Stimulus-stimulus Transfer Based on Time-frequency-joint Representation in SSVEP-based BCIs

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

***Abstract*— Objective: Brain-computer interfaces (BCIs) based on steady-state visual evoked potential (SSVEP) require extensive and costly calibration to achieve high performance. Using transfer learning to re-use existing calibration data from old stimuli is a promising strategy, but finding commonalities in the SSVEP signals across different stimuli remains a challenge. Method: This study presents a new perspective, namely time-frequency-joint representation, in which SSVEP signals corresponding to different stimuli can be synchronized, and thus can emphasize common components. According to this time-frequency-joint representation, an adaptive decomposition technique based on the multi-channel adaptive Fourier decomposition (MAFD) is proposed to adaptively decompose SSVEP signals of different stimuli simultaneously. Then, common components can be identified and transformed across stimuli. Results: A simulation study on public SSVEP datasets demonstrates that the proposed stimulus-stimulus transfer method has the ability to extract and transfer these common components across stimuli. By using calibration data from eight source stimuli, the proposed stimulus-stimulus transfer method can generate SSVEP templates of other 32 target stimuli. It boosts the stimulus-stimulus transfer based recognition method's ITR from 95.966 bits/min to 123.684 bits/min. Conclusion: By extracting and transfer common components across stimuli in the proposed time-frequency-joint representation, the proposed stimulus-stimulus transfer method produces good classification performance without requiring calibration data of target stimuli. Significance: This study provides a synchronization standpoint to analyze and model SSVEP signals. In addition, the proposed stimulus-stimulus method shortens the calibration time and thus improve comfort, which could facilitate real-world applications of SSVEP-based BCIs.**

***Index Terms*— Adaptive Fourier decomposition, brain-computer interface, multi-channel signal analysis, steady-state visual evoked potential, stimulus-stimulus transfer.**

This work was supported in part by the Science and Technology Development Fund, Macau SAR under Grant 055/2015/A2, Grant 0045/2019/AFJ, Grant 0018/2019/AKP, Grant 0022/2021/APD; in part by the University of Macau under Grant MYRG2016-00240-FST, and Grant MYRG2017-00207-FST.

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

Agostinho Rosa is with the Department of Bioengineering, LaSEEB-System and Robotics Institute, Instituto Superior Tecnico, University of Lisbon, Lisbon, Portugal.

Tao Qian is with the Macao Center for Mathematical Sciences, Macau University of Science and Technology, Macau.

Tzyy-Ping Jung is with the Institute for Neural Computation and Institute of Engineering in Medicine, University of California, La Jolla, CA 92093, USA.

I. Introduction

The brain-computer interfaces (BCIs) enable communication between the brain and the external world by directly measuring brain activities [1]. Because of high signal-to-noise ratio (SNR) and high information transfer rate (ITR), steady-state visual evoked potential (SSVEP)-based BCIs have gained increasing attention [2]–[13]. A good recognition method is critical in SSVEP-based BCIs. Currently, the most widely used SSVEP recognition methods are based on the canonical correlation analysis (CCA) [2], [4], [6], [10]. The CCA-based methods measure the similarity between the recorded EEG signals and the corresponding set of the reference or template signals. The stimulus with the highest similarity is regarded as the target [8]. The conventional CCA method uses combinations of the sine-cosine signals as the reference signals [2]. Although the robustness of the conventional CCA method in the SSVEP recognition has been proved in the literature, it is susceptible to interference from spontaneous EEG activities [6]. To reduce the interference from the spontaneous EEG signals, calibration-based algorithms were introduced, including the extended CCA (eCCA) [4], ensemble task-related component analysis (eTRCA) [6], multi-stimulus eCCA (ms-eCCA) [10], multi-stimulus eTRCA (ms-eTRCA) [10], and task-discriminant component analysis (TDCA) [14]. Although these methods can provide the state-of-the-art recognition performance, the critical issue of these methods is that they require a large number of calibration data to optimize spatial filters and construct suitable template signals for high performance. The calibration process of SSVEP-based BCIs is time-consuming. Users can easily become fatigued by such a long calibration process, discouraging them from using SSVEP-based BCIs [15].

To solve this critical issue of these prominent SSVEP recognition methods, the transfer learning technique was adopted to

re-use calibration data across domains [9], [16]–[19]. The early studies focus on transferring SSVEP templates across subjects, including the transfer-template-based CCA (ttCCA) proposed in [16] and the inter-subject transfer method proposed in [18]. In these studies, the weighted averages of source subjects' SSVEP signals are used to form common templates and then transferred to target subjects. Then, several studies extend the transfer learning to more domains: 1) The least-square transformation methods proposed in [9], [17] estimate the transformation matrices to minimize the differences between signals in the target and source sessions, subjects, and devices; 2) The align and pool for EEG headset domain adaptation (ALPHA) proposed in [19] aligns the spatial patterns and covariances across target and source devices. Although these methods care effective at transferring signals across different domain for SSVEP signals, the source and target signals are corresponding to same stimuli. Recently, two stimulus-stimulus transfer methods were proposed in [12], [20]. The transfer-extended canonical correlation analysis (t-eCCA) [20] models SSVEP responses as a Volterra filter, and transfers frequency-domain features across stimuli using the spectrum linear relationships of SSVEP signals with adjacent stimulus frequencies. The transfer learning CCA (tlCCA) proposed in our previous work [12] models SSVEP signals as the convolution of the impulse response and the periodic stimulus following the superposition theory [21], and uses the similarity of the impulse responses with adjacent stimulus frequencies to transfer SSVEP templates across stimuli. However, because the visual stimuli in the SSVEP-based BCIs are designed to induce SSVEP signals as different as possible [22], the commonality of SSVEP signals corresponding to various stimuli are naturally not obvious in both time and frequency domains. Therefore, these features or components extracted from the t-eCCA and tlCCA are still stimulus-specific, resulting these stimulus-stimulus transfer technique is limited to adjacent frequencies. In order to develop a more general cross-stimulus transfer learning method that can extract common components of SSVEP signals corresponding to various stimuli, two key issues in existing stimulus-stimulus transfer methods must be addressed: 1) Existing SSVEP signal models are not designed for the stimulus-stimulus transfer and thus cannot adequately describe the commonality of SSVEP signals across stimuli; 2) Conventional signal decomposition methods are not suitable for extracting common components of SSVEP signals.

For the first issue, there are two well-known and widely-used SSVEP signal models in the literature. One is the sine-cosine-based SSVEP model proposed with the conventional CCA method and commonly applied to generate reference signals in SSVEP recognition methods [2], [4], [5], [18], [23], [24]. The sine-cosine-based model describes SSVEP signals as the weighted summation of the sinusoidal and cosinusoidal signals with corresponding harmonic stimulus frequencies. According to the properties of Fourier basis, the sinusoidal signals with different frequencies are orthogonal, which makes the sine-cosine-based model naturally emphasize the differences between the SSVEP signals of different stimuli. The other is the superposition-theory-based SSVEP model [12], [21], in which SSVEP signals are described as the convolution of periodic impulses and impulse responses. The impulse responses and the periodic impulses are defined in the time domain, making them still stimulus-specific. To solve this issue, this study revisits these two models and develops a general signal model that directly describes the commonality of SSVEP signals across stimuli.

For the second issue, the conventional non-adaptive signal decomposition methods typically use the pre-defined basis to conduct the signal decomposition [25]. For example, the conventional Fourier decomposition uses the complex exponential basis and provides a useful frequency-domain analysis. However, when the expected signal components cannot be separated in the frequency domain, or contain the time-varying frequencies, the conventional Fourier decomposition cannot provide good performance. Another example is that the wavelet decomposition uses the pre-defined mother wavelet to provide a good localization in both time and frequency domains simultaneously. However, pre-defining proper mother wavelets requires enough prior knowledge. Due to the varying properties of SSVEP signals on different channels, sessions, subjects, and paradigms as well as the lack of prior knowledge about common components across stimuli, it is difficult to find or design proper pre-defined basis components for conventional decomposition methods. Therefore, this study considers the adaptive decomposition to solve this issue. Most adaptive decomposition methods, such as the empirical mode decomposition (EMD), the variation mode decomposition (VMD), and the nonlinear chirp model decomposition (NCMD), lack explicit mathematical explanations, thereby making their decomposition results difficult to understand and unsuitable for modeling [26]–[28]. To avoid these drawbacks, this study adopts the multi-channel adaptive Fourier decomposition (MAFD) to decompose the SSVEP signals. The MAFD, which we proposed in previous research [29], [30], is the multi-channel extension of the conventional AFD. Specifically, the MAFD finds same adaptive basis components to decompose a set of signals in different channels simultaneously. Common components can be extracted by analyzing decomposition coefficients. However, the MAFD is designed to analyze general signal, not dedicated to the SSVEP signal. Hence, in this study, the MAFD is deliberately modified to the MAFD with different phase (DP-MAFD) for the SSVEP signal.

In order to develop a more general cross-stimulus transfer learning method in the SSVEP-based BCIs, we firstly propose the DP-MAFD to decompose periodic components among the existing SSVEP templates corresponding to the known stimuli, and then recommends three rules to find the shared knowledge across stimuli. Then, based on the shared knowledge, we can construct the new SSVEP templates for recognizing unknown stimuli. These constructed SSVEP templates are used to recognize of the target stimuli. The following is a summary of the novelty of proposed method: 1) A new time-frequency-joint representation is proposed, which emphasize commonality and provides an effective way to describe SSVEP signals of various stimuli together by synchronizing SSVEP signals; 2) Inspired by the sine-cosine-based and the superposition-theory-based SSVEP signal models in the time-frequency-

joint representation, this study proposes a new SSVEP signal model, which directly describes SSVEP signals as the summation of common components; 3) This study proposes the DP-MAFD, which incorporates the proposed SSVEP model into the MAFD, allowing it adaptively decomposes periodic SSVEP components of different stimuli simultaneously; 4) This study recommends three rules to select common components across stimuli based on the design of SSVEP-based BCIs and characteristics of SSVEP signals; 5) The transferability of common components extracted by the proposed method is evaluated through simulations on widely-used and reliable datasets. For SSVEP-based BCIs, this study is the first time to model SSVEP signals using common components across stimuli, which promotes the cross-stimulus transfer learning. Section II and Section III provide the preliminaries and the proposed stimulus-stimulus transfer learning method. After we present the performance evaluation in Section IV and discussions in Section V, the paper concludes in Section VI.

II. Preliminaries

A. Notations

For a clear introduction to the time-frequency-joint representation and the proposed stimulus-stimulus transfer method, Table I summarizes the key notations used in this study. In general, the variables or functions with right subscripts c and n denote that they are corresponding to the c-th EEG channel and n-th decomposition level, respectively. In addition, the functions with t and θ as variables denote signals in the time-domain and time-frequency-joint representations, respectively. It should be noted that, the following contents of this paper assume that there are total M stimuli, with the first J stimuli being source stimuli that have calibration data, and the rest Q stimuli being target stimuli that do not have calibration data. The variables and functions with right subscripts j and q denote that they are corresponding to source stimulus and target stimuli, respectively, where $j = 1, \cdots, J$, $q = 1, \cdots, Q$, and $M = J + Q$.

TABLE I
KEY NOTATIONS

Notation	Description
$x_{\text{sine-cosine}}(t)$ and $s_{\text{sine-cosine}}(\theta)$	Sine-cosine-based SSVEP signal model in the time-domain and time-frequency-joint representations
$x_{\text{superpos}}(t)$ and $s_{\text{superpos}}(\theta)$	Superposition-theory-based SSVEP signal model in the time-domain and time-frequency-joint representations
$x_{\text{syn}}(t)$	Proposed SSVEP signal model based on the time-frequency representation
f_{stim}, T_{stim}, and θ_{stim}	Frequency, period, and initial phase of stimulus
T_j, f_j and θ_j $(j = 1, \cdots, J)$	Stimulus period, frequency, and phase of j-th source stimulus
T_q, f_q and θ_q $(q = 1, \cdots, Q)$	Stimulus period, frequency, and phase of q-th target stimulus
$x_{c,j}(t)$ and $g_{c,j}(\theta)$	SSVEP signal in the time-domain representation and corresponding analytic signal in the time-frequency-joint representation for the c-th channel and the j-th stimulus
$B_{c,n}(\theta)$	AFD basis component of the c-th channel and the n-th decomposition level
$A_{c,j,n}$	Decomposition coefficient of the c-th channel, the j-th source stimulus, and the n-th decomposition level
$a_{n,\text{AFD}}$ and $a_{n,\text{MAFD}}$	Parameters of basis components of AFD and MAFD in the n-th decomposition level
$a_{c,n,\text{DP-MAFD}}$	Parameter of basis component of DP-MAFD in the c-th channel and the n-th decomposition level
$\text{e}_{\{a\}}(\theta)$	Evaluator of the basis component with the parameter a
$\mathbf{S}_c$	Averaged SSVEP signals across trials of all source stimuli for the c-th channel
$\hat{\mathbf{S}}_c$	SSVEP template of all target stimuli constructed by the proposed stimulus-stimulus transfer method for the c-th channel
$\mathbf{B}_c$	AFD basis components for the c-th channel
$\mathbf{A}_c$	Matrix of decomposition coefficients of source stimuli for the c-th channel
$\mathbf{A}_c^{\text{sel}}$	Matrix of selected decomposition coefficients of common components for the c-th channel
$\mathbf{A}_c^{\text{re}}$	Matrix of re-arranged decomposition coefficients that are applied to construct SSVEP templates of target stimuli for the c-th channel

B. SSVEP dataset

This study used the benchmark dataset collected by the Tsinghua group from 35 subjects participating in the SSVEP-based BCI experiment [31]. This experiment uses 40-target BCI speller and a sampled sinusoidal stimulation method to present visual stimuli with the luminance of the screen is controlled by the stimulus sequence sampled from $0.5 + 0.5\sin(2\pi f_{\text{stim}} t + \theta_{\text{stim}})$ where f_{stim} and θ_{stim} denote the corresponding stimulus frequency and phase, respectively. The stimulus frequencies start from 8 Hz to 15.8 Hz with 0.2 Hz interval. The relevant stimulus phases range from 0 to 1.5π with a 0.5π interval and are repeated for stimulus frequencies. For each subject, this experiment included 6 blocks, with each block consisting of 40 trials corresponding to 40 stimuli. Every trial started with a 0.5-s target cue. Then, all stimuli were flickered on the screen for 5 s. Finally, the screen went blank for 0.5 s before next trial.

EEG signals from 64 channels were recorded and down-sampled to 250 Hz. A notch filter at 50 Hz was applied to remove the power-line noise. Following [31], the EEG signals from 9 selected channels (Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, and O2) are used in this study. Furthermore, EEG data in each 6-s trial were filtered using a band-pass filter with low and high cut-off frequencies of 7 Hz and 90 Hz, respectively, based on the ranges of stimulus frequencies.

C. SSVEP signal models based on sine-cosine signals and superposition theory

According to the working hypothesis of the conventional CCA method [2], the SSVEP signals can be considered as the output of a linear system with the corresponding stimulus signals as the input, which is well-known and also widely used in SSVEP recognition studies [4], [6], [10]. The idea

behind this hypothesis is that the fundamental frequencies and phases of the SSVEP signals follow the corresponding stimulus and thus differ among different stimuli. Therefore, the conventional SSVEP signal model proposed in [2] uses the sine-cosine signals to model the SSVEP signals as shown in

$$x_{\text{sine-cosine}}(t) = \sum_{k=1}^{K} \{p_k \sin(2\pi k f_{\text{stim}} t + k\theta_{\text{stim}}) + q_k \cos(2\pi k f_{\text{stim}} t + k\theta_{\text{stim}})\}, \quad (1)$$

where p_k and q_k are coefficients of sine-cosine signals with the k-th harmonic stimulus frequency and normally calculated by maximizing the correlation between EEG signals and these sine-cosine signals.

Besides the sine-cosine SSVEP signal model, Capilla et al. used the superposition theory to model SSVEP signals as the convolutions of the periodic impulses and the impulse responses following

$$x_{\text{superpos}}(t) = r(t) * h(t), \quad (2)$$

where $*$ denotes the convolution computation, $h(t)$ denotes the periodic impulse, i.e., $h(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_{\text{stim}})$, $k \in \mathbb{Z}$, $\delta(t)$ is the Dirac delta function, T_{stim} is the period of corresponding stimulus, i.e., $T_{\text{stim}} = 1/f_{\text{stim}}$, and $r(t)$ denotes the impulse response [21]. Our previous work [12] adopts the least-square method to estimate the impulse responses.

D. AFD and MAFD

The AFD is an adaptive signal decomposition method. It generates adaptive basis $\{B_n\}_{n=1}^{\infty}$ based on a matching pursuit process, where

$$B_n(\theta) = \frac{\sqrt{1-|a_n|^2}}{1-\overline{a}_n e^{j\theta}} \prod_{k=1}^{n-1} \frac{e^{j\theta} - a_k}{1-\overline{a}_k e^{j\theta}}, \quad (3)$$

$a_n \in \mathbb{D}$, $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, $\mathbb{C}$ is the complex plane, and n denotes the decomposition level [32]. It should be noted that the basis components in $\{B_n\}_{n=1}^{\infty}$ are periodic with period of 2π and orthogonal to others. Because $B_n(\theta)$ is determined solely by the sequence of a_n, selecting a suitable a_n is a key procedure in each decomposition level of the AFD. The value of a_n should be able to guarantee a fast signal decomposition in terms of the energy convergence and thus make the decomposition components match the processed signal well, which can be achieved based on the attainability of

$$a_{n,\text{AFD}} = \arg\max_{a \in \mathbb{D}} \left\{ E_{\{B_n\}}(a) \right\}, \quad (4)$$

where $E_{\{B_n\}}(a)$ denotes the extracted energy from using a as the basis parameter of $B_n(\theta)$ at the n-th decomposition level. To consider the common oscillations shared by a set of signals, our previous work [29], [30] proposes the MAFD. The MAFD considers a set of signals together and generates common adaptive basis components for all signals. At each decomposition level, the MAFD finds the basis parameter by maximizing the total extracted energy, which can be expressed as

$$a_{n,\text{MAFD}} = \arg\max_{a \in \mathbb{D}} \left\{ \sum_{c=1}^{C} E_{\{c,B_n\}}(a) \right\}, \quad (5)$$

where $E_{\{c,B_n\}}(a)$ denotes the extracted energy of the c-th channel, and C is the total number of channels.

III. Proposed Method

A. Time-frequency-joint representation based SSVEP signal model

The fundamental difficulty in defining the commonality of the SSVEP signals elicited by multiple stimuli stems from the fact that the SSVEP signals elicited by different stimuli are distinctive in the time domain. Fig. 1(a) uses the 2D plots to illustrate the dynamic information and the inter-stimulus differences of the SSVEP signals in the time-domain represenation where the x-axis presents the time, the y-axis presents the stimulus frequencies, and different colors present signal amplitudes. The first 1s SSVEP signals elicited by each stimulus are averaged across six trials (each block contains one trial for every stimulus). It can be seen that the colors in different rows have different distributions. The high amplitudes (red pixels) and the low amplitudes (blue pixels) of different rows are distributed in different columns, indicating that the signals from different stimuli have different patterns. Signals of the first subject are selected as examples depicted in the conventional view. It can be seen that these signals are mingled together. The averaged signals across all stimuli, as shown by the thick red curve, are around 0. Obviously, it is difficult to distinguish the commonality of SSVEP signals in the time-domain representation.

To tackle this problem, we propose to present the SSVEP signals of different stimuli using both time and the stimulus frequencies. This time-frequency-joint representation is inspired by the sine-cosine-based and superposition-theory-based SSVEP signal models. Assuming that the signal lengths of the impulse responses are equal to the periods of corresponding stimuli, the superposition-theory-based signal model can be simplified as $x_{\text{superpos}}(t) = R(t)$ where $R(t)$ is a periodic signal, and $R(t) = R(t + kT_{\text{stim}})$, $k \in \mathbb{Z}$. Considering both the sine-cosine-based model and the superposition-theory-based model after regulating the signal length of the impulse response, we can assume that

1) The main components in SSVEP signals are periodic;
2) The periods of the main components in SSVEP signals are the same as the periods of corresponding stimuli.

Let us define the single-period signals (SPSs) as signals that occur during the stimulus presentation. According to the two assumptions mentioned above, comparing SPSs from different stimuli is critical for creating the SSVEP signal model of the stimulus-stimulus transfer and extracting the common SSVEP components from different stimuli. By converting time values to phase values based on the stimulus frequencies and phases, SSVEP signals can be synchronized to make the SPSs of different stimuli comparable. Such transformations

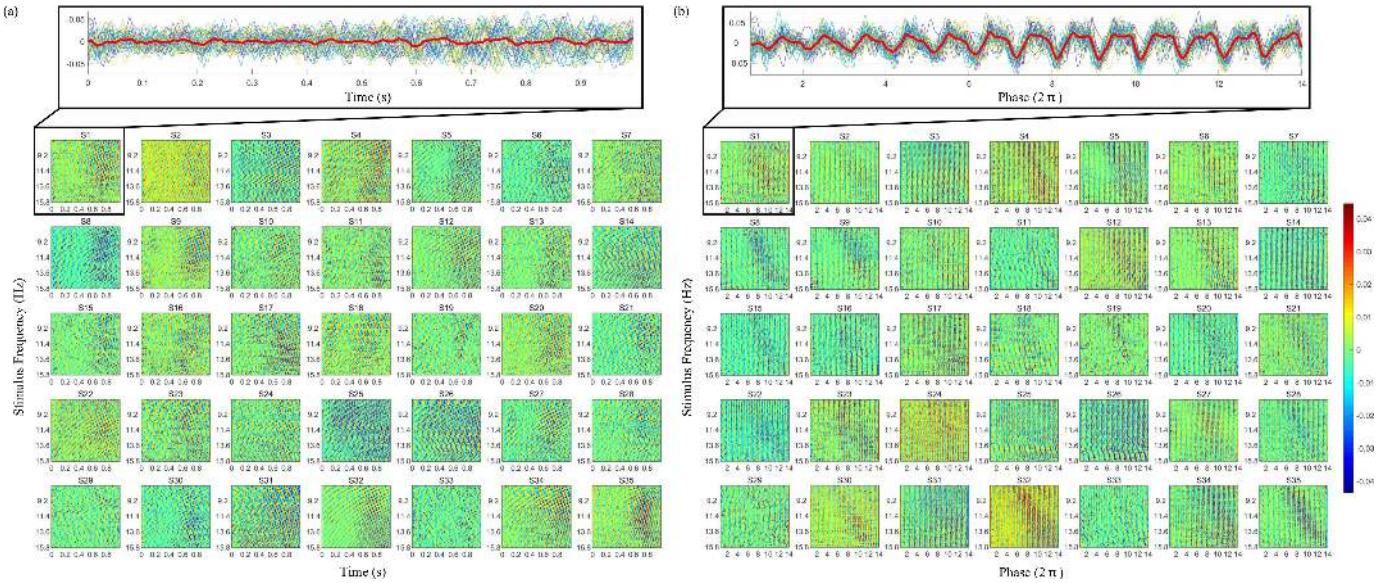

Fig. 1. Dynamic information and inter-stimulus differences of the SSVEP signals in the time-domain representation and the time-frequency-joint representation (averaged SSVEP signals of Oz channel as examples). (a) shows the SSVEP signals in the time-domain representation. (b) shows the SSVEP signals in the time-frequency-joint representation. Different small subfigures show signals from different subjects. Each row in the subfigures depicts the SSVEP signal of a specific stimulus. The y-axis presents the stimulus frequencies and denotes the corresponding stimuli of rows. The x-axis denotes the time in (a) and the phase in (b), respectively. The z-axis presented by colors denotes the signal amplitudes. Signals of the first subject are selected as an example and displayed in the conventional view. The averaged signals across all stimuli are shown by thick red lines.

between signals in the time-domain and time-frequency-joint representations are defined as follows:

$$s(\theta) = x(\theta\frac{T_{\text{stim}}}{2\pi} - \theta_{\text{stim}}\frac{T_{\text{stim}}}{2\pi}) \text{ and } x(t) = s(t\frac{2\pi}{T_{\text{stim}}} + \theta_{\text{stim}}). \tag{6}$$

In the time-frequency-joint representation, the sine-cosine-based model and the superposition-theory-based model are

$$\begin{aligned} &s_{\text{sine-cosine}}(\theta) = \sum_{k=1}^{K} p_k \sin(\theta) + q_k \cos(\theta), \text{ and} \\ &s_{\text{superpos}}(\theta) = s_R(\theta), \end{aligned} \tag{7}$$

where $s_R(\theta)$ is the signal $R(t)$ in the time-frequency-joint representation and $s_R(\theta) = s_R(\theta + 2k\pi)$, $k \in \mathbb{Z}$. According to (7), the sine-cosine-based and superposition-theory-based SSVEP models both indicate that the SPSs in the time-frequency-joint representation are similar across stimuli. Fig. 1(b) illustrates the dynamic information and inter-stimulus differences of SSVEP signals in the time-frequency-joint representation where the x-axis presents the phase, the y-axis presents the stimulus frequencies, and different colors present signal amplitudes. SSVEP signals from the 2nd to the 14th periods of each stimulus are averaged across six trials. Because all the SSVEP signals have the same sampling rate, the SSVEP signals in response to stimuli whose stimulus frequencies are high are up-sampled to ensure that the sampling numbers of SPSs from different stimuli are the same in one period. It can be seen that the colors in different rows show the similar distribution. The high amplitude values (red pixels) and the low amplitude values (blue pixels) of different rows are distributed around the same columns. Signals from the first subject are selected as examples displayed in the conventional view. It can be seen that the signals are all distributed around the averaged signal depicted by the thick red curve. In contrast, the commonality of SSVEP signals of various stimuli are evident in the time-frequency-joint representation. In summary, another major assumption can be made, namely that

3) The SPSs of different stimuli share commonality in the time-frequency-joint representation.

With these three assumptions listed above, the SSVEP signals can be modeled as

$$x_{\text{syn}}(t) = s_{\text{stim}}(\frac{2\pi}{T_{\text{stim}}}t + \theta_{\text{stim}}) + w_{\text{stim}}(\frac{2\pi}{T_{\text{stim}}}t + \theta_{\text{stim}}) + n_{\text{stim}}(t), \tag{8}$$

where $s_{\text{stim}}(\theta)$ denotes the common periodic component in the time-frequency-joint representation, $w_{\text{stim}}(\theta)$ denotes the uncommon periodic component in the time-frequency-joint representation, and $n_{\text{stim}}(t)$ denotes the noisy component in the time-domain representation. In this signal model, the components $s_{\text{stim}}(\theta)$ are periodic with the period of 2π and are the same for different stimuli, which can be regarded as the SSVEP-task signal. The components $w_{\text{stim}}(\theta)$ are also periodic with the same period as $s_{\text{stim}}(\theta)$, but they are not shared between different stimuli. The components $n_{\text{stim}}(t)$ are non-periodic or do not have the same period as the stimulus. The components $w_{\text{stim}}(\theta)$ and $n_{\text{stim}}(t)$ can be regarded as the signals unrelated to the SSVEP task.

It should be noted that this study focuses on using the proposed signal model to conduct the stimulus-stimulus transfer. Therefore, the key issues of the following sections are decomposing the major periodic component in the SSVEP signals and extracting the common SSVEP component $s_{\text{stim}}(\theta)$ from the decomposition components. Although analyzing components $w_{\text{stim}}(\theta)$ and $n_{\text{stim}}(t)$ in detail is also very helpful for

understanding the SSVEP signals, it is not directly related to the stimulus-stimulus transfer and thus are not covered in this study.

B. DP-MAFD of SSVEP signals

To estimate the components of (8), the SPSs must be decomposed and analyzed. The MAFD is used to adaptively decompose the SPSs of all source stimuli together. As the MAFD uses same adaptive basis components to decompose a set of signals, it can be applied to analyze common components of these signals. The MAFD is designed as a general multi-channel signal decomposition method, which assumes that the processed signals are in a single period and defines $B_n(\theta)$ from 0 to 2π. This study modifies the MAFD and proposes the DP-MAFD to allow the MAFD focus on analyzing the SPSs of different stimuli.

Assume the SSVEP signals are from total J source stimuli, which are denoted as $\{\mathbf{X}_j\}_{j=1}^{J}$ where $\mathbf{X}_j$ presents the SSVEP signals of j-th stimulus with the stimulus frequency f_j and the stimulus phase θ_j and contains total C channels, then we can select all SSVEP signals of c-th channel from $\{\mathbf{X}_j\}_{j=1}^{J}$ to construct a new set of signals $\mathbf{S}_c$ ($\mathbf{S}_c \in \mathbb{R}^{J\times L}$ where L is the total sample number). According to (8), $\mathbf{S}_c$ can be written as

$$\mathbf{S}_c = \begin{bmatrix} x_{c,1}(t) \\ x_{c,2}(t) \\ \vdots \\ x_{c,J}(t) \end{bmatrix} + \mathbf{N}_c, \tag{9}$$

where $x_{c,j}(t)$ is

$$x_{c,j}(t) = s_c(2\pi f_j t + \theta_j) + w_{c,j}(2\pi f_j t + \theta_j), \tag{10}$$

and $\mathbf{N}_c$ represents the set of $n_{c,j}(t)$ for the c-th channel. The MAFD can be modified to decompose $\mathbf{S}_c$ to extract the common periodic component $s_c(\theta)$ across stimuli.

To apply the AFD related methods, these signals must be transferred to their analytic signals [33]. The analytic signal $g_{c,j}(t)$ of $x_{c,j}(t)$ can be denoted as

$$g_{c,j}(t) = x_{c,j}(t) + j\mathcal{H}\{x_{c,j}(t)\}, \tag{11}$$

where $\mathcal{H}\{\cdot\}$ represents the Hilbert transform. Assuming that the SSVEP-task signal $s_c(\theta)$ and the non-SSVEP-task signals $\{w_{c,j}(\theta)\}_{j=1}^{J}$ are independent, $g_{c,j}(t)$ can be represented by the basis of the AFD. Moreover, to focus on analyzing SPSs, the basis components $B_n(\theta)$ are limited in one period from θ_j to $\theta_j + 2\pi$ for the j-th stimulus and repeatedly applied to all periods. Then, $g_{c,j}(t)$ can be written as

$$g_{c,j}(t) = \sum_{n\in\mathbb{S}_c} A_{c,j,n}B_{c,n}(\widetilde{\theta}) + \sum_{n\notin\mathbb{S}_c} A_{c,j,n}B_{c,n}(\widetilde{\theta}) + n_{c,j}(t), \tag{12}$$

where $\mathbb{S}_c$ denotes the set of decomposition levels related to the common periodic component $s_c(\theta)$, $\widetilde{\theta} = 2\pi f_j t - 2\pi d + \theta_j$, $d = \lfloor t/T_j \rfloor$, and $\lfloor\cdot\rfloor$ is the floor operation that returns the largest integer less than or equal to the input. The decomposition coefficient $A_{c,j,n}$ can be computed as

$$A_{c,j,n} = \left\langle G_{c,j,n}(t), \mathrm{e}_{\{a_{c,n}\}}(\widetilde{\theta}) \right\rangle. \tag{13}$$

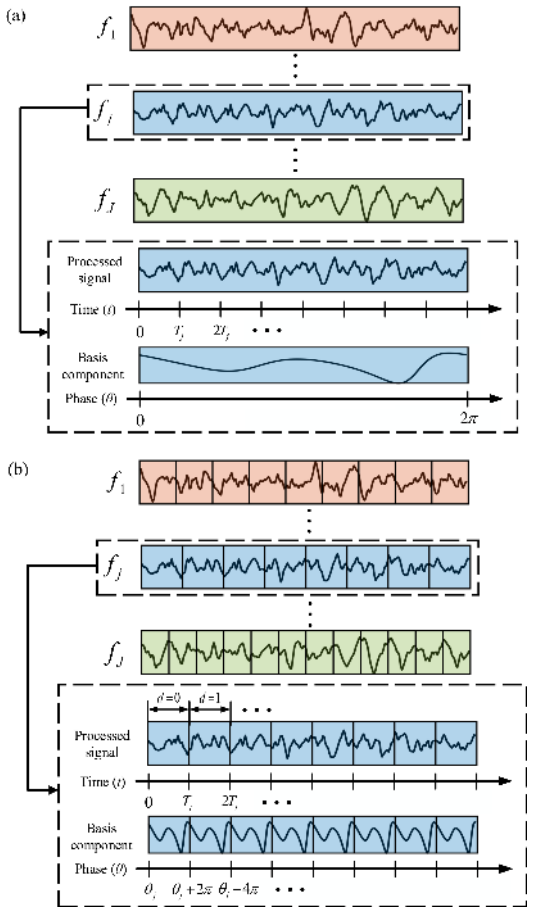

Fig. 2. Differences of (a) decomposing general signals in the conventional MAFD and (b) decomposing SSVEP signals in the DP-MAFD. The conventional MAFD assumes that processed signals are all in one period from 0 to 2π. The DP-MAFD assumes that processed signals contain several periods and divides them according to the stimulus frequencies and the stimulus phases.

$G_{c,j,n}(t)$ is called the reduced remainder and can be obtained through the following recursive procedure:

$$\begin{aligned} G_{c,j,n+1}(t) &= R_{c,j,n}(t)\prod_{k=1}^{n}\frac{1-\overline{a}_{c,k}e^{j\widetilde{\theta}}}{e^{j\widetilde{\theta}}-a_{c,k}} \\ &= \left(G_{c,j,n}(t) - A_{c,j,n}\cdot \mathrm{e}_{\{a_{c,n}\}}(\widetilde{\theta})\right)\frac{1-\overline{a}_{c,n}e^{j\widetilde{\theta}}}{e^{j\widetilde{\theta}}-a_{c,n}}, \end{aligned} \tag{14}$$

where $\mathrm{e}_{\{a\}}(\theta)$ is the evaluator of searching a_n and is defined as

$$\mathrm{e}_{\{a\}}(\theta) = \frac{\sqrt{1-|a|^2}}{1-\overline{a}e^{j\theta}}. \tag{15}$$

The basis parameter a_n should make the decomposition component $A_{c,j,n}B_{c,n}(\widetilde{\theta})$ match all SPSs, which can be achieved by solving

$$a_{c,n,\text{DP-MAFD}} = \arg\max_{a\in\mathbb{D}}\left\{\sum_{j=1}^{J}\sum_{d=0}^{D_j}\left|\left\langle G_{c,j,n}(t), \mathrm{e}_{\{a\}}(\widetilde{\theta})\right\rangle\right|^2\right\}, \tag{16}$$

where t is defined in $[dT_j, (d+1)T_j)$ for each value of d, $D_j = \lfloor T_w/T_j \rfloor$, and T_w is total signal length of the

processed signal. Unlike the AFD and the MAFD, the DP-MAFD searches for the basis parameters by splitting the SSVEP signals into SPSs according to the stimulus frequencies and then maximizing the total extracted energy of all SPSs. Fig. 2 illustrates the differences of the conventional MAFD and the proposed DP-MAFD.

C. Extraction of common components across source stimuli

For each stimulus, the recorded SSVEP signals are averaged across trials to enhance the SSVEP-task signals, which is denoted as $\overline{\mathbf{X}}_j$ of j-th source stimulus ($\overline{\mathbf{X}}_j \in \mathbb{R}^{C\times L}$ where C is the total number of channels and L is the total sample number). Then, the DP-MAFD is used to decompose the averaged SSVEP signals of different stimuli, as shown in Section III-B. The DP-MAFD analyzes the averaged SSVEP signals channel by channel. After the decomposition, the averaged SSVEP signals in c-th channel denoted as $\mathbf{S}_c$ in (9) can be represented as

$$\mathbf{S}_c = \text{Real}\left\{\mathbf{A}_c \times \mathbf{B}_c + \mathbf{R}_c\right\}, \tag{17}$$

where $\text{Real}\{\cdot\}$ denotes the real part of the analytic signal, $\mathbf{A}_c$ denotes the matrix of decomposition coefficients ($\mathbf{A}_c \in \mathbb{C}^{J\times N}$), $\mathbf{B}_c$ denotes the matrix of basis components ($\mathbf{B}_c \in \mathbb{C}^{N\times L}$), $\mathbf{R}_c$ denotes the remainder, J is the total number of source stimuli, and N is the maximum decomposition level. Then, because the basis components are same for all stimuli, the decomposition components related to the common periodic component $s_c(\theta)$ can be selected by analyzing the decomposition coefficients. According to the prior knowledge of the properties and the relationships of the SSVEP-task and non-SSVEP-task signals, the following selection rules are recommended.

1) According to the experimental settings in Section II-B, the processed EEG signals were all recorded while subjects were performing the SSVEP tasks. Furthermore, the processed signals were already prepossessed and averaged across trials to weaken non-SSVEP-task signals. As a result, the SSVEP-task signals should have more energy than non-SSVEP-task signals. Because the DP-MAFD extracts the components that can best match SPSs first, the remainder $\mathbf{R}_c$ will contain components of $\{n_j(t)\}_{j=1}^{J}$ and parts of components of $\{w_j(\theta)\}_{j=1}^{J}$ with tiny energy when N is large enough.
2) Following the energy relationship between the SSVEP-task and the non-SSVEP-task signals mentioned in 1), the energy of $\sum_{n\in\mathbb{S}_c} A_{c,j,n}B_{c,n}(\widetilde{\theta})$ should be greater than that of $\sum_{n\notin\mathbb{S}_c} A_{c,j,n}B_{c,n}(\widetilde{\theta})$. As a result, the decomposition levels in $\mathbb{S}_c$ should be the first several N_{sel} decomposition levels where N_{sel} satisfies
$$\sum_{n=1}^{N_{\text{sel}}} |A_{c,j,n}|^2 - \alpha \sum_{n=1}^{N} |A_{c,j,n}|^2 < \epsilon \;\forall\; j = 1, \cdots, J, \tag{18}$$
where ϵ denotes a small positive number, and the selection parameter α can be set based on the noise level of the processed signals.
3) For each decomposition level n in $\mathbb{S}_c$, the elements in $\{A_{c,j,n}\}_{j=1}^{J}$ are coefficients of the common SSVEP components and thus should be distributed closely. The phases and the amplitudes of the decomposition coefficients determine the time shifting and shapes of decomposition components, respectively. Common sense dictates that the shapes of the common components should be similar for different stimuli, even if the time locations of these common components are allowed to vary. Therefore, this study selects the decomposition levels based on the distributions of the decomposition coefficient amplitudes. The interquatile range (IQR) is used to measure the statistical dispersion of the amplitudes of the decomposition coefficients [34], [35]. If the decomposition level n is in $\mathbb{S}_c$, then n should satisfy
$$\Omega\left\{|A_{c,j,n}| \notin [l_{\text{lower}}, l_{\text{higher}}] \;\forall\; j = 1, \cdots, J\right\} < \beta \cdot J \tag{19}$$
where $l_{\text{lower}} = Q_1 - \kappa \cdot \text{IQR}$, $l_{\text{higher}} = Q_3 + \kappa \cdot \text{IQR}$, $\text{IQR} = Q_3 - Q_1$, Q_3 and Q_1 denote the upper and lower quartiles of $\{|A_{c,j,n}|\}_{j=1}^{J}$, $\Omega\{\cdot\}$ denotes the operation of counting the element number, and the selection parameter β also can be adjusted by the noise levels of processed signals.

According to these three rules, the decomposition coefficients of $n \in \mathbb{S}_c$, denoted as $\mathbf{A}_c^{\text{sel}}$, can be determined and transferred across stimuli to construct SSVEP templates of the target stimuli.

D. Construction of SSVEP templates for target stimuli

After the selection process mentioned in Section III-C, the decomposition coefficients that are not common for different stimuli in certain decomposition levels were rejected. Although the remaining decomposition coefficients have similar distributions for different stimuli, the decomposition coefficients of the stimuli with closer stimulus frequencies share more commonalities. Therefore, for one target stimulus with the stimulus frequency f_q, this study constructs SSVEP templates using the decomposition components of the source stimulus whose stimulus frequency $\hat{f}_q$ is closest to f_q, implying that the chosen source stimulus for the q-th target stimulus can be searched by

$$\hat{q} = \arg \min_{j\in\{1,\cdots,J\}} \{|f_j - f_q|\}. \tag{20}$$

Following (12), the SSVEP template of the c-th channel for the q-th target stimulus can be constructed by

$$\hat{X}_{c,q}(t) = \text{Real}\left\{\sum_{n\in\mathbb{S}_c} A_{c,\hat{q},n}B_{c,n}(\widetilde{\theta})\right\}, \tag{21}$$

where $\widetilde{\theta} = 2\pi f_q t - 2\pi d + \theta_q$, and $d = \lfloor t/T_q \rfloor$. Fig. S1 in the supplementary material shows the entire processes from the DP-MAFD decomposition of the source SSVEP templates to the construction of the target SSVEP templates.

The SSVEP templates of the target stimuli built from the selected decomposition coefficients of source stimuli can be used to find the spatial filters of target stimuli and then

TABLE II
PEARSON CORRELATION COEFFICIENTS BETWEEN THE SSVEP SIGNALS OF DIFFERENT STIMULI IN THE TIME-DOMAIN AND TIME-FREQUENCY-JOINT REPRESENTATIONS, ORIGINAL AND RECONSTRUCTED SSVEP SIGNALS OF THE DP-MAFD, AS WELL AS THE RECONSTRUCTED SPSS OF ALL PAIRS OF STIMULI⋆.

SSVEP signals of different stimuli in time-domain representation	SSVEP signals of different stimuli in time-frequency-joint representation
-0.003 ± 0.168	0.522 ± 0.315

Original and reconstructed SSVEP signals of the DP-MAFD	Reconstructed SPSs of all pairs of stimuli
0.588 ± 0.154	0.775 ± 0.112

⋆: More detailed individual results are in the supplementary material.

perform the SSVEP recognition. This study considers two types of spatial filters, i.e., the class-nonspecific and class-specific spatial filters following [4] and [18]. The details are shown in Section S.I of the supplementary material.

IV. RESULTS

The impact of each procedure in the proposed method is illustrated through the simulations in this section. Firstly, Section IV-A verifies the proposed time-frequency-joint representation's ability to synchronize SSVEP signals across stimuli, laying the groundwork for the decomposition of the DP-MAFD. Then, Section IV-B demonstrates that the DP-MAFD has the ability to extract major periodic components. In addition, Section IV-C verifies that the selected decomposition components can be transferred across stimuli, which is the groundwork for constructing proper SSVEP templates. Finally, the recognition performance of constructed SSVEP templates for target stimuli is further demonstrated in Section IV-D. The related simulation codes are available at `https://stimulus-stimulus-transfer-ssvep-bcis.readthedocs.io`.

A. Stressing common components in time-frequency-joint representation

Whether the proposed time-frequency-joint representation can synchronize the SSVEP signals of different stimuli is verified by the similarities between the SSVEP signals of different stimuli before and after presenting the SSVEP signals in the time-frequency-joint representation. The time-frequency-joint representation is based on (6). The similarity between each pair of 40 stimuli is evaluated by the Pearson correlation coefficient. Before presenting the SSVEP signals in the time-frequency-joint representation, signals are in the time domain. The first 1s signals are averaged across trials for each stimulus. After presenting SSVEP signals in the time-frequency-joint representation, the first 14-period signals are averaged across trials for each visual stimulus. As listed in Table II, the averaged correlation coefficient of 35 subjects in the time-domain is around -0.003, indicating that the common knowledge shared across stimuli is difficult to be identify in the time domain. In the time-frequency-joint representation, the averaged correlation coefficient of 35 subjects is improved to around 0.522. This considerable improvement indicates that the proposed time-frequency-joint representation can effectively synchronize SSVEP signals of different stimuli. The individual results of these similarities are illustrated in Section S.II of the supplementary material.

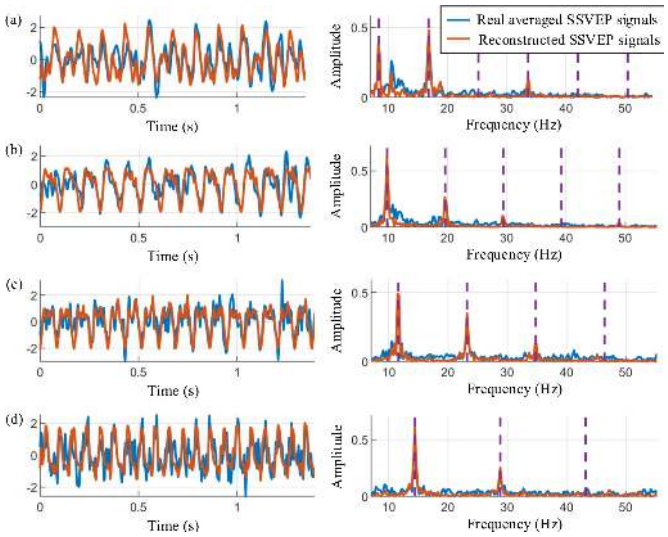

Fig. 3. The averaged original SSVEP signals (blue lines) and the SSVEP signals reconstructed by the decomposition components in the first 50 decomposition levels (red lines) at Oz. The time-domain and frequency-domain signals are shown in the left and right figures, respectively. (a) Subject 1's SSVEP signals induced by the 8.4 Hz stimulus; (b) Subject 1's SSVEP signals induced by 9.8 Hz stimulus; (c) Subject 22's SSVEP signals induced by 11.6 Hz stimulus; (d) Subject 22's SSVEP signals induced by 14.4 Hz stimulus. The purple dotted lines denote the harmonic stimulus frequencies.

B. Decomposition of major periodic components by DP-MAFD

Whether the DP-MAFD can extract major periodic components in the SPSs in response to different stimuli simultaneously is validated by comparing the original and reconstructed SSVEP signals. The first 2.5s SSVEP signals are averaged across trials, and are regarded as the original SSVEP signals. Then, for each channel, the averaged SSVEP signals of 40 stimuli are decomposed by the DP-MAFD together. Finally, the decomposed components of the first 50 decomposition levels are used to reconstruct the SSVEP signals.

Fig. 3 compares original signals shown in blue and the reconstructed signals shown in red, with signals from the Oz channel are selected as examples. The signals in (a) and (b) are responses to the stimuli flickering at 8.4 Hz and 9.8 Hz from Subject 1. The signals in (c) and (d) are responses to the stimuli flickering at 11.6 Hz and 14.4 Hz from Subject 22. Comparisons in the time and frequency domains both show that, while some abrupt or non-periodic changes as well as periodic components whose major frequencies are different from the corresponding stimulus frequencies cannot be recovered in the reconstructed signals, periodic components whose major frequencies are same as the corresponding stimulus frequencies can be extracted by the DP-MAFD.

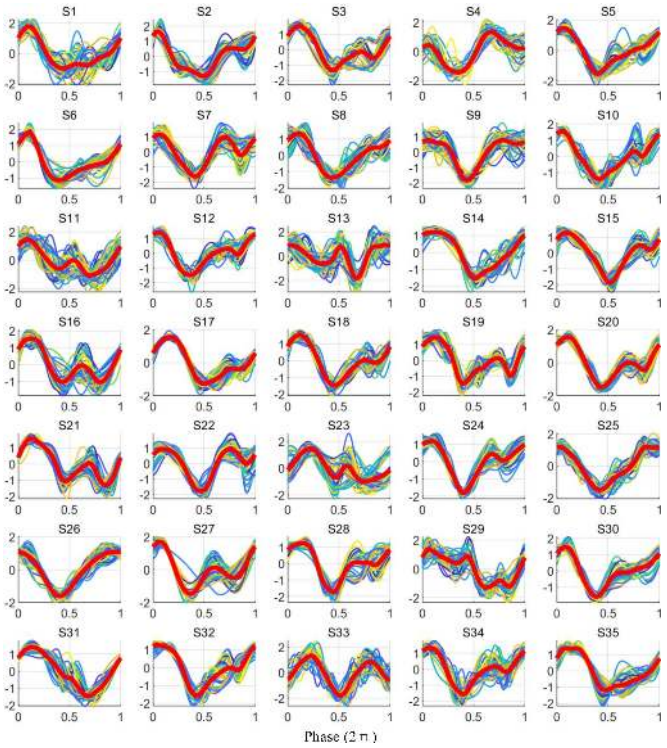

Fig. 4. Reconstructed SPSs of Oz channel. The thin lines with different colors show reconstructed SPSs of different stimuli. The thick red line shows the signal of reconstructed SPSs averaged over 40 stimuli.

Then, the similarities between the original and reconstructed signals are evaluated by the Pearson correlation coefficients. The averaged Pearson correlation coefficient is around 0.588 as shown in Table II. Such high correlation indicates that, after modifying the MAFD, the DP-MAFD can extract the major periodic components in SSVEP signals. Individual Pearson correlation coefficients and mean square errors (MSEs) are shown in Section S.III of the supplementary material.

C. Transferability of selected decomposition components

The transferability of the common components selected by three rules recommended in Section III-C is verified by comparing SPSs constructed by selected the decomposition coefficients of different stimuli. Firstly, the DP-MAFD decomposes averaged SSVEP signals channel by channel. The SPSs are then reconstructed after the decomposition coefficients are selected based on the three recommended rules. Fig. 4 compares reconstructed SPSs of different stimuli, with examples drawn from the Oz channel. The reconstructed SPSs of different stimuli, represented by thin lines of different colors, are all distributed around the averaged SPS, represented by the thick red curve, and are all very similar.

The similarities of reconstructed SPSs are then quantitatively validated by the Pearson correlation coefficients and shown in Table II. The averaged Pearson correlation coefficients between reconstructed SPSs of all pairs of stimuli are around 0.775, which is much higher than the correlations between the SSVEP signals of all pairs of stimuli before the decomposition. This result indicates that the transferability is improved from the SSVEP signals before the decomposition to the decomposition components selected using recommended rules. More individual results can be found in Section S.IV of the supplementary material.

D. Classification performance of constructed SSVEP templates

The leave-one-out cross validation is used to test the classification performance of the SSVEP templates created using the proposed stimulus-stimulus transfer method. In six blocks, the SSVEP signals of five blocks are selected for constructing the templates and training the spatial filters. The remaining block is used for testing. The entire evaluation process is repeated to test all six blocks. From forty stimuli, source stimuli are randomly chosen. To avoid the stochastic effects, the source stimuli are randomly selected five times. In the proposed method, the parameters in the selection rules stated in Section III-C are set to $\alpha = 0.9$, $\kappa = 1.5$, and $\beta = 0.2$. The classification performance is evaluated by the classification accuracy and the ITR. The ITR is computed as

$$\mathrm{ITR} = \frac{60}{T_{\mathrm{det}}} \left[\log_2 M + P \log_2 P + (1-P) \log_2 \frac{1-P}{M-1} \right], \tag{22}$$

where P is the classification accuracy, M is the number of stimuli, i.e., 40 ($M = J + Q$) in this study, and T_{det} is the total time for each detection [36]. T_{det} is equal to the sum of the time required for shifting visual attention and the time required for recognizing the target. Because this study focuses on verifying the performance of the stimulus-stimulus transfer, the following analysis of the classification performance only considers the performance of the target stimuli even though all stimuli are considered during the classification. More results from the BETA Dataset [37] are included in Section S.V of the supplementary material to further verify classification performance of SSVEP templates constructed using the proposed stimulus-stimulus transfer method.

Firstly, the proposed method is compared to the calibration-free method, namely the conventional CCA based on the sine-cosine reference signals as well as another stimulus-stimulus transfer based method, namely the tlCCA proposed in [12]. These three methods do not use the calibration data of the target stimuli to train the spatial filters and the SSVEP templates. Fig. 5(a) and (b) show the classification performance of these three methods. The proposed method clearly outperforms the other two methods, especially for the short signal lengths. In addition, the proposed method requires a shorter signal length to achieve its highest ITR. The proposed method achieves its highest ITR when the signal length is around 1 s. The conventional CCA and the tlCCA achieve their highest ITRs when the signal lengths are around 1.75 s and 1.25 s, respectively.

Secondly, the proposed method is compared to the eCCA proposed in [4]. The eCCA, like the proposed method, uses the averaged SSVEP signals to generate the SSVEP templates and uses the sine-cosine signals as the reference signals to train the class-specific spatial filters and perform the stimulus frequency recognition. However, the eCCA requires the calibration data

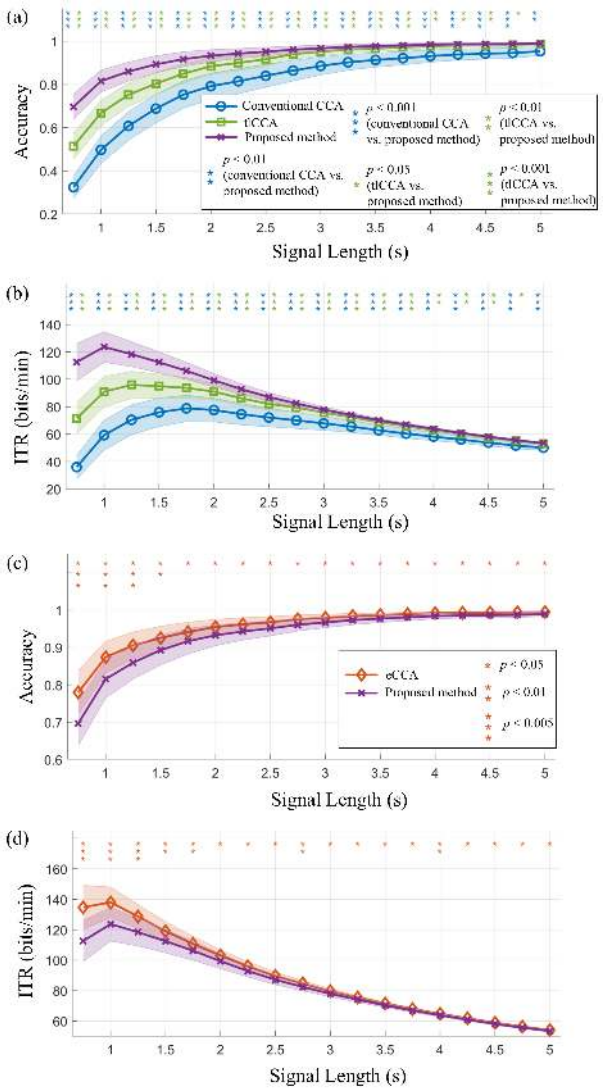

Fig. 5. Comparisons of classification accuracy and ITR for the benchmark dataset. (a) and (b) compare the classification accuracy and ITR of the methods that do not use the calibration data of target stimuli, such as the conventional CCA, the tlCCA, and the proposed method. (c) and (d) compare the classification accuracy and ITR of the proposed method and the method with calibration data of target stimuli, i.e., the eCCA.

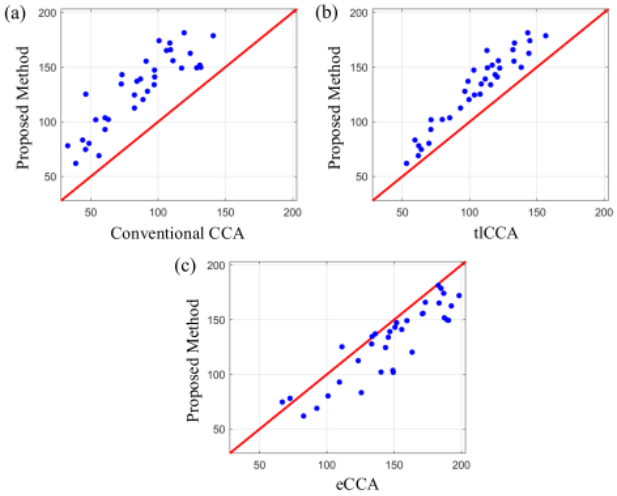

Fig. 6. Comparisons of individual maximum ITRs (bits/min) for the benchmark dataset. (a) ITRs of the proposed method and the conventional CCA. (b) ITRs of the proposed method and the tlCCA. (c) ITRs of the proposed method and the eCCA. Each point represents the maximum ITR of a single subject.

TABLE III
COMPARISONS OF THE MAXIMUM AVERAGED ITR AND REQUIRED CALIBRATION TIMES FOR ONE SUBJECT.

Method	Number of stimuli for training	Maximum averaged ITR (bits/min)	Required calibration time (s)
Conventional CCA	0	78.805 (1.75 s)	0
tlCCA	8	95.966 (1.25 s)	90
Proposed method	8	123.684 (1.00 s)	80
eCCA	32	138.021 (1.00 s)	320

of the target stimuli to build the SSVEP templates. The classification performance of the proposed method and the eCCA is illustrated in Fig. 5(c) and (d). The proposed method cannot perform better than the eCCA because the proposed method does not use the calibration data of the target stimuli. However, the differences in performance between the proposed method and the eCCA are small.

Individual differences cause differences in classification performance for different subjects. Fig. 6 shows the maximum ITRs of 35 subjects. Each blue dot represents one subject's maximum ITR. In Fig. 6(a) and (b), all points are in the upper left region and far away from the diagonal red line, indicating that the proposed method outperforms the conventional CCA and the tlCCA for all subjects. Fig. 6(c) shows that most points are in the lower right region and very close to the diagonal red curve, indicating that, for most subjects, while the proposed method cannot perform better than the eCCA, the classification performance of the proposed method and the eCCA is very close.

Although the performance of the proposed method is slightly inferior to that of eCCA, we have to point out that the proposed method does not require the calibration data of target stimuli and thus can significantly shorten the calibration time. Table III summarizes the maximum ITRs and required calibration times of the conventional CCA, tlCCA, eCCA, and the proposed method for a single subject. The required calibration time is calculated by

$$t_{\text{cal}} = N_{\text{trial}} \times N_{\text{stim}} \times T_{\text{cal}}, \tag{23}$$

where N_{trial} denotes the number of training trials for each stimulus, which is 5 ($N_{\text{trial}} = J$) in this study, N_{stim} denotes the number of stimuli for training, which is 0 for the conventional CCA, 8 for the tlCCA and the proposed method, and 32 for the eCCA, and T_{cal} denotes the required calibration time of each trial and is computed as the sum of the time for subject to shift visual attention (0.5 s in this study), the time for displaying the target cue (0.5 s in this study), and the signal length required to achieve the highest ITR. Although the eCCA has the highest ITR, it requires considerably longer calibration time than the

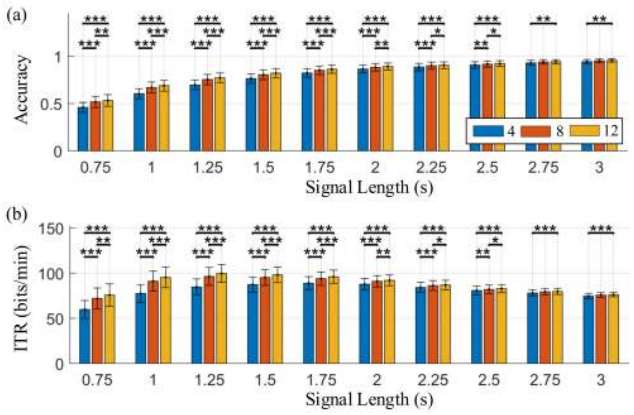

Fig. 7. Comparisons of classification accuracy and ITR of the proposed method with three numbers of source stimuli for the benchmark dataset (blue: 4 source stimuli, red: 8 source stimuli, and yellow: 12 source stimuli). (a) the classification accuracy. (b) the ITR. The statistical results are indicated as $\star$ ($p < 0.05$), $\star\star$ ($p < 0.01$), and $\star\star\star$ ($p < 0.001$).

proposed method.

E. Effects of source-stimulus number on classification performance

According to Section III-D, the proposed stimulus-stimulus transfer method requires calibration data from a little number of source stimuli (at least one) to construct different SSVEP templates to a greater number of target stimuli. Fig. 7 shows the effects of source-stimulus number on classification. The source stimuli are chosen at random for each source-stimulus number and repeated five times. Because the number of the SPSs that can be analyzed increases with the number of source stimuli, the estimations of the common periodic components become more accurate, and thus the classification performance enhances, especially for the short signal lengths.

V. Discussions

The proposed method extracts the common components of different stimuli to build the SSVEP templates without calibration data. From the signal modeling point of view, because the different stimuli are same in one period, we assume that the responses of our brains are the same as well. Fig. 8 uses examples to demonstrate this key assumption of the proposed method. The blue curve in the left polar plot represents the SSVEP-related component, which is assumed to be stable for all stimuli. The black straight arrow denotes the current signal state that rotates counterclockwise and follows the blue curve. The differences between the EEG signals elicited by different stimuli are caused by the rotation speeds of the signal states and the disturbance of the non-SSVEP-related components as shown in the right part. From the perspective of the transfer learning, the common components extracted by the proposed method can be regarded as being in a high-level common domain that links sub-domains of different stimuli. The proposed method can be considered as an asymmetric transfer [38]. As illustrated in Fig. 9(a), if the calibration data in the target domain is sufficient, the classification model of the target domain can be directly obtained by the calibration-based

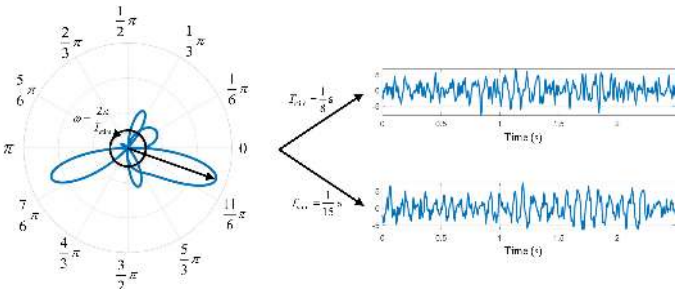

Fig. 8. Examples to show the fundamental assumption of the proposed stimulus-stimulus transfer based on the time-frequency-joint representation. The left side shows the signal state that follows the SSVEP-related component for all stimuli. The right side shows the real EEG signals with signal states rotating at different speeds for different stimuli and being disturbed by non-SSVEP-related components.

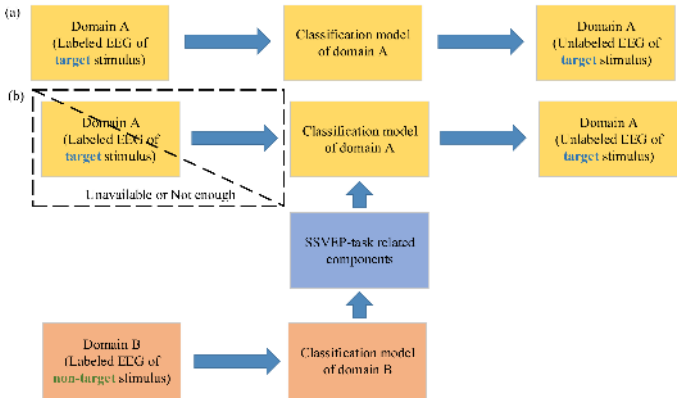

Fig. 9. Diagrams of (a) the classification with calibration data of target stimuli and (b) the classification based on the proposed stimulus-stimulus transfer without calibration data of target stimuli.

methods, such as the eCCA. However, when the calibration data in the target domain is unavailable or insufficient, the proposed method can be used to generate the classification model of the target domain by transferring the classification model of the non-target domain through the time-frequency-joint representation as shown in Fig. 9(b). The differences between features before and after the transformation are discussed in Section S.VI of the supplementary material. This is the first study to extract the general non-stimulus-specific components in SSVEP signals. The comparisons between the previous and proposed transfer schemes are discussed in Section S.VII of the supplementary material.

Although the proposed method can produce good classification performance, it still has some practical limitations. Firstly, because the proposed method is based on the assumption that the stimulus-specific and -nonspecific components are linearly combined, the non-linear components in SSVEP signals mentioned in [39] are not considered in this study. The performance may be further improved by taking into account the non-linearity of SPSs. Secondly, the selection method of decomposition coefficients shown in Section III-C is empirical and preliminary, particularly for parameter settings in the selection rules. The qualities of generated classification models may be improved further in the future by delving deeper into these decomposition coefficients and considering novel

transfer learning methods. Third, this study mainly focuses on how to extract and transfer commonality across stimuli, so some advanced recognition techniques that are not directly related to the stimulus-stimulus transfer are not included in this study, such as the filter-bank technique and the classification methods based on the artificial neural network [6], [40]. The performance of the proposed method may also benefit from these techniques.

VI. Conclusions

This study proposes a cross-stimulus transfer learning framework to investigate the commonality of SSVEP signals. Based on the time-frequency-joint representation, a new SSVEP signal model is proposed, which directly uses the common periodic components to describe SSVEP signals of different stimuli. Moreover, the DP-MAFD incorporates the characteristics of SSVEP signals to adaptively decompose periodic components of source stimuli simultaneously. The major common periodic components can be selected from these decomposed components. The simulation results show that these common components extracted by the proposed method can be transferred across stimuli. The proposed method outperforms other stimulus-stimulus transfer methods in term of practical utility. This is the first study to analyze SSVEP signals from a synchronization standpoint and model SSVEP signals using SPSs, which will inspire further research on the transfer learning of the SSVEP signals. In addition, the proposed stimulus-stimulus transfer method shortens the calibration time and thus improve comfort. In this regard, this research could facilitate real-world applications of SSVEP-based BCIs.

References

[1] J. R. Wolpaw *et al.*, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[2] Z. Lin *et al.*, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2610–2614, 2006.

[3] Y.-T. Wang *et al.*, "A cell-phone-based braincomputer interface for communication in daily life," *J. Neural Eng.*, vol. 8, no. 2, p. 025018, 2011.

[4] M. Nakanishi *et al.*, "A high-speed brain speller using steady-state visual evoked potentials," *Int. J. Neur. Syst.*, vol. 24, no. 06, p. 1450019, 2014.

[5] X. Chen *et al.*, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based braincomputer interface," *J. Neural Eng.*, vol. 12, no. 4, p. 046008, 2015.

[6] M. Nakanishi *et al.*, "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, 2018.

[7] X. Chen *et al.*, "Control of a 7-DOF Robotic Arm System With an SSVEP-Based BCI," *Int. J. Neur. Syst.*, vol. 28, no. 08, p. 1850018, 2018.

[8] C. M. Wong *et al.*, "Spatial filtering in SSVEP-based BCIs: Unified framework and new improvements," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 11, pp. 3057–3072, 2020.

[9] K.-J. Chiang *et al.*, "Boosting template-based SSVEP decoding by cross-domain transfer learning," *J. Neural Eng.*, 2020.

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

[11] Y. Ke *et al.*, "An online SSVEP-BCI system in an optical see-through augmented reality environment," *J. Neural Eng.*, vol. 17, no. 1, p. 016066, 2020.

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

[13] Y. Chen *et al.*, "A novel training-free recognition method for SSVEP-based BCIs using dynamic window strategy," *J. Neural Eng.*, vol. 18, no. 3, p. 036007, 2021.

[14] B. Liu *et al.*, "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.

[15] T. Cao *et al.*, "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.

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

[17] M. Nakanishi, Y.-T. Wang, C.-S. Wei, K.-J. Chiang, and T.-P. Jung, "Facilitating calibration in high-speed BCI spellers via leveraging cross-device shared latent responses," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 4, pp. 1105–1113, 2020.

[18] 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, 2020.

[19] 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.*, early access.

[20] J. Lin, L. Liang, X. Han, C. Yang, X. Chen, and X. Gao, "Cross-target transfer algorithm based on the volterra model of SSVEP-BCI," *Tsinghua Sci. Technol.*, vol. 26, no. 4, pp. 505–522, 2021.

[21] A. Capilla *et al.*, "Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses," *PLoS ONE*, vol. 6, no. 1, p. e14543, 2011.

[22] X. Chen *et al.*, "High-speed spelling with a noninvasive braincomputer interface," *PNAS*, vol. 112, no. 44, pp. E6058–E6067, 2015.

[23] O. Friman *et al.*, "Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 742–750, 2007.

[24] Y. Zhang *et al.*, "Multivariate synchronization index for frequency recognition of SSVEP-based braincomputer interface," *J. Neurosci. Methods*, vol. 221, pp. 32–40, 2014.

[25] R. L. Allen and D. W. Mills, *Signal Analysis: Time, Frequency, Scale, And Structure*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2003.

[26] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, pp. 531–544, 2014.

[27] S. Chen, X. Dong, Z. Peng, W. Zhang, and G. Meng, "Nonlinear chirp mode decomposition: A variational method," *IEEE Trans. Signal Process.*, vol. 65, pp. 6024–6037, 2017.

[28] A. Stallone, A. Cicone, and M. Materassi, "New insights and best practices for the successful use of empirical mode decomposition, iterative filtering and derived algorithms," *Sci. Rep.*, vol. 10, no. 1, p. 15161, 2020.

[29] Z. Wang *et al.*, "Adaptive Fourier decomposition for multi-channel signal analysis," *IEEE Trans. Signal Process.*, vol. 70, pp. 903–918, 2022.

[30] T. Qian, "Sparse representations of random signals," accepted to appear in *Math. Meth. Appl. Sci.*, 2021. arXiv:2008.10473 [math.PR].

[31] Y. Wang *et al.*, "A benchmark dataset for SSVEP-based braincomputer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1746–1752, 2017.

[32] T. Qian *et al.*, "Algorithm of adaptive Fourier decomposition," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5899–5906, 2011.

[33] L. Marple, "Computing the discrete-time "analytic" signal via FFT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2600–2603, 1999.

[34] D. Zwillinger and S. Kokoska, *CRC Standard Probability and Statistics Tables and Formulae*. CRC Press, 2000.

[35] H. P. Vinutha *et al.*, "Detection of outliers using interquartile range technique from intrusion dataset," in *Information and Decision Sciences*. Singapore: Springer Singapore, 2018, pp. 511–518.

[36] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Braincomputer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[37] B. Liu, X. Huang, Y. Wang, X. Chen, and X. Gao, "BETA: A large benchmark database toward SSVEP-BCI application," *Front. Neurosci.*, vol. 14, p. 627, 2020.

[38] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, p. 9, 2016.

[39] A. Notbohm *et al.*, "Modification of brain oscillations via rhythmic light stimulation provides evidence for entrainment but not for superposition of event-related responses," *Front. Hum. Neurosci.*, vol. 10, 2016.

[40] O. B. Guney *et al.*, "A deep neural network for SSVEP-based brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 932–944, 2022.