High-Speed Spelling in Virtual Reality with Sequential Hybrid BCIs

Zhaolin YAO†,††, Member, Xinyao MA†,††, Yijun WANG†,††, Xu ZHANG†,††, Ming LIU†,††, Weihua PEI†,††, and Hongda CHEN†,††, Nonmembers

SUMMARY A new hybrid brain-computer interface (BCI), which is based on sequential controls by eye tracking and steady-state visual evoked potentials (SSVEPs), has been proposed for high-speed spelling in virtual reality (VR) with a 40-target virtual keyboard. During target selection, gaze point was first detected by an eye-tracking accessory. A 4-target block was then selected for further target selection by a 4-class SSVEP BCI. The system can type at a speed of 1.25 character/sec in a cue-guided target selection task. Online experiments on three subjects achieved an averaged information transfer rate (ITR) of 360.7 bits/min.

key words: brain-computer interface, steady-state visual evoked potentials, eye tracking, virtual reality, speller

1. Introduction

Newly emerged commercial virtual reality (VR) headsets such as HTC Vive [1] and Oculus Rift pose new challenges in intuitive human-computer interaction since the traditional mouse-keyboard configuration is no longer the best method. In contrast, gaze interaction, which only requires the movement of the eyes, is a good solution to interact with VR environments. Currently, eye tracking and brain-computer interface (BCI) are the two most widely used approaches in gaze-based human-computer interaction [2]. Eye tracking has been widely used for people with disabilities to control their computers. For example, eye tracking based text entry systems, which can type with speed from 5-10 words/min, have been well established [3]. Gaze-based visual BCIs have also been widely used as assistive methods for communication and control [4]. For example, high-speed BCI spellers, which detect gaze direction by steady-state visual evoked potentials (SSVEP), can achieve typing speed up to 10 words/min [5]. Recently, both eye tracking [6] and SSVEP BCIs [7] have been developed and implemented separately in VR head-mounted displays (VRHMD). The combination of the two methods is therefore potential for high-performance gaze interaction in VR.

Several recent studies have combined SSVEP and eye tracking to develop hybrid BCIs for spelling [8], [9]. However, due to performance gap between the two modalities, the hybrid systems did not achieve large performance improvement. In a 30-target speller [8], a webcam-based eye tracking method was only able to detect binary gaze direction (left vs right). Another hybrid BCI obtained an information transfer rate (ITR) similar to SSVEP, but much lower than eye tracking [9].

According to high performance reported for eye tracking and SSVEP BCI [3], [5], the hybrid method is expected to achieve much higher performance than the existing systems. However, in VR environments, accurate and precise gaze detection remains a big challenge to both eye tracking and SSVEP BCI.

In this Letter we proposed a new 40-class BCI speller in VR based on hybrid gaze controls with eye tracking and SSVEP BCI. Character input was facilitated by a sequential two-stage control strategy, which included a 27-point block selection with eye tracking and a 4-point target selection with SSVEP. In a cue-guided target selection task, the system obtained an average ITR of 360.7 bits/min at a typing speed of 1.25 character/sec.

2. System Design

The hybrid BCI speller consisted of four major components (Fig. 1 (a)). An HTC Vive system (HTC, Corp.) presented stimulus and feedback in VR with a 90Hz refresh rate. A
Synamps2 system (Neuroscan, Inc.) acquired EEG signals from nine electrodes over the parietal and occipital areas (Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, O2) at 250Hz. An aGlass DKI infrared eye-tracking module (7invensun Technology Co., Ltd.) mounted into the VRHMD measured eye position data at 90Hz. A computer (Intel Core i7 processor, an NVIDIA GTX1080 graphic card) supported the VR system and performed data analysis. Figure 1 (b) shows the 40-target virtual keyboard (4×10 matrix), which contains 26 English alphabet letters, 10 digits, and 4 symbols. The size of a target and intervals between two neighbouring targets were 4.8 and 0.48 degrees respectively. Figure 2 illustrates the diagram of the sequential two-stage controls of selecting a character. A 4-target block was first selected with eye tracking. Next, four targets in the block started to flicker at different frequencies. At last, the gazed target was determined by analyzing SSVEPs. Detections of eye position and SSVEPs were realized by template-based classification approaches, which included separate training and testing procedures.

The user interface program for stimulus/feedback presentation and eye-tracking data collection was written in C# in the Unity3D engine and SteamVR platform. The online data processing program, which also received EEG data from the EEG system through TCP/IP, was developed in an Anaconda Python environment. Communications between the user interface and data processing programs were performed through TCP/IP.

3. Data Analysis Algorithms

3.1 Block Selection with Eye Tracking

The 4-target block selection design can effectively reduce the precision requirement of eye tracking. Figure 3 shows the 27-point block design for eye tracking (Fig. 3 (a)) and subsequent 4-target stimulus design for SSVEP (Fig. 3 (b)). According to the 4×10 matrix layout, each of the 40 targets can be assigned to 1 (4 corners), 2 (other 20 border targets) or 4 blocks (16 inner targets). For example, four different blocks, depending on the location of gaze point detected by eye tracking, are all valid (correct) for character ‘T’ (Fig. 3 (b)). Towards fast eye tracking, eye data were measured within a 50 ms duration before the stimulus onset of SSVEP. In the training procedure, template coordinates for 40 eye-gaze points corresponding to the 40 characters were first calculated by averaging across multiple training trials, and then template coordinates for the 27 blocks were calculated as the mean of the corresponding template coordinates of the 4 targets in each block. In the testing procedure, distances between the gaze point and the 27 template coordinates were calculated and the nearest block was selected to present SSVEP stimuli for further target selection.

3.2 Target Selection with SSVEP

The 4-target SSVEP BCI adopted a joint frequency-phase modulation (JFPM) method (frequency range: 12.4–15.4 Hz with an interval of 1 Hz, phase interval: 0.25π) [5]. To ensure fast and accurate detection, the stimulus duration was set to 300 ms. Stimulus onsets from the user interface program were sent to the Neuroscan system via a parallel port. After applying a 140 ms delay [5], data epochs from 140–440 ms after stimulus onset were extracted for analysis. A task-related component analysis (TRCA) algorithm [10] was implemented to detect SSVEP. In SSVEP detection, TRCA have been used to design spatial filters for enhancing the SNR of SSVEPs. TRCA optimizes the spatial filters to maximize inter-trial covariance of SSVEPs at the same stimulation frequency. The details of the TRCA-based SSVEP detection algorithm can be found in [10]. Spatial filters and EEG templates corresponding to the four frequencies were obtained in the training procedure. In the testing procedure, the data processing program implemented the template-matching approach to determine the target frequency, which showed the maximal correlation coefficient between test data and EEG template.

3.3 Performance Evaluation

The BCI performance was evaluated by calculating target identification accuracy and ITR. The accuracy is defined as the ratio between the number of correctly identified trials and the total number of trials in cue-guided tasks. The ITR
in bits/min is calculated as follows [11]:

$$ITR = \left( \log_2 M + P \log_2 P + (1 - P) \log_2 \left( \frac{1 - P}{M - 1} \right) \right) \times \frac{60}{T}$$ (1)

where $M$ is the number of targets (40 in this study), $P$ is the accuracy of target identification, and $T$ (seconds/selection) is the average time for a selection (0.8s in this study, including 0.3s for visual stimulation and 0.5s for gaze switching).

4. Experiment

Three healthy subjects (mean age: 27 years old), who were experienced in eye tracking and SSVEP BCI-based gaze control, participated in an online cue-guided target selection experiment. Each participant was asked to read and sign an informed consent form approved by the Institutional Review Board of the Chinese Academy of Sciences before the experiment.

The subjects performed a 9-point eye-tracking calibration before the experiment. This study adopted the 40-target SSVEP BCI (frequency range: 8–15.8 Hz with an interval of 0.2 Hz, phase interval: 0.35π) [5] to collect training data for eye-tracking and 40-class SSVEPs. The training procedure of eye tracking included 8 repetitions of fixing all 40 characters indicated by cues (red squares) in a random order (in total 320 trials). At the same time, the 40-class SSVEP signals (8 repetitions for each target) were also recorded for offline comparison. In the 4-class SSVEP training, stimuli were applied to the 4-target block in the center of the virtual keyboard (‘T’, ‘Y’, ‘G’, ‘H’). The training data comprised 8 repetitions of all 4 targets indicated by cues (in total 80 trials). In the testing stage, subjects performed 5 runs each containing 40 trials corresponding to all 40 characters (in total 200 trials). Each trial lasted 1 second including 50 ms for eye tracking, 300 ms for SSVEP stimulation, and 450 ms for gaze switching. Target cues appeared for 500 ms during gaze switching and eye tracking. Online feedbacks (blue squares) were provided to subjects during testing.

Classification accuracy and ITR were used to evaluate BCI performance. We also estimated offline testing performance of two single-modality methods (i.e., 40-class eye tracking, and 40-class SSVEP). The 40-class eye-tracking method employed a similar template-based distance measurement approach. For the 40-class SSVEP method, the similar TRCA-based method [10] was used for SSVEP detection and the offline BCI performance was evaluated using a leave-one-out cross validation.

5. Results

Table 1 lists the testing performance of all subjects. The online hybrid method achieved an averaged ITR of 360.7 bits/min with mean accuracy of 95.2%. The two single-modality methods showed largely reduced ITRs (40-class eye-tracking, and 40-class SSVEP) were also recorded for each target (in total 320 trials). At the same time, the 40-class SSVEP classification obtained mean accuracy of 97.4% (S1: 98.0%, S2: 96.5%, S3: 98.5%, mean: 97.7%), which is not convertible to ITR. Note that, the 4-class SSVEP classification obtained mean accuracy of 97.4% (S1: 98.0%, S2: 94.8%, S3: 99.5%), which was largely higher than the 40-class SSVEP classification (S1: 90.0%, S2: 63.4%, S3: 93.7%).

6. Conclusion and Discussion

We propose a hybrid BCI for high-speed spelling in VR by combining eye tracking and SSVEP. A 40-target virtual keyboard was designed for two-stage sequential controls of eye-based block selection and SSVEP-based target selection. Online experiments with a cue-guided target selection task show that the proposed system can type at a speed around 15 words/min (1.25 character/sec), achieving an averaged ITR of 360.7 bits/min, to our knowledge, the highest ITR in the VR-based BCIs. The results also demonstrate the superiority of the sequential hybrid method over the single-modality methods.

As shown in Table 1, the proposed hybrid method achieved higher accuracy and ITR compared with the single-modality methods (i.e., the 40-class eye-tracking method and the 40-class SSVEP method). Compared with the 40-class SSVEP method, the 4-class SSVEP classification showed much higher accuracy (97.4% vs 82.4%) because of the largely reduced task difficulty related to the number of classes in target detection. The high accuracy in the eye-tracking based 27-point block selection (97.7%) ensured the overall accuracy of the sequential hybrid system. Compared with the 27-point block selection, the 40-class eye-tracking method obtained much lower accuracy (63.3%) because the 40-class eye-position identification required much higher precision. The separation of the overall task into two sequential tasks facilitates the 40-class classification, and thereby leading to higher classification accuracy. Although the hybrid method showed advantage in all subjects in this study, it is necessary to point out that if one of the single-modality methods is good enough to obtain high accuracy, the sequential hybrid method will not provide further performance improvement. Future work will focus on real spelling applications with the proposed method in VR.

Table 1  Classification accuracy (%) and ITR (bits/min) of three subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>40-class Eye (Offline)</th>
<th>40-class SSVEP (Offline)</th>
<th>40-class Hybrid (Online)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>ACC 58.0 ITR 160.1</td>
<td>ACC 90.0 ITR 326.5</td>
<td>ACC 96.0 ITR 365.3</td>
</tr>
<tr>
<td>S2</td>
<td>ACC 74.0 ITR 236.8</td>
<td>ACC 63.4 ITR 184.3</td>
<td>ACC 91.5 ITR 335.8</td>
</tr>
<tr>
<td>S3</td>
<td>ACC 58.0 ITR 160.1</td>
<td>ACC 93.7 ITR 350.2</td>
<td>ACC 98.0 ITR 381.1</td>
</tr>
<tr>
<td>Mean</td>
<td>ACC 63.3 ITR 185.7</td>
<td>ACC 82.4 ITR 287.0</td>
<td>ACC 95.2 ITR 360.7</td>
</tr>
</tbody>
</table>

Eye: 185.7 bits/min, 40-class SSVEP: 287.0 bits/min). The 40-class eye-tracking method obtained low mean accuracy of 63.3% due to the requirement of high precision. The mean accuracy of the 40-class SSVEP method was 82.4%. For all subjects, the hybrid method consistently achieved the highest accuracy and ITR among the three methods. The eye-tracking based 27-point block selection obtained very high target-specific accuracy (S1: 98.0%, S2: 96.5%, S3: 98.5%, mean: 97.7%), which is not convertible to ITR. Note that, the 4-class SSVEP classification obtained mean accuracy of 97.4% (S1: 98.0%, S2: 94.8%, S3: 99.5%), which was largely higher than the 40-class SSVEP classification (S1: 90.0%, S2: 63.4%, S3: 93.7%).
Acknowledgments

This work is supported in part by the National Key Research and Development Plan (2017YFA0205903), the National Natural Science Foundation of China (61671424, 61335010, and 61634006), National Science Foundation of China (AWS16J028), Beijing S&T planning task (Z161100002616019), and the Recruitment Program of Young Professionals.

References


