency of practical applications [9]. Thus far, many spatial filtering methods have been proposed to enhance the performance of SSVEP-based BCIs [10], [11]. One type of spatial filtering methods, such as canonical correlation analysis (CCA) [12], filter bank CCA (FBCCA) [13], and multivariate variational mode decomposition CCA (MVMD-CCA) [14], are called training-free methods, which identify SSVEP frequency with artificial sine-cosine reference signals without pre-training. However, such easily-used methods would obtain limited information transfer rates (ITRs) due to the interference of spontaneous brain activities. To address this issue, another type of spatial filtering method based on individual training data is further developed based on CCA. By incorporating individual calibration data, L1-regularized multi-way CCA (L1-MCCA) [15], multi-set CCA (MsetCCA) [16], and extended CCA (eCCA) [17] were proposed to improve the SSVEP frequency detection performance. Furthermore, spatial filtering methods as the task-related component analysis (TRCA) [18] and the sum of squared correlation (SSCOR) method [19] were reported to detect SSVEP frequency only using individual calibration data of the corresponding target stimulus, which significantly improved the SSVEP detection performance. To further improve the TRCA method, the correlated component analysis (CORCA) [20] was proposed by incorporating individual calibration data from other subjects; the spatial filtering method incorporating the data from neighboring-location stimuli [21] was introduced, which further enhanced the target identification performance of SSVEP-based BCIs.

Despite these training-based spatial filtering methods have effectively boosted the performance of SSVEP-based BCIs, the time-consuming calibration session needs to be conducted for each one of the subjects, which would cause visual fatigue of the subject and affect the applicability of SSVEP-based BCIs. To alleviate the impact of the variability of different subjects, the transfer learning (TL) strategy has been introduced to SSVEP-based BCIs to transfer common knowledge from source subjects to target subjects [22], [23]. One direction of transfer learning in SSVEP-based BCIs is to directly transfer templates across subjects to boost the SSVEP detection performance of new subjects by augmenting the calibration data [24]. For example, Yuan et al. proposed a transfer template-based CCA (tt-CCA) method [25] to transfer SSVEP templates from source subjects to a target subject for enhancing the SSVEP detection performance. Waytowich et al. introduced the Adaptive Combined-CCA (Adaptive-C3A) [26] to extend CCA by incorporating the SSVEP templates computed from previously collected subjects. Such methods would extract insufficient SSVEP features from the transferred templates which were obtained by averaging across trials from source subjects.

In addition to transferred templates, common spatial filters were also learned from source subjects to transfer across different subjects. For instance, Wang et al. proposed a novel inter- and intra-subject maximal correlation (IISM) method [27] to obtain transferred spatial filters and templates via employing the similarity and variability within and between source and target subjects. Zhang et al. proposed an inter-subject transfer learning method [28] to train the transferred templates and spatial filters by maximizing the correlations among the training data, the individual template, and the artificial reference signal simultaneously. Wong et al. proposed a subject transfer-based CCA (stCCA) method [29] which utilizes the knowledge within the target subject and between source and target subjects to reduce the calibration effort. All these methods significantly boosted the SSVEP detection performance, but individual calibration data are still required from the target subject for spatial filter training. By contrast, some subject-independent methods extended by CCA were introduced to further reduce the impact of calibration sessions. For instance, Yan et al. successively proposed a cross-subject spatial filter transfer (CSSFT) method [30] and an improved CSSFT method [31] that transfer the templates and spatial filters from the existing users to the new user test data. However, in the aforementioned methods, except for transferred spatial filters and templates learned from source subjects, at least one spatial filter or template was learned from the data provided by the target subject to detect the SSVEP frequency. It means that these models have to get access to target domain data, and barely consider an unseen test domain.

In this paper, considering the unseen target subject, we aim to develop a cross-subject transfer framework based on domain generalization strategy [32] that can generalize to an unseen test domain by learning domain-invariant features, including the internally- and mutually-invariant features [33], [34]. First, the internally-invariant template for each source subject is constructed by learning the spatial filter which maximizes the

intra-subject correlation to extract common frequency information from neighboring stimuli. Second, the mutually-invariant template is obtained by learning a spatial filter that maximizes the inter-subject correlation to learn common knowledge shared across all source subjects. Third, a test-trial spatial filter is trained by maximizing the correlation between one-trial data from the source subject and the two types of domain-invariant templates to improve the signal-to-noise (SNR) of test-trial data. Finally, all spatial filters and templates learned from source subjects are transferred to detect the SSVEP frequency of test data from the target subject by constructing a four-dimensional feature vector, which is comprised of the four different types of correlation coefficients between the test data spatially filtered by the transferred spatial filters and transferred templates. For the performance evaluation, two publicly available datasets including the dataset from the University of California San Diego (UCSD) [35] and the benchmark dataset from Tsinghua University [36] and a self-collected SSVEP dataset are utilized to conduct extensive comparisons with state-of-art methods such as FBCCA, tt-CCA, and CSSFT. The experimental results have demonstrated the feasibility and efficiency of the proposed cross-subject transfer method.

The organization of the article is arranged as follows: Section II introduces the materials and methods. In section III, the experimental results are reported. The discussions are shown in Section IV. Finally, the conclusion is presented in Section V.

## II. MATERIALS AND METHODS

### A. Dataset Descriptions

In this study, two public SSVEP datasets and a self-collected dataset, namely Dataset I, II, and III, are utilized to evaluate the proposed cross-subject transfer method. The details of the three datasets are described as follows.

1) *Dataset I*: The first public dataset is the 12-target SSVEP dataset collected by UCSD [35], which is freely downloaded at [https://github.com/mnakanishi/12JFPM\\_SSVEP](https://github.com/mnakanishi/12JFPM_SSVEP). The UCSD dataset is collected from 10 healthy subjects at 8 electrodes (PO7, PO3, POz, PO4, PO8, O1, Oz, and O2) covering the occipital area. Each subject was instructed to gaze at the 12 stimuli, and the whole SSVEP-BCI experiment was repeated 15 times, forming a 15-block dataset. Each block includes 12 trials corresponding to the 12 targets, and each trial contains a 1-s gaze shifting cue and a 4-s gazing time. The SSVEP data were collected at a sampling rate of 2048 Hz and downsampled to 256 Hz. The 12 stimuli were coded by the joint frequency and phase modulation (JFPM) method [37], where the frequency ranged from 9.25 Hz to 14.75 Hz with an interval of 0.5 Hz, and the phase range was from  $0\pi$  to  $1.5\pi$  with an interval of  $0.5\pi$ , shown as Fig. 1(a).

2) *Dataset II*: The second public dataset is the benchmark dataset proposed by Tsinghua University [36], which is available at <http://bci.med.tsinghua.edu.cn/download.html>. The benchmark dataset consists of 64-channel EEG recordings of 35 healthy subjects by gazing at 40 characters. As shown in Fig. 1(b), the 40 targets are stimulated at 8–15.8 Hz frequencies with an interval of 0.2 Hz and  $0-1.5\pi$  phases with an interval of  $0.5\pi$  using the JFPM method. The SSVEP

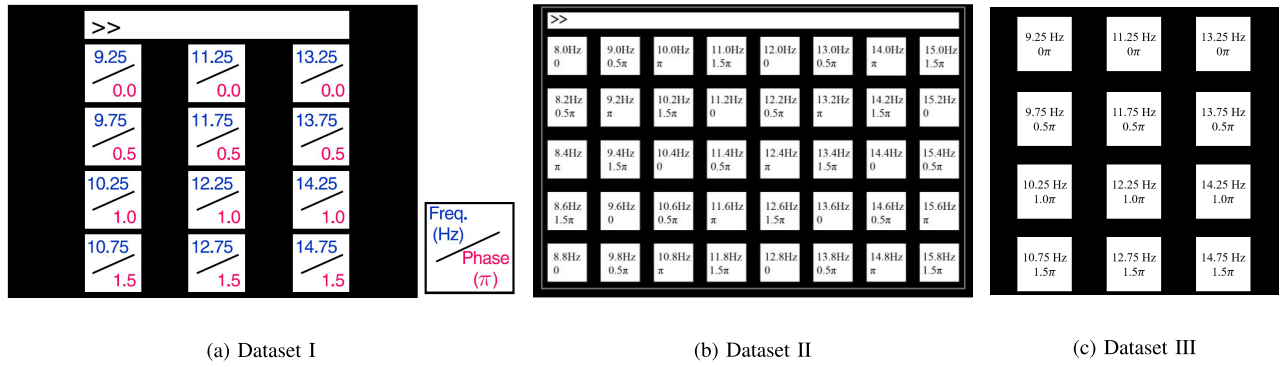


Fig. 1. The stimulation interfaces of the three SSVEP datasets (Dataset I (a), Dataset II (b), and Dataset III (c)) and the corresponding frequency and phase values.

data were recorded with a sampling rate of 1000 Hz and then downsampled to 250 Hz. For each subject, the dataset contains 6-block data, and each block is comprised of 40 trials corresponding to the 40 targets, and each trial contains a 0.5-s gaze shifting cue, a 5-s stimulation, and a 0.5-s rest. The 9 channels (Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, and O2) located at the occipital area were selected for SSVEP signal analysis in this study.

3) *Dataset III*: In addition, a self-collected SSVEP dataset with 12 stimuli is also used for performance evaluation. 9-channel EEG data were collected from 11 healthy participants (four females and seven males, aged from 24 to 27). All subjects were informed of the experimental process and protocols and signed the informed consent before the experiment. And the experiment is approved by the Research Ethics Committee of Xidian University. The 9 Ag/AgCl electrodes (Pz, PO7, PO3, POz, PO4, PO8, O1, Oz, and O2) were selected from the parietal and occipital regions. The ground and reference electrodes were placed at FPz and right earlobe respectively. For each subject, 10-block EEG data were collected in the BCI experiments. Each block contains 0.5-s cue, 5-s stimulation, and 0.5-s rest. During the stimulation, subjects were asked to avoid eye blinks. To prevent visual fatigue, there was a two-minute rest between two successive blocks. The interface comprises a  $4 \times 3$  matrix of visual stimuli coded by the JFPM method as shown in Fig. 1(c). The frequency ranged from 9.25 Hz to 14.75 Hz with an interval of 0.5 Hz. And the phase range was from  $0\pi$  to  $1.5\pi$  with an interval of  $0.5\pi$ .

### B. Data Preprocessing

According to the 0.14-s visual latency [36], [38], the first 0.14-s data were removed for SSVEP signal analysis. And then, all the extracted data were filtered by a sixth-order Butterworth filter with the 7-90 Hz band. A notch filter at 50 Hz is utilized to eliminate the power-line noise. After the data preparation was completed, all the data processing and target detection were then performed.

### C. Transferred Spatial Filters and SSVEP Templates

Assume that the individual calibration data from one source subject corresponding to the  $n$ -th stimulus is defined as  $\mathbf{X}_n = [\mathbf{X}_n^1, \mathbf{X}_n^2, \dots, \mathbf{X}_n^{N_b}] \in \mathbb{R}^{N_c \times N_d \times N_b}$ , and  $n = 1, 2, \dots, N_f$ .

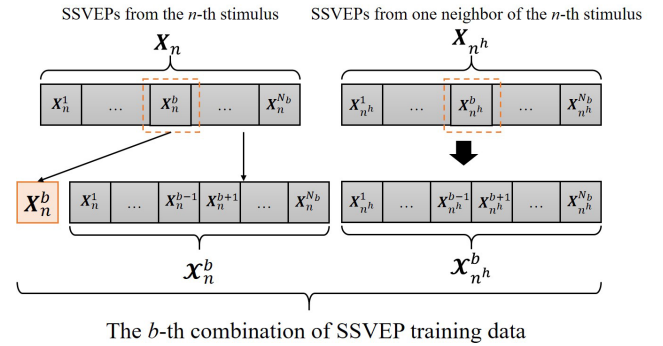


Fig. 2. The illustration of the  $b$ -th combination of SSVEP data consists of the target and its neighboring stimuli data.

Here,  $N_c$ ,  $N_d$ ,  $N_b$ , and  $N_f$  represent the number of channels, the number of sampling points, the number of blocks, and the number of stimuli respectively.  $\mathbf{X}_n^b \in \mathbb{R}^{N_c \times N_d}$  ( $b = 1, 2, \dots, N_b$ ) denotes the  $b$ -th block data of  $\mathbf{X}_n$ .

In order to extract accurate common knowledge across subjects for transferring, we construct internally- and mutually-invariant templates and a test-trial spatial filter by using SSVEP data from different blocks. Therefore, we first separate the EEG data  $\mathbf{X}_n$  into two parts according to each block  $b$ : the one-trial data  $\mathbf{X}_n^b$  from block  $b$  and the multi-trial data  $\mathcal{X}_n^b \in \mathbb{R}^{N_c \times N_d \times (N_b - 1)}$  from the other blocks.  $\mathbf{X}_n^b$  is used to simulate the unseen test-trial data to calculate the spatial filter for improving the SNR of the test signal.  $\mathcal{X}_n^b$  is used for transferred template training, defined as:

$$\mathcal{X}_n^b = [\mathbf{X}_n^1, \dots, \mathbf{X}_n^{b-1}, \mathbf{X}_n^{b+1}, \dots, \mathbf{X}_n^{N_b}]. \quad (1)$$

Based on previous studies [18], [21], SSVEPs from the neighboring-location stimuli share a common spatial pattern and contain common frequency information. The neighboring stimuli data are also used for transferred template training. Here, define the neighbors of the  $n$ -th stimulus as the horizontally and vertically adjacent stimuli, as the  $n^1$ -th,  $n^2$ -th,  $\dots$ ,  $n^{N_h}$ -th stimuli. Therefore, the collection of the neighbors of the  $n$ -th stimulus is denoted as:

$$\Theta_n = \{\mathbf{X}_{n^1}, \mathbf{X}_{n^2}, \dots, \mathbf{X}_{n^{N_h}}\}, \quad (2)$$

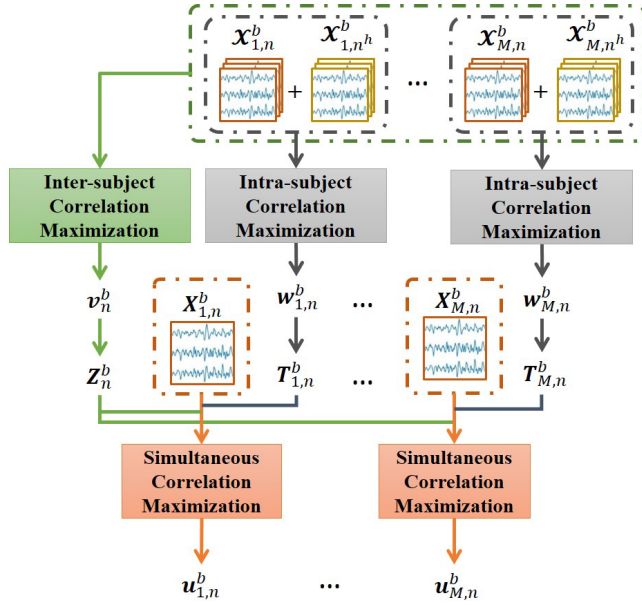


Fig. 3. The training procedure to obtain the transferred spatial filters and templates.

where  $\mathbf{X}_{n^h} \in \mathbb{R}^{N_c \times N_d \times N_b}$  ( $h = 1, 2, \dots, N_h$ ) is the all-block SSVEP data of one neighboring stimulus,  $h$  is the index of the neighbor of the  $n$ -th stimulus,  $n^h$  represents the stimulus index of the  $h$ -th neighbor to the  $n$ -th stimulus, and  $N_h$  represents the number of neighboring stimuli. Consistent with  $\mathbf{X}_n^b$ , the multi-trial data of each neighbor  $\mathbf{X}_{n^h}^b \in \mathbb{R}^{N_c \times N_d \times (N_b-1)}$  is denoted as:

$$\mathbf{X}_{n^h}^b = [\mathbf{X}_{n^h}^1, \dots, \mathbf{X}_{n^h}^{b-1}, \mathbf{X}_{n^h}^{b+1}, \dots, \mathbf{X}_{n^h}^{N_b}]. \quad (3)$$

As shown in Fig. 2, the  $b$ -th combination of SSVEP data consists of the target and its neighboring stimuli data, (i.e.,  $\mathbf{X}_n^b$ ,  $\mathbf{X}_{n^h}^b$ , and  $\mathbf{X}_{n^h}^b$  ( $h = 1, 2, \dots, N_h$ )). Since  $b$  ranges from 1 to  $N_b$  for each stimulus  $n$ , there are  $N_b$  combinations of SSVEP training data.

Using the  $b$ -th combination of SSVEP training data, the whole training procedure to obtain the transferred spatial filters and templates mainly contains three steps:

- 1) Learning the internally-invariant spatial filter and template for each source subject to extract common frequency information across neighboring stimuli;
- 2) Calculating the mutually-invariant spatial filter and template from all source subjects to learn common knowledge shared across subjects;
- 3) Training a test-trial spatial filter to improve the SNR of test-trial data by incorporating the internally- and mutually-invariant templates.

The flowchart of the training process is illustrated in Fig. 3. We will elaborate on the three steps in detail below.

**1) Internally-Invariant Template:** To obtain the internally-invariant template for each source subject  $m$ , the spatial filter corresponding to the  $n$ -th stimulus  $\mathbf{w}_{m,n}^b \in \mathbb{R}^{N_c}$  is calculated by maximizing the intra-subject correlation using SSVEPs corresponding to the target and its neighboring stimuli

$\mathbf{X}_{m,n}^b \in \mathbb{R}^{N_c \times N_d \times (N_b-1)}$  and  $\mathbf{X}_{m,n^h}^b \in \mathbb{R}^{N_c \times N_d \times (N_b-1)}$  ( $h = 1, 2, \dots, N_h$ ) from each source subject  $m$ .

To simplify the expression of  $\mathbf{X}_{m,n}^b$  and  $\mathbf{X}_{m,n^h}^b$  defined as Eq. (1) and (3), we redefine the multi-trial data corresponding to the target stimulus as:

$$\mathbf{X}_{m,n}^b = [\mathbf{X}_{m,n}^{b_1}, \mathbf{X}_{m,n}^{b_2}, \dots, \mathbf{X}_{m,n}^{b_{N_t}}], \quad (4)$$

where  $\mathbf{X}_{m,n}^{b_k} \in \mathbb{R}^{N_c \times N_d}$  ( $k = 1, 2, \dots, N_t$ ).  $N_t$  is the number of trials of the multi-trial data and  $N_t = N_b - 1$ . In the same way,

$$\mathbf{X}_{m,n^h}^b = [\mathbf{X}_{m,n^h}^{b_1}, \mathbf{X}_{m,n^h}^{b_2}, \dots, \mathbf{X}_{m,n^h}^{b_{N_t}}], \quad (5)$$

where  $\mathbf{X}_{m,n^h}^{b_k} \in \mathbb{R}^{N_c \times N_d}$  ( $k = 1, 2, \dots, N_t$ ).

Therefore, the internally-invariant spatial filter  $\hat{\mathbf{w}}_{m,n}^b$  is calculated as:

$$\hat{\mathbf{w}}_{m,n}^b = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{S} \mathbf{w}}{\mathbf{w}^\top \mathbf{Q} \mathbf{w}}. \quad (6)$$

The summation of the auto-covariances of  $\bar{\mathbf{X}}_{m,n}^b \in \mathbb{R}^{N_c \times N_d}$  and  $\bar{\mathbf{X}}_{m,n^h}^b \in \mathbb{R}^{N_c \times N_d}$  is denoted as:

$$\mathbf{S} = \text{cov}(\bar{\mathbf{X}}_{m,n}^b) + \sum_{h=1}^{N_h} \text{cov}(\bar{\mathbf{X}}_{m,n^h}^b) \quad (7)$$

where

$$\bar{\mathbf{X}}_{m,n}^b = \frac{1}{N_t} \sum_{k=1}^{N_t} \mathbf{X}_{m,n}^{b_k}, \quad (8)$$

$$\bar{\mathbf{X}}_{m,n^h}^b = \frac{1}{N_t} \sum_{k=1}^{N_b} \mathbf{X}_{m,n^h}^{b_k}. \quad (9)$$

And the sum of covariances of all-trial data from  $n$ -th stimulus and its neighboring stimuli is defined as:

$$\mathbf{Q} = \sum_{k=1}^{N_t} \text{cov}(\mathbf{X}_{m,n}^{b_k}) + \sum_{h=1}^{N_h} \sum_{k=1}^{N_t} \text{cov}(\mathbf{X}_{m,n^h}^{b_k}). \quad (10)$$

By using generalized eigendecomposition of  $\mathbf{Q}^{-1} \mathbf{S}$  to solve Eq. (6), the spatial filter  $\hat{\mathbf{w}}_{m,n}^b$  is determined as the eigenvector corresponding to the largest eigenvalue. With the internally-invariant spatial filter  $\mathbf{w}_{m,n}^b$ , the internally-invariant template  $\mathbf{T}_{m,n}^b \in \mathbb{R}^{N_d}$  for each source subject  $m$  is obtained as:

$$\mathbf{T}_{m,n}^b = \mathbf{w}_{m,n}^{b \top} \bar{\mathbf{X}}_{m,n}^b. \quad (11)$$

**2) Mutually-Invariant Template:** To obtain the mutually-invariant template, the mutually-invariant spatial filter is learned from the SSVEP data from all  $M$  source subjects. The mutually-invariant spatial filter  $\mathbf{v}_n^b \in \mathbb{R}^{N_c}$  corresponding to the  $n$ -th stimulus is estimated by maximizing the correlation between different subjects [20], [27]. In this method, instead of only using the data of the target stimulus, the inter-subject maximal correlation is calculated using the SSVEP data of the target and its neighboring stimuli.

First, the  $b$ -th multi-trial data of  $n$ -th stimulus from two different subjects  $m_1$  and  $m_2$  are respectively denoted as  $\mathbf{X}_{m_1,n}^b \in \mathbb{R}^{N_c \times N_d \times N_t}$  and  $\mathbf{X}_{m_2,n}^b \in \mathbb{R}^{N_c \times N_d \times N_t}$ . Then,

define  $\mathbf{C}_{12}$  and  $\mathbf{C}_{21}$  as the inter-subject cross-covariances,  $\mathbf{C}_{11}$  and  $\mathbf{C}_{22}$  as the intra-subject auto-covariances. Assuming that  $\mathbf{v}^\top \mathbf{C}_{11} \mathbf{v} = \mathbf{v}^\top \mathbf{C}_{22} \mathbf{v}$ , the term  $\mathbf{C}_{21}$  is the transposition of  $\mathbf{C}_{12}$ . And the optimization problem can be solved as:

$$\hat{\mathbf{v}}_n^b = \arg \max_{\mathbf{v}} \frac{\mathbf{v}^\top \mathbf{P} \mathbf{v}}{\mathbf{v}^\top \mathbf{R} \mathbf{v}}, \quad (12)$$

where  $\mathbf{P} = \mathbf{C}_{12} + \mathbf{C}_{21}$  and  $\mathbf{R} = \mathbf{C}_{11} + \mathbf{C}_{22}$ . Therefore, the matrices  $\mathbf{P}$  and  $\mathbf{R}$  are respectively calculated as:

$$\mathbf{P} = \frac{1}{M(M-1)} \left[ \sum_{m_1} \sum_{\substack{m_2 \\ m_2 \neq m_1}} \text{cov}(\bar{\mathbf{x}}_{m_1,n}^b, \bar{\mathbf{x}}_{m_2,n}^b) + \sum_{h=1}^{N_h} \sum_{m_1} \sum_{\substack{m_2 \\ m_2 \neq m_1}} \text{cov}(\bar{\mathbf{x}}_{m_1,n^h}^b, \bar{\mathbf{x}}_{m_2,n^h}^b) \right]. \quad (13)$$

$$\mathbf{R} = \frac{1}{M} \sum_{m=1}^M \left[ \text{cov}(\bar{\mathbf{x}}_{m,n}^b) + \sum_{h=1}^{N_h} \text{cov}(\bar{\mathbf{x}}_{m,n^h}^b) \right]. \quad (14)$$

The optimal estimation of  $\hat{\mathbf{v}}_n^b$  in Eq. (12) can be obtained from the eigenvector  $\mathbf{v}$  of  $\mathbf{P}^{-1} \mathbf{R}$  with the maximal eigenvalue. With the mutually-invariant spatial filter  $\mathbf{v}_n^b$ , the mutually-invariant template  $\mathbf{Z}_n^b \in \mathbb{R}^{N_d}$  is obtained as:

$$\mathbf{Z}_n^b = \mathbf{v}_n^{b\top} \tilde{\mathbf{x}}_n^b, \quad (15)$$

where  $\tilde{\mathbf{x}}_n^b = \frac{1}{M} \sum_{m=1}^M \bar{\mathbf{x}}_{m,n}^b$  is the averaged template across all source subjects.

**3) Test-Trial Spatial Filter:** With the internally-invariant and mutually invariant templates, a test-trial spatial filter  $\mathbf{u}_{m,n}^b \in \mathbb{R}^{N_c}$  for each source subject  $m$  is trained using the one-trial data  $\mathbf{X}_{m,n}^b$ . The spatial filter  $\mathbf{u}_{m,n}^b$  is obtained by simultaneously maximize the correlation between  $\mathbf{X}_{m,n}^b$  and  $\mathbf{T}_{m,n}^b$  and the correlation between  $\mathbf{X}_{m,n}^b$  and  $\mathbf{Z}_n^b$ . Therefore, the estimation of  $\mathbf{u}_{m,n}^b$  is formulated by the multi-objective optimization problem:

$$\begin{aligned} \hat{\mathbf{u}}_{m,n}^b &= \arg \max_{\mathbf{u}} \mathcal{F}(\mathbf{u}_{m,n}^b), \\ &\text{subject to } \sum_{c=1}^{N_c} u^c = 0, \end{aligned} \quad (16)$$

where

$$\mathcal{F}(\mathbf{u}_{m,n}^b) = \begin{bmatrix} \rho(\mathbf{u}_{m,n}^{b\top} \mathbf{X}_{m,n}^b, \mathbf{T}_{m,n}^b) \\ \rho(\mathbf{u}_{m,n}^{b\top} \mathbf{X}_{m,n}^b, \mathbf{Z}_n^b) \end{bmatrix}, \quad (17)$$

and  $c$  is the index of channel, and  $c = 1, 2, \dots, N_c$ .  $u^c$  is the weight value corresponding to the  $c$ -th channel in  $\mathbf{u}_{m,n}^b$ , that is,  $\mathbf{u}_{m,n}^b = [u^1, u^2, \dots, u^{N_c}]^\top$ .  $\rho(s_1, s_2)$  is the Pearson's correlation coefficient between  $s_1$  and  $s_2$  [39]. The constrained multi-objective optimization problem described in Eq. (16) can be solved by the function `fgoalattain()` in MATLAB.

Given  $N_b$  combinations of training data, the training process from 1) to 3) would repeat  $N_b$  times. The final transferred

spatial filters and templates are calculated by averaging across  $N_b$  as follows:

$$\bar{\mathbf{w}}_{m,n} = \frac{1}{N_b} \sum_{b=1}^{N_b} \mathbf{w}_{m,n}^b, \quad (18)$$

$$\bar{\mathbf{v}}_n = \frac{1}{N_b} \sum_{b=1}^{N_b} \mathbf{v}_n^b, \quad (19)$$

$$\bar{\mathbf{u}}_{m,n} = \frac{1}{N_b} \sum_{b=1}^{N_b} \mathbf{u}_{m,n}^b, \quad (20)$$

$$\bar{\mathbf{T}}_{m,n} = \frac{1}{N_b} \sum_{b=1}^{N_b} \mathbf{T}_{m,n}^b, \quad (21)$$

$$\bar{\mathbf{Z}}_n = \frac{1}{N_b} \sum_{b=1}^{N_b} \mathbf{Z}_n^b. \quad (22)$$

#### D. SSVEP Detection With Transferred Parameters

With all transferred spatial filters and templates, single-trial data from one target subject  $\mathbf{Y} \in \mathbb{R}^{N_c \times N_d}$  will be recognized. For SSVEP detection, four different types of correlation coefficients between spatially filtered test data and transferred templates are computed by incorporating the transferred spatial filters and templates trained from source subjects, which are described as follows:

- (i)  $\rho(\bar{\mathbf{v}}_n^\top \mathbf{Y}, \bar{\mathbf{Z}}_n)$  with the mutually-invariant spatial filter and template;
- (ii)  $\rho(\bar{\mathbf{w}}_{m,n}^\top \mathbf{Y}, \bar{\mathbf{T}}_{m,n})$  with the internally-invariant spatial filter and template;
- (iii)  $\rho(\bar{\mathbf{u}}_{m,n}^\top \mathbf{Y}, \bar{\mathbf{Z}}_n)$  with the test-trial spatial filter and the mutually-invariant template;
- (iv)  $\rho(\bar{\mathbf{u}}_{m,n}^\top \mathbf{Y}, \bar{\mathbf{T}}_{m,n})$  with the test-trial spatial filter and the internally-invariant template.

For each source subject  $m$ , the correlation coefficients (ii)-(iv) can be estimated. The correlation values can be obtained by averaging across  $M$  source subjects. The correlation feature vector  $\lambda_n$  between the spatially filtered test data and  $n$ -th templates is defined as:

$$\lambda_n = \begin{bmatrix} \lambda_n(1) \\ \lambda_n(2) \\ \lambda_n(3) \\ \lambda_n(4) \end{bmatrix} = \begin{bmatrix} \rho(\bar{\mathbf{v}}_n^\top \mathbf{Y}, \bar{\mathbf{Z}}_n) \\ \frac{1}{M} \sum_{m=1}^M \rho(\bar{\mathbf{w}}_{m,n}^\top \mathbf{Y}, \bar{\mathbf{T}}_{m,n}) \\ \frac{1}{M} \sum_{m=1}^M \rho(\bar{\mathbf{u}}_{m,n}^\top \mathbf{Y}, \bar{\mathbf{Z}}_n) \\ \frac{1}{M} \sum_{m=1}^M \rho(\bar{\mathbf{u}}_{m,n}^\top \mathbf{Y}, \bar{\mathbf{T}}_{m,n}) \end{bmatrix}. \quad (23)$$

And then the correlation values  $\lambda_n(\alpha)$  ( $\alpha = 1, 2, 3, 4$ ) in Eq. (23) are combined as the correlation value  $\gamma_n$  corresponding to the  $n$ -th stimulus:

$$\gamma_n = \sum_{\alpha=1}^4 \text{sign}(\lambda_n(\alpha)) (\lambda_n(\alpha))^2, \quad (24)$$

where `sign()` is to retain discriminative information from negative correlation coefficients.

### E. Filter Bank Processing

The filter bank analysis [13] is applied to decompose SSVEPs into subband components, which extracts accurate harmonic information from SSVEP data. With the filter bank technique, the SSVEP detection performance can be further boosted. Here, each subband  $j$  ( $j = 1, 2, \dots, N_j$ ) of the filter bank is at the frequency range of  $[j \times 8 \text{ Hz}, 88 \text{ Hz}]$ , which is implemented by zero-phase Chebyshev type I infinite impulse response (IIR) filters. After that, the feature  $\gamma_n^j$  is calculated for each subband via Eq. (24). By integrating  $\gamma_n^j$  from all subbands, the final correlation feature  $\Gamma_n$  is obtained as:

$$\Gamma_n = \sum_{j=1}^{N_j} \beta(j) \cdot \left(\gamma_n^j\right)^2, \quad (25)$$

where  $\beta(j) = j^{-1.25} + 0.25$  is the weight function [13]. Finally, the target frequency  $\hat{f}$  with the maximal correlation coefficient is described as:

$$\hat{f} = \underset{n}{\operatorname{argmax}} \Gamma_n, n = 1, 2, \dots, N_f. \quad (26)$$

### F. Performance Evaluation

In this study, the classification accuracy and ITR estimates were computed to evaluate the SSVEP detection performance of the proposed method. The classification accuracy is defined as the percentage of the correct predictions out of all predictions. ITR is the amount of information transferred per minute, defined as:

$$\text{ITR} = \frac{60}{T} \times \left[ \log_2 N_f + P \times \log_2 P + (1-P) \times \log_2 \left( \frac{1-P}{N_f-1} \right) \right], \quad (27)$$

where  $T$  is the selection time for each target, including gazing time and 0.5-s gaze-shifting time,  $N_f$  is the number of stimuli, and  $P$  represents the classification accuracy.

The estimates were calculated by using holdout cross-validation. For Dataset I and III, the transferred spatial filters and templates were trained with 5 source subjects and then tested on the other subjects. For dataset II, the data from 10 source subjects was used for training transferred spatial filters and templates, while data from the other subjects were used as test data. In the proposed method, source subjects are selected randomly from the datasets. The settings of the number of source subjects will be discussed in Section III-C. To get the general performance of the proposed cross-subject transfer method, the whole training and test process was conducted 10 times for Dataset I and III and 5 times for Dataset II respectively.

In the training stage, since Dataset I contains 15-block data ( $N_b = 15$ ), the whole dataset can be divided into 15 combinations. Therefore, the training procedure using Dataset I was repeated 15 times for each source subject. The final transferred spatial filters and templates were computed by averaging across 15 runs. In the same manner, the training procedure for each source subject in Dataset II was conducted for 6 runs according to the number of blocks ( $N_b = 6$ ). Therefore, for each source subject, the transferred spatial filters and templates were obtained by averaging across 6 runs. For

Dataset III with 10 blocks, the training procedure for each source subject was conducted for 10 runs, so the transferred spatial filters and templates were obtained by averaging across 10 runs.

In the test stage, all-block data from the target subjects were used for SSVEP detection. Therefore, for Dataset I ( $N_b = 15$ ), the test process of each target subject was repeated 15 times; for Dataset II ( $N_b = 6$ ), the test process of each target subject was conducted 6 times; for Dataset III ( $N_b = 10$ ), the test process of each target subject was conducted 10 times. The classification accuracy and ITR of each target subject were estimated by averaging across blocks.

## III. EXPERIMENTAL RESULTS

### A. Baseline Methods and Parameter Settings

To verify the efficiency of the proposed method, extensive comparisons of SSVEP frequency detection performance evaluated by the classification accuracy and ITR were implemented using three datasets between the proposed method and state-of-the-art methods, FBCCA, tt-CCA, and CSSFT. The parameter settings of each baseline method are described below.

1) *FBCCA*: For all three datasets, the number of harmonics of the reference signal in FBCCA was set to 5. The numbers of subband filters of FBCCA were set to 3, 5, and 3 for Dataset I, II, and III respectively [30]. The final classification accuracy and ITR obtained by FBCCA were calculated by averaging across all subjects.

2) *Tt-CCA*: The number of harmonics and subband filters were set the same as FBCCA. The target subject was sequentially selected one from all subjects, and the remaining subjects were treated as the source subjects.

3) *CSSFT*: Here, the CSSFT was applied to FBCCA, which referred to the FBCCA-based CSSFT method. The number of harmonics and subband filters in CSSFT were the same as in FBCCA. The source subjects in CSSFT were selected from all subjects who achieved the highest recognition accuracies with the FBCCA method, and the remaining subjects are used as the target subjects. Specifically, the highest 2 for Dataset I and III, the highest 5 for Dataset II are selected to be the source subjects.

### B. SSVEP Detection Performance

The overall SSVEP detection performance is illustrated as the averaged accuracies and ITRs at different data lengths shown in Fig. 4. The data lengths ranged from 0.2 s to 2.0 s with an interval of 0.2 s for all three datasets. It is shown in the figures that the proposed method reaches the highest averaged accuracy and ITR among all the compared methods at any data length with both datasets. To further verify the significance of the proposed method on SSVEP detection performance, paired t-tests of accuracy and ITR were conducted among the methods on three datasets. The statistical analysis results show that the proposed method outperformed the other competing methods by a significant margin, especially with Dataset I. And the smaller the data length is, the more significance can be shown between the proposed method and the other competing

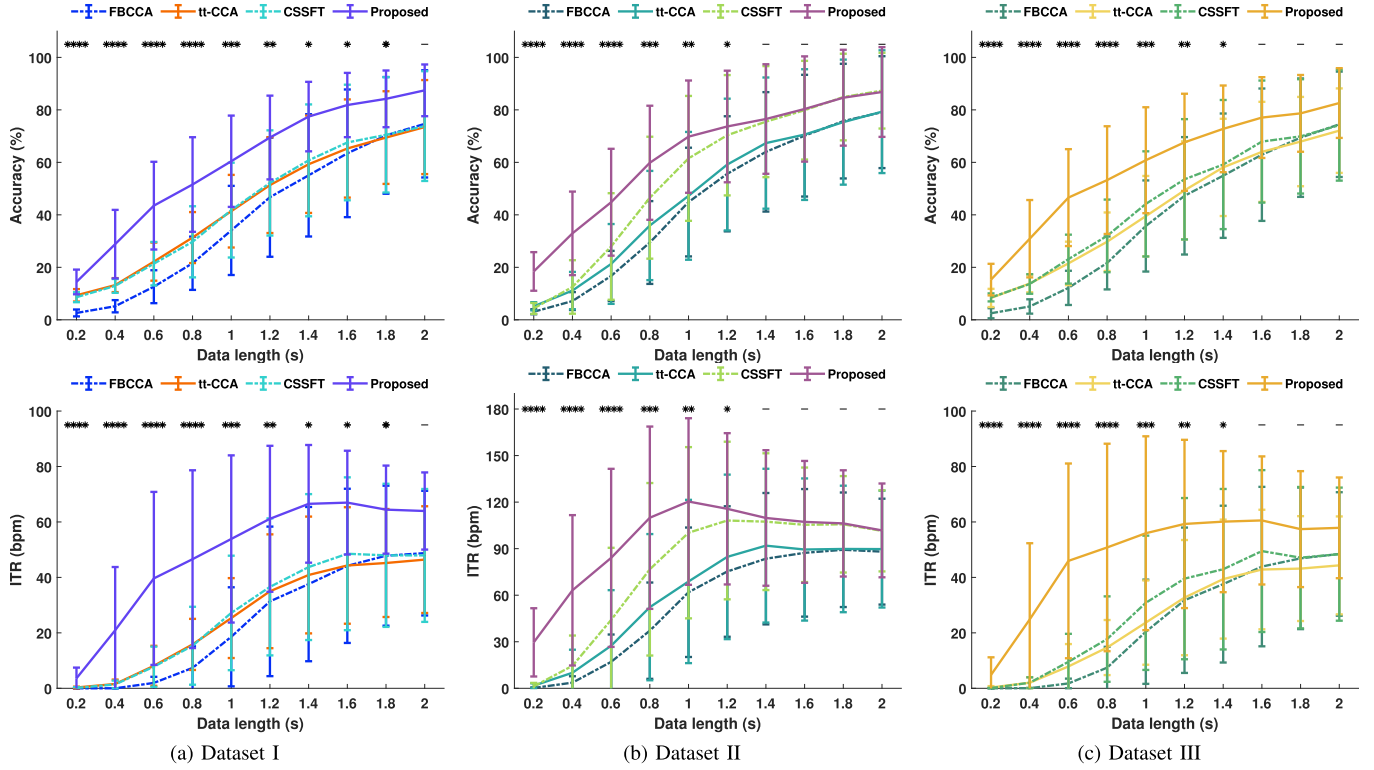


Fig. 4. The averaged accuracies and ITRs obtained by FBCCA, tt-CCA, CSSFT, and the proposed method at different data lengths on Dataset I (a), II (b) and III (c). The data lengths range from 0.2 s to 2.0 s with a step of 0.2 s. The error bars represent standard deviations. The asterisks indicate significant differences between the four methods obtained by paired t-tests (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ).

methods. The comparison results show that the proposed cross-subject transfer method can effectively generalize to the unseen target subjects for SSVEP frequency detection.

The final target SSVEP frequency is directly determined by the feature values calculated by the constructed feature vector as Eq. (24)-(26). To intuitively present the contribution of the proposed method, Fig. 5 presents the feature values corresponding to each stimulus in Dataset I from example subject S1 (Fig. 5(a)), Dataset II from example subject S9 (Fig. 5(b)), and Dataset III from example subject S9 (Fig. 5(c)) obtained by the proposed method and FBCCA. The feature values in each subfigure were calculated by averaging the feature values across blocks and then normalized to 1. The corresponding SSVEP frequency of the target stimulus is presented with the black dashed line. As shown in the figures, the proposed method obtained more accurate decisions by selecting the target stimulus with the largest feature value. It is observed that the feature values obtained by FBCCA show a slight difference between the target and non-target stimulus frequencies, causing a high false rate. By contrast, the proposed method can obtain more discriminative features to detect the true SSVEP target frequency. The comparison result indicates that the proposed feature vector can effectively extract discriminant features to distinguish between the target and non-target stimuli.

### C. The Impact of Parameters

1) *The Number of Source Subjects*: Since the SSVEP data from source subjects play a vital role in this cross-subject

transfer method, the impact of the number of source subjects ( $M$ ) on SSVEP detection performance was first explored in this section. Fig. 6 presents the averaged classification accuracies of tt-CCA, CSSFT, and the proposed method with the varying  $M$ . For Dataset I shown in Fig. 6(a), the averaged classification accuracies were obtained using 1.5-s SSVEP data, where  $M$  varied from 1 to 9 with an interval of 2. In Fig. 6(b), the accuracy estimates were calculated using 1.0-s SSVEP data with  $M$  varying from 5 to 25 with an interval of 5. For Dataset III (Fig. 6(c)), the data length was set as 1.5 s, and the number of source subjects increased from 2 to 10 with a step of 2. As shown in the figures, the averaged accuracies of the state-of-art methods CSSFT remain stable with different numbers of source subjects. Distinct from the CSSFT method, the averaged classification accuracies obtained by the proposed method and tt-CCA increase with the number of source subjects. Compared with the tt-CCA and CSSFT methods, the averaged classification accuracies of the proposed method show superior performance with sufficient source subjects, that is,  $M \geq 5$  for Dataset I and III and  $M \geq 10$  for Dataset II.

2) *The Number of Training Blocks*: In the proposed cross-subject transfer method, the transferred spatial filters and templates were learned using individual calibration data from source subjects. The impact of training blocks on SSVEP detection performance should also be investigated. Fig. 7 provides the averaged classification accuracies across target subjects obtained with different numbers of training blocks ( $N_b$ ) on three datasets. Here, in the proposed method, the

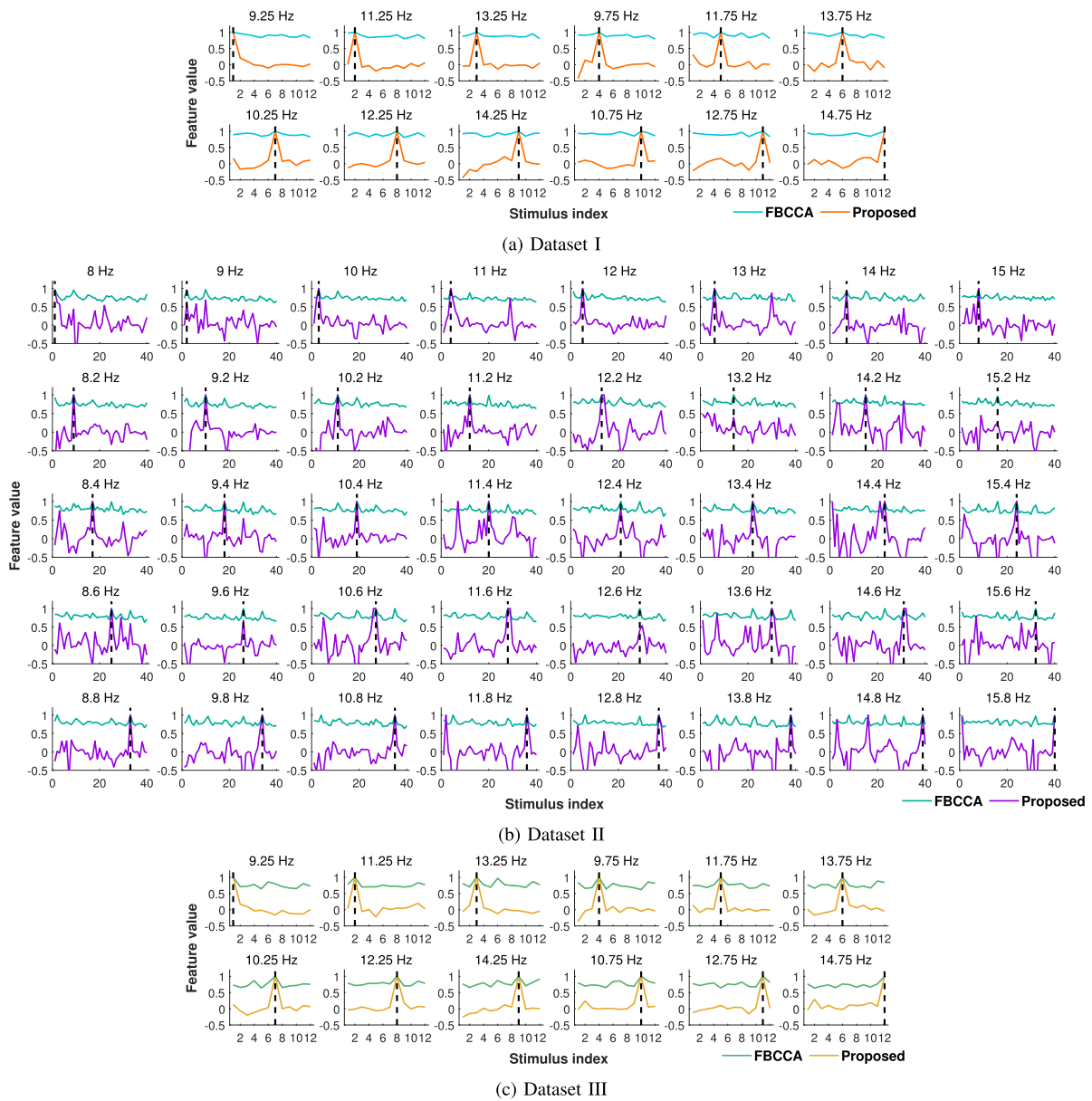


Fig. 5. The feature values corresponding to all stimuli obtained by FBCCA and the proposed method from an example subject. For Dataset I with 12 stimuli (a), Dataset II with 40 stimuli (b), and Dataset III with 12 stimuli (c), the data lengths were set as 1.5 s, 1.0 s, and 1.5 s respectively. The source subjects were randomly selected. The dashed line represents the corresponding SSVEP frequency of the target stimulus.

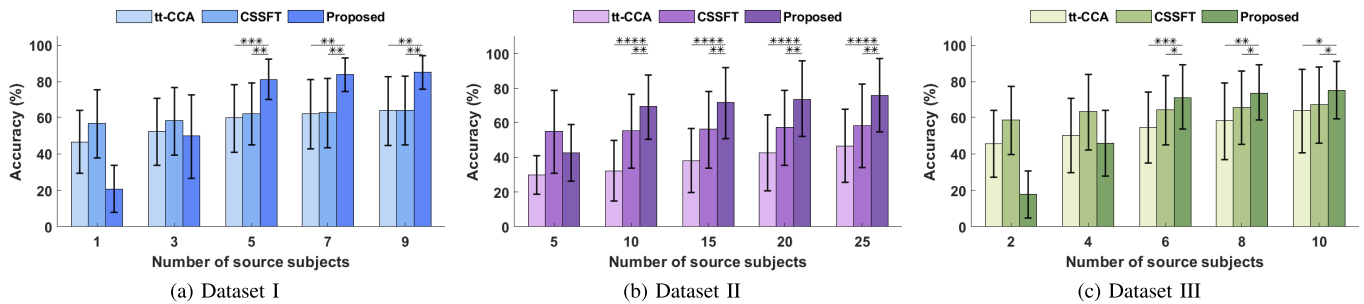


Fig. 6. Averaged classification accuracies across target subjects of Dataset I and III at 1.5 s data length (a,c) and Dataset II at 1.0 s data length (b) with different numbers of source subjects. The vertical error bars represent standard deviations. The asterisks indicate significant differences between every two methods obtained by paired t-tests (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ).

$N_b$  blocks were split into  $N_t$  trials as Eq. (1) and one trial for calculating the three types of transferred spatial filters.

While for the tt-CCA and CSSFT methods, the  $N_b$ -block data were used to calculate the transferred templates by



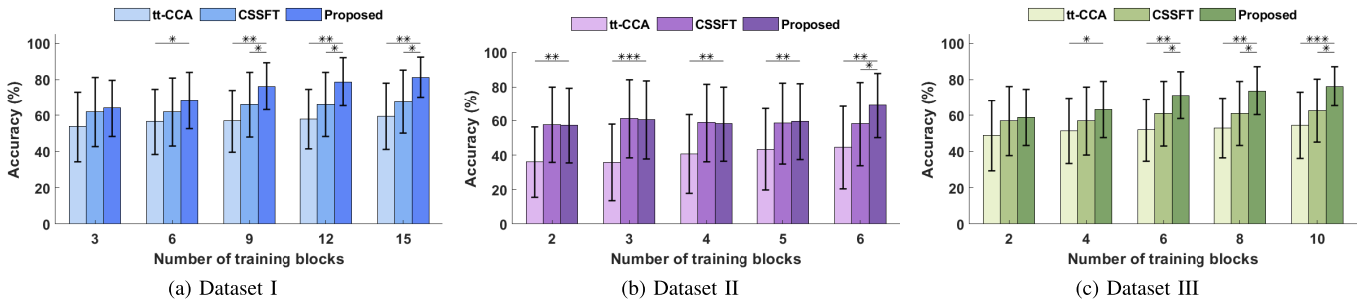


Fig. 7. Averaged classification accuracies across target subjects of Dataset I and III at 1.5 s data length (a,c) and Dataset II at 1.0 s data length (b) with different numbers of training blocks. The vertical error bars represent standard deviations. The asterisks indicate significant differences between each two methods obtained by paired t-tests (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

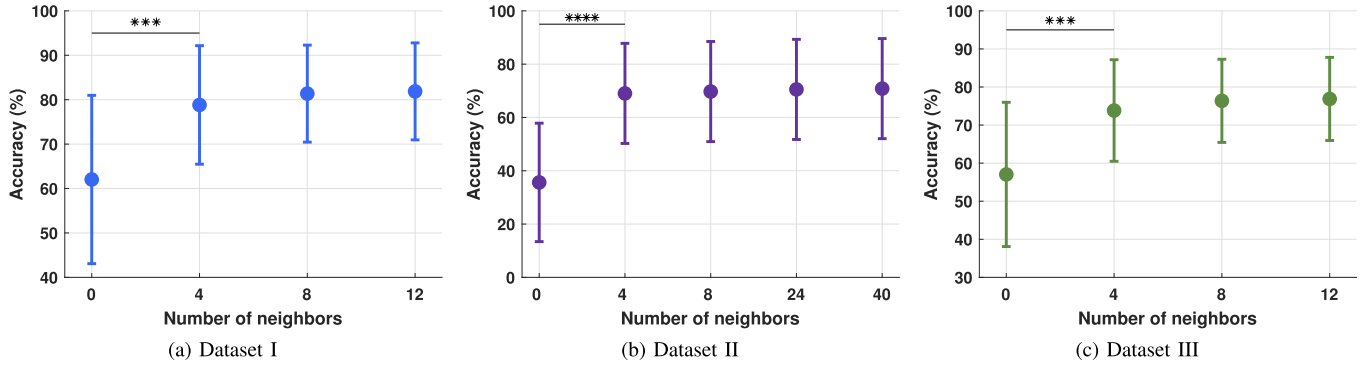


Fig. 8. The averaged classification accuracies across all target subjects from Dataset I at 1.5 s data length (a), Dataset II at 1.0 s data length (b), and Dataset III at 1.5 s data length (c) with different numbers of neighboring stimuli. In the figures, the vertical error bars represent standard deviations. The asterisks indicate significant differences between every two conditions obtained by paired t-tests (\*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ).

averaging across blocks. In Fig. 7(a) for Dataset I, the data length was set as 1.5 s, and the number of training blocks varied from 3 to 15 with an interval of 3. In Fig. 7(b) for Dataset II, the classification accuracies were obtained using 1.0-s SSVEP data with the number of training blocks varying from 2 to 6. In Fig. 7(c) for Dataset III, the classification accuracies were obtained using 1.5-s SSVEP data with the number of training blocks increasing from 2 to 10 with an interval of 2. As can be seen from the graph, with the increment of the number of training blocks, the classification accuracies of the three competing methods increase with a 5% step. Moreover, the proposed method obtained higher accuracies than the other state-of-art methods regardless of the number of training blocks. Therefore, the number of training blocks has shown a positive impact on the SSVEP detection performance.

3) *The Number of Neighboring Stimuli*: The proposed method obtained enhanced performance on SSVEP detection by employing SSVEPs from source subjects corresponding to both target and its neighboring stimuli. Finally, we further explored how the number of neighbors influences the SSVEP frequency detection performance. The results of the classification accuracies with the different numbers of neighbors on three datasets are provided in Fig. 8 by error bars. The comparison was conducted with the number of neighbors ( $N_h$ ) increasing from 0 to the total number of stimuli. For Dataset I and III, the total number of stimuli is 12; while for Dataset II, the total number of stimuli is 40. In the proposed method, the neighbors are defined as the neighboring-location

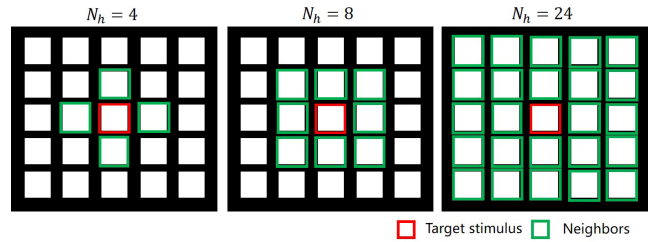


Fig. 9. The illustration of the neighboring stimuli at different numbers of neighbors.

stimuli. Therefore, as shown in Fig. 9,  $N_h = 4$  represents the horizontal and vertical neighbors of the target stimulus.  $N_h = 8$  indicates all neighbors surrounding the target stimulus, including horizontal, vertical, and diagonal neighbors,  $N_h = 24$  represents two-layer neighbors surrounding the target stimulus, including the two-layer horizontal, vertical, and diagonal neighbors. It is noted that  $N_h = 4$  or 8 or 24 only applies to those stimuli in the middle, and the stimuli on the border or corner will have fewer neighbors for both cases. From the graph, we can see that as the number of neighbors increases from 0 to 4, the accuracies of the proposed method show obvious increments in classification accuracies. While the number of neighbors is larger than 4, the accuracy of the proposed method remains stable with a slight difference. Therefore, the incorporation of neighboring stimuli data does contribute to improving SSVEP detection performance, and the horizontal and vertical neighbors are adequate for

learning the transferred spatial filters and templates from source subjects.

#### D. Performance Comparison With Other Transfer Learning Methods

In this proposed method, SSVEP data from the source subject are utilized to obtain spatial filters and templates for the SSVEP detection of the target subject. In other words, the target detection of SSVEPs from the target subject was independent of the SSVEP data of the target subject himself. In this subsection, the proposed method was further compared with three CCA-based subject-independent transfer learning methods, respectively incorporating TRCA [18], multi-stimulus TRCA (ms-TRCA) [40], and task-discriminant component analysis (TDCA) [41]. Here, the transferred spatial filters and templates were obtained by TRCA, ms-TRCA, or TDCA using SSVEPs from source subjects, and then the correlation coefficients between the spatially filtered test data from the target subject and the artificial sine-cosine reference signal and the transferred template were calculated to detect the SSVEP frequency. The selection of source subjects is consistent with the proposed method. The numbers of training blocks were set to 15, 6, and 10 for the three validation datasets in all methods. For TDCA, the number of subspaces and the number of delayed points were set as eight and one respectively. According to the comparison results in Fig. 10, the proposed method achieved the highest accuracies among all four methods with data length increasing from 0.2 s to 2.0 s. The paired t-tests result revealed that the proposed method showed statistically significant superiority compared with the other three methods, which further demonstrated the effectiveness and feasibility of the proposed method in SSVEP detection.

## IV. DISCUSSIONS

### A. Model's Performance

In this paper, we propose a cross-subject transfer scheme based on domain generalization that learns transferred spatial filters and templates from source subjects and then generalizes them to the unseen target subject. In the proposed method, the transferred spatial filters and templates are obtained by maximizing the intra- and inter-subject correlations only from the source domain. Although numerous cross-subject methods in SSVEP-based BCIs utilized intra- and inter-subject correlation maximization to learn transferred parameters, such as stCCA [29] and IISMIC [27], the transferred parameters were learned from both the source and target domain, which indicates that individual calibration data are still required from target subjects. Instead of using individual calibration data from the target subject, domain adaption methods using the unlabeled data from target domain were proposed to improve the detection performance and reduce the calibration effort, such as ALign and Pool for EEG Headset domain Adaptation (ALPHA) [42]. Compared to these closely-related transfer learning methods, the proposed method learns transferred parameters without access to the data from target domain. From the users' viewpoint, there is no initial training period

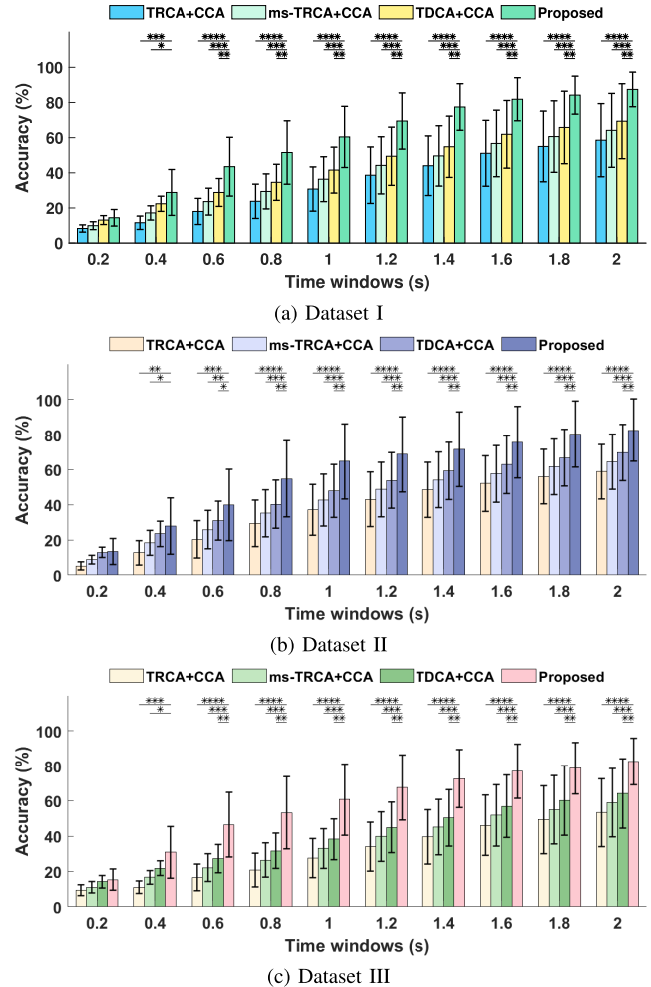


Fig. 10. The averaged classification accuracies obtained by different transfer learning methods across target subjects from three datasets with time windows increase from 0.2 s to 2.0 s. The vertical error bars represent standard deviations. The asterisks indicate significant differences between every two methods obtained by paired t-tests (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ).

for data collection and analysis of the new users. Therefore, the training process in the proposed method is implicit, which makes the SSVEP-based BCI a plug-and-play system. By conducting extensive comparisons with the state-of-art methods, the experimental results present the superiority on SSVEP detection performance by at least 10% (Fig. 4), meaning that the proposed method can provide a satisfactory classification accuracy and ITR for practical applications. In addition, the comparison results in Fig. 7 illustrate that the proposed method can reach relatively high SSVEP detection accuracy in the case of small calibration data from source subjects.

### B. Transferred SSVEP Templates

In this proposed method, the accurate SSVEP frequency detection mainly depends on the transferred SSVEP templates containing accurate SSVEP features, which are constructed by the SSVEPs from source subjects. To reveal the contributions of the transferred templates, the target stimulus of 14.75-Hz frequency and  $1.5\pi$  phase was taken as an example. Fig. 11 provides the averaged spatially filtered SSVEP templates in

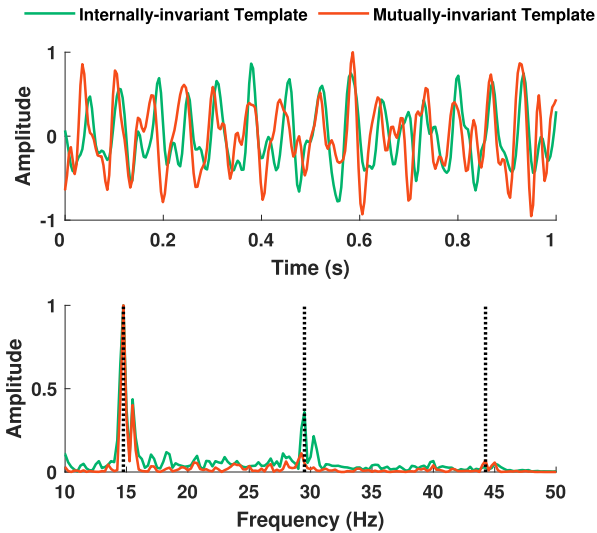


Fig. 11. The averaged spatially filtered SSVEP templates in the time domain (top) and frequency domain (bottom) from constructed using 1.0-s SSVEP data from source subjects in Dataset I. Here, the example target stimulus is set as the one at the 14.75-Hz frequency and  $1.5\pi$  phase. The dotted lines represent the target stimulus frequency and its corresponding harmonics.

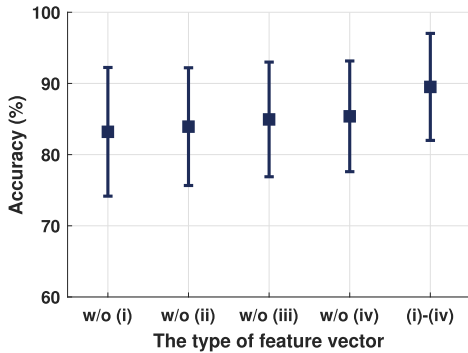


Fig. 12. The averaged classification accuracies with different types of feature vectors using Dataset I. The vertical error bars represent standard deviations. The data length was set to 1.5 s.

the time domain (top) and frequency domain (bottom) with 1.0-s SSVEP data from source subjects in Dataset I. As illustrated in the figures, the internally-invariant template extracts relatively strong responses corresponding to the fundamental and harmonic frequencies, while the two types of templates can extract accurate periodic impulse responses in the time domain. In conclusion, both internally- and mutually-invariant templates can capture accurate features in the aspect of time and frequency domain. By combining two types of templates for SSVEP detection, the performance can be significantly boosted for the SSVEP-based BCIs.

### C. Feature Vector Construction

In this method, the feature vector (Eq. (23)) is constructed with four different types of correlation coefficients as (i)-(iv) described in Section II, D. We further explored the contribution of each correlation feature by conducting the ablation experiments. In Fig. 12, the averaged classification accuracies across target subjects with different types of feature

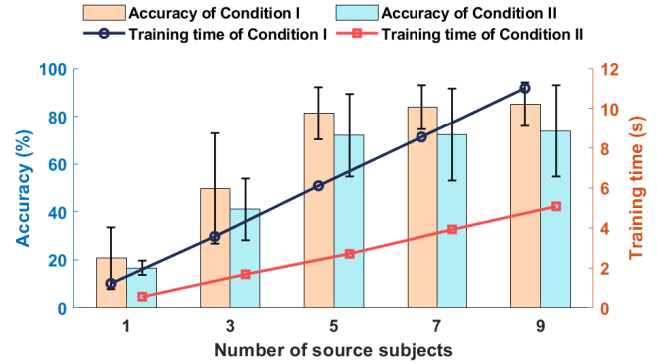


Fig. 13. The SSVEP detection accuracy and training time for two conditions with different numbers of source subjects using Dataset I are illustrated in a double y-axis graph. The bars represent the classification accuracies with the left y-axis, and the lines represent the training time with the right y-axis. The vertical error bars represent standard deviations of classification accuracy. The data length was set to 1.5 s.

vectors were compared using Dataset I. The different types of feature vectors were respectively defined as: 1) without (w/o) (i); 2) w/o (ii); 3) w/o (iii); 4) w/o (iv); 5) (i)-(iv). As can be seen from the figure, with all four correlations as (i)-(iv), the proposed method reached the highest classification accuracies among the five types of feature vectors. Compared to the feature vector 5) in the proposed method, the accuracies achieved by the other four types of feature vectors showed at least a 5% difference. The comparison result demonstrates that each correlation coefficient used in Eq. (23) contributes to improving the SSVEP recognition performance of the target subject.

### D. Algorithm Implementation Complexity

It can be seen from the experimental results in Fig. 6 that the classification accuracies of the proposed method increased as the number of source subjects. However, such improvement came at the expense of more computational costs. Moreover, the  $N_b$  times of training also requires massive computation time. To further investigate the computation overhead, we compared the detection performance and training time between the two settings, one is training with  $N_b$  combinations of SSVEPs from source subjects as the proposed method (termed as Condition I), and the other is training with all-block data from the source subjects without the leave-one-out splitting (termed as Condition II). Figure 13 compares the classification accuracies and the computation costs under these two settings with the number of source subjects increasing from 1 to 9 using Dataset I. The cross-subject transfer algorithm was implemented on a Lenovo PC with the Intel(R) Xeon(R) Silver 4116 CPU @ 2.10GHz, 32 GB RAM, and 64-bit Windows 10 OS using Matlab 2022a. From the graph, we can see that the detection performance and the training time for each target stimulus of the proposed method both increase as the number of source subjects. Compared with the all-block training condition, the proposed method requires more computation overhead but achieves higher SSVEP detection performance. Although the training time of the proposed method was relatively high with sufficient source subjects, the

SSVEP detection stage only costs 0.03 s for each time window, which indicates that it won't affect the computational speed of real-time SSVEP detection in practical applications.

### E. Limitations and Future Directions

As a subject-independent transfer method, the proposed cross-subject transfer method has shown satisfactory performance on SSVEP detection, which provides the potential of constructing plug-and-play SSVEP-based BCIs. However, there still exists abundant room for further progress on the proposed method. First, according to the analysis of implementation complexity as Fig. 13, the training process of the proposed transfer method causes relatively high computation overhead due to the requirements of sufficient source subjects and  $N_b$ -run repetitive training. To further boost the SSVEP detection performance with fewer source subjects, the feature extraction of SSVEP signals could be further improved by incorporating sine-cosine reference signals as [28]. Second, as the possible targets increase in the SSVEP-based BCIs for practical applications [43], [44], the training process corresponding to all targets still consumes massive time. Therefore, the cross-target transfer scheme [45], [46], [47] should be further considered to improve the usability of the SSVEP-based BCIs. Finally, the proposed method is a proof-of-concept that verified the effectiveness and feasibility of offline experiments. To meet the requirements of practical applications, the dynamic window strategy [48], [49] can be incorporated for the robust control of the SSVEP-based BCIs.

## V. CONCLUSION

In this paper, a cross-subject transfer method based on domain generalization was proposed, which transferred the spatial filters and templates learned from source subjects to the target subject with no access to the SSVEP data from the target subject. The transferred spatial filters and templates are obtained by maximizing the intra- and inter-subject correlations using the SSVEP data corresponding to the target and its neighboring stimuli. For SSVEP detection, four types of correlation coefficients based on the transferred spatial filters and templates were calculated to construct the feature vector. The effectiveness and feasibility of the proposed method were demonstrated through experimental evaluations on three SSVEP datasets.

## REFERENCES

- [1] J. R. Wolpaw et al., "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000.
- [2] S. He et al., "EEG- and EOG-based asynchronous hybrid BCI: A system integrating a speller, a web browser, an e-mail client, and a file explorer," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 519–530, Feb. 2020.
- [3] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an EEG-based brain-computer interface (BCI)?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 145–147, Jun. 2003.
- [4] W. Pei et al., "A pre-gelled EEG electrode and its application in SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 843–850, 2022.
- [5] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1181–1186, Oct. 2002.
- [6] Y. 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.
- [7] J. Huang, P. Yang, B. Xiong, B. Wan, K. Su, and Z.-Q. Zhang, "Latency aligning task-related component analysis using wave propagation for enhancing SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 851–859, 2022.
- [8] N. Shi, X. Li, B. Liu, C. Yang, Y. Wang, and X. Gao, "Representative-based cold start for adaptive SSVEP-BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1521–1531, 2023.
- [9] Y. Zhang, S. Q. Xie, H. Wang, and Z. Zhang, "Data analytics in steady-state visual evoked potential-based brain-computer interface: A review," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1124–1138, Jan. 2021.
- [10] 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.
- [11] Y. Zhang, Z. Li, S. Q. Xie, H. Wang, Z. Yu, and Z.-Q. Zhang, "Multi-objective optimization-based high-pass spatial filtering for SSVEP-based brain-computer interfaces," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.
- [12] 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.
- [13] 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.
- [14] K. Wang, D.-H. Zhai, Y. Xiong, L. Hu, and Y. Xia, "An MVMD-CCA recognition algorithm in SSVEP-based BCI and its application in robot control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 5, pp. 2159–2167, May 2022.
- [15] Y. Zhang, G. Zhou, J. Jin, M. Wang, X. Wang, and A. Cichocki, "L1-regularized multiway canonical correlation analysis for SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 6, pp. 887–896, Nov. 2013.
- [16] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in SSVEP-based BCI using multisets canonical correlation analysis," *Int. J. Neural Syst.*, vol. 24, no. 4, Jun. 2014, Art. no. 1450013.
- [17] Y. Wang, M. Nakanishi, Y.-T. Wang, and T.-P. Jung, "Enhancing detection of steady-state visual evoked potentials using individual training data," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 3037–3040.
- [18] M. Nakanishi, Y. Wang, X. Chen, Y.-T. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, Jan. 2018.
- [19] G. R. K. Kumar and M. R. Reddy, "Designing a sum of squared correlations framework for enhancing SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2044–2050, Oct. 2019.
- [20] Y. Zhang et al., "Correlated component analysis for enhancing the performance of SSVEP-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 5, pp. 948–956, May 2018.
- [21] J. Huang et al., "Incorporating neighboring stimuli data for enhanced SSVEP-based BCIs," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.
- [22] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 1, pp. 4–19, Mar. 2022.
- [23] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, Jan. 2021.
- [24] K.-J. Chiang, C.-S. Wei, M. Nakanishi, and T.-P. Jung, "Boosting template-based SSVEP decoding by cross-domain transfer learning," *J. Neural Eng.*, vol. 18, no. 1, Feb. 2021, Art. no. 016002.
- [25] P. Yuan, X. Chen, Y. Wang, X. Gao, and S. Gao, "Enhancing performances of SSVEP-based brain-computer interfaces via exploiting inter-subject information," *J. Neural Eng.*, vol. 12, no. 4, Aug. 2015, Art. no. 046006.

- [26] N. R. Waytowich, J. Faller, J. O. Garcia, J. M. Vettel, and P. Sajda, "Unsupervised adaptive transfer learning for steady-state visual evoked potential brain-computer interfaces," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 4135–4140.
- [27] H. Wang et al., "Cross-subject assistance: Inter- and intra-subject maximal correlation for enhancing the performance of SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 517–526, 2021.
- [28] Y. Zhang, S. Q. Xie, C. Shi, J. Li, and Z.-Q. Zhang, "Cross-subject transfer learning for boosting recognition performance in SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1574–1583, 2023.
- [29] 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.
- [30] W. Yan, Y. Wu, C. Du, and G. Xu, "Cross-subject spatial filter transfer method for SSVEP-EEG feature recognition," *J. Neural Eng.*, vol. 19, no. 3, Jun. 2022, Art. no. 036008.
- [31] W. Yan, Y. Wu, C. Du, and G. Xu, "An improved cross-subject spatial filter transfer method for SSVEP-based BCI," *J. Neural Eng.*, vol. 19, no. 4, Aug. 2022, Art. no. 046028.
- [32] J. Wang et al., "Generalizing to unseen domains: A survey on domain generalization," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 8, pp. 8052–8072, Aug. 2022.
- [33] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 10–18.
- [34] W. Lu, J. Wang, H. Li, Y. Chen, and X. Xie, "Domain-invariant feature exploration for domain generalization," *Trans. Mach. Learn. Res.*, 2022.
- [35] M. Nakanishi, Y. Wang, Y.-T. Wang, and T.-P. Jung, "A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials," *PLoS One*, vol. 10, no. 10, Oct. 2015, Art. no. e0140703.
- [36] Y. Wang, X. Chen, X. Gao, and S. Gao, "A benchmark dataset for SSVEP-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1746–1752, Oct. 2017.
- [37] 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.
- [38] F. Di Russo and D. Spinelli, "Electrophysiological evidence for an early attentional mechanism in visual processing in humans," *Vis. Res.*, vol. 39, no. 18, pp. 2975–2985, Sep. 1999.
- [39] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction speech Process.* Berlin, Germany: Springer, 2009, pp. 1–4.
- [40] 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.
- [41] B. Liu, X. Chen, N. Shi, Y. Wang, S. Gao, and X. Gao, "Improving the performance of individually calibrated SSVEP-BCI by task-discriminant component analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1998–2007, 2021.
- [42] B. Liu, X. Chen, X. Li, Y. Wang, X. Gao, and S. Gao, "Align and pool for EEG headset domain adaptation (ALPHA) to facilitate dry electrode based SSVEP-BCI," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 795–806, Feb. 2022.
- [43] 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.*, Jun. 2021, Art. no. 046094.
- [44] X. Chen, B. Liu, Y. Wang, and X. Gao, "A spectrally-dense encoding method for designing a high-speed SSVEP-BCI with 120 stimuli," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2764–2772, 2022.
- [45] 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.
- [46] Z. Wang, C. M. Wong, A. Rosa, T. Qian, T.-P. Jung, and F. Wan, "Stimulus-stimulus transfer based on time-frequency-joint representation in SSVEP-based BCIs," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 2, pp. 603–615, Feb. 2023.
- [47] X. Wang, A. Liu, L. Wu, C. Li, Y. Liu, and X. Chen, "A generalized zero-shot learning scheme for SSVEP-based BCI system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 863–874, 2023.
- [48] J. Jiang, E. Yin, C. Wang, M. Xu, and D. Ming, "Incorporation of dynamic stopping strategy into the high-speed SSVEP-based BCIs," *J. Neural Eng.*, vol. 15, no. 4, Aug. 2018, Art. no. 046025.
- [49] P. Gaur, H. Gupta, A. Chowdhury, K. McCreddie, R. B. Pachori, and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.