IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 28, 2018, accepted December 8, 2018, date of publication December 14, 2018,
date of current version January 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2886759

A Single-Channel SSVEP-Based BCI Speller Using Deep Learning

TRUNG-HAU NGUYEN AND WAN-YOUNG CHUNG, (Senior Member, IEEE)
Department of Electronic Engineering, Pukyong National University, Busan 48513, South Korea

Corresponding author: Wan-Young Chung (wychung@pknu.ac.kr)

This work was supported by the Mid-Career Researcher Program through an NRF grant, Korean Government (MSIT), under Grant NRF-2016R1A2B4015818.

ABSTRACT This paper aims to develop a speller system based on a bipolar single-channel electroencephalogram with sufficient accuracy. The proposed system consists of a custom-designed headset, a new virtual keyboard with 58 characters, special symbols, and digits, and a five-target steady-state visual-evoked potential (SSVEP)-based brain–computer interface (BCI) utilizing one-dimensional convolutional neural network (1-D CNN) for SSVEP frequency detection. The deep learning model is implemented and trained under the training mode before being applied in the operation mode of the system. To validate the proposed model, we acquire the training dataset with numerous testing conditions, including different frequency resolutions of the feature and different time-window lengths of analysis. Two types of features based on the frequency domain are investigated to compare their performances in terms of classification accuracy of the model. The experimental results from eight subjects shows that on average, the proposed model can classify five-class SSVEP data with a high accuracy of 99.2%. The proposed BCI is then employed in an online experiment of spelling the word "SPELLER" using 2-s time window. Consequently, the system achieves an average accuracy of 97.4% and an information transfer rate of 49 ± 7.7 bpm, showing the practicality and feasibility of implementing a reliable single-channel SSVEP-based speller utilizing 1-D CNN.

INDEX TERMS Brain–computer interface (BCI), electroencephalogram (EEG), bipolar single channel, speller, one-dimensional convolutional neural network (1-D CNN), steady-state visual evoked potential (SSVEP).

I. INTRODUCTION

Brain–computer interface (BCI) has recently set up a new path way for direct communication between human and machine in addition to conventional communication. BCI plays an important role in assisting people with disabilities to enable them to communicate effectively with the whole community [1]–[3]. The common electroencephalogram (EEG) signals used in a BCI-based system are event-related synchronization/desynchronization (ERS/ERD), event-related potential (ERP), and steady-state visual evoked potential (SSVEP) [4]–[7]. SSVEP is a periodic response of the brain to a periodic visual stimulus modulated at a frequency higher than 6 Hz [8]. In a SSVEP-based BCI, the stimulus is used to induce an SSVEP response on the scalp, which oscillates at the same frequency of the stimulus and its higher harmonics [9]. Most studies verified that the strongest SSVEP can be observed in the visual cortex [10]–[12]. SSVEP has been generally used because of its advantages, including high signal-to-noise ratio (SNR) and high information transfer rate (ITR). Moreover, a SSVEP-based system requires little training time or even no training process [13]. Therefore, SSVEP has attracted more attention and has been exploited in many studies on BCI systems in the past few years [8], [9], [13]–[16]. Most of these studies mainly focused on the usage of a multi-channel headset to develop BCI systems with high accuracy [13], [14], [17]. However, in real-life applications, a multi-channel system could be an inefficient device because of its complicated setup. This study presents a simple and convenient BCI based on a single channel SSVEP signal which might increase the usability as well as reduce the complexity of the system while maintaining the wearing comfort over time. Two dry electrodes were used to form a bipolar montage to acquire the EEG signal from the O1–Oz pair and improve the SNR. These two electrodes were attached to a 3D-printed holder, which was custom designed to improve the contacting quality between the electrodes and the scalp. The prototype was totally fabricated through a 3D printing technique with a flexible material, thereby allowing to achieve a highly suitable wearability.

Many recent studies have employed multivariate statistical analysis including canonical correlation analysis (CCA) and least absolute shrinkage and selection operator (LASSO) as classifiers to detect SSVEP frequencies [14], [17], [18]. The CCA-based approach has proven its powerful performance in the cases of multi-channel-based applications by exploring the correlation between two sets of variables. Following those works, modified CCA-based classifiers have been introduced [19], [20] in which the performance of those models have been significantly improved. Meanwhile, the power spectral density analysis (PSDA) is known as a conventional approach most widely used in detecting SSVEP frequencies. Nevertheless, PSDA has drawbacks because of its sensitivity to noise resulting in a low accuracy in the SSVEP frequency detection. The CCA and LASSO-based approaches generally achieve higher recognition accuracy compared to the PSDA-based technique [14], [17], [18]. Recently, deep learning has been considered to explore the feasibility of implementing a SSVEP-based BCI. One significant benefit of neural network is that it does not strictly require feature extraction before processing compared to other machine learning techniques (i.e., linear discriminant analysis (LDA), Support Vector Machine (SVM), k-nearest neighbor (KNN), CNN) in which the feature extraction play a critical role to contribute to the performance of the model. One of the earliest applications of deep learning in a SSVEP-based BCI is presented in [21]. In that work, the network is designed to functions a spatial and time filter in the first two hidden layers, a signal transformation in the frequency domain in the next two hidden layers, and a classifier in the output layer. In another work shown in [22], a CNN model has been introduced as a robust classifier of the SSVEP frequencies. The work acquires a 2-D map (channels x frequencies) of SSVEP data as the input to classify up to five SSVEP frequencies using a multi-channel EEG headset. In this study, we present a novel 1-D CNN-based approach applied to the frequency recognition strategy to improve the accuracy of the current single channel-based BCI. The current work takes 1-D Fast Fourier transform (FFT)-based data as the input data instead of 2-D maps as usual, consequently could reduce the computational time of the system.

Aside from accuracy, the ITR is also an important factor in evaluating a SSVEP-based speller. Many virtual keyboard designs have been developed in the past few years. For instance, Bremen-BCI GUI, which is composed of 32 characters (i.e.., letters and special symbols), could offer a high spelling speed of 18.09 bpm on average. This interface requires five stimuli (i.e., corresponding to the commands "left," "right," "up," "down," and "select") for operation [23]. In another work [13], an ITR of 37.62 bpm was achieved with a keyboard consisting of 27 characters using five stimuli. Meanwhile, a boosting in spelling speed of up to 267 bpm shown in [11], in which a novel SSVEP decoding method based on phase and harmonics was introduced. In that work, 40 characters were displayed on a screen with 40 different corresponding flickering frequencies. On the other hand, a comparable high spelling speed of 41.08 bpm was achieved in [24], with up to 36 items on the interface using six SSVEP frequencies. Albeit offering a high ITR, the complicated interface associated with many stimuli could be a challenge to the user. The current study introduces a new keyboard design that offers high-speed spelling with less needed stimuli. This keyboard can offer up to 58 characters, which mostly include all the characters of a standard keyboard. The current user interface requires five stimuli corresponding to five SSVEP commands for operation.

The contribution of this work is to develop an efficient and feasible BCI in several aspects, including a simple headset design, a robust CNN-based classifier for frequency recognition, and a new spelling layout.

II. METHODS

A. EEG HEADSET DESIGN

Two dry sensors from Cognionics, Inc. (Santa Fe, San Diego, CA, USA) are used to build a single-channel headset. Compared to a wet sensor, the dry sensor has some advantages, including the exclusion of conductive gels or glues needed during operation, being easily attached to the brain scalp through the hair, and being re-usable for many times. Fig. 1(a) shows the block diagram of the proposed headset. The current headset adopted a bipolar montage to acquire the EEG signal from the O1–Oz pair because of its better noise cancellation ability compared to a unipolar montage [25]. The potential from each electrode is first passed to an active-shield circuit to minimize the electromagnetic interference from the external sources. This active circuit serves as a unity gain buffer, which help convert the high output impedance of the dry sensor to the low output impedance of the active circuit. A detailed design of the active circuit can be found in [26]. The reference voltage is set to 1.65 V, which was half the value of an analog-to-digital converter (ADC) voltage span (i.e., 3.3 V), to center the signal around it. The differential potential between the two electrodes is then fed into a differential amplifier (INA128, Texas Instrument, Texas, USA). This differential amplifier has a high common mode rejection ratio (i.e., ~100 dB) for

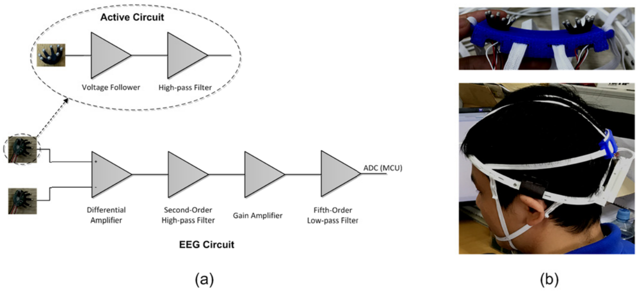

FIGURE 1. **EEG headset: (a) schematic of the EEG circuit with dry active electrodes (top), (b) Photograph of the active electrodes with the flexible cap (top) and headset (bottom).**

better line–noise rejection. Its gain is set to approximately 10. The differential voltage signals are first high-pass-filtered at 4 Hz to remove the DC offset and the low-frequency noises using an active high-pass filter. The output voltage of this stage is mainly amplified at the gain of approximately 3,000 to ensure that the output voltage matches the input range of the 12-bit ADC (i.e., 3.3 V span and 0.8 mV resolution) of the microcontroller. An active low-pass filter with a corner frequency of 40 Hz is applied to limit the frequency band of interest (i.e., 4–40 Hz) and ensure that the Nyquist sampling rate theory was satisfied. Moreover, the effect of the 50/60 Hz line noise is minimized because the proposed headset was powered by a 3.7 V battery. Finally, the analog signal is sampled at a 128 Hz sampling rate by a 12-bit ADC (STM32F103CB microcontroller, STMicroelectronics, Geneva, Switzerland) to convert the analog voltage to digital prior to further processes.

A completely wearable headset is fabricated based on the 3D printing technique using a flexible material to minimize the preparation time for use and improve the system reliability. The model is first designed using SolidWorks software (SolidWorks Corp., Massachusetts, USA). The G-code of the model generated by the MakerBot MakerWare software is then transferred to a 5th generation 3D printer (MakerBot, New York, USA) for fabrication. The printing process took almost 2 h. Fig. 1(b) shows a photograph of the proposed wearable headset. It is noted that the proposed headset was typically equipped with several elastic ropes and a flexible material-made electrode holder useful to maintain a good contact between the electrodes and the brain scalp through the hair.

B. VISUAL STIMULATOR AND VIRTUAL KEYBOARD DESIGN

The stimulus frequencies normally used in the SSVEP can be divided into three frequency bands: low (i.e., 1–12 Hz), medium (12–30 Hz), and high (30–60 Hz). The strongest SSVEP was particularly observed at approximately 10 Hz followed by 16–18 Hz [27]. Another study showed that the peak of the SSVEP amplitude occurred near 15 Hz in the 5–25 Hz range [8]. Therefore, in this study, five frequencies (i.e., 6.67, 7.5, 8.57, 10, and 12 Hz) in the lower range were selected as stimulus frequencies. These frequencies belonged to the alpha band of the brain waves, resulting in their equal interferences with this band. A 60 Hz-refresh rate LCD monitor is employed as the visual stimulator that consisted of five boxes, called stimuli, which flick at 6.67, 7.5, 8.57, 10, and 12 Hz corresponding to 9, 8, 7, 6, and 5 frames per period of the monitor, respectively (Fig. 2(a)). One box labeled ‘‘Undo’’ is used to cancel the previous action when it comes with a wrong command. The stimulus ‘‘Undo’’ is located at the top-left corner of the screen, while the other four were located at the top, right, bottom, and bottom left corners of the screen. Each stimulus is a simple square of 6 cm × 6 cm, which toggled the color between black and red. The toggling cycle determines the flickering frequency of the stimulus. The color toggling between black and red is done using the square function as a transparent factor with a period that changed its value at the stimulus frequency:

$$s\left(f_i\right) = square(2\pi f_i t), \tag{1}$$

where f_i is the stimulus frequency i^{th} ($i = 1, \ldots, 5$); t is the period of flickering; and $s(f_i)$ is the transparent factor at frequency f_i.

The flickering frequencies are selected as the integer factors of the refresh rate of the screen (i.e., 60 Hz) because of stability and precision [1]. A red light is chosen as the stimulation light source because of its strong SSVEP response in the low and medium frequency bands [2]. The stimulation application is developed in MATLAB using the Psychtoolbox toolbox.

Recent studies have shown various virtual keyboard designs [13], [28]–[30], which exhibited good performance with high typing speed and a friendly user interface. The present study designs a new virtual keyboard to combine the strengths of the previous works. Thus, the keyboard can provide high-speed spelling, a user-friendly interface, and less number of required stimuli. Fig. 2 illustrates the proposed interface, where the characters and the visual stimuli were merged. The interface is composed of three layers (i.e., first, second, and third layers). Each layer corresponds to a menu with four choices. Fig. 2(a) depicts the first layer of the virtual keyboard. Once a choice in this layer is selected, its content is split into four new choices that belong to the second layer. Each choice is a stimulation box embedded with 16 characters. The top-middle box contains letters ‘‘A’’ to ‘‘P’’ of the English alphabet. The right box contains letters ‘‘Q’’ to ‘‘Z’’ and six special symbols. The bottom-middle box contains other 16 special symbols, while the bottom-left box consists of 10 digits. Each choice of the menu in the second layer (Fig. 2(b)) contains four characters. When a choice is selected, its content is split into four new choices that belong to the third layer (Fig. 2(c)). Meanwhile, each choice in the third layer is composed of a single character. The selection of a character is done when a choice is selected in this layer. Once the character is selected, the system automatically turns back to the first layer for the next spelling process. After each command, an animation is generated by changing the color of the choice from red to green as a visual feedback to the user. Furthermore, the status of the text in comparison with the target text is automatically updated after each trial and displayed on the top-right corner of the interface. The stimulation application in combination with the keyboard layout was implemented in MATLAB (The MathWorks, Inc., Massachusetts, USA) using the Psychtoolbox toolbox.

C. DATA ACQUISITION OF OPTICAL POWER SIGNALS DURING VISUAL STIMULATION

Fig. 3(a) shows the experimental setup for measuring the optical power of the stimulation light. An optical sensor (918d-sl-od3, Newport, California, USA) was placed at a

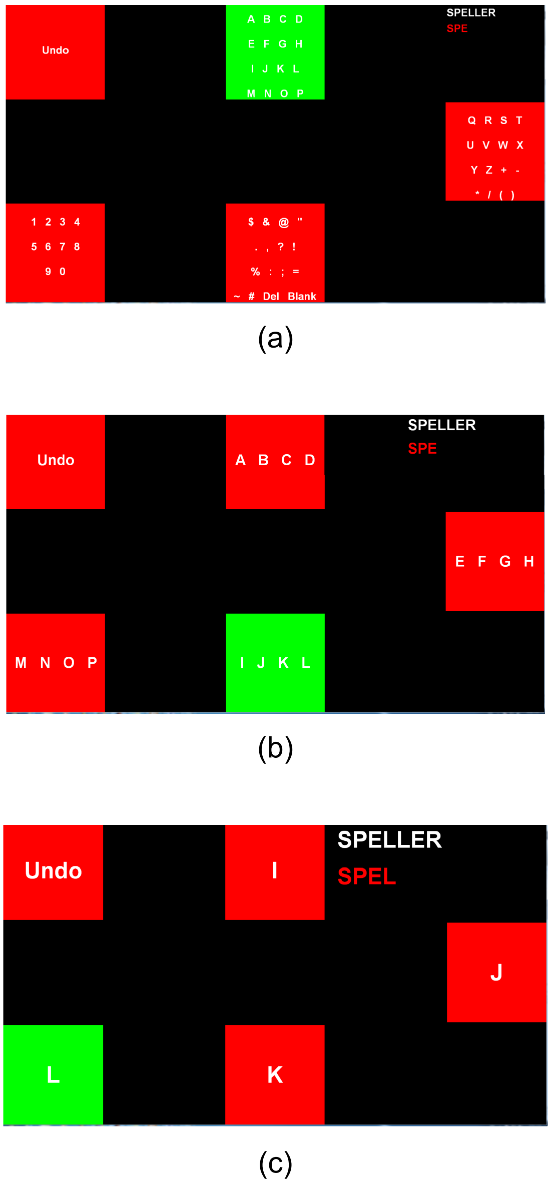

FIGURE 2. Demonstration of the 58-character virtual keyboard in combination with five stimulation boxes (6 cm x 6 cm). (a) First layer. (b) Second layer. (c) Third layer of the interface. Note that the top right corner of the GUI indicates the target word and the current status of the spelling process.

5 mm distance in front of the LCD screen, where the on-screen stimuli generated the optical power. The sensor combined with a power meter (1918-R, Newport, California, USA) continuously acquires the stimulus-induced optical power. This sensor is sampled at 200 Hz, which is twice higher than the highest frequency of interest (i.e., the frequency band of the SSVEP ranges from 6 Hz to 16 Hz). The data from the power meter is then transferred to a computer via a universal serial bus for further analyses. The measurement is conducted for all five stimuli to verify their flickering frequencies. All the experiments are implemented in a dim room to minimize the ambient light noise.

D. EXPERIMENTAL DESIGN AND SETUP

Offline experiment was conducted to verify the optimal conditions of the system and to collect data for training the CNN model prior to the online experiment. Eight healthy volunteers, who have no problem with visual impairment, participated in this study (i.e., six males and two females with an age range of 24 to 32 years). The subjects were seated on a comfortable chair at a distance of approximately 40 cm in front of the visual stimulator in a dim room (Fig. 3(b)). In order to avoid any random error and to maintain the consistence of the system, the experiment was design to be consisted of 20 trials. In each trial, the participants were asked to gaze on five stimuli (i.e., five choices of the menu) in a pre-defined order. Moreover, they were given a 0.5-s interval to shift their gaze between two consecutive targets. The gazing time of the subjects on each target was set to 10 s. During the experiments, EEG signals were recorded and further processed for training step. All five targets keep flickering until the end of each trial. To prevent the subjects from visual fatigue, 2-min break was given after each trial.

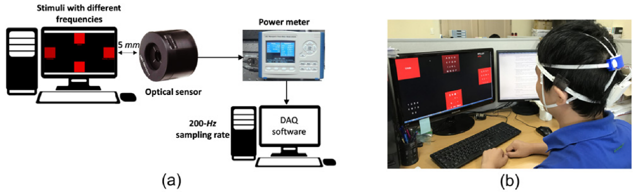

FIGURE 3. Experimental set-up. (a) Measurement of stimulus-induced optical power. (b) The spelling process.

In the online experiment, the subjects were asked to input the desired word "SPELLER" shown on the screen. They were instructed to focus their gaze on the stimuli resulting in inducing corresponding commands. The subjects were given 0.5 s to shift their gaze between two stimuli. A set of commands issued from the proposed system help the subject to navigate the "path" leading to the desired characters. The user must produce three commands to write a character. This number of commands is equal to the number of layers of the interface fixed and independent of the character. Thus, 21 commands must be produced to finish spelling the word "SPELLER." The number of correct counts out of the 21 commands is used to evaluate the system accuracy. Once an incorrect command is generated, the user can immediately correct it by choosing the button "Undo" of the menu to bring the user interface back to previous state. The ITR is also calculated as another factor for evaluating the performance of the proposed BCI. The ITR is defined as follows according to [7]:

$$ITR = \frac{n}{T} * \left(P\log_2 P + (1-P)\log_2\left(\frac{1-P}{N-1}\right) + \log_2 N\right), \tag{2}$$

where P is the probability to correctly generate a command; N is the number of stimuli ($N = 5$); and T (in minute) is the time needed to produce n commands.

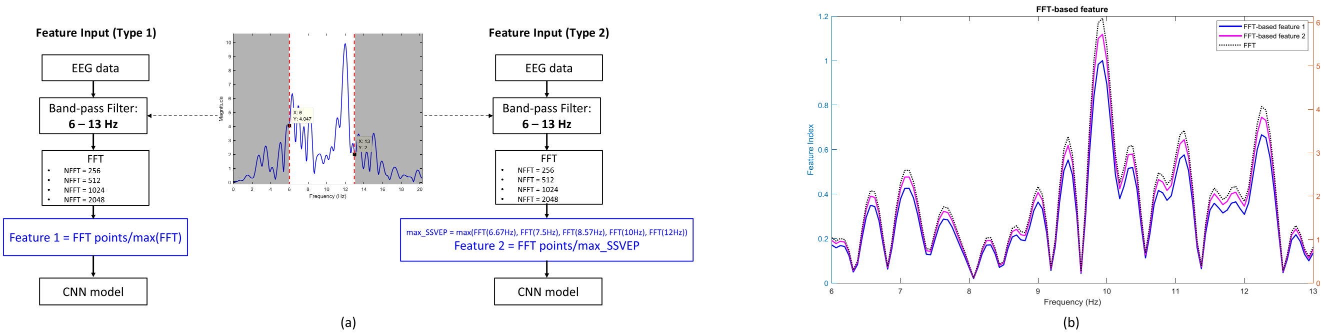

FIGURE 4. FFT-based feature: (a) extraction procedure of two type of features, (b) feature visualization of 10 Hz SSVEP signals.

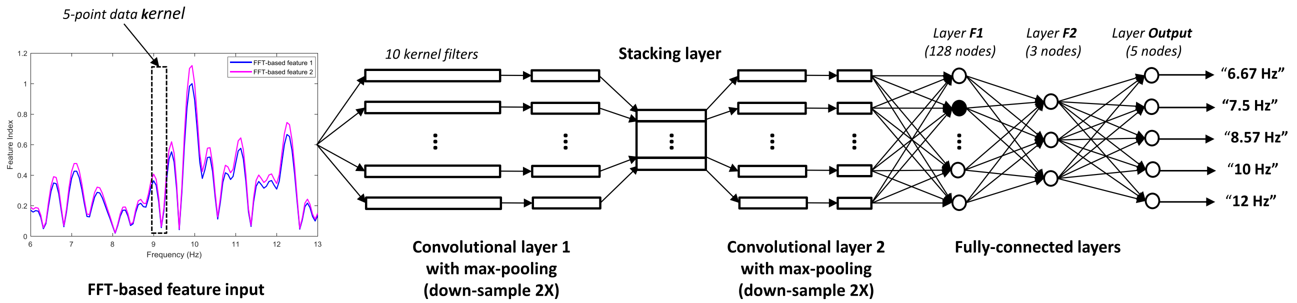

FIGURE 5. The architecture of the proposed 1-D CNN model.

E. FFT-BASED FEATURE EXTRACTION OF EEG SIGNALS

The current study applies a bipolar montage to record the EEG signal from the O1–Oz pair at a sampling rate of 128 Hz. The EEG data are first band pass-filtered from 6 to 13 Hz to explore the characteristics of the fundamental component of the SSVEP signals. FFT is performed with various number of points (i.e., 256, 512, 1024 and 2048) resulting in different frequency resolutions of 0.5, 0.25, 0.125 and 0.0625 Hz in the frequency domain, respectively. Moreover, a Hamming window of a similar length to the time window length is applied to improve the signal SNR. To evaluate the system performance with respect to analyzed window of the signal, three kinds of time windows (i.e., 1-s, 2-s, and 3-s EEG data) are explored. Two kinds of FFT-based features are introduced in the current work. Fig. 4(a) shows the block diagram of feature implementation. The first feature (''Feature 1'') is computed by divided all FFT power points over the range 6-13 Hz by the maximum value in this range. As a result, all magnitude values of "Feature 1" are normalized between 0 and 1 as feature index vector. To compute the second feature ("Feature 2"), the highest FFT power among five SSVEP frequencies is picked (called "max_SSVEP"). Then the feature index vector of "Feature 2" are obtained by divided all FFT power points over the range 6-13 Hz by the "max_SSVEP." The combinations of these two kinds of features with different time window lengths are investigated to find out the optimal conditions for the proposed system. In the next step, the computed features are fed into the CNN model for training. Fig. 4(b) demonstrates the feature index curves of above-mentioned features along with the Fourier transform of 2-s EEG signal utilizing 2048-point NFFT.

F. CNN STRUCTURE AND SYSTEM BLOCK DIAGRAM

Fig. 5 illustrates the 1-D CNN structure being applied in SSVEP frequency detection. In general, the proposed CNN model consists of three main layers: input layer, hidden layer and output layer. The input layer comprises two convolutional layers named "convolutional layer 1" and "convolutional layer 2." It is input with FFT-based feature of windowed EEG data using 5-point kernel filter. In total, ten kernel filters are used. To reduce computational time, a max-pooling layers is applied right after each convolutional layer. In the hidden layer, two fully connected layers are included. The first fully connected layer consists of 128 neurons. To prevent the network from overfitting, dropping-out units with the rate of 25% are used. The second fully connected layer comprises three nodes. ReLu is used in the first fully connected layer as activation function while tanh/ softplus is used in the second one. There are five nodes involving five SSVEP frequencies in the output layer. In the current work, softmax is applied for each node in the output layer. As a result, an input signal is predicted to be belonged to a class such that its mapping output value obtained highest among five outputs of output layer. The general form of softmax function is given by

$$f_j(z) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \quad j = 1, \ldots, K. \tag{3}$$

The function takes a real-valued input vector z and maps it to a vector of real values in the range (0, 1).

In conjunction with the softmax classifier, cross-entropy is chosen as the loss function to evaluate the quality of the neural network. Cross-entropy loss can be expressed as follows:

$$L\left(X,Y\right)=-\frac{1}{n}\sum\nolimits_{i=1}^{n}\left[y^{(i)}\ln a\left(x^{(i)}\right)\right.\\ \left.+\left(1-y^{(i)}\right)\ln\left(1-a(x^{(i)})\right)\right],\quad(4)$$

where $X=\left\{x^{(1)},\ldots,x^{(n)}\right\}$ is the set of input samples in the training dataset, and $Y=\left\{y^{(1)},\ldots,y^{(n)}\right\}$ is the corresponding set of labels. The $a(x)$ represents the output of the neural network given input x.

To obtain better convergence rate of the network, Adam optimization algorithm is used among with a classic stochastic gradient descent (SGD) algorithm as network weight update rule.

The network is trained during training mode of the proposed system. Five-fold cross validation is conducted to statistically evaluate performance of the proposed model as well as preventing the model from overfitting [31]. Average error and accuracy of 5-fold cross validation during testing phase is evaluated.

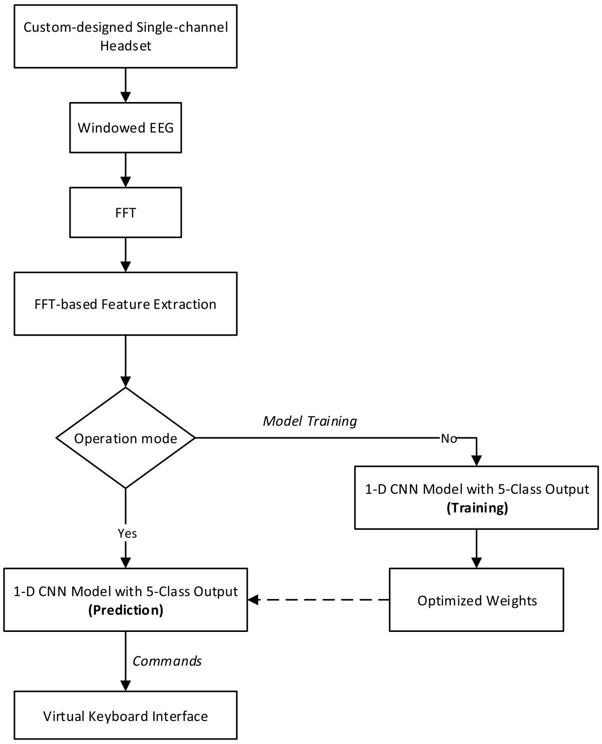
Custom-designed Single-channel Headset
Windowed EEG
FFT
FFT-based Feature Extraction
Operation mode
Model Training
No
1-D CNN Model with 5-Class Output (Training)
Yes
Optimized Weights
1-D CNN Model with 5-Class Output (Prediction)
Commands
Virtual Keyboard Interface

FIGURE 6. **Block Diagram of the proposed BCI speller based on 1-D CNN.**

The block diagram of the proposed spelling system, as shown in Fig. 6, consists of training mode and operation mode. In the training mode, the CNN model is trained with given dataset to obtain the optimized weight set. First, the captured EEG signals are segmented with different time window lengths (i.e. 1 s, 2 s, 3 s) by sliding the windows along the signal. The overlapping samples between two consecutive windows are set to 16 samples (i.e. 0.125-s data). Next, the windowed EEG data are band-pass filtered between 6 and 13 Hz. The, discrete Fourier transform is performed on the filtered EEG data. Finally, frequency domain-based feature extraction is conducted. Depend upon difference setting conditions, the number of training data samples can be slightly different. For example, with the settings of 2-s time window and 0.125 s overlapping time interval, the numbers of training data samples of class 1 (“6.67 Hz”), class 2 (“7.5 Hz”), class 3 (“8.57 Hz”), class 4 (“10 Hz”), and class 5 (“12 Hz”) for subject 1 were 1584, 1,584, 1,584, 1,584, and 1,424, respectively. In total, there were 7,760 data samples being used for model training. Among them, the 5-fold cross validation takes 1552 data samples for validation of the model in each fold. After successfully trained with the given dataset, the well-trained model is applied in the operation mode as prediction machine for the system. Each predicted output of the CNN model is associated with one of five SSVEP frequencies. The command consequently, generated by the CNN model is used to operate the virtual keyboard module.

G. FREQUENCY RECOGNITION USING CCA AND LASSO

To evaluate the performance of the current CNN-based model, we compare its detection accuracy with those obtained from some state-of-art methods including CCA and LASSO. The current study adopts CCA and LASSO techniques which presented in [17] and [18], respectively. The dataset consists of 20 trials. In each trial, for each stimulus frequency, 4-s single-channel EEG data are extracted to investigate the effect of time window duration on the recognition accuracy. The first two harmonics of the frequency spectrum of EEG signals are taken into account during frequency recognition.

III. RESULTS

A. STIMULATION RESPONSE

Fig. 7 shows an example of optical stimulation signal at 10 Hz and its SSVEP response. Fig. 7(a) illustrates the

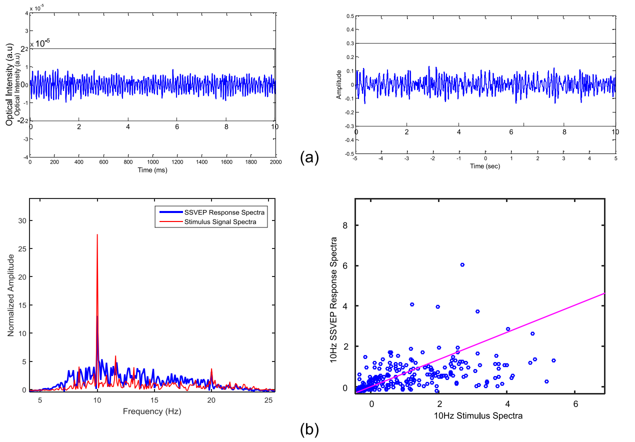

FIGURE 7. **(a) Stimulus-induced optical power signal at 10 Hz and its 10-s SSVEP response from the O1–Oz channel. (b) Power spectra of those signals and the correlation measurement between them.**

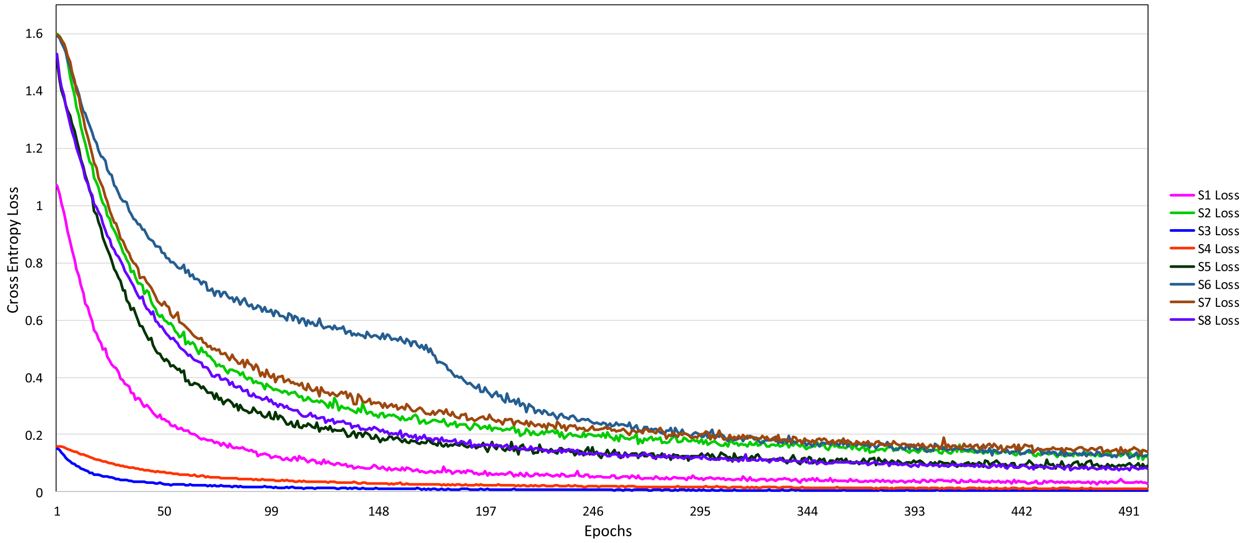

FIGURE 8. Learning curve of 5-fold cross validation for eight subjects in 500 training epochs.

optical power of the stimulation light from the stimulus at the left portion of the figure and its SSVEP response waveform from the O1-Oz bipolar channel (the right portion of the figure). Both signals are band pass-filtered from 5 to 20 Hz, which covered the SSVEP frequency band being used in this study, to reject the 60 Hz component from the refresh rate of the screen. Frequency content of both signals discovered by discrete Fourier transform is shown in Fig. 7(b). The PSD peak (“red” curve) occurs at 10 Hz (Fig. 7(b) (left)), which definitely confirms the stimulation frequency. On the other hand, the “blue” curve represents a typical response in frequency domain of the brain to its 10-Hz optical stimulation. Its Fourier transform indicates that a PSD peak is also obtained at 10 Hz. A measurement of linear correlation between these two signals in frequency domain is shown in Fig. 7(b) (right). The correlation coefficient is calculated to be 0.7 which typically confirmed a comparable correlation between two signals. Moreover, as demonstrated in Fig. 7(b) (left), the major correlation is resulted from the signals at 10 Hz (1st harmonic) and 20 Hz (2^{sd} harmonic). The abovementioned results verified the high precision visual stimuli using the Psychtoolbox and their clear SSVEP responses.

B. CNN MODEL LEARNING

Learning process of the CNN model is shown in Fig. 8 in terms of cross entropy loss during testing phase for 500 iterations across eight subjects. It is noted that these results are obtained given the following setting of the CNN model: 2-s windowed input data, 2048-point NFFT in combination with feature type 2. Apparently, for all subjects, the model fits the data at highest rate in the first 100 epochs of training. Then, the validation loss is approaching zero as it become saturated at around epoch 500 showing the convergence of the model. Subject 3 and 4 achieve best average convergence rate among eight participants. The lowest convergence rate occurs in the case of subject 6. In addition, this learning curve shows an abrupt change in convergence rate at epoch 170 consequently it can keep pace with other learning curves within the next 100 epochs. All learning curves obtain a cross entropy loss of lower than 0.2 after 300 epochs as the model get converged. These learning curves (i.e., cross entropy loss) reflect the trends of the corresponding validation accuracies which were exhibited in Fig. 9. As a results, the lower cross entropy loss the higher accuracy is obtained. For all subjects, the validation accuracy curves are approaching 100% with different converging rates due to different learning rates of the model. Although with a bad learning rate at starting (the first 170 epochs), the accuracy for subject 6 is slightly higher than that of subject 7 after 500 training epochs (97.28% vs. 97.11%, respectively).

C. FEATURED-DATA 3-D SCATTER PLOT

As mentioned earlier, the fully connected layer F2 is composed of three nodes. Thus, in order to observe the convergence of the model during training, 3-D data associated with these three nodes are visualized in 3-D coordinate in which each node output is according to an axis direction. Fig. 10 illustrates 3-D scatter plot of three output node data at layer F2 after 500 training epochs. With softplus is chosen as activation function at layer F2, the 3-D scatter data are distributed around the origin of the coordinate (Fig. 10(a)). Moreover, data points which belonged to a specific class gather in a densely volume and well distinguished from those of other classes. On the other hand, Fig. 10(b) shows the 3-D data distribution of three node

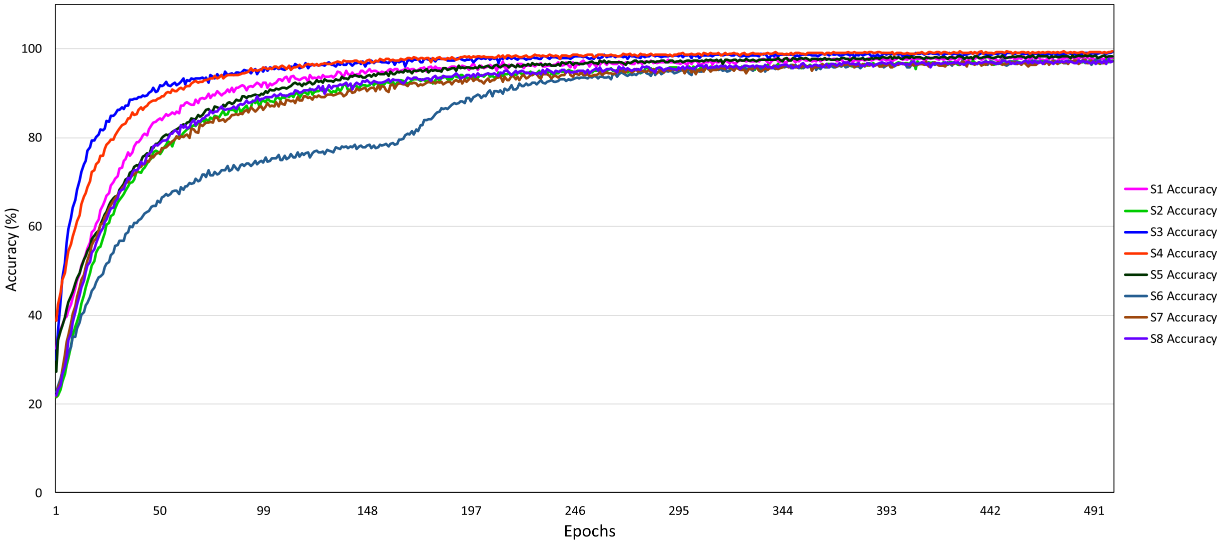

FIGURE 9. Average accuracy of 5-fold cross validation for eight subjects in 500 training epochs.

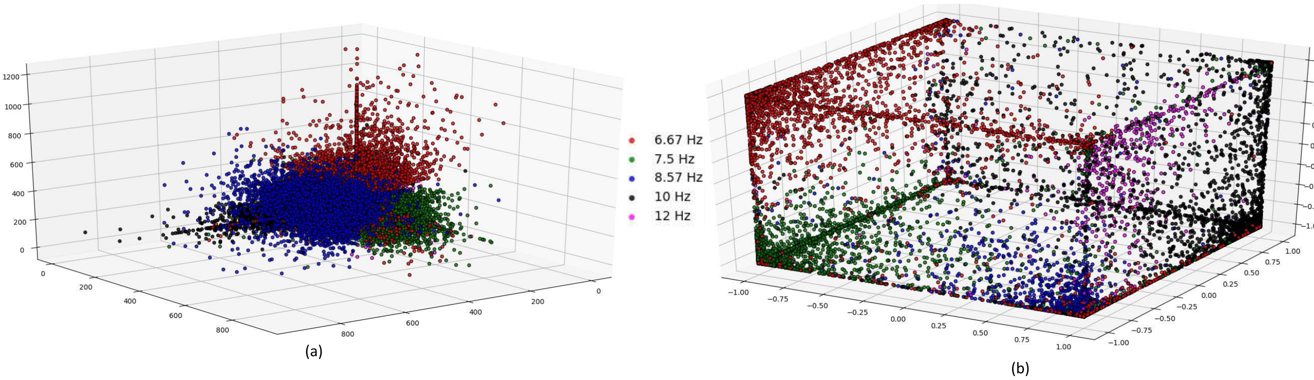

FIGURE 10. 3-D scatter visualization of featured data at layer F2, after 500 training epochs: (a) with softplus activation function, (b) with tanh activation function.

outputs in the case of tanh activation is selected in this layer. Unlike the softplus-induced features, the tanh-induced features distributed in a 2x2x2 cube. Each group among five-group data are mainly distributed in a specific corner of the cube. Although, there are conflict data samples between two different groups, in general, all classes are distinguished nicely. It is noted that, for both cases, "6.67 Hz," "7.5 Hz," "8.57 Hz," "10 Hz," and "12 Hz" data samples are represented with red, green, blue, black, and pink colors, respectively. Apparently, the proposed CNN model is able to distinguish different SSVEP frequencies as it can extract the meaningful information from the frequency spectral of the signal.

D. CNN MODEL EVALUATION

Fig. 11 compares the performance of the model between two types of features with respect to time windows and frequency resolutions. Obviously, the test accuracy of the CNN model increases as the frequency resolution and the analyzed time window increase for both kinds of features. Five-fold cross validation results show that "Feature 1" and "Feature 2"-based models achieved their highest accuracies in the case of 3-s time window and 0.0625-Hz resolution (i.e., NFFT is equal to 2048) at 99.5% and 99.8%, respectively. Likewise, the lowest accuracy of 39.6% and 45.2% (i.e., "Feature 1" and "Feature 2"-based models, respectively) are obtained in the case of 1-s time window and 0.5-Hz resolution (i.e., NFFT is equal to 256). An accuracy above 95% occurs in the cases of 2-s and 3-s time windows in conjunction with 0.125-Hz and 0.0625-Hz frequencies for both models. Another finding is that less variance of the accuracy is obtained with higher frequency resolution for all investigating time windows. For instance, with an exception of 0.5-Hz spectra case, the standard deviation of the accuracy of all other cases are lower than 6%. "Feature 2"-based model outperforms the "Feature 1"-based model in terms of accuracy for all cases except for the case of 1-s time window and 0.25-Hz resolution (i.e., NFFT is equal to 512). This

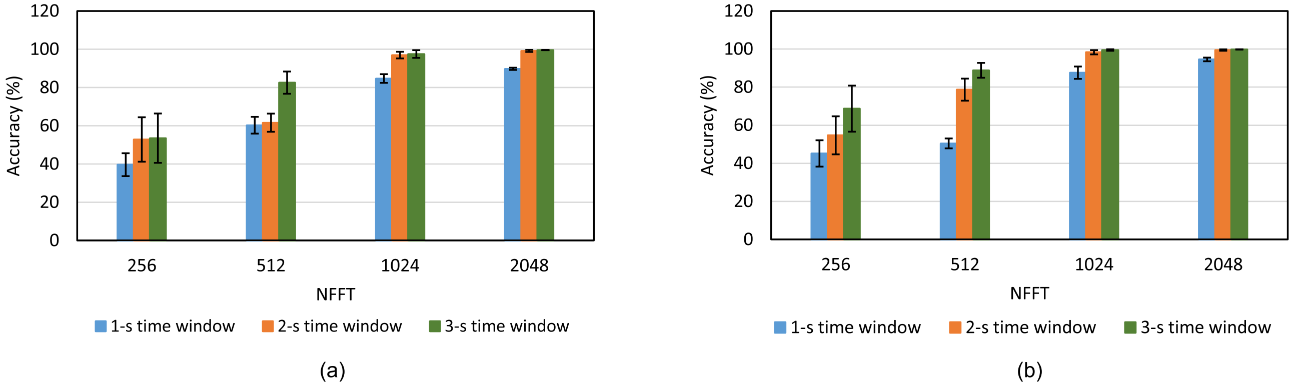

FIGURE 11. Five-fold cross validation of the 1-D CNN model for subject 1 with respect to different numbers of FFT point and different time windows, utilizing: (a) "Feature 1," (b) "Feature 2."

is reasonable due to the less reliability of 1-s time window application.

According to equation (2), detection accuracy and time-window of analysis are two main contributed factors to the ITR. The shorter detection time and higher detection accuracy, the higher ITR can be achieved. Considering the information transfer rate of the proposed system, we pick 2-s time window and 0.0625-Hz frequency resolution in conjunction with "Feature 2" as the best setting parameters of the proposed SSVEP-based virtual keyboard.

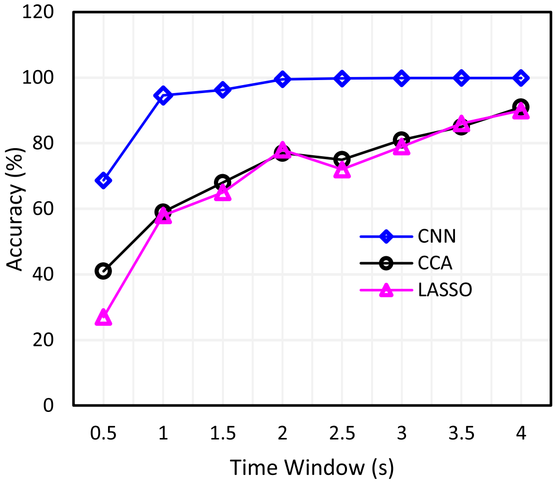

FIGURE 12. Comparison of detection accuracy of five-class SSVEP data among three detection methods CCA, LASSO and CNN with respect to time window length. Note that in the case of CNN classifier, "Feature 2" utilizing 2048-point FFT was used and the model was trained for 1000 iterations.

Fig. 12 compares the detection accuracy of five-class SSVEP data among three methods including CCA, LASSO, and the proposed CNN. Apparently, the proposed CNN-based classifier outperforms CCA and LASSO techniques in terms of SSVEP frequency detection accuracy for all time window lengths. There is no significant difference of accuracy between CCA and LASSO techniques (i.e. CCA obtains higher detection accuracy for most of time window lengths except for the cases of 2 and 3.5 s). In general, the detection accuracy of SSVEP frequencies increases as the time window length increases for all three classification methods (except for the case of 2.5-s time window for CCA and LASSO techniques). The CCN-based classifier achieves its highest detection accuracy of 99.87% for 3.5 s and 4 s time windows whereas CCA and LASSO methods achieve their highest accuracy of 91% and 90%, respectively, in the case of 4-s time window.

Fig. 13 compares the 5-fold cross-validation results between two types of features for eight subjects after 1000 training epochs. It is noted that the test model is based on 2-s windowed signal and with highest investigating resolution (0.0625 Hz). For all subjects, "Feature 2" obtains its higher accuracy in comparison to "Feature 1." An accuracy above 95% was obtained for all subjects. For "Feature 2" application, the model achieves its highest accuracy (99.7 ± 0.2%) in the case of subject 3. In contrast, the lowest accuracy occurs in the case of subject 7 (98.4 ± 0.4%).

Table 1 lists the online testing results of the proposed SSVEP-based virtual keyboard across eight subjects. The average accuracy of the proposed speller (based on the ratio of number of correct commands to total commands being generated by the system to input the text "SSVEP") over all the subjects was 97.36 ± 2.86%. All subjects achieved an accuracy over 95% except for subject 6 (92.00%). Moreover, subjects 2, 4, 5 and 8 were successfully input the text "SPELLER" without any mistake (i.e., 21 correct commands out of total 21 commands need to be generated). The proposed system sent a correct control command in an average of 2.5 s and achieved an average ITR of 48.99 ± 7.67 bit/min. Moreover, participants completed their typing the word "SPELLER" for an average of 0.92 ± 0.06 min, thereby resulting to the mean spelling speed of 7.6 letters/min.

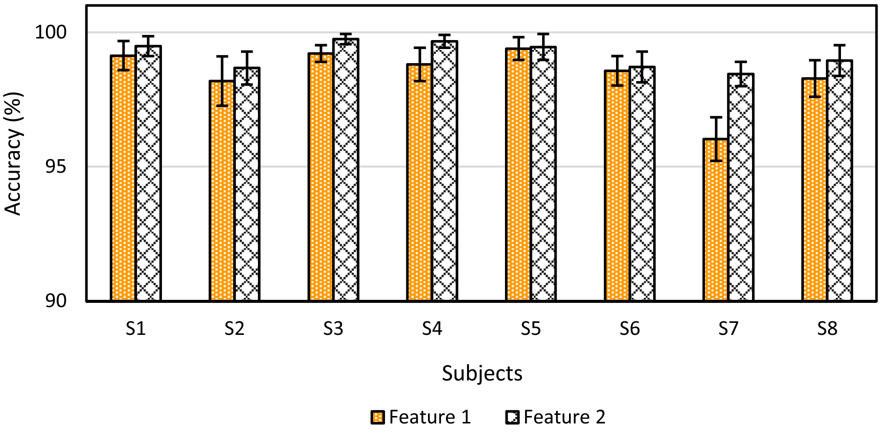

FIGURE 13. Comparison of performance of proposed CNN classifier between two features using 5-fold cross validation for eight subjects, after 1000 training epochs.

TABLE 1. Online results.

Subjects	No. of commands	Complete time (min)	Accuracy (%)	ITR (bit/min)
S1	23	0.95	95.65	43.70
S2	21	0.87	100.00	56.26
S3	23	0.95	95.65	43.70
S4	21	0.87	100.00	56.26
S5	21	0.87	100.00	56.26
S6	25	1.03	92.00	35.76
S7	23	0.95	95.65	43.70
S8	21	0.87	100.00	56.26
Mean ± STD	22.25 ± 1.39	0.92 ± 0.06	97.37 ± 2.86	48.99 ± 7.7

IV. DISCUSSION AND CONCLUSION

BCIs based on multi-channel EEG signals could provide high accuracy, but requires a complicated set-up prior to use, which could be a challenge in real-world applications.

The current study aimed to develop a practicable and simple spelling BCI considering real-life feasibility based on the SSVEP signals. Thus, the headset was designed to be a single dry-electrode with only one channel (bipolar O1-Oz) that acquired the SSVEP signals from the visual cortex. The proposed headset was composed of several separated parts fabricated using the 3D printing technique with a flexible material. Thus, it allowed a fast self-applicability of the cap and electrodes. The new virtual keyboard design was also introduced along with a visual stimulator.

The current work, tending to simplify the BCI in real-work applications, used only one channel; hence, a novel SSVEP frequency detection approach based on 1-D CNN was introduced. The proposed system was divided into two modes which are training mode and operation mode. The CNN model is trained during training mode before it can be employed to predict the SSVEP frequencies in the operation mode. Numerous testing experiments were conducted to find out the optimal setting parameters of the model which can be applied in the real-time operation mode. Particularly, two kinds of FFT-based features have been investigated along with different frequency resolutions and different lengths of time windows of the EEG signals. From the analysis, considering the reliability of the system, the fast speed of spelling and the high accuracy, “Feature 2” in conjunction with 0.0625-Hz frequency resolution and 2-s time window were picked as the optimized setting parameters for the system.

To evaluate the effectiveness of the proposed CNN-based method, we also compared its performance with the well-known CCA and LASSO-based techniques. As shown in Fig. 12, the proposed CNN-based classifier obtained higher accuracy than CCA and LASSO-based techniques in classifying five classes involving five SSVEP frequencies, for all different time windows. Through many studies [14], [17]–[20], CCA and LASSO methods have shown their powerful performance in detecting SSVEP frequencies in multi-channel BCI applications. Therefore, a relative low accuracy of CCA and LASSO-based techniques might be resulted from the single channel EEG data in the current work.

Experimental results from eight subjects in spelling the text “SPELLER” showed that, on average, the proposed speller achieved an accuracy of 97.37% with a relative high ITR of 48.99 ± 7.67 bit/min. On average, the offline accuracy under training mode across eight subjects was slightly higher than the online accuracy under operation mode (i.e., 99.15% vs. 97.37%, respectively). The difference might be resulted from the variations in the experimental conditions between training mode and operation mode, such as light intensity of the room, SNR of the EEG signal, subject's readiness, etc. Another explanation is that there was a low resolution in accuracy during evaluating the online test resulted from 100%/21 commands = 4.76%/command.

Our study tends to simplify the BCI system design and implementation in the aspects of practicability and feasibility to real-life applications. Moreover, the accuracy of the proposed system remains high in comparison with those in several recent studies. For example, [17] obtained around 75% accuracy on average, whereas [14] acquired around 80% accuracy with the 2-s window length. Another work achieved a nearly 100% average accuracy [32]. The recognition accuracy does not only depend on the number of channels being used, but also on the frequency detection algorithm and the SNR of the EEG signal [17], [32]. Thus, in comparison with [32], the lower accuracy of the current study might have resulted from the different frequency recognition methods and the quality of the EEG signal (i.e., SNR) acquired from the custom headset.

The average information transfer rate of the proposed speller for an SSVEP-based speller with five commands was higher than that reported in [33] (i.e., 48.99 vs. 25.89 bpm, respectively) and slightly higher than that in [13] (i.e., 48.99 vs. 37.62 bpm, respectively). Moreover, the proposed BCI speller showed a good practicability in terms of the number of characters of the interface compared with those in several recent studies (i.e., 58 vs. 27 characters in [13], 58 vs. 32 characters in [33], and 58 vs. 26 characters in [34]) while the ITR of the system remains high.

Although the current work employed only single channel EEG signals and with 1-D input data for the CNN, the performance of the model remains comparable high in comparison with that in the work [22], in which 99.28% classification rate was obtained with same 2-s time window. However, the training process might take longer time (500 epochs vs. 10 epochs, respectively) to obtain an optimized network with the same performance as [22]. It might not matter with the proposed system since the network need to be trained at once prior to operation.

Through the study, results showed the feasibility of the proposed BCI in providing spelling assistance to people with disabilities in terms of high accuracy and practicability in real-life applications. Further studies will be performed to investigate the limit of number of SSVEP frequencies that the proposed BCI can detect with a relative high performance to extend its feasible applications.

APPENDIX

Algorithm 1 Feature Extraction and CNN Model Training

```
Input: A visual cue for starting the training process
Output: The optimized weight set of the CNN model
for each trial
    Start Visual Stimulation Program
    Set the gazing time to 50 s (5 stimuli x 10 s gazing
     duration for each stimulus)
    while (current time < gazing time)
            Run EEG recording program
    end while
    Stop Visual Stimulation Program
    Stop EEG Recording program
end for
Compute segmented EEG data
Set current EEG segment to 1
for each EEG segment
    Compute 6-13 Hz band-pass filter
    Compute Fast Fourier Transform (FFT)
    Find maximum value in FFT magnitude vector
  (max(FFT))
    Find maximum value of FFT magnitude involving
  5 SSVEP frequencies (max(FFT of SSVEP))
    Compute "Feature 1": FFT magnitude
  vector/max(FFT)
    Compute "Feature 2": FFT magnitude
  vector/max(FFT of SSVEP)
    Update the feature and label vector
    INCREMENT current EEG segment
end for
Save "feature" and "label" vectors
Import "feature" and "label" vectors to CNN model
Initialize number of epoch, target loss, CNN weight
  set
repeat
    Run CNN model training process
until number of epoch, target loss
Return optimizedCNN weight set
```

REFERENCES

[1] N. Birbaumer, "Breaking the silence: Brain–computer interfaces (BCI) for communication and motor control," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, 2006.

[2] F. Nijboer *et al.*, "A P300-based brain–computer interface for people with amyotrophic lateral sclerosis," *Clin. Neurophysiol.*, vol. 119, no. 8, pp. 1909–1916, 2008.

[3] T. M. Vaughan *et al.*, "The wadsworth BCI research and development program: At home with BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 229–233, Jun. 2006.

[4] A. Nijholt *et al.*, "Brain-computer interfacing for intelligent systems," *IEEE Intell. Syst.*, vol. 23, no. 3, pp. 72–79, May/Jun. 2008.

[5] G. Pfurtscheller and F. L. Da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[6] H. Vaughan *et al.*, "Brain-computer interface technology: A review of the second international meeting," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 94–109, Jun. 2003.

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

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

[9] G. R. Müller-Putz, R. Scherer, C. Brauneis, and G. Pfurtscheller, "Steady-state visual evoked potential (SSVEP)-based communication: Impact of harmonic frequency components," *J. Neural Eng.*, vol. 2, no. 4, p. 123, 2005.

[10] 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.

[11] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P, Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 44, pp. E6058–E6067, 2015.

[12] Y.-T. Wang, Y. Wang, C.-K. Cheng, and T.-P. Jung, "Measuring steady-state visual evoked potentials from non-hair-bearing areas," in *Proc. EMBC*, San Diego, CA, USA, Aug./Sep. 2012, pp. 1806–1809.

[13] H. Cecotti, "A Self-paced and calibration-less SSVEP-based brain–computer interface speller," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 127–133, Apr. 2010.

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

[15] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1181–1186, Oct. 2002.

[16] O. Friman, I. Volosyak, and A. Graser, "Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 742–750, Apr. 2007.

[17] 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. 6, pp. 1172–1176, Jun. 2007.

[18] Y. Zhang, J. Jin, X. Qing, B. Wang, and X. Wang, "LASSO based stimulus frequency recognition model for SSVEP BCIs," *Biomed. Signal Process. Control*, vol. 7, no. 2, pp. 104–111, 2012.

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

[20] 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. 4, pp. 1450013-1–1450013-14, 2014.

[21] H. Cecotti, "A time–frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses," *Pattern Recognit. Lett.*, vol. 32, no. 8, pp. 1145–1153, 2011.

[22] 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, pp. e0172578, 2017.

[23] I. Volosyak, H. Cecotti, D. Valbuena, and A. Graser, "Evaluation of the Bremen SSVEP based BCI in real world conditions," in *Proc. ICORR*, Kyoto, Japan, Jun. 2009, pp. 322–331.

[24] E. Yin, Z. Zhou, Y. Yu, and D. Hu, "A dynamically optimized SSVEP brain–computer interface (BCI) speller," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1447–1456, Jun. 2015.

[25] P. F. Diez, V. Mut, E. Laciar, and E. Avila, "A comparison of monopolar and bipolar EEG recordings for SSVEP detection," in *Proc. EMBC*, Buenos Aires, Argentina, Aug./Sep. 2010, pp. 5803–5806.

[26] Y. M. Chi, P. Ng, E. Kang, J. Kang, J. Fang, and G. Cauwenberghs, "Wireless non-contact cardiac and neural monitoring," in *Proc. Wireless Health*, San Diego, CA, USA, 2010, pp. 15–23.

[27] D. Regan, *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. New York, NY, USA: Elsevier, 1989, p. 672.

[28] T. D'Albis, "A predictive speller for a brain-computer interface based on motor-imagery," M.S. thesis, Dept. Comp. Eng., Politecnico di Milano, Milan, Italy, 2009.

[29] R. Scherer, G. R. Müller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 979–984, Jun. 2004.

[30] D. Valbuena, I. Sugiarto, and A. Gräser, "Spelling with the bremen brain-computer interface and the integrated SSVEP stimulator," presented at the 4th Int. BCI Workshop Training Course, Graz, Austria, 2008, pp. 291–296.

[31] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, no. 2, pp. 387–399, 2011.

[32] Y.-T. Wang, M. Nakanishi, Y. Wang, C.-S. Wei, C.-K. Cheng, and T.-P. Jung, "An online brain-computer interface based on SSVEPs measured from non-hair-bearing areas," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 1, pp. 14–21, Jan. 2017.

[33] I. Volosyak, H. Cecotti, and A. Gräser, "Impact of frequency selection on LCD screens for SSVEP based brain-computer interfaces," in *Proc. IWANN*, Salamanca, Spain, 2009, pp. 706–713.

[34] T. H. Nguyen, D.-L. Yang, and W.-Y. Chung, "A high-rate BCI speller based on eye-closed EEG signal," *IEEE Access*, vol. 6, pp. 33995–34003, 2018.

TRUNG-HAU NGUYEN received the B.E. degree in mechatronics engineering from the Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2008, and the M.E. degree in biomedical engineering from Pukyong National University, Busan, South Korea, in 2015.

He is currently pursuing the Ph.D. degree in electronics engineering with Pukyong National University. His research interests include the development of wearable healthcare sensors, signal processing, and electroencephalogram-based driver-assistance systems.

WAN-YOUNG CHUNG (SM'17) received the B.S. and M.S. degrees in electronic engineering from Kyungpook National University, Daegu, South Korea, in 1987 and 1989, respectively, and the Ph.D. degree in sensor engineering from Kyushu University, Fukuoka, Japan, in 1998. From 1993 to 1999, he was an Assistant Professor with Semyung University. From 1999 to 2008, he was an Associate Professor with Dongseo University. Since 2008, he has been a Full Professor with the Department of Electronic Engineering, Pukyong National University, South Korea. His areas of interest include ubiquitous healthcare, wireless sensor network applications, and gas sensors.