Filter Bank Convolutional Neural Network for Short Time-Window Steady-State Visual Evoked Potential Classification

Wenlong Ding, Jianhua Shan, Bin Fang, *Member, IEEE*, Chengyin Wang, Fuchun Sun, and Xinyue Li

***Abstract*— Convolutional neural network (CNN) has been gradually applied to steady-state visual evoked potential (SSVEP) of the brain-computer interface (BCI). Frequency-domain features extracted by fast Fourier Transform (FFT) or time-domain signals are used as network input. In the frequency-domain diagram, the features at the short time-window are not obvious and the phase information of each electrode channel may be ignored as well. Hence we propose a time-domain-based CNN method (tCNN), using the time-domain signal as network input. And the filter bank tCNN (FB-tCNN) is further proposed to improve its performance in the short time-window. We compare FB-tCNN with the canonical correlation analysis (CCA) methods and other CNN methods in our dataset and public dataset. And FB-tCNN shows superior performance at the short time-window in the intra-individual test. At the 0.2 s time-window, the accuracy of our method reaches 88.36 ± 4.89% in our dataset, 77.78 ± 2.16% and 79.21 ± 1.80% respectively in the two sessions of the public dataset, which is higher than other methods. The impacts of training-subject number and data length in inter-individual or cross-individual are studied. FB-tCNN shows the potential in implementing inter-individual BCI. Further analysis shows that the deep learning method is easier in terms of the implementation of the asynchronous BCI system than the training data-driven CCA. The code is available for reproducibility at https://github.com/ DingWenl/FB-tCNN.**

***Index Terms*— Asynchronous brain-computer interface (BCI) system, convolutional neural network (CNN), cross-individual, inter-individual, short time-window, steady-state visual evoked potential (SSVEP).**

I. Introduction

BRAIN-COMPUTER interface (BCI) can realize the information exchange between the human brain and external devices [1], [2], it can transform human's intention into control command of external devices. Disabled people who have difficulties in oral communication or movement can still generate electroencephalogram (EEG) signals that can be detected and recognized by the BCI, which can be classified into invasive and non-invasive types nowadays [3]–[6].

Manuscript received August 3, 2021; revised October 27, 2021; accepted November 23, 2021. Date of publication December 1, 2021; date of current version December 22, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 91848206 and Grant 62173197 and in part by the Natural Science Foundation of Anhui Province under Grant 2108085MF224. *(Wenlong Ding and Jianhua Shan contributed equally to this work.) (Corresponding author: Bin Fang.)*

Wenlong Ding, Jianhua Shan, and Chengyin Wang are with the Department of Mechanical Engineering, Anhui University of Technology, Maanshan, Anhui 243032, China.

Bin Fang and Fuchun Sun are with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: fangbin@tsinghua.edu.cn).

Xinyue Li is with the Department of Mathematics, New York University, Abu Dhabi, United Arab Emirates.

Digital Object Identifier 10.1109/TNSRE.2021.3132162

A. Invasive BCI

Invasive BCI refers to the need for craniotomy, the probe implanted into the cerebral cortex, which is used to obtain high-quality brain activity signals.

B. Non-Invasive BCI

Non-invasive BCI means that EEG signals can be obtained by sticking electrodes on the scalp without craniotomy.

The quality of the EEG signal obtained by non-invasive BCI is relatively poor [7] compared with the signal obtained by the invasive BCI, but the operation is less risky medically and more convenient. Therefore, the EEG obtained by the non-invasive BCI is more widely used [8]–[12]. At present, the main research directions of EEG are motor imagery (MI) [13]–[15], P300 [16], [17], and steady-state visual evoked potential (SSVEP) [18]–[20].

SSVEP has been widely applied and proved to be the most promising EEG paradigm in BCI technology [21]. It is more stable and reliable since its frequency-domain features are evident and easy to identify. Several stimulus targets flicker at different fixed frequencies to induce specific SSVEP signals. By analyzing the induced SSVEP signals, which stimulus target the user is focusing on can be determined.

Various methods have been proposed to aid the recognition of SSVEP. Canonical correlation analysis (CCA) is the mainstream recognition algorithm [22]–[26]. It recognizes SSVEP by determining the correlation between the collected data and the stimulus target signal template. But the CCA method (Lin *et al*. [22]) performs poorly at the short time-window. Wei *et al*. [24] propose a training data-driven CCA method (which is named CCA-M3 in [24]) overcomes the problem. But CCA-M3 has the problem of time-window alignments. It requires a specific-position time-window, such as [0.14, $0.14 + d$] s, which starts 0.14 s after the onset time of the stimulus, and d s is the data length of time-window. Thus

triggers with high temporal precision are required, and the subjects' attention needs to be highly focused to ensure the good performance of CCA-M3. Asynchronous BCI is difficult to implement via the CCA-M3. Deep learning [27], which allows end-to-end automatic learning of preprocessing, feature extraction, and classification modules and can often simplify the processing process. Compared with CCA-M3, deep learning is an easier method in terms of implementing asynchronous BCI. Recently, the convolutional neural networks (CNN) method has been applied to SSVEP recognition. While most of these methods use the frequency-domain feature as the input of the network after fast Fourier Transform (FFT) [28]–[30] because of its obvious features (spikes) at the long time-window (FFT-CNN).

However, realizing the recognition of SSVEP signal at a short time-window is expected in practical application. But the spikes at the frequency of stimulus target on the frequency-domain would become less obvious at the short time-window. Therefore, the FFT-CNN performs poorly at short time-window. In addition, the frequency-domain as input would result in the loss of other task-relevant information, such as time-difference which is represented by phase information in the time-domain. The difference in the distribution of these electrodes on the EEG cap causes the time of action potential in the occipital region reaching these electrodes to be different too. The time-difference of SSVEP data between electrode channels possesses great value since it contains the spatial information of these electrode channels. Furthermore, FFT would greatly lengthen the training time of a model. Therefore, we propose a time-domain-based CNN method (tCNN), which does not need FFT and requires a much simpler operation, while being able to recognize more effectively at the short time-window.

Aznan *et al*. [31] and Waytowich *et al*. [32] respectively proposed the CNN architecture based on time-domain data. But their network architectures are different from our tCNN. There are fewer subject-quantities in [31], resulting in the lack of persuasion. For [32], their network architecture (Compact-CNN) contains pooling layers, which result in a relatively complex architecture. There is no pooling layer in our network architecture, which makes the network architecture more concise. And the dynamic convolution kernels are added to the network architecture to make it more suitable for different input sizes. Referring to the FBCCA proposed by Chen *et al*. [26], the filter bank is used as the pre-filtering to make the network architecture extract the task-related features (such as the frequency-features of fundamental and harmonic) more effectively at the short time-window. Therefore, based on the tCNN method, we propose a filter bank tCNN (FB-tCNN) method.

The main contributions of this article can be summarized as follows.

1) FB-tCNN method is proposed that shows superior performance in the intra-individual test at the short time-window (such as 0.1 s, 0.2 s, and 0.3 s) in our dataset and the public dataset [36].
2) The impacts of training-subject number and data length in inter-individual (cross-individual) are studied, and FB-tCNN shows superior performance at all data lengths in the range of short time-window ([0.1, 1.0] s). FB-tCNN shows the potential to implement inter-individual BCI.
3) The frame-by-frame recognition method is used to determine whether the algorithm is suitable for implementing the asynchronous BCI system. At the same time, we use the frame-by-frame recognition method to test some trials that contain misclassified samples and propose the possibility of error labels.

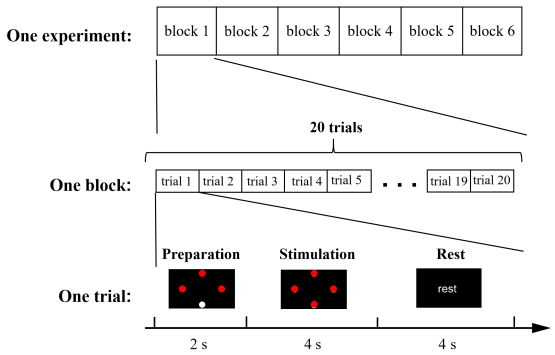

Fig. 1. Experimental design of our dataset.

The architecture of this paper is as follows. Section I gives a brief introduction to BCI and some SSVEP-recognition methods. Section II presents our methods. Section III makes a comparison between our method with others and conducts inter-individual studies. In section IV, the frame-by-frame recognition method is used to explore whether these methods can realize asynchronous BCI. Meanwhile, some trials that contain misclassified samples in section III are recognized frame by frame. In section V, we make a summary and put forward the prospects.

II. Methods

A. Dataset

1) Our Dataset: Seven healthy subjects (ages 23-30, all-male) participated in the experiment. Four subjects have never used BCI, while others have previous experience with BCI experiments. The study was approved by the local Ethics Committee at the Department of Psychology, Tsinghua University (IRB202138).

Referring to [33]–[35], we adopt a sampled sinusoidal stimulation method to present visual flickers on the LCD monitor. The red-color stimulus target brightness $s(n, f_i, \varphi)$ is expressed as

$$s(n, f_i, \varphi) = \{1 + sin[2\pi f_i(n/R_s) + \varphi]\}/2 \quad (1)$$

where $sin()$ generates a sine wave, n indicates the frame index in the sequence, R_s is the screen refresh rate, φ is the phase, and f_i is the stimulus frequency of the i-th target, $i = 1, 2, 3, 4$.

The frequencies of the four SSVEP targets are designed to flicker at 9 Hz, 11 Hz, 13 Hz, and 15 Hz. The phases of the four SSVEP targets are designed to flicker at 0, 0.5π, π, and $1.5\ \pi$. The four SSVEP targets display at four positions (up, right, down, and left) on the monitor. The experimental design of our dataset is shown in Fig. 1. Each trial began with

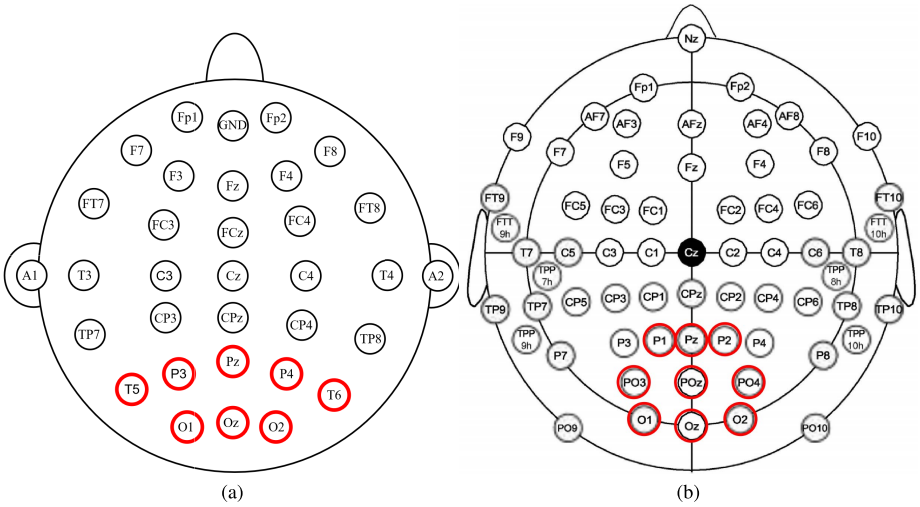

Fig. 2. The position distribution of the selected electrodes on the EEG cap is marked with a red circle. (a) The position distribution of electrodes in our dataset; (b) The position distribution of electrodes in the public dataset [36].

a visual cue for 2 s (preparation), prompting the location of the stimulus target that the subject needs to focus on. Then each red-color stimulus target flickers for 4 s (stimulation), followed by a rest period lasting for 4 s. In the visual cue stage, the stimulus target at the specified location would be displayed in white to facilitate the subjects to find the stimulus target at the specified location. Each stimulus frequency is randomly displayed five times in one block, there is a total of six blocks for each experiment. Therefore, the collected EEG data has 120 trials (4 classes $\times$ 5 times $\times$ 6 blocks). The subject would get enough time for rest after each block. The EEG acquisition device (Neuroscan Inc.) is used to record EEG data at a sampling rate of 1000 Hz. For necessity and simplicity, the EEG data of only eight electrodes are recorded. The reference electrode is Cz. Electrode impedances are kept below 10 kΩ during recording. The channel configuration of the International 10-20 system.

2) Public Dataset: The public dataset published by Lee *et al*. [36] is used in this study. Fifty-four healthy subjects (ages 24-35, 25 females) participated in the experiment. Thirty-eight subjects have never used BCI, while others have previous experience with BCI experiments.

Four SSVEP targets are designed to flicker at 5.45 Hz, 6.67 Hz, 8.57 Hz, and 12 Hz, and display at four positions (up, down, right, and left) on the monitor. Each trial contains 4 s target-notice stage, 4 s stimulus-flickering stage, and 2 s rest stage. Each stimulus target is displayed 25 times. Therefore, the EEG data has 100 trials (4 classes $\times$ 25 times) in the offline training phase ('EEG_SSVEP_train') and another 100 trials in the online test phase ('EEG_SSVEP_test'). There is a total of two sessions ('session01' and 'session02') in the dataset. 'session01' and 'session02' are the two experiments on different days for the same people. The sampling frequency is 1000 Hz. The channel configuration of the International 10-20 system. More details refer to [36].

B. Pre-Processing

The data of some electrodes close to the occipital region is used because of the higher signal-to-noise ratio (SNR) [22]–[26], [28]–[32]. Therefore, eight electrodes (O1, O2, Oz, P3, P4, Pz, T5, and T6) in our dataset and nine electrodes (O1, O2, Oz, POZ, PZ, PO3, PO4, P1, and P2) in the public dataset are selected in this study. The distribution of these electrodes on the EEG cap is shown in Fig. 2. Since the EEG cap used in our dataset is different from the public dataset, different electrodes are selected. To reduce the amount of calculation, the data of these electrode channels are downsampled to 250 Hz. To make the network architecture more effective in learning frequency-related information at the short time-window, the data is filtered.

1) For FB-tCNN: To make better use of the fundamental and harmonic information, the filter bank is designed with different bandpass-range sixth-order Butterworth filters. The design of the filter bank mainly considers two factors, (1) the range of fundamental and harmonic and (2) the SNR of harmonics. Take our dataset as an example to illustrate. The fundamental range is 9-15 Hz, the second harmonic range is 18-30 Hz, and the third harmonic range is 27-45 Hz. Harmonic information above 50 Hz not be used, because the data with relatively low SNR would have negative effects in our study. To keep important information, the bandpass range of the sub-filters is designed to be slightly larger (1-3 Hz) than the harmonic range. Therefore, three sub-filters with different bandpass ranges are designed as 7-17 Hz, 16-32 Hz, and 25-47 Hz respectively. And these three sub-filters make up the filter bank. Similarly, for the public dataset, the filter bank contains 4 bandpass-range sub-filters is designed, which are 3-14 Hz, 9-26 Hz, 14-38 Hz, and 19-50 Hz respectively. A sub-filter should completely contain the information of a specific harmonic of all stimulus targets, and the low SNR data above 50 Hz should be abandoned. Therefore, two datasets with different fundamental ranges generate filter banks with different numbers of sub-filters.

2) For tCNN: To better compare FB-tCNN with tCNN, sixth-order Butterworth filters with frequency ranges of 7-47 Hz and 3-50 Hz are used in our dataset and public dataset respectively.

Considering a latency (delay) in the visual system [37], the data range we use in each trial is $[0.14, 0.14 + L]$ s, which

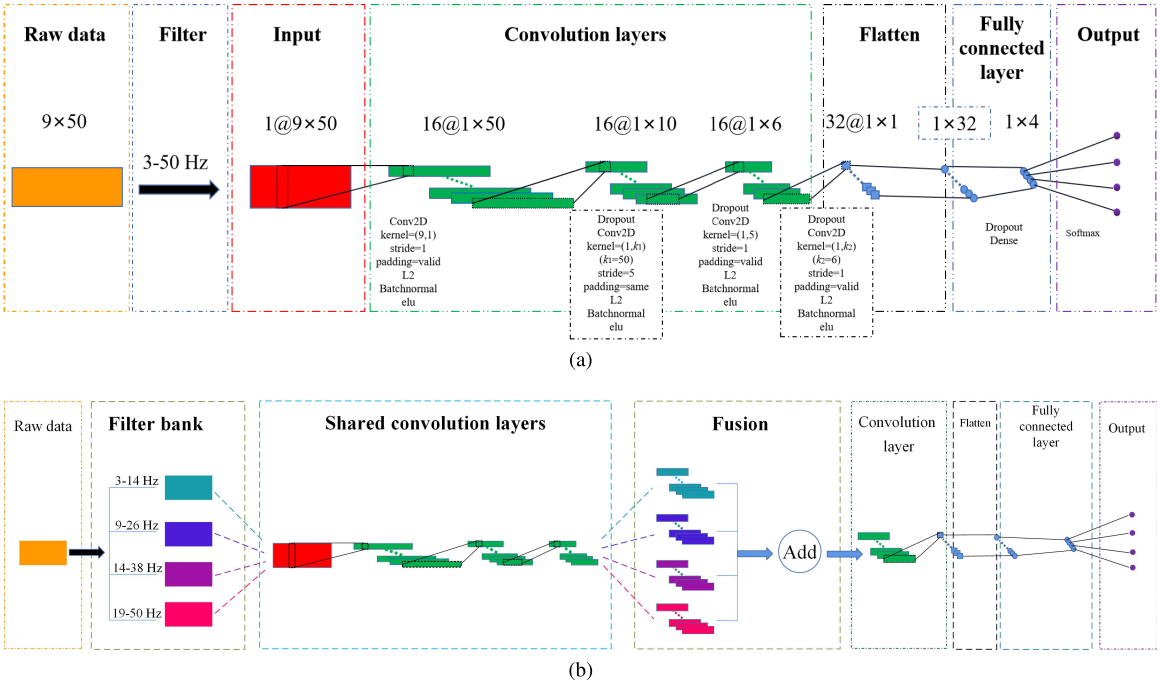

Fig. 3. (a) The network architecture of tCNN, 'Conv2D' means two-dimensional convolution, 'kernel' means convolution kernel size, 'stride' means step size, 'padding = same' means padding is used, 'padding = valid' means padding is not used, 'L2' means L2 regularization, 'Batchnormal' means batch normalization, 'elu' means 'eLU' activation function; (b) The network architecture of FB-tCNN, the network hyper-parameters are the same as tCNN, which will not be described here.

start 0.14 s after the onset time of the stimulus, and L s is the lasted time of the stimulus.

C. SSVEP Classification by FB-tCNN

To better explain the network architecture of FB-tCNN, the network architecture of tCNN is introduced before introducing FB-tCNN. The 0.2 s time-window of the public dataset is taken as an example to illustrate.

1) Network Architecture of tCNN: The network architecture of the tCNN is shown in Fig. 3(a). The input data is the 0.2 s time-window of 9 electrode channels, and the sampling frequency is moderated to 250 Hz by downsampling. So the input size is 9×50 (be reshaped into $1@9 \times 50$). 9 electrode channels contain spatial information, and 50 frames contain time information, so the input can be regarded as spatial-dimension $\times$ time-dimension (9×50). The convolution kernel size of the first convolution is 9×1, the step size is 1 without padding, the number of output channels is 16, and the output is $16@1 \times 50$. The first convolution kernel can make each frame of output contain the spatial information of 9 electrodes. The convolution kernel of the second convolution is $1 \times k_1$, k_1 is the time-frame length of the input (here k_1 is 50), the step size is 5 with padding, the number of output channels is 16, and the output is $16@1 \times 10$. The second convolution kernel can make each frame of output contain enough time information. The convolution kernel of the third convolution is 1×5, the step size is 1 without padding, the number of output channels is 16, and the output is $16@1 \times 6$. The convolution kernel of the fourth convolution is $1 \times k_2$ ($1 \times k_2$ is the size of a channel of the front layer network, here is 1×6), the step size is 1 without padding, the number of output channels is 32, and the output is $32@1 \times 1$. Finally, through flatten (here we only use the flatten to reduce the dimension of the features) and a fully connected layer, we can obtain the scores of four classes through softmax and the class with the highest score is considered the prediction class. All the convolution kernels are one-dimensional in this study. But to express the length direction of convolution kernels more intuitively, the two-dimensional form of convolution kernel is used. For example, the first layer convolution kernel 9×1 and the third layer convolution kernel 1×5 are both one-dimensional convolution kernels. Their convolution kernels are 9 and 5 in size respectively, but the length of the convolution kernel goes in a different direction, which can be simply understood as vertical and horizontal (spatial-dimension and time-dimension).

The model is implemented in Tensorflow [38], using the Keras API [39]. The L2 regularization and dropout parameters used in this study are set to 0.01 and 0.4 respectively.

2) Network Architecture of FB-tCNN: The network architecture of the FB-tCNN is shown in Fig. 3(b). For the network hyper-parameters of FB-tCNN are the same as tCNN, here we only illustrate the network architecture of FB-tCNN simply. The raw data is filtered by four sub-filters with different filter ranges in the filter bank module to obtain four sub-inputs. These four sub-inputs are respectively passed through three convolution layers to obtain four sub-features. And these four sub-inputs' three convolution layers share weights. Next, these four sub-features are fused (added) in the fusion module. Then passed through the fourth convolution layer. After that, the feature dimension is reduced by flattening. Finally, through a fully connected layer, the output is obtained through softmax.

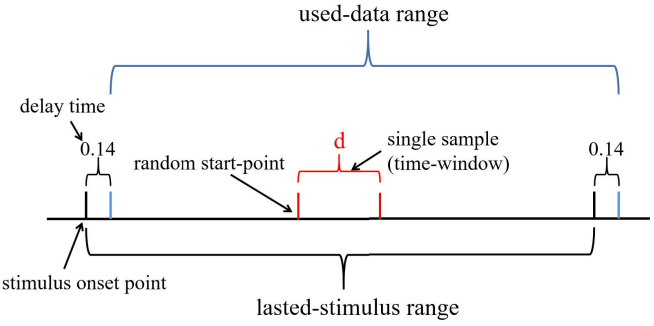

Fig. 4. The random selection process of a single sample in a single trial.

3) Training Process: During the training process, a fixed-length time-window is randomly selected as a single sample for training. The fixed-length time-window range is $[0.14+r,\ 0.14+r+d]$ s, which start $0.14+r$ s after the onset time of the stimulus, r is a random number between $[0,\ L-d]$, L s is the lasted time of the stimulation, and d s is the data length of time-window. The selection process is shown in Fig. 4.

'categorical cross' and 'Adam' are chosen as loss function and optimization algorithm respectively. A model is trained for 4000 iterations with a minibatch size of 256 trials and recorded with the minimum loss of validation set during the training process.

D. SSVEP Classification by CCA-Related Methods

In this section, CCA and CCA-M3 are introduced.

1) CCA: CCA is a standard statistical technique [40], [41], which is used to find the linear combination coefficients (weight vectors) of two variables (matrixes). Using the method of a linear transformation of matrixes $\boldsymbol{X}$ and $\boldsymbol{Y}$ into linear combinations (vectors), CCA seeks to obtain $\boldsymbol{x}=\boldsymbol{w}_x^T\boldsymbol{X}$ and $\boldsymbol{y}=\boldsymbol{w}_y^T\boldsymbol{Y}$, where vectors $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$ are linear combination coefficients of matrixes $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. Then, we maximize the correlation coefficients of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ to obtain $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$. The correlation coefficient ρ is expressed as

$$\begin{aligned}\rho &= \max_{\boldsymbol{w}_x,\boldsymbol{w}_y}\frac{E[\boldsymbol{x}\boldsymbol{y}^T]}{\sqrt{E[\boldsymbol{x}\boldsymbol{x}^T]E[\boldsymbol{y}\boldsymbol{y}^T]}}\\ &= \max_{\boldsymbol{w}_x,\boldsymbol{w}_y}\frac{E[\boldsymbol{w}_x^T\boldsymbol{X}\boldsymbol{Y}^T\boldsymbol{w}_y]}{\sqrt{E[\boldsymbol{w}_x^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{w}_x]E[\boldsymbol{w}_y^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{w}_y]}}\end{aligned}\tag{2}$$

where superscript T denotes the transpose operation. $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$ in Eq. (2) can be solved by singular value decomposition (SVD) or other methods.

In SSVEP recognition [22], $\boldsymbol{X}\in R^{C\times N_s}$ is a matrix composed of multi-channel EEG data, C is the number of electrode channels, N_s indicates the number of sampling points, $N_s=d\times f_s$, d s is the data length of time-window, f_s is the sampling rate. Each row of matrix $\boldsymbol{X}$ is one electrode channel data. Matrix $\boldsymbol{Y}\in R^{(2\cdot N_h)\times N_s}$ is the reference signal (template) composed of the sine-cosine of the stimulus target frequency and its harmonics, N_h is the number of harmonics. Each row of the matrix $\boldsymbol{Y}$ is a sine or cosine signal of length N_s. For the specific stimulus frequency f_i $(i=1,2,\cdots,K,\ K$ is the number of stimulus targets), the reference signal matrix $\boldsymbol{Y}_{(i)}\in R^{(2\cdot N_h)\times N_s}$ is expressed as

$$\boldsymbol{Y}_{(i)}=\begin{pmatrix}sin(2\pi f_i t)\\ cos(2\pi f_i t)\\ \vdots\\ sin(2\pi N_h f_i t)\\ cos(2\pi N_h f_i t)\end{pmatrix},\quad t=\frac{1}{f_s},\frac{2}{f_s},\cdots,\frac{N_s}{f_s}\tag{3}$$

In the process of target recognition, a correlation coefficient ρ_i is calculated between a single-trial testing signal $\boldsymbol{X}$ and each reference signal $\boldsymbol{Y}_{(i)}$. Thus, the frequency f_t of the testing signal is decided as the frequency f_i of the reference signal (template) with the maximum correlation as follows:

$$f_t=\max_{f_i}\rho_i,\quad i=1,2,\cdots,K\tag{4}$$

2) CCA-M3: In 2020, Wei *et al.* [24] proposed a CCA method that creating spatial filters ($\boldsymbol{w}_x$ and $\boldsymbol{w}_y$) using training data only, which is named CCA-M3 in [24]. Assume that the training data has N_r trials for each stimulus target. For a specific target with stimulus frequency f_i $(i=1,2,\cdots,K)$, matrix $\boldsymbol{X}^{(i)}\in R^{C\times(N_s\cdot N_r)}$ is the continuous training signal yielded by concatenating the N_r training trials, $\boldsymbol{X}^{(i)}=[\boldsymbol{X}^{i1},\boldsymbol{X}^{i2},\cdots,\boldsymbol{X}^{iN_r}]$. Matrix $\boldsymbol{X}^{ih}\in R^{C\times N_s}$ is the multi-channel data of a single trial for the specific stimulus target in the training set, $h=1,2,\cdots,N_r$. Matrix $\boldsymbol{Y}^{(i)}\in R^{C\times(N_s\cdot N_r)}$ is the splicing of N_r repetitions of the average of N_r trials for the specific stimulus target, $\boldsymbol{Y}^{(i)}=[\bar{\boldsymbol{X}}^i,\bar{\boldsymbol{X}}^i,\cdots,\bar{\boldsymbol{X}}^i]$, matrix $\bar{\boldsymbol{X}}^i=\frac{1}{N_r}\sum_{h=1}^{N_r}\boldsymbol{X}^{ih}\in R^{C\times N_s}$.

In the training phase, the two spatial filters $\boldsymbol{w}_{xi}$ and $\boldsymbol{w}_{yi}$ for i-th target are estimated by Eq. (2) between the training signal $\boldsymbol{X}^{(i)}$ and the individual template reference signal $\boldsymbol{Y}^{(i)}$.

In the test phase, only the spatial filter $\boldsymbol{w}_{xi}$ is used for subsequent feature extraction. Specifically, for a single-trial testing signal $\tilde{\boldsymbol{X}}\in R^{C\times N_s}$, it is spatially filtered with $\boldsymbol{w}_{xi}$ as $\boldsymbol{x}_i=\boldsymbol{w}_{xi}^T\tilde{\boldsymbol{X}}$. In the same way, the multi-channel template signal $\bar{\boldsymbol{X}}^i$ is spatially filtered as $\boldsymbol{y}_i=\boldsymbol{w}_{xi}^T\bar{\boldsymbol{X}}^i$. Then the correlation between $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ is used as the feature signal for target recognition and K correlation coefficients can be obtained for all the K stimulus frequencies as

$$r_i=corr(\boldsymbol{w}_{xi}^T\tilde{\boldsymbol{X}},\boldsymbol{w}_{xi}^T\bar{\boldsymbol{X}}^i),\quad i=1,2,\cdots,K\tag{5}$$

where $corr(\boldsymbol{a},\boldsymbol{b})$ denotes the Pearson correlation coefficient between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Finally, the stimulus frequency f_t corresponding to the testing trial can be decided as:

$$f_t=\arg\max_i r_i,\quad i=1,2,\cdots,K\tag{6}$$

$\tilde{\boldsymbol{X}}$ and $\boldsymbol{X}^{ih}$ in the same specific-position time-window (such as $[0.14,\ 0.14+d]$ s) is needed for good performance.

The correlation coefficient is directly calculated between the single-trial testing signal and template in CCA. While in CCA-M3, spatial filters are firstly estimated from the training data, and then one of these spatial filters is used to calculate the correlation coefficient between single-trial testing signal and template. It makes CCA-M3 a training data-driven method.

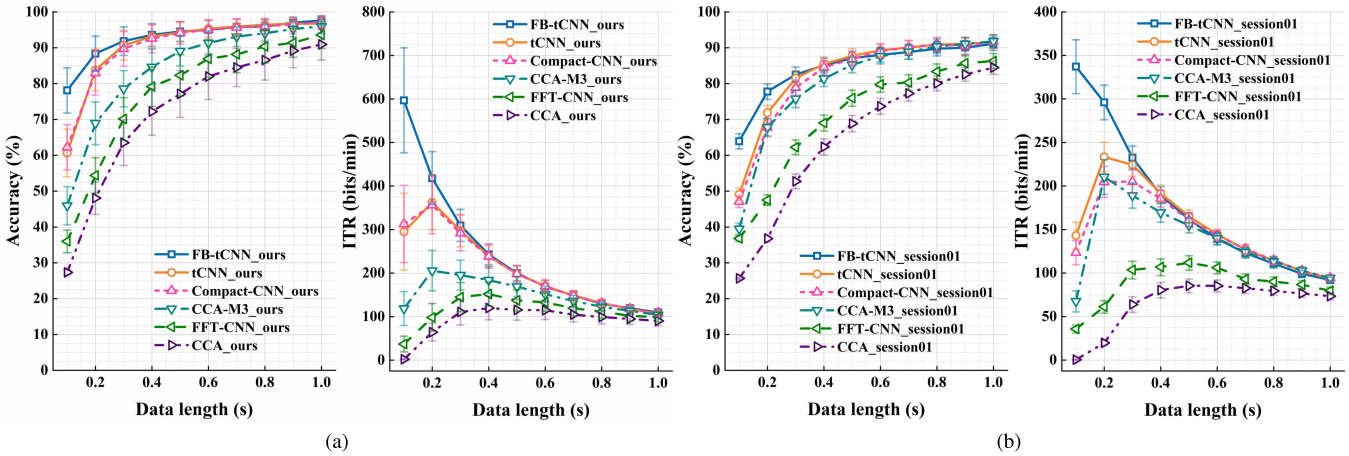

Fig. 5. The average accuracy and ITR across all subjects yielded by these methods at different data lengths. The error bars denote SEM (the standard error of the mean). (a) In our dataset; (b) In the 'session01' of the public dataset.

E. Performance Evaluation

Two parameters are used to evaluate the performance of the BCI: accuracy and information transfer rate (ITR) [42].

In most studies, accuracy has been defined as the ratio of the number of correct samples to the number of total samples, and the accuracy P is expressed as

$$P = m_1/m \tag{7}$$

where m_1 represents the number of correct samples, and m represents the number of total samples. Accuracy can be used to evaluate the performance of the classification algorithm.

For our dataset, leave-one-out cross-validation is used. Using each of the six blocks as a testing set, and the remaining five blocks as training data. The training data is randomly divided into the training set and validation set according to the ratio of 9:1. The classification process repeats six times, and the classification accuracy is taken as the average of the six classification results. The public dataset is divided into 'EEG_SSVEP_train' and 'EEG_SSVEP_test'. To facilitate the comparison of other researchers in this dataset, only the accuracy of the 'EEG_SSVEP_test' is used as the classification accuracy. The 'EEG_SSVEP_train' is randomly divided into the training set and validation set according to the ratio of 9:1, and the entire 'EEG_SSVEP_test' is used as the testing set.

Although accuracy is an important index, the ITR of the BCI system should be also considered in the application. ITR considers the trade-off between accuracy and data length. The ITR (bits/min) can be written as follows:

$$ITR = 60 \times [log_2 K + P log_2 P + (1-P) log_2 \frac{1-P}{K-1}]/d \tag{8}$$

where K is the number of stimulus targets, P is the accuracy, and d is the data length of time-window.

III. Results

In this section, the methods that are compared contain not only CCA [22] and CCA-M3 [24], but also other deep learning methods such as FFT-CNN [29] (the network architectures of [28]–[30] are similar) and Compact-CNN [32]. We also compare FB-tCNN with tCNN to demonstrate the advantages of the filter bank at the short time-window. Specific-position time-window (such as $[0.14, 0.14 + d]$ s) for CCA-M3 and random-position time-window (such as $[0.14 + r, 0.14 + r + d]$ s) for other methods.

A. Impact of the Data Length in the Intra-Individual Test

Fig. 5 illustrates the average accuracy and ITR across all subjects in our dataset and public dataset yielded by these methods at ten data lengths, which ranged from 0.1 to 1.0 s with an interval of 0.1 s. The result of 'session02' is not shown, because it is extremely similar to 'session01'. In Fig. 5, the accuracy of these methods is generally positively correlated with the data length. But for FB-tCNN, tCNN, Compact-CNN, and CCA-M3, the improvement of accuracy is not obvious after 0.6 s. The ability of feature extraction at the short time-window of the four methods may be enough strong, a high accuracy has already been achieved at 0.6 s. By contrast, FFT-CNN and CCA have poor performance at the short time-window. Thus, the accuracy of the two methods can still improve with the increase of data length after 0.6 s. And there is no distinct difference between FB-tCNN and tCNN after 0.4 s. The filter bank has the effect of denoising before 0.4 s, which can improve the ability of SSVEP-related feature extraction. But when the data length is long enough, the SNR increases, one-filter-based tCNN can extract relevant features well enough. This might be the reason why FB-tCNN has no advantage over tCNN after 0.4 s (further analysis in section III-B). It can be noted that FB-tCNN shows the best performance of the ITR in all data sets at the 0.1 s time-window. The maximum ITR reaches 596.86 ± 120.48 bits/min and 337.18 ± 30.92 bits/min, respectively, in our dataset and the 'session01' of the public dataset.

Considering the trade-off between the accuracy and the ITR, Table I illustrates the average accuracy across all the subjects in our dataset and the public dataset yielded by these methods at 0.2 s to show the performance of FB-tCNN more specifically. At 0.2 s, the paired t-test ($tails = 1$) shows that

TABLE I
THE AVERAGE ACCURACY (MEAN ± SEM %) OF THESE METHODS IN OUR DATASET AND PUBLIC DATASET AT 0.2 S

Method	Our dataset	public-session01	public-session02
CCA	48.05 ± 4.53	36.76 ± 1.02	37.23 ± 1.14
FFT-CNN	54.39 ± 4.94	47.49 ± 1.49	48.68 ± 1.62
CCA-M3	68.93 ± 5.96	67.80 ± 2.53	68.22 ± 2.50
Compact-CNN	82.98 ± 6.22	67.79 ± 2.34	69.59 ± 2.14
tCNN	83.83 ± 5.78	71.83 ± 2.07	73.51 ± 1.85
FB-tCNN	**88.36 ± 4.89**	**77.78 ± 2.16**	**79.21 ± 1.80**

the average accuracy of FB-tCNN is significantly higher than other methods (all $p < 0.05$) in the public dataset. And in our dataset, FB-tCNN is significantly higher than CCA-M3, FFT-CNN, and CCA ($p < 0.05$), but not significantly higher than tCNN and Compact-CNN ($p > 0.05$).

B. Inter-Individual Performance

Inter-individual performance is important. The model trained with the existing subjects as the training set is expected to show good performance on the new subjects (unseen subjects). At present, various algorithms for improving inter-individual performance are developed [43]–[45]. For the number of subjects in our dataset is insufficient, the inter-individual performance is studied only in 'session01' of the public dataset.

1) Impact of the Training-Subject Number: Ten subjects (subject 1, 5, 10, 13, 24, 27, 31, 32, 45, 51) are randomly selected as the testing-subjects, then randomly select 5, 10, 15, 20, 25, 30, 35, and 40 subjects from the remaining subjects as the training-subjects (subject 23 and 47 have been removed because of their problematic data). Leave-one-subject-out cross-validation not be used because of enough number of subjects. Three representative data lengths (0.2 s, 0.5 s, and 1.0 s) are selected. Fig. 6 illustrates the average inter-individual accuracy across the selected subjects based on FB-tCNN yielded by three data lengths at the different numbers of training-subjects, which ranged from 5 to 40 with an interval of 5. Table II illustrates the inter-individual accuracy of each selected subject based on FB-tCNN at 0.5 s to show more details. The intra-individual test accuracy is also shown in Fig. 6 and Table II to be convenient to compare. Overall, average inter-individual accuracy is positively correlated with the number of training-subjects at the three data lengths. However, when the number of training-subjects is more than 20, the number seems to have less impact on inter-individual performance. The gap of accuracy between inter-individual and intra-individual would not distinctly decrease with the increase of number after 20 training-subjects. For a specific individual (subject), SSVEP recognition is not only related to the common SSVEP-related features (such as frequency features) from different individuals but also related to individual-related information from the specific individual. This might cause a gap. The importance of individual information in the intra-individual test can be seen in the CCA-related methods. Some CCA-related methods [23]–[25] introduce the individual

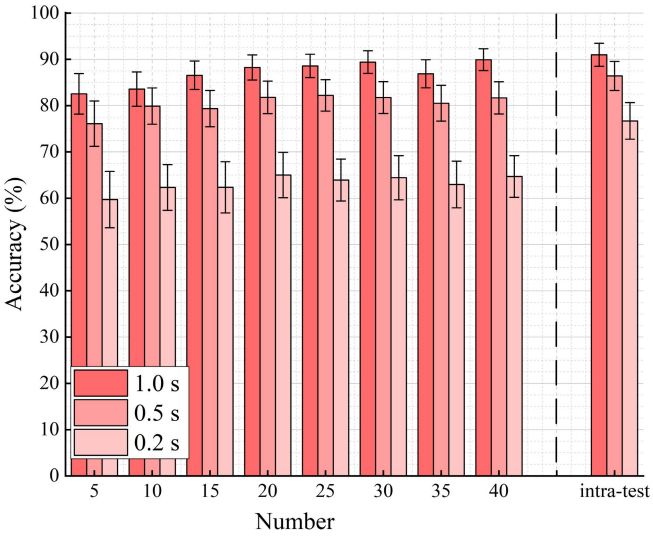

Fig. 6. The average inter-individual accuracy across the selected subjects based on FB-tCNN is yielded by the three data lengths at the different numbers of training-subjects. The average intra-individual accuracy is shown at the right. The error bars denote SEM.

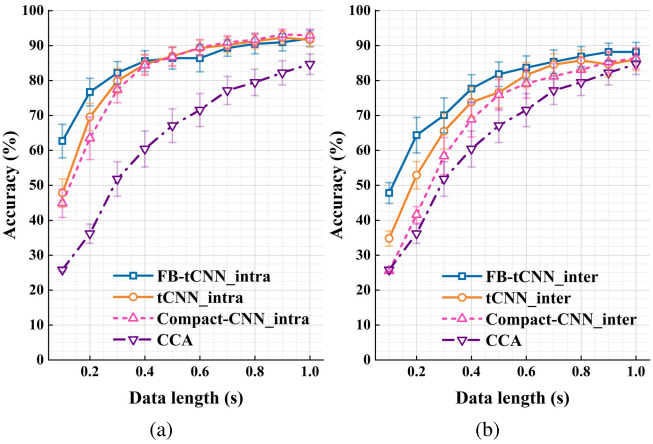

Fig. 7. The average accuracy across the selected subjects is yielded by these methods at different data lengths. The error bars denote SEM. (a) Intra-individual test; (b) Inter-individual test.

information to improve the performance in the intra-individual test for a specific individual (for example, CCA-M3 performs better than CCA in section III-A). The gap can decrease with the increase of data length. It might be because the common SSVEP-related features become more apparent as the data length increases.

2) Impact of the Data Length: The inter-individual performance of the selected 10 subjects for deep learning methods based on 20 training-subjects is studied in this section. CCA is used as the baseline for its training-free, FFT-CNN is removed for its bad performance at the short time-window. Fig. 7 illustrates the average intra-individual and inter-individual accuracy across the selected subjects yielded by FB-tCNN, tCNN, Compact-CNN, and CCA at ten data lengths. The paired t-test ($tails = 2$) is used between intra-individual and inter-individual tests of the selected subjects for each method (except CCA for its training-free), and Fig. 8 illustrates the corresponding p-value across the selected subjects

TABLE II
THE INTER-INDIVIDUAL AND INTRA-INDIVIDUAL ACCURACY (%) OF EACH SELECTED SUBJECT BASED ON FB-TCNN AT 0.5 S WITH DIFFERENT NUMBERS OF TRAINING-SUBJECTS

Subject	5	10	15	20	25	30	35	40	Intra-test
1	73.38	70.74	65.85	74.29	69.76	71.52	71.00	70.94	76.44
5	84.42	83.35	83.94	83.28	82.08	83.74	84.16	83.97	77.19
10	56.44	65.40	72.30	69.73	75.40	69.63	67.32	72.40	88.14
13	61.19	68.07	74.36	78.66	78.89	79.11	77.22	79.31	93.97
24	69.76	70.77	69.24	74.03	75.14	72.99	69.60	72.66	82.37
27	99.02	99.09	99.28	99.51	99.45	99.38	99.22	98.89	97.59
31	96.12	98.14	97.95	98.04	98.50	98.47	98.99	98.60	99.35
32	90.75	90.29	88.69	91.53	91.53	89.08	89.74	90.78	95.99
45	69.05	79.73	75.07	79.54	80.32	81.04	80.35	78.20	81.13
51	60.83	73.35	66.93	69.31	70.97	72.47	67.61	71.00	71.98
Mean	**76.10**	**79.89**	**79.36**	**81.79**	**82.20**	**81.74**	**80.52**	**81.67**	**86.42**
SEM	**4.89**	**3.92**	**3.92**	**3.51**	**3.40**	**3.45**	**3.87**	**3.47**	**3.13**

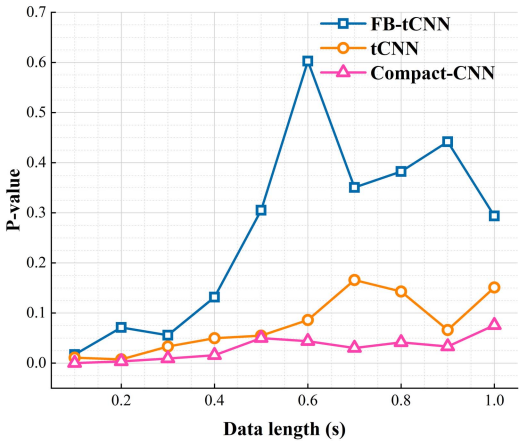

Fig. 8. The p-value between intra-individual and inter-individual tests across the selected subjects is yielded by the three methods at different data lengths.

yielded by these methods at ten data lengths. Fig. 7(a) is similar to Fig. 5(b), FB-tCNN has no advantage over tCNN and Compact-CNN after 0.4 s. But in Fig. 7(b), FB-tCNN performs better than other methods at all data lengths. In Fig. 8, the p-value of FB-tCNN is bigger than tCNN and Compact-CNN, which means that the difference between intra-individual and inter-individual tests of FB-tCNN is smaller than tCNN and Compact-CNN. It seems that FB-tCNN tends to extract more common SSVEP-related features and less individual-related information than tCNN and Compact-CNN in the intra-individual test. It might demonstrate that for each sub-filter of the filter bank, the frequency information of a fundamental or harmonic gets more attention than individual-related information. Part individual-related information is lost for each narrower sub-filter of the filter bank. While one-filter-based tCNN can better extract the features of the individual-related information. When the SNR increases with the increase of the time-window, the missing individual-related information will have a relatively large impact. This might be a reason why FB-tCNN has no advantage after 0.4 s in the intra-individual test.

Generally, FB-tCNN can perform well with enough training-subjects and data length. For example, using 20 training-subjects at 0.5 s, the inter-individual accuracy and ITR reach $81.79 \pm 3.51\%$ and 131.51 ± 18.31 bits/min respectively.

IV. DISCUSSION

In this section, 1 s is used as the data length of the fixed-length time-window to show the performance of CCA-M3 better. The start-point range of the fixed-length time-window is $[0.14, 0.14 + L - d]$ s, which starts 0.14 s after the onset time of the stimulus, L s is the lasted time of the stimulus, and d s is the data length of the fixed-length time-window.

'f1', 'f2', 'f3', and 'f4' represent stimulus targets with frequencies of 12 Hz, 8.57 Hz, 6.67 Hz, and 5.45 Hz, respectively.

A. Asynchronous BCI Analyses

According to the different processing methods of input data and the different forms of interaction, BCI systems can be divided into synchronous and asynchronous. Synchronous BCI means that the system has a fixed-position time-window mechanism (such as $[0.14, 0.14 + d]$ s). The system can only process and decode the signal in the specific time-window, while the information outside this time-window is considered meaningless, and thus ignored by the system. While the asynchronous BCI does not require the fixed-position time-window (such as $[0.14 + r, 0.14 + r + d]$ s).

The CCA-M3 algorithm has a limitation of the time-window alignment. It only performs well at the specific-position time-window. It can be difficult to achieve an asynchronous BCI system with CCA-M3. The fixed-length time-window moving curve of CCA-M3 is shown in Fig. 9(a). The results are obtained by moving a fixed-length time-window over a trial frame by frame. The frequency of the fixed-length time-window data is decided as the frequency of the stimulus target frequencies with maximum correlation coefficient. It can be seen that with the movement of the fixed-length time-window, the correlation coefficient of correct-class-'f2' is not always bigger than that of other classes. To be more precise, their correlation coefficient changes periodically, which is the

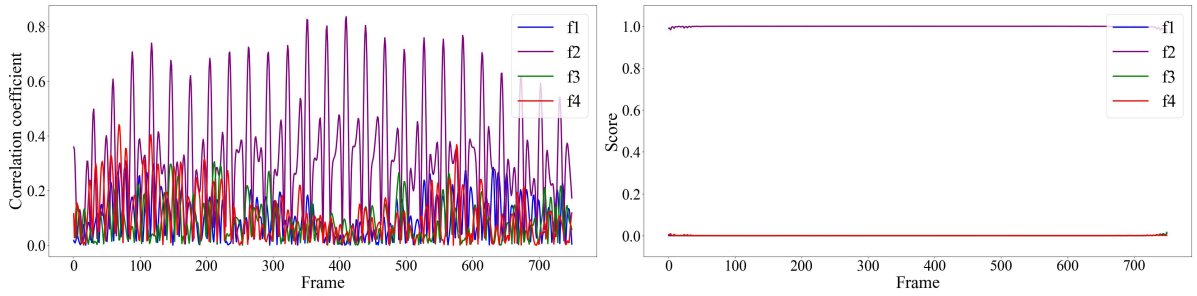

Fig. 9. (a) CCA-M3 based frame-by-frame recognition of a trial, the y-axis represents the correlation coefficient; (b) FB-tCNN based frame-by-frame recognition of a trial, the y-axis represents the score. The x-axis represents the number of frames (250 frames = 1 s), 'f2' is the correct class.

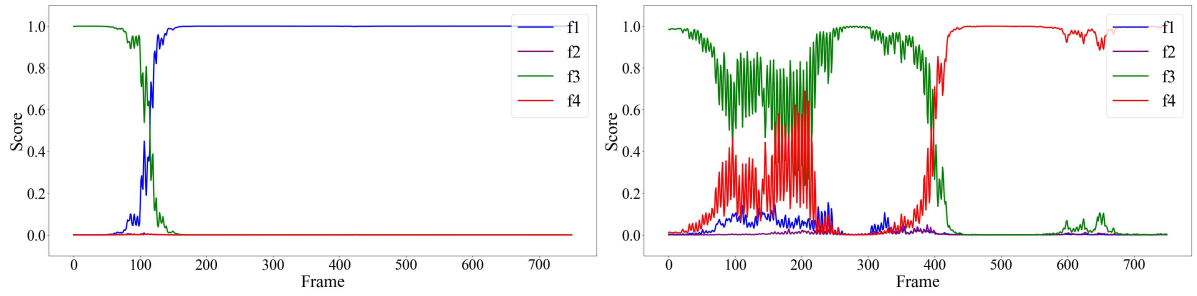

Fig. 10. FB-tCNN based frame-by-frame recognition of some trials. (a) 'f1' is the correct class; (b) 'f4' is the correct class.

problem of time-window alignment. While deep learning has no such limitation. Fig. 9(b) is a frame-by-frame recognition diagram of the same trial based on FB-tCNN. The scores of four classes are calculated by using the model proposed in Section II (score is the output of softmax). The higher the score of a class is, the higher the probability of the class is. The score of class-'f2' is close to 1, while the other three classes are close to 0. The score of class-'f2' is always higher than that of other classes, which is very stable compared with CCA-M3. Therefore, FB-tCNN is easier in terms of the implementation of the asynchronous BCI system than CCA-M3.

B. Error Label Analyses

In the field of BCI, good algorithms can certainly improve the recognition accuracy, but good datasets also play a decisive role. It is very difficult to collect high-quality data in the BCI field. The low-quality data limits the development of the algorithm. Sometimes the error of recognition comes from the problem of the data, instead of the algorithm. For some misclassified samples, we propose two possibilities.

1) The algorithm is not generalizable enough to deal with noisy signals.
2) These subjects did not concentrate when collecting data. They did not pay attention to the stimulus target corresponding to the prompt (label) on the monitor in time, the label is inconsistent with the stimulus target that the subject is staring at, resulting in the error label.

Some trials that contain misclassified samples are recognized frame by frame based on FB-tCNN. The results of two of these trials are shown in Fig. 10. The recognition of the fixed-length time-window is wrong at the beginning but finally became correct. The situation might not arise from the classification inaccuracy of the algorithm, but the problem of the data collection process. During the experiment, it is possible that the subjects pay attention to other stimulus targets (the stimulus target does not correspond to the prompt) or did not concentrate on their tasks fully. It can be said that there are some error labels in the data.

V. Conclusion

In this study, we propose a time-domain-based CNN (tCNN) and use a filter bank to improve its performance at the short time-window (FB-tCNN). We compare FB-tCNN with other methods at different data lengths in the intra-individual test and FB-tCNN shows the significant advantage at 0.1 s, 0.2 s, and 0.3 s. Performance on inter-individual is also studied. FB-tCNN shows the potential to implement inter-individual BCI. And we find there is always a gap of performance between inter-individual and intra-individual might because of the lack of individual-related information, intra-individual test might have overfitted the individual-related information. Then we discuss whether CCA-M3 or deep learning methods (FB-tCNN as an example) can better implement the asynchronous BCI system, and draw the conclusion that it is easier to implement the asynchronous BCI system with deep learning. Finally, some trials that contain misclassified samples are recognized frame by frame. We find that some recognition errors might not be caused by the algorithm, but the subjects' poor attention while collecting data. Some labels are not consistent with the actual prompt.

In future work, how to find samples with error labels and removing (or correcting) these are important to improve the algorithm performance. And our method can be used to build an asynchronous SSVEP-BCI control system, which may lead to a faster and more flexible human-computer interaction system.

REFERENCES

[1] J. J. Vidal, "Toward direct brain-computer communication," *Annu. Rev. Biophys. Bioeng.*, vol. 2, no. 1, pp. 157–180, Jun. 1973.

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

[3] G. Dornhege *et al.*, *Toward Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, 2007.

[4] D. J. Krusienski and J. R. Wolpaw, "Brain–computer interface research at the Wadsworth Center: Developments in noninvasive communication and control," *Int. Rev. Neurobiol.*, vol. 86, pp. 147–157, Mar. 2009.

[5] F. Vogel, "The genetic basis of the normal human electroencephalogram (EEG)," *Humangenetik*, vol. 10, no. 2, pp. 91–114, 1970.

[6] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, p. R1, 2007.

[7] J. P. Cunningham, P. Nuyujukian, V. Gilja, C. A. Chestek, S. I. Ryu, and K. V. Shenoy, "A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces," *J. Neurophysiol.*, vol. 105, no. 4, pp. 1932–1949, 2011.

[8] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, "Affective EEG-based person identification using the deep learning approach," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 3, pp. 486–496, Sep. 2020.

[9] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.

[10] B. J. Edelman, B. Baxter, and B. He, "EEG source imaging enhances the decoding of complex right-hand motor imagery tasks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 4–14, Jan. 2016.

[11] J. Cao, J. Zhu, W. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked CNNs," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 4, pp. 709–722, Dec. 2020.

[12] Y. Li *et al.*, "A novel Bi-hemispheric discrepancy model for EEG emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, vol. 13, no. 2, pp. 354–367, Jun. 2021.

[13] A. Khasnobish, A. Konar, D. N. Tibarewala, and A. K. Nagar, "Bypassing the natural visual-motor pathway to execute complex movement related tasks using interval type-2 fuzzy sets," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 1, pp. 91–105, Jan. 2017.

[14] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 566–576, Jun. 2017.

[15] C.-R. Phang and L.-W. Ko, "Intralobular and interlobular parietal functional network correlated to MI-BCI performance," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2671–2680, Dec. 2020.

[16] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: Assessing the speed of a P300-based brain-computer interface," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 174–179, Jun. 2000.

[17] T. Bhattacharjee, R. Kar, A. Konar, A. Lekova, and A. K. Nagar, "A general type-2 fuzzy set induced single trial P300 detection," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2017, pp. 1–6.

[18] A. Rakshit, A. Konar, and A. K. Nagar, "A hybrid brain–computer interface for closed-loop position control of a robot arm," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 5, pp. 1344–1360, Sep. 2020.

[19] G. R. K. Kumar and M. R. Reddy, "Designing a sum of squared correlations framework for enhancing SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2044–2050, Oct. 2019.

[20] X. Mao, W. Li, H. Hu, J. Jin, and G. Chen, "Improve the classification efficiency of high-frequency phase-tagged SSVEP by a recursive Bayesian-based approach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 561–572, Mar. 2020.

[21] Y. Zhang, P. Xu, T. Liu, J. Hu, R. Zhang, and D. Yao, "Multiple frequencies sequential coding for SSVEP-based brain–computer interface," *PLoS ONE*, vol. 7, no. 3, Mar. 2012, Art. no. e29519.

[22] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2610–2614, Dec. 2006.

[23] G. Bin, X. Gao, Y. Wang, Y. Li, B. Hong, and S. Gao, "A high-speed BCI based on code modulation VEP," *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025015.

[24] 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. 5, May 2020, Art. no. 2050020.

[25] X. Chen *et al.*, "High-speed spelling with a noninvasive brain–computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 44, pp. E6058–E6067, 2015.

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

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[28] N. S. Kwak, K. R. Müller, and S. W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS ONE*, vol. 12, no. 2, 2017, Art. no. e0172578.

[29] X. Zhang *et al.*, "A convolutional neural network for the detection of asynchronous steady state motion visual evoked potential," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1303–1311, Jun. 2019.

[30] A. Ravi, N. H. Beni, J. Manuel, and N. Jiang, "Comparing user-dependent and user-independent training of CNN for SSVEP BCI," *J. Neural Eng.*, vol. 17, no. 2, Apr. 2020, Art. no. 026028.

[31] N. K. Nik Aznan, S. Bonner, J. Connolly, N. Al Moubayed, and T. Breckon, "On the classification of SSVEP-based dry-EEG signals via convolutional neural networks," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 3726–3731.

[32] N. Waytowich *et al.*, "Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials," *J. Neural Eng.*, vol. 15, no. 6, Dec. 2018, Art. no. 066031.

[33] N. V. Manyakov *et al.*, "Sampled sinusoidal stimulation profile and multichannel fuzzy logic classification for monitor-based phase-coded SSVEP brain–computer interfacing," *J. Neural Eng.*, vol. 10, no. 3, 2013, Art. no. 036011.

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

[35] 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, Nov. 2016.

[36] M.-H. Lee *et al.*, "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, May 2019, giz002.

[37] F. Di Russo and D. Spinelli, "Electrophysiological evidence for an early attentional mechanism in visual processing in humans," *Vis. Res.*, vol. 39, no. 18, pp. 2975–2985, Sep. 1999.

[38] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Conf. Operating Syst. Design Implement.* Berkeley, CA, USA: USENIX Association, 2016, pp. 265–283.

[39] F. Chollet. 2015. *Keras*. [Online]. Available: https://github.com/keras-team/keras

[40] H. Hotelling, "Relations between two sets of variates," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 162–190.

[41] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York, NY, USA: Wiley, 1962.

[42] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: Improved accuracy by response verification," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 326–333, Sep. 1998.

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

[44] P. Yuan *et al.*, "Enhancing performances of SSVEP-based brain–computer interfaces via exploiting inter-subject information," *J. Neural Eng.*, vol. 12, no. 4, 2015, Art. no. 046006.

[45] K.-J. Chiang, C.-S. Wei, M. Nakanishi, and T.-P. Jung, "Cross-subject transfer learning improves the practicality of real-world applications of brain–computer interfaces," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Mar. 2019, pp. 424–427.