Multi-Objective Optimisation-based High-pass Spatial Filtering for SSVEP-based Brain-Computer Interfaces

Yue Zhang, *Student Member, IEEE,* Zhenhong Li, *Member, IEEE,* Sheng Quan Xie, *Senior Member, IEEE,* He Wang, *Member, IEEE,* Zhibin Yu, and Zhi-Qiang Zhang, *Member, IEEE*

***Abstract*—Many spatial filtering methods have been proposed to enhance the target identification performance for the steady-state visual evoked potential (SSVEP)-based brain-computer interface (BCI). The existing approaches tend to learn spatial filter parameters of a certain target using only the training data from the same stimulus, and they rarely consider the information from other stimuli or the volume conduction problem during the training process. In this paper, we propose a novel multi-objective optimisation-based high-pass spatial filtering method to improve the SSVEP detection accuracy and robustness. The filters are derived via maximising the correlation between the training signal and the individual template from the same target whilst minimising the correlation between the signal from other targets and the template. The optimisation will also be subject to the constraint that the sum of filter elements is zero. The evaluation study on two self-collected SSVEP datasets (including 12 and 4 frequencies, respectively) shows that the proposed method outperformed the compared methods such as CCA, MsetCCA, SSCOR, and TRCA. The proposed method was also verified on a public 40-class SSVEP benchmark dataset recorded from 35 subjects. The experimental results have demonstrated the effectiveness of the proposed approach for enhancing the SSVEP detection performance.**

***Index Terms*—Brain-computer interface (BCI), electroencephalography (EEG), steady-state visual evoked potential (SSVEP), multi-objective optimisation, high-pass spatial filter.**

I. Introduction

The brain-computer interface (BCI) establishes a direct communication path between the human brain and external devices without relying on normal motor output pathways [1]–[3]. The steady-state visual evoked potential (SSVEP)-based BCI has been widely explored in various applications, such as robot arm [4], communication [5] and augmented reality (AR) glasses [6] [7] due to its high signal-to-noise ratio (SNR), minimal training time, and fast communication rate [8]–[10].

The main task of the SSVEP-based BCI system is to identify target stimuli and then translate them into corresponding commands for subsequent control actions. Thus far, various spatial filters applied transformations in the channel domain, which enhanced the SSVEP identification effectively via removing background artifacts [10]–[12]. As one of the most popular spatial filtering methods, canonical correlation analysis (CCA) seeks a pair of weights to maximise the correlation between SSVEP signals and sine-cosine reference signals [13]. Some extended versions of the standard CCA were proposed to further improve the performance of SSVEP detection. In these methods, one branch starts with brain activity processing, such as filter bank canonical correlation analysis (FBCCA) [14]. As it decomposes SSVEP signals into various sub-band components, the specific information contained in harmonic components can be employed for frequency recognition. Another branch is to optimise the predefined artificial reference signal by incorporating individual calibration data. For example, Zhang *et al.* presented multiway CCA (MwayCCA) [15] and L1-regularized MwayCCA [16] to construct the new reference template via the third-order EEG data tensor. They later proposed the Multiset CCA (MsetCCA) which extracts common features shared by multiple calibration signals to create optimised references, and it outperforms MwayCCA in classification [17] [18]. The latest SSVEP detection methods tend to learn the spatial filter by incorporating a training stage. Besides, these methods are commonly based on template matching, in which the individual template is acquired by averaging multiple training trials [19]. For instance, Wei *et al.* [20] proposed a training data-driven CCA algorithm that yields more robust spatial filters to enhance the SNR of SSVEP signals. However, the inter-trial relationship from the same target was not fully considered. In another study, Nakanishi *et al.* [5] presented task-related component analysis (TRCA) based on the idea that the source signal can be reconstructed through a linear summation of multi-channel EEG signals. The TRCA-based spatial filter is learned by inter-trial covariance maximisation. However, during the spatial filter training process, the aforementioned methods only utilised the signal from a single stimulus, and they rarely considered the information from other stimuli.

Although multi-channel scalp EEG activities are beneficial to achieve high-quality spatial filters, the volume conduction artifact may also be introduced due to the coherence among various EEG channels [21]. It is well known that the measurement of each EEG electrode is a linear mixture of concurrently multiple brain source activities, not just the activity of brain source in its vicinity [22]. In other words,

This work was supported in part by Engineering and Physical Sciences Research Council (EPSRC) (Grant No. EP/S019219/1) and in part by China Scholarship Council (CSC) (Grant No. 201906460007). (Corresponding author: Zhi-Qiang Zhang.)

Yue Zhang, Zhenhong Li, Sheng Quan Xie, and Zhi-Qiang Zhang are with the Institute of Robotics, Autonomous System and Sensing, School of Electrical and Electronic Engineering, University of Leeds, Leeds LS2 9JT, U.K. (e-mail: elyzh@leeds.ac.uk; z.h.li@leeds.ac.uk; s.q.xie@leeds.ac.uk; z.zhang3@leeds.ac.uk). He Wang is with School of Computing, University of Leeds, LS2 9JT, UK (email: H.E.Wang@leeds.ac.uk). Zhibin Yu is with School of Electrical Engineering, Southwest Jiaotong University, 611756, Chengdu, Sichuan, China (zbyu@swjtu.edu.cn).

since the signal from each brain source spreads among various EEG channels when passing through the intracerebral area and the scalp, it is infeasible to connect each electrode with a single brain region [23]. Therefore, the volume conduction decreases the spatial precision of EEG signals when spatially broad features are shared among multiple electrodes [24]. The high-pass spatial filter is commonly used to reduce the effect of spatial blurring for EEG signals [25]. For example, the surface Laplacian has been implemented to diminish volume conduction effectively by attenuating low-spatial-frequency activities whereas highlighting high-spatial-frequency signals [24] [26]. The surface Laplacian filter has some advantages in EEG signal analysis including ease of use and conceptually simple. However, it is still a kind of "stationary" spatial filter since it assumes that the EEG activity is not variable across time [27] [28]. In order to solve this problem, Lu *et.al* [29] designed an adaptive Laplace filter for the sensorimotor rhythms (SMR) analysis, which employs a Gaussian kernel to construct parameters of the spatial filter. Currently, most existing research on volume conduction focused on other paradigms of EEG, like event-related potential (ERP) [27] [26] and SMR [29] [30], rather than SSVEP signals. Therefore, there is still a lack of sufficient studies about the high-pass spatial filter on the volume conduction phenomenon in SSVEP signals. The feasibility of this filter to enhance target detection in the SSVEP paradigm remains unclear.

In this study, we proposed a multi-objective optimisation-based high-pass spatial filtering method to improve the SSVEP identification performance. The filter was derived by maximising the correlation between the training signal and the individual template from the same target, whilst minimising the correlation between signals from other targets and the template. Meanwhile, the constraint condition is that the sum of spatial filter elements is zero. This setting allows the high spatial-frequency SSVEP signals to pass and attenuates low spatial-frequency signals. Therefore, the proposed approach has the potential to extract the target-relevant features, reject the target-irrelevant information, and reduce the effect of volume conduction simultaneously. The method's performance was evaluated on two self-collected SSVEP datasets containing 12 and 4 visual stimulation, respectively. The effectiveness of the proposed approach was verified with an averaged SSVEP recognition accuracy of 87.58% in Dataset I and 87.5% in Dataset II using a 1s-long data epoch. Many well-known and state-of-the-art methods such as CCA, MsetCCA, sum of squared correlations (SSCOR) [31], and TRCA were implemented for extensive comparisons. In addition, a 40-class SSVEP benchmark dataset [32] recorded from 35 subjects was also employed to evaluate the feasibility of the proposed model. The experimental results demonstrate the outperformance of the proposed method in terms of classification accuracy and information transfer rates (ITRs), which is particularly true when the number of stimuli is small. As the number of stimuli increases, the performance of our proposed method will downgrade slightly, but it could still achieve better performance than compared methods.

The remaining paper is arranged as follows. The experiment description and multi-objective optimisation-based high-pass spatial filtering methodology are described in Section II. The results and discussion are provided in Section III. Section IV presents the conclusion.

II. Our methods

A. SSVEP experiment

A public benchmark SSVEP dataset [32] and two self-collected SSVEP datasets (termed as Dataset I and Dataset II hereafter) are employed in this study. The benchmark dataset was collected from 35 healthy subjects with 40 visual stimulation modulated at different frequencies (8-15.8 Hz with an interval of 0.2 Hz), while the phase difference between two neighboring targets is 0.5 π. More information about this dataset can be found in [32]. For the two self-collected datasets, the detailed information is elaborated below.

1) Participants: In Dataset I, SSVEP signals were recorded from 11 subjects (four females and six males, mean age: twenty-five years). In Dataset II, eight subjects (four females and four males, mean age: twenty-seven years) participated in the SSVEP-based BCI experiment. All subjects were in good health, with normal or corrected to normal vision. Each participant seated in a comfortable chair facing the visual simulation. The experiment has been approved by the Research Ethics Committee of the University of Leeds. All subjects have read and signed the participant consent form before the SSVEP experiment.

2) Stimulus Design: In Dataset I, the visual stimulation is designed as a 4×3 matrix coded by the joint frequency and phase modulation (JFPM) method on a 23.6-inch LCD monitor. Its resolution and refresh rate are 1920×1080 pixels and 60 Hz. The 12 visual stimuli flashed at frequencies ranging from 9.25 Hz to 14.75 Hz with an interval of 0.5 Hz. The phase started from 0π to 1.5π in steps of 0.5π. The experiment consisted of 5 blocks for each subject. Each block includes 12 trials corresponding to 12 stimuli. Each trial began with a 0.5 s red dot cue which indicates the target flicker. Later, 12 stimuli started to flash for 5 s simultaneously, during which the subject should stare at the target stimulus and avoid eye movements. In Dataset II, we adopted the g.SSVEPbox for stimulation design. It contains 4 LEDs flickering at various frequencies, namely 14 Hz, 15 Hz, 16 Hz, and 17 Hz. The experiment of each participant also consists of 5 blocks, and there are 4 trials corresponding to four visual stimulation in each block. A small green light appeared about 2 mm away from one of the LEDs for cue-guided purpose. In each trial, four stimuli flash in white for 5 s concurrently. There was a short break of a few seconds between two adjacent blocks in both datasets.

3) EEG Recording: In this study, SSVEP signals were recorded by the experiment equipment from g.tec medical engineering GmbH. The g.USBamp amplifier was applied to sample data at 256 Hz. For Dataset I and II, nine channels including Pz, PO3, POz, PO4, PO7, O1, Oz, O2, and PO8 were selected for EEG data collection. The ground electrode was located over electrode FPz, and the right earlobe served as the reference channel.

B. Data preprocessing

Considering the latency delay in the visual system of each participant, the data epoch after 0.14 s is extracted for analysis. This study incorporated the Chebyshev Type I Infinite Impulse Response (IIR) filter to provide a band-pass filter with a band [8 40] Hz for Dataset I and [13 40] Hz for Dataset II. The zero-phase digital filtering was performed in both the forward and reverse directions.

C. Data processing and target identification method

In this subsection, we introduce a novel multi-objective optimisation-based high-pass spatial filter method. The proposed algorithm aims to extract maximally target-relevant features and minimally target-irrelevant information while reducing the volume conduction artifact.

Denote the single-trial individual calibration signal as $\boldsymbol{\chi}_i^h \in \mathbb{R}^{N_c \times N_s}$ $(h = 1, 2, ..., N_t,\ i = 1, 2, ..., N_f)$, where h and i refer to the index of training trial and the stimulus, respectively. Hereafter, N_c represents the number of channels, N_s is the number of samples, N_t is the number of training trials, and N_f is the number of visual stimulation. Therefore, the continuous training signal of i-th stimulus yielded by concatenating N_t training trials is represented as $\boldsymbol{\chi}_i = [\boldsymbol{\chi}_i^1, \boldsymbol{\chi}_i^2, ..., \boldsymbol{\chi}_i^{N_t}] \in \mathbb{R}^{N_c \times (N_s \cdot N_t)}$. The single-trial individual template signal is denoted as $\overline{\boldsymbol{\chi}}_i = \frac{1}{N_t} \sum_{h=1}^{N_t} \boldsymbol{\chi}_i^h \in \mathbb{R}^{N_c \times N_s}$ which obtained by averaging multiple training trials. Therefore, the continuous individual template signal is defined as $\overline{\boldsymbol{X}}_i = [\overline{\boldsymbol{\chi}}_i, \overline{\boldsymbol{\chi}}_i, ..., \overline{\boldsymbol{\chi}}_i] \in \mathbb{R}^{N_c \times (N_s \cdot N_t)}$. Compared with the artificial reference with sine and cosine waves, the averaged SSVEP signal can effectively enhance the SNR of EEG data [18].

The multi-objective optimisation-based high-pass spatial filtering method designs objective functions for each visual stimulation by means of training signals from not only the same stimulus but also others. To be specific, as the model trains the spatial filter $\hat{\boldsymbol{w}}_i \in \mathbb{R}^{N_c}$, i-th stimulus is regarded as "aim" and the remaining $(N_f - 1)$ stimuli represent "non-aim". For ease of reference, signals of $(N_f - 1)$ "non-aim" form a new dataset $[\gamma_1, \gamma_2, ..., \gamma_{N_f - 1}]$. The proposed method aims to maximise the correlation coefficient between the continuous training signal $\boldsymbol{\chi}_i$ from "aim" and its continuous individual template $\overline{\boldsymbol{X}}_i$, whereas minimise the correlation between the continuous training signal from each "non-aim" γ_j $(j = 1, 2, ..., N_f - 1)$ and $\overline{\boldsymbol{X}}_i$. Therefore, N_f objective functions are designed as:

$$\begin{aligned} f_1(\boldsymbol{w}_i) &= -\rho(\boldsymbol{\chi}_i^{\mathsf{T}} \boldsymbol{w}_i, \overline{\boldsymbol{X}}_i^{\mathsf{T}} \boldsymbol{w}_i), \\ f_2(\boldsymbol{w}_i) &= \rho(\boldsymbol{\gamma}_1^{\mathsf{T}} \boldsymbol{w}_i, \overline{\boldsymbol{X}}_i^{\mathsf{T}} \boldsymbol{w}_i), \\ f_3(\boldsymbol{w}_i) &= \rho(\boldsymbol{\gamma}_2^{\mathsf{T}} \boldsymbol{w}_i, \overline{\boldsymbol{X}}_i^{\mathsf{T}} \boldsymbol{w}_i), \\ &\vdots \\ f_{N_f}(\boldsymbol{w}_i) &= \rho(\boldsymbol{\gamma}_{N_f - 1}^{\mathsf{T}} \boldsymbol{w}_i, \overline{\boldsymbol{X}}_i^{\mathsf{T}} \boldsymbol{w}_i) \end{aligned} \tag{1}$$

where $\rho(\boldsymbol{a}, \boldsymbol{b})$ refers the Pearson correlation coefficient between vector $\boldsymbol{a}$ and vector $\boldsymbol{b}$. In this case, all objective functions are to be minimised. Therefore, the feature related to the i-th stimulus is maximally extracted. On the contrary,

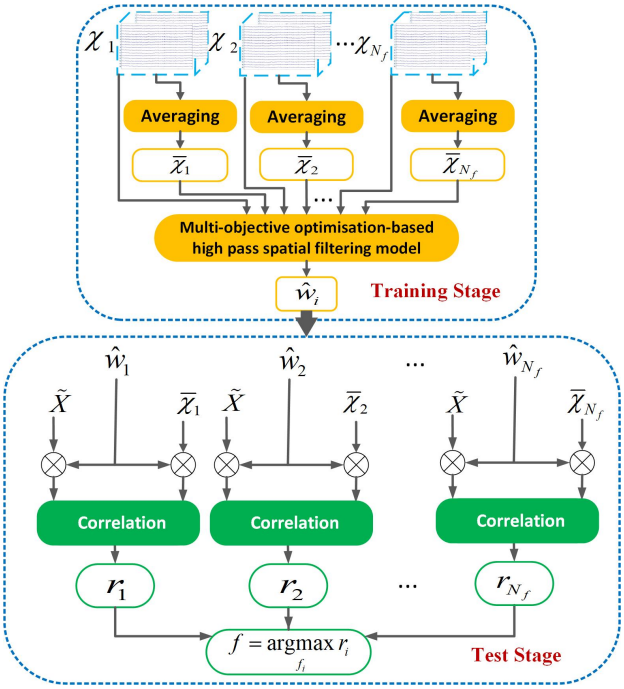


Fig. 1. Flowchart of the proposed SSVEP recognition model. In the training stage, the spatial filter for each stimulus $\hat{\boldsymbol{w}}_i, i = 1, 2, ..., N_f$, is generated with formulas (1)-(3), and the new reference signal, i.e. $\overline{\boldsymbol{\chi}}_1, \overline{\boldsymbol{\chi}}_2, ..., \overline{\boldsymbol{\chi}}_{N_f}$, is obtained by averaging across multiple training trials. In the test stage, with spatial filters $\hat{\boldsymbol{w}}_1, \hat{\boldsymbol{w}}_2, ..., \hat{\boldsymbol{w}}_{N_f}$, the correlation between a test trial $\widetilde{\boldsymbol{X}}$ and each individual template $\overline{\boldsymbol{\chi}}_i, i = 1, 2, ..., N_f$, is computed by formula (4). The frequency of template signal with the maximum correlation coefficient is determined as that of the test sample by formula (5).

the information relevant to "non-aim" is minimally included in the spatial filter $\hat{\boldsymbol{w}}_i$. Simultaneously, considering the effect of volume conduction in SSVEP signals, the model is subject to a linear equality constraint, that is, the sum of spatial filter elements is zero. It aims to reduce the low spatial-frequency signal whilst retaining the high spatial-frequency. Therefore, the multi-objective optimisation problem is formulated by the following statement:

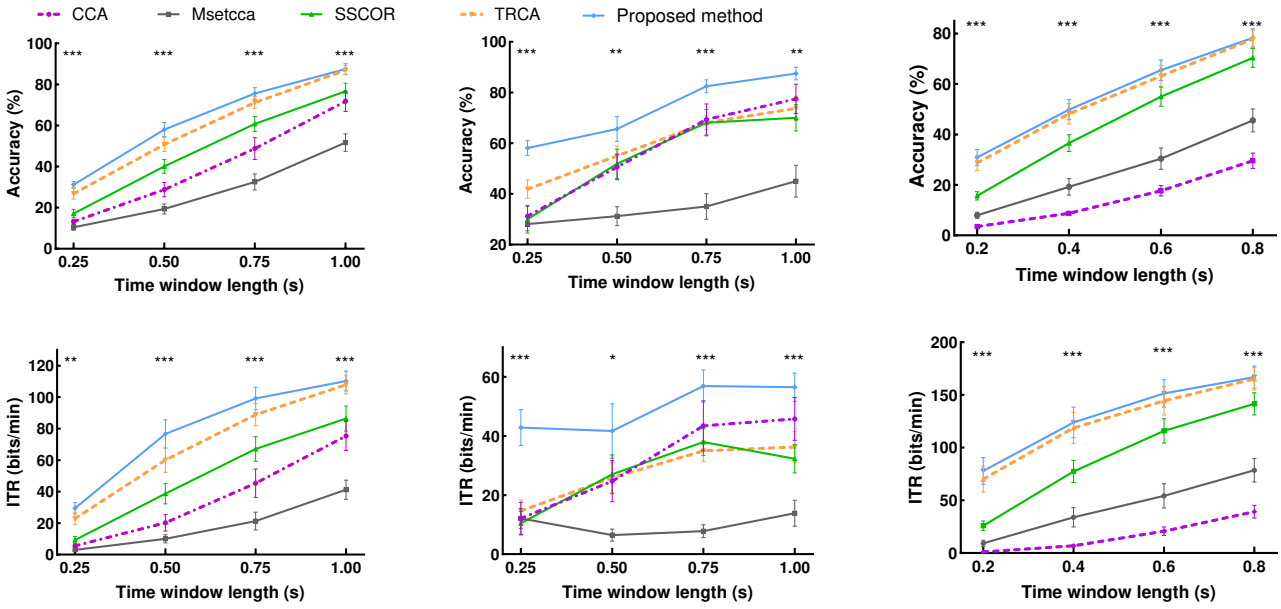
$$\begin{aligned} &\underset{\boldsymbol{w_i}}{\text{minimise}}\ F(\boldsymbol{w}_i) = [f_1(\boldsymbol{w}_i), f_2(\boldsymbol{w}_i), ..., f_{N_f}(\boldsymbol{w}_i)], \\ &\text{subject to}\ \sum_{c=1}^{N_c} w_i^c = 0, \boldsymbol{w}_i \in \boldsymbol{W} \end{aligned} \tag{2}$$

where c is the channel index and w_i^c refers to the c-th element of spatial filter vector $\boldsymbol{w}_i$. The $\boldsymbol{W} \subseteq \mathbb{R}^{N_c}$ is the feasible set of solution vectors. Therefore, the spatial filter $\hat{\boldsymbol{w}}_i$ can be acquired as follows:

$$\hat{\boldsymbol{w}}_i = \arg\min_{\boldsymbol{w}_i} F(\boldsymbol{w}_i). \tag{3}$$

This constrained multi-objective optimisation problem can be solved by the fgoalattain() function in Matlab. In the test phase, the signal-trial test SSVEP data $\widetilde{\boldsymbol{X}} \in \mathbb{R}^{N_c \times N_s}$ and the single-trial template signal $\overline{\boldsymbol{\chi}}_i$ are both spatially filtered with the optimal solution vector $\hat{\boldsymbol{w}}_i$. The correlation coefficient calculated between two optimised signals is shown as follows:

$$r_i = \rho(\widetilde{\boldsymbol{X}}^{\mathsf{T}} \hat{\boldsymbol{w}}_i, \overline{\boldsymbol{\chi}}_i^{\mathsf{T}} \hat{\boldsymbol{w}}_i),\ i = 1, 2, ..., N_f \tag{4}$$



(a) Dataset I (b) Dataset II (c) Benchmark dataset

Fig. 2. Averaged recognition accuracy and ITRs across subjects of various methods using different time windows on (a) Dataset I, (b) Dataset II and (c) benchmark dataset. The error bars represent standard errors. The asterisks indicate significant difference between the five methods obtained by one-way repeated-measures ANOVA ($*$ p$<$0.05, $**$p$<$0.01, $***$p$<$0.001).

The N_f coefficients can be calculated by this formula. The frequency of the template signal corresponding to the maximal correlation coefficient value is determined as the frequency f of the SSVEP test trial $\widetilde{\boldsymbol{X}}$, and it is can be represented as:

$$f = \arg\max_{f_i} r_i, i = 1, 2, ..., N_f \tag{5}$$

The diagram of the whole SSVEP recognition model is shown in Fig.1. It includes two stages, namely training and test. For each stimulus, the training part provides its spatial filter and corresponding reference signal. The test part aims to classify the single-trial test signal to a specific visual stimulus based on outputs of the training phase.

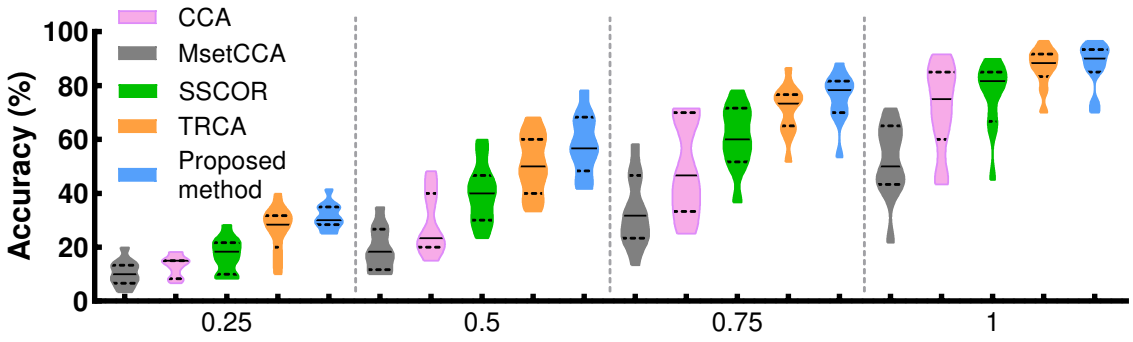
III. Results and Discussion

In this section, we first evaluated the performance of the proposed multi-objective optimisation-based high-pass spatial filtering method on two self-collected SSVEP datasets and a 40-target benchmark dataset [32]. Extensive comparisons were implemented between the proposed method and many state-of-the-art SSVEP recognition methods. The ensemble-based methods were also implemented in this section for comparison purposes. The influences of different parameters such as the number of electrodes, the number of training blocks, and the number of frequencies on the performance were also discussed.

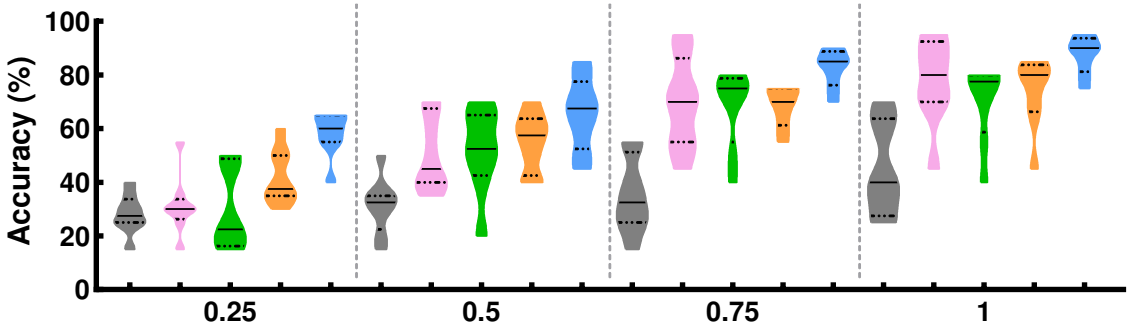
A. Performance Evaluation

The performance comparison was conducted between the proposed model and many recognition methods such as CCA, MsetCCA, SSCOR, and TRCA. The classification accuracy and ITRs were calculated by leave-one-out cross-validation to evaluate the recognition performance of these algorithms. Specifically, the SSVEP signals from four (in Dataset I and II) or five (in benchmark dataset) blocks were employed as the training data, and the left-out block was used as test data. Fig. 2 illustrates the averaged classification accuracy and ITRs across all subjects in three datasets with different time windows (TWs). It is obvious that the proposed method can in general achieve the highest accuracy and ITRs with different data lengths. One-way repeated-measures ANOVA was conducted to investigate the similarity of accuracy or ITRs of different SSVEP detection approaches. The results indicate that their accuracy and ITRs have a significant difference in most TWs on three datasets. For example, in Fig. 2(a), the proposed method improved by 27.24% for CCA, 38.64% for MsetCCA, 17.88% for SSCOR, and 7.27% for TRCA with 1s data length.

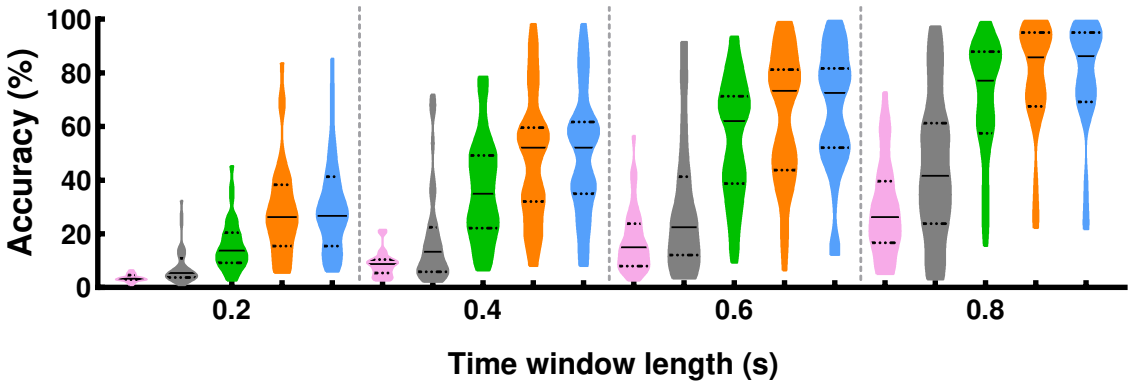
In addition to the averaged values, we further explored the distribution of numeric data across multiple methods. Fig. 3 adopts the violin plot to show the probability density of recognition accuracy of all subjects on (a) Dataset I, (b) Dataset II, and (c) benchmark dataset, respectively. The plots are based on five methods with different data lengths. Each density curve can be compared to see the similarities or differences between the methods. Fig. 3(a) and Fig. 3(b) show that our method achieves the highest median and most concentrated distribution with most data lengths. Therefore, to a large extent, the proposed method provides higher and more consistent accuracy across subjects. In Fig. 3(c), all methods show a more scattered distribution compared with Fig. 3(a) and Fig. 3(b), because there are significantly more subjects in the



(a) Dataset I



(b) Dataset II



(c) benchmark dataset

Fig. 3. Violin plots represent the distributions of classification accuracy of all subjects achieved by the five methods with various TWs on (a) Dataset I, (b) Dataset II, and (c) benchmark dataset. Black solid line in each violin indicates median, and two black dotted lines represent interquartile ranges (25% and 75% percentiles).

benchmark dataset. The proposed method provides medians similar to TRCA, but obviously higher than the values of CCA, MsetCCA, and SSCOR.

Aside from the overall comparison of the five approaches, we further carried out a pairwise analysis between the proposed method and each comparison method on three datasets. Table I shows the multiple comparison results in terms of SSVEP recognition accuracy with different TWs. The multiple comparison correction method is Dunnett's test. The proposed method is regarded as the control group. The results show that in most TWs, the proposed method is statistically superior to MsetCCA, CCA, and SSCOR, and provides similar accuracy to TRCA.

The above experiment results were carried out on a DELL laptop with a 1.8GHz quad-core CPU, 8 GB RAM, using Matlab 2019a and running on Windows 10. The first step was to train the spatial filter, and it took up to 6.36 s in Dataset I, 2.33 s in Dataset II, and 26.81 s in benchmark dataset. Once the spatial filter is trained, the averaged recognition time per time window for performing the proposed method on Dataset I, Dataset II, and benchmark dataset was 0.044s, 0.012s, and 0.193s, respectively. We would like to point out that although

TABLE I
MULTIPLE COMPARISON ON CLASSIFICATION ACCURACY OF FIVE METHODS BY DUNNETT'S TEST

Methods	Time windows											
	Dataset I				Dataset II				benchmark dataset			
	0.25s	***0.5s***	***0.75s***	***1s***	***0.25s***	***0.5s***	***0.75s***	***1s***	***0.2s***	***0.4s***	***0.6s***	***0.8s***
MsetCCA vs Proposed method*	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CCA vs Proposed method	<0.0001	<0.0001	<0.0001	0.0158	<0.0001	0.0982	0.1534	0.4514	<0.0001	<0.0001	<0.0001	<0.0001
SSCOR vs Proposed method	<0.0001	0.0012	0.0299	0.1378	<0.0001	0.1436	0.1042	0.0672	<0.0001	0.0178	0.1567	0.3664
TRCA vs Proposed method	0.2428	0.3339	0.8348	0.9998	0.0193	0.3312	0.1042	0.1932	0.8771	0.9893	0.9778	0.9999

*The proposed methods serves as the control group.

the training time is a bit long, it won't affect the performance of the proposed method. We could always train the spatial filter offline before we move on to real-time SSVEP recognition.

B. Ensemble-based method evaluation

This work extends the proposed method with an ensemble version. Ideally, the transformation coefficients from SSVEP source to scalp could be treated similarly within the used frequency range [5] [33] [31]. Hence, N_f different spatial filters should be similar. The ensemble-based method concatenates all N_f spatial filters to construct an ensemble spatial filter $\boldsymbol{W} \in \mathbb{R}^{N_c \times N_f}$,

$$\boldsymbol{W} = [\hat{\boldsymbol{w}}_1, \hat{\boldsymbol{w}}_1, ..., \hat{\boldsymbol{w}}_{N_f}] \tag{6}$$

Then the (4) can be re-defined as follows:

$$r_i = \rho(\widetilde{\boldsymbol{X}}^{\mathsf{T}}\boldsymbol{W}, \overline{\boldsymbol{\chi}}_i^{\mathsf{T}}\boldsymbol{W}),\ i = 1, 2, ..., N_f \tag{7}$$

The intended target is determined via (5). Fig. 4 shows the performance comparison between different standard methods and ensemble methods on (a) Dataset I, (b) Dataset II, and (c) benchmark dataset. As shown by the blue and purple lines in Fig. 4, the performance of the proposed method is improved via the ensemble version in three datasets. Meanwhile, in Fig. 4(a) and Fig. 4(c), the classification accuracy of the ensemble proposed method is slightly better than that of ensemble TRCA, but the gap with ensemble SSCOR is even greater. In Fig. 4(b), the standard proposed method shows superiority compared with ensemble TRCA and ensemble SSCOR with all data lengths.

C. The Influence of Parameters on Performance

In order to further evaluate the performance of the five approaches, we explored the impact of the number of electrodes, training blocks, and stimulation on the SSVEP recognition accuracy.

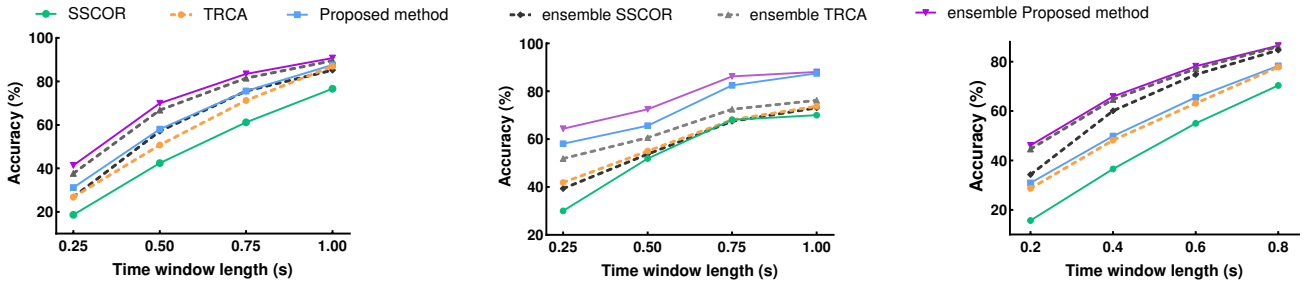
1) The number of electrode: Fig. 5 depicts the averaged SSVEP classification accuracy with various numbers of channels using 0.75s-long data on (a) Dataset I, (b) Dataset II, and (c) benchmark dataset. In Fig. 5, the accuracy of each method generally declines as the number of channels decreases. It indicates that the number of channels affects the performance of various SSVEP recognition methods. Specifically, the method with more channels can commonly achieve better classification accuracy. It is worth mentioning that for channels = 5, 6, 7, 8, and 9, the proposed model (bars in blue) always outperforms the other four recognition methods on all datasets.

2) The number of training blocks: We further investigated how the number of training blocks affects the target identification performance of five algorithms on three datasets. The heat maps in Fig. 6 illustrate comparison results with 0.75 s TWs. The x-axis refers to the method with a different number of training blocks. The y-axis is the subject index, and color indicates the corresponding performance. The maximal SSVEP detection accuracy shows the deepest color. For all methods in the three datasets, the color obtained with more training blocks is usually darker than that obtained with fewer training blocks. In addition, the proposed method generally shows deeper color compared with the other four methods for most subjects under the different number of training blocks. The results indicate that, to some extent, our method is superior to other methods regardless of the number of training blocks.

3) The number of frequencies: In this study, we also investigated the influence of the number of targets on the proposed method. In order to explore more types of target numbers, we employed the 40-class benchmark dataset in this subsection. SSCOR and TRCA significantly outperform CCA and MsetCCA on the benchmark dataset as shown in Fig. 2, so these two methods are utilized for performance comparison. Fig. 7 shows the averaged classification accuracy of three methods to classify 8, 16, 24, 32, and 40 stimulation, respectively. The choice of targets is random. The comparison result indicates that the performance of our method tends to decrease slightly with the increasing number of stimulation. Under the same setting, TRCA and SSCOR also show a similar declining trend. However, even with a large number of targets, the recognition accuracy of the proposed method is always better than other models. Similar comparison results can also be found in Fig. 2(a) with 12-class Dataset I and Fig. 2(b) with 4-class Dataset II.

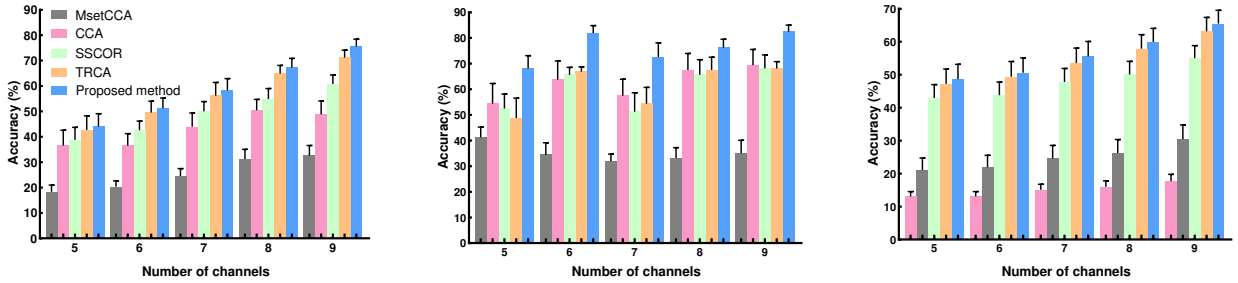
D. Discussion

In order to learn the class-specific spatial filter, the traditional SSVEP recognition scheme generally considers the correlation among training trials from the same class [5] [17] [31]. For example, TRCA trains the spatial filter by maximizing the sum of covariance of all possible combinations of



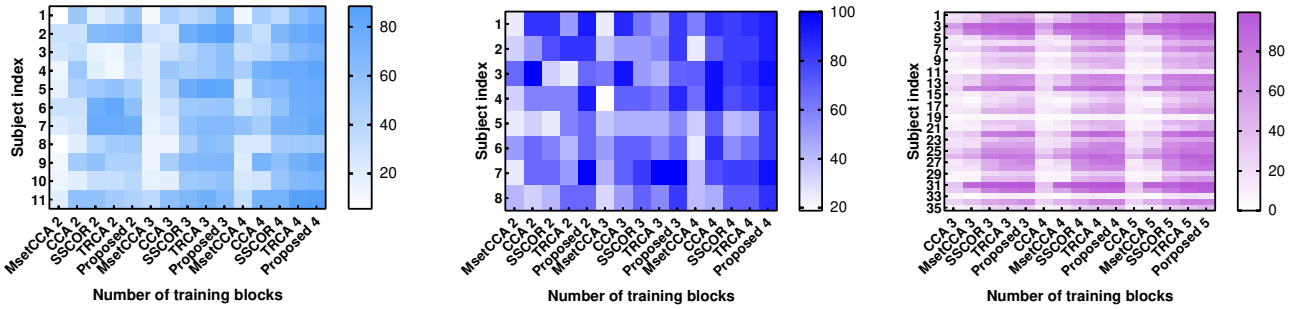
(a) Dataset I (b) Dataset II (c) benchmark dataset

Fig. 4. Averaged recognition accuracy comparison between ensemble methods and standard methods on (a) Dataset I, (b) Dataset II, and (c) benchmark dataset.



(a) Dataset I (b) Dataset II (c) benchmark dataset

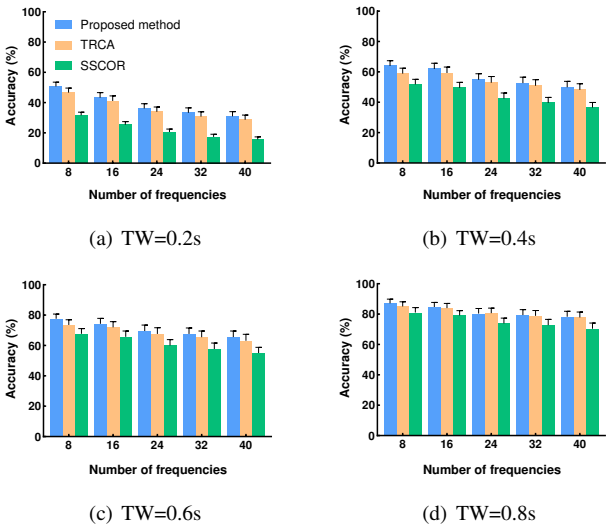
Fig. 5. Barcharts of the five methods' classification accuracy with different number of channels on (a) Dataset I, (b) Dataset II, and (c) benchmark dataset. The error bars represent standard errors.



(a) Dataset I (b) Dataset II (c) benchmark dataset

Fig. 6. Heat maps of the classification accuracy of five methods under different number of training blocks on (a) Dataset I, (b) Dataset II, and (c) benchmark dataset.

training trials from the same stimulation [5]. For the proposed model, it constructs multi-objective functions creatively not only to decrease the distance between the training signal and the template from the same class but also to increase the distance between SSVEP data from other classes and the template, as shown in formulas (1)-(3). Therefore, the trained spatial filters have the potential to extract target-related features and decrease target-unrelated information. Besides, the optimisation problem is subject to the constraint that the sum of filter elements is zero so that low spatial-frequency signals can be alleviated. With the spatial filters, the proposed model could better distinguish between target and non-target stimulation. The false positive rate (FPR) for a multi-target system is the percentage of non-targets that are wrongly categorised as the true target [20]. As a representative example, the confusion matrix in Fig. 8 provides the averaged FPR across all subjects and targets. Clearly, the averaged FPRs of the proposed method and TRCA are 12.5% and 26.25%, respectively. Further, we explored the probability density of detection accuracy of the proposed method in Fig. 3. The frequency of data points in each region correlates to the width of each curve. The thicker part means that the corresponding value has a higher frequency. As shown in Fig. 3, when the median value among methods is approximated (e.g. 1 s data length in Dataset II), the proposed method has a more concentrated distribution. It implies that the proposed method could achieve a more stable and consistent performance across subjects.



(a) TW=0.2s (b) TW=0.4s

(c) TW=0.6s (d) TW=0.8s

Fig. 7. Performance comparison with various numbers of targets, i.e. 8, 16, 24, 32, and 40 using different data lengths (a) TW=0.2s, (b) TW=0.4s, (c) TW=0.6s, (d) TW=0.8s. The error bars represent standard errors.

In order to further clarify the effect of stimulation mechanisms on the performance, visual stimuli are displayed in different ways, i.e. LCD monitor in Dataset I and LEDs in Dataset II, respectively. As shown by the blue lines in Fig. 2, the proposed method always exhibits the best classification accuracy with various TWs. Blue violins in Fig. 3 reflect the same result via another form. It illustrates that different choices of stimulation mechanisms do not affect the performance of the proposed method too much.

As a classical method, CCA has been widely validated by many studies in detecting SSVEPs [13] [14]. Traditionally, its reference signals, constructed by sine-cosine waves, are employed to model the visual stimuli [34]. Fig. 2 shows that CCA could reach an averaged recognition accuracy of 71.81% in Dataset I, 77.5% in Dataset II, and 29.57% in the Benchmark dataset. It does not achieve as high accuracy as other methods at some TWs. A potential problem with this method is that the artificially predefined reference signal cannot reflect subject-specific features [17] [35]. Our method adopts the individual training data to construct subject-specific templates. As illustrated in [18], one of the factors in the target recognition method that leads to class and non-class SSVEP signals being more distinguishable is individual templates.

Although our proposed method provides good recognition performance in SSVEP-based BCIs, it still has some potential improvement directions. First, as shown in Fig. 7, as the number of stimuli increases, the performance of our approach tends to decline slightly. In fact, as the number of objectives in a multi-objective optimisation problem increases, the dimensionality of the model also increases, which makes it difficult to converge [36]. In addition, the interaction among various objective functions makes the optimisation problem more intricate. It has been observed that the multi-objective optimisation technique experiences performance degradation when applied to a problem with a high number of objective functions [37]. Therefore, the performance comparisons do not show a huge difference when the proposed approach is applied to an optimisation problem based on a large number of objective functions. Future work will thus be required to 1) optimise the selection of non-aim stimuli to reduce the number of minimisation objectives and 2) explore other algorithms to solve complex multi-objective optimisation problems with a large number of objectives for the SSVEP-based BCI system. Second, as Fig. 6 demonstrated, the training blocks may not be sufficient for some subjects, resulting in relatively low classification accuracy. Therefore, future work would transfer information from other subjects to solve the data insufficiency problem.

IV. Conclusion

In this study, a novel multiple-objective optimisation-based high-pass spatial filtering method was proposed to improve the target recognition performance for the SSVEP-based BCI system. Our method defined multi-objective functions for each target by the correlation between the training signal and the individual template from the same stimulation maximisation, whereas the correlation between SSVEP data from others and the template minimisation. It aims to effectively extract target-relevant information while decreasing target-irrelevant features. Meanwhile, the proposed model has the constraint that the sum of filter elements is zero. This setting could reduce the negative effect of volume conduction by alleviating low spatial-frequency signals. Experimental results based on two self-collected SSVEP datasets and a public benchmark database showed that the proposed model achieved better classification performance than some state-of-the-art methods.

References

[1] X. Yu, M. Z. Aziz, M. T. Sadiq, Z. Fan, and G. Xiao, "A new framework for automatic detection of motor and mental imagery EEG signals for robust BCI systems," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[2] S. Chevallier, E. K. Kalunga, Q. Barthélemy, and E. Monacelli, "Review of Riemannian distances and divergences, applied to SSVEP-based BCI," *Neuroinformatics*, vol. 19, no. 1, pp. 93–106, 2021.

[3] P. Stegman, C. S. Crawford, M. Andujar, A. Nijholt, and J. E. Gilbert, "Brain–computer interface software: A review and discussion," *IEEE T. Hum-Mach. Syst.*, vol. 50, no. 2, pp. 101–115, 2020.

[4] X. Chen, X. Huang, Y. Wang, and X. Gao, "Combination of augmented reality based brain-computer interface and computer vision for high-level control of a robotic arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2020.

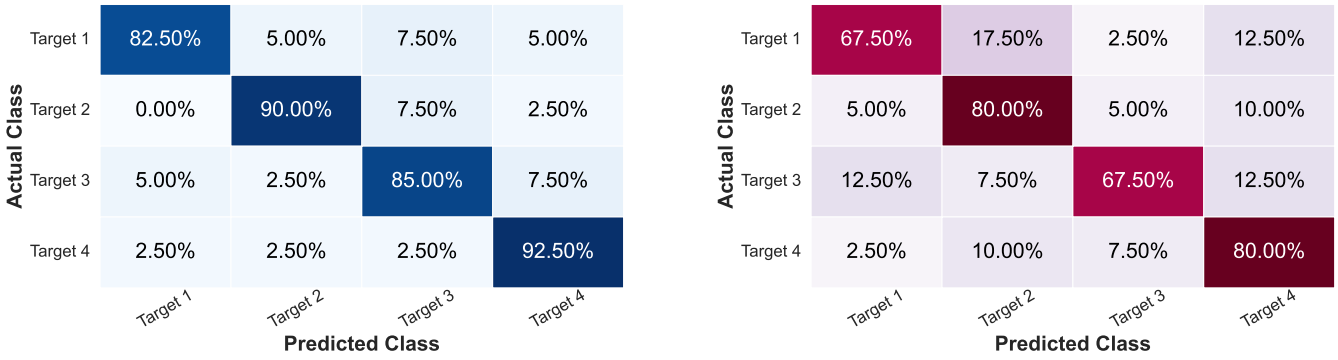
[5] M. Nakanishi, Y. Wang, X. Chen, Y.-T. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, 2017.

[6] L. Angrisani, P. Arpaia, A. Esposito, and N. Moccaldi, "A wearable brain–computer interface instrument for augmented reality-based inspection in industry 4.0," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1530–1539, 2019.

[7] P. Arpaia, L. Duraccio, N. Moccaldi, and S. Rossi, "Wearable brain–computer interface instrumentation for robot-based rehabilitation by augmented reality," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6362–6371, 2020.

[8] L. Angrisani, P. Arpaia, D. Casinelli, and N. Moccaldi, "A single-channel SSVEP-based instrument with off-the-shelf components for trainingless brain-computer interfaces," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3616–3625, 2018.

[9] A. Liu, Q. Liu, X. Zhang, X. Chen, and X. Chen, "Muscle artifact removal towards mobile SSVEP-based BCI: A comparative study," *IEEE Trans. Instrum. Meas.*, 2021.



(a) Confusion matrix generated by the proposed model

(b) Confusion matrix generated by TRCA

Fig. 8. Confusion matrices of target recognition on Dataset II achieved by (a) multi-objective optimisation-based high-pass spatial filtering model and (b) TRCA with 1s data length. Each cell shows the ratio of the number of observations to the total number of test trials per target.

[10] 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, 2020.

[11] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, 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.

[12] 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, 2020.

[13] 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, 2006.

[14] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface," *J. Neural Eng.*, vol. 12, no. 4, p. 046008, 2015.

[15] Y. Zhang, G. Zhou, Q. Zhao, A. Onishi, J. Jin, X. Wang, and A. Cichocki, "Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCIs," in *Proc. Int. Conf. Neural. Inf. Process.*, 2011, pp. 287–295.

[16] 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, 2013.

[17] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis," *Int. J. Neural Syst.*, vol. 24, no. 04, p. 1450013, 2014.

[18] 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, 2015.

[19] T. Tanaka and M. Arvaneh, *Signal processing and machine learning for brain-machine interfaces*. Institution of Engineering & Technology, 2018.

[20] Q. Wei, S. Zhu, Y. Wang, X. Gao, H. Guo, and X. Wu, "A training data-driven canonical correlation analysis algorithm for designing spatial filters to enhance performance of SSVEP-Based BCIs," *Int. J. Neural. Syst.*, vol. 30, no. 05, p. 2050020, 2020.

[21] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from EEG signals: a review," *Sensors*, vol. 19, no. 5, p. 987, 2019.

[22] A. Khadem and G.-A. Hossein-Zadeh, "Quantification of the effects of volume conduction on the EEG/MEG connectivity estimates: an index of sensitivity to brain interactions," *Physiol Meas.*, vol. 35, no. 10, p. 2149, 2014.

[23] A. Anzolin, P. Presti, F. Van de Steen, L. Astolfi, S. Haufe, and D. Marinazzo, "Effect of head volume conduction on directed connectivity estimated between reconstructed EEG sources," *BioRxiv*, p. 251223, 2018.

[24] R. Debnath, V. C. Salo, G. A. Buzzell, K. H. Yoo, and N. A. Fox, "Mu rhythm desynchronization is specific to action execution and observation: Evidence from time-frequency and connectivity analysis," *Neuroimage*, vol. 184, pp. 496–507, 2019.

[25] N. Caramia, F. Lotte, and S. Ramat, "Optimizing spatial filter pairs for EEG classification based on phase-synchronization," in *Proc. Int. Conf. Acoust. Speech. Signal. Process.* IEEE, 2014, pp. 2049–2053.

[26] J. Kayser and C. E. Tenke, "Issues and considerations for using the scalp surface Laplacian in EEG/ERP research: A tutorial review," *Int. J. Psychophysiol.*, vol. 97, no. 3, pp. 189–209, 2015.

[27] R. J. Bufacchi, C. Magri, G. Novembre, and G. D. Iannetti, "Local spatial analysis: an easy-to-use adaptive spatial EEG filter," *Journal of Neurophysiology*, vol. 125, no. 2, pp. 509–521, 2021.

[28] F. P. Kalaganis, E. Chatzilari, S. Nikolopoulos, N. A. Laskaris, and Y. Kompatsiaris, "A collaborative representation approach to detecting error-related potentials in SSVEP-BCIs," in *Proc. Thematic. Workshop. ACM. Multimedia.*, 2017, pp. 262–270.

[29] J. Lu, D. J. McFarland, and J. R. Wolpaw, "Adaptive laplacian filtering for sensorimotor rhythm-based brain–computer interfaces," *J. Neural Eng.*, vol. 10, no. 1, p. 016002, 2012.

[30] S. E. Kober, C. Neuper, and G. Wood, "Differential effects of up- and down-regulation of SMR coherence on EEG activity and memory performance: A neurofeedback training study," *Front. Hum. Neurosci.*, vol. 14, 2020.

[31] K. K. GR and 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, 2019.

[32] 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, 2016.

[33] R. Srinivasan, F. A. Bibi, and P. L. Nunez, "Steady-state visual evoked potentials: distributed local sources and wave-like dynamics are sensitive to flicker frequency," *Brain. Topogr.*, vol. 18, no. 3, pp. 167–187, 2006.

[34] C. M. Wong, F. Wan, B. Wang, Z. Wang, W. Nan, K. F. Lao, P. U. Mak, M. I. Vai, and A. Rosa, "Learning across multi-stimulus enhances target recognition methods in SSVEP-based BCIs," *J. Neural Eng.*, vol. 17, no. 1, p. 016026, 2020.

[35] 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 Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 3037–3040.

[36] S. Mane and M. N. Rao, "Many-objective optimization: Problems and evolutionary algorithms–a short review," *Int. J. Appl. Eng. Res.*, vol. 12, no. 20, pp. 9774–9793, 2017.

[37] A. Prajapati, "A comparative study of many-objective optimizers on large-scale many-objective software clustering problems," *Complex & Intelligent Systems*, vol. 7, no. 2, pp. 1061–1077, 2021.