Combining Brain-Computer Interface and Eye Tracking for High-Speed Text Entry in Virtual Reality

Xinyao Ma, Zhaolin Yao, Yijun Wang, Weihua Pei, Hongda Chen
Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China
University of Chinese Academy of Sciences, Beijing, China
Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China
mxy1108@gmail.com; {yaozhaolin, wangyj, peiwh, hdchen}@semi.ac.cn

ABSTRACT
Gaze interaction provides an efficient way for users to communicate and control in virtual reality (VR) presented by head-mounted displays. In gaze-based text-entry systems, eye tracking and brain-computer interface (BCI) are the two most commonly used approaches. This paper presents a hybrid BCI system for text entry in VR by combining steady-state visual evoked potentials (SSVEP) and eye tracking. The user interface in VR designed a 40-target virtual keyboard using a joint frequency-phase modulation method for SSVEP. Eye position was measured by an eye-tracking accessory in the VR headset. Target-related gaze direction was detected by combining simultaneously recorded SSVEP and eye position data. Offline and online experiments indicate that the proposed system can type at a speed around 10 words per minute, leading to an information transfer rate (ITR) of 270 bits per minute. The results further demonstrate the superiority of the hybrid method over single-modality methods for VR applications.

Author Keywords
Brain-computer interface; steady-state visual evoked potentials; eye tracking; virtual reality; text entry

ACM Classification Keywords
Multi-modal interfaces; Intelligent wearable, mobile and ubiquitous interfaces

INTRODUCTION
Gaze interaction is one of the most common ways for people with motor disabilities to navigate and control their computer with their eyes [18]. Gaze interaction only requires the movement of the eyes without any need for other additional muscle control, which makes it a perfect solution for those with disabilities such as spinal cord injury (SCI), repetitive strain injury (RSI), and amyotrophic lateral sclerosis (ALS) to communicate with their environments. Gaze interaction approaches including eye tracking, electrooculogram (EOG), and brain-computer interface (BCI) have been widely used in augmentative and alternative communication (AAC) systems [5].

Eye tracking is the predominant method used for gaze interaction nowadays. Eye-tracking devices detect eye positions by processing visible light or infrared images of the eyes captured by a camera. Towards different applications, desktop-mounted and head-mounted eye trackers have been developed by eye-tracking companies such as Tobii [19]. Eye tracking based text entry techniques have been well established and an eye typing system can be implemented by combining an eye tracker with a virtual keyboard interface on the computer screen [10]. In general, eye tracking based typing speed ranges from 5-10 words per minute (wpm) [11]. Although high-performance eye trackers can detect eye movements with high accuracy and precision, current applications of eye tracking are still limited because the system is always expensive and the tracking performance is sensitive to environmental factors such as light condition.

BCIs provide a very promising technique for interaction with computers by establishing a direct link between human brain and computer [20]. Using frequency or time domain coding methods, gaze detection can be realized by visual BCIs based on visual evoked potentials (VEPs) and P300 potentials recorded by scalp electroencephalogram (EEG) [4]. In BCI based text-entry systems, steady-state visual evoked potentials (SSVEP) have shown high performance and usability. For example, Chen et al. [2] developed an SSVEP-based BCI speller with a typing speed around 10 wpm, which was comparable to the eye-tracking approach. Furthermore, combining BCIs with other assistive technologies have received increasing attention in recent years [12]. A hybrid BCI system that combines a traditional BCI with other physiological signals has shown great advantages in BCI performance and system flexibility [16, 13]. For example, eye tracking and BCI methods have been combined to develop hybrid BCIs [8, 9].

Recently, with the proliferation of virtual reality (VR) headsets such as HTC Vive and Oculus Rift, human-computer interface design in VR opens a new frontier, as

IUI'18, March 7–11, 2018, Tokyo, Japan
 ACM ISBN 978-1-4503-4945-1/18/03...$15.00
https://doi.org/10.1145/3172944.3172988

the traditional mouse-keyboard configuration is no longer the best interaction method inside VR environments. HTC Vive [15] uses active motion-tracked handheld controllers with input buttons and triggers to interact with the environment, which is reliable but not intuitive and may cause fatigue. Alternatively, gaze input provides an efficient way for interaction in VR applications because of its intuitive and immersive style. On one hand, eye-tracking accessories in VR head-mounted displays (VRHMD) have been proposed and developed [17]. On the other hand, gaze control has been realized in recent VR-based BCI studies [6]. Technically, for applications such as text entry in VRHMDs, eye tracking and BCI methods can be easily combined towards improved accuracy and precision. However, to our knowledge, gaze-based text entry in consumer VR headsets has not been reported yet.

In this paper, we designed and implemented a hybrid gaze-based text-entry system in VR for high-speed typing. The system combined an SSVEP-based BCI with a low-cost eye-tracking module in VRHMD. With a virtual keyboard, gaze direction was detected by combining simultaneously recorded SSVEP and eye position data. To evaluate system performance, this study designed offline and online experiments to simulate text entry tasks. In addition, classification results corresponding to the hybrid method and the single-modality methods were estimated separately for comparison. The contributions of this paper include: a) a viable solution for robust and intuitive high-speed gaze interaction in VR, b) a hybrid method to combine SSVEP BCIs and low-cost eye-tracking devices, and c) a general experimental platform for VR-based BCI research [7, 1].

METHODS

System Design

This study designed a hybrid gaze-based text-entry system using an HTC Vive VR platform. Figure 1 shows the system architecture with closed-loop data flow. EEG data were acquired using a Synamps2 system from Neuroscan (https://compumedicsneuroscan.com/). Eye positions were measured with an embedded infrared eye-tracking module: aGlass DKI from 7invensun (https://www.7invensun.com). All software programs ran on a computer (with an Intel core i7 processor and an NVIDIA GTX1080 graphic card) supporting the HTC Vive platform. The VRHMD used two screens (one per eye) and each screen had a resolution of 1080×1200 pixels. The text-entry software was written in C# in Unity 3D engine and SteamVR platform, which rendered 3D scene inside HMD to give the user an immersive 3D perception at a 90Hz refresh rate. The software included programs for stimulus presentation, data communication, data analysis, and feedback presentation. Eye position data were collected using aGlass SDK in real time. A separate computer was used to record and transfer real-time EEG data to the VR computer through TCP/IP. Event triggers (i.e., stimulus onset) from the VR computer were sent to the EEG system through a parallel port for synchronizing event and EEG data.

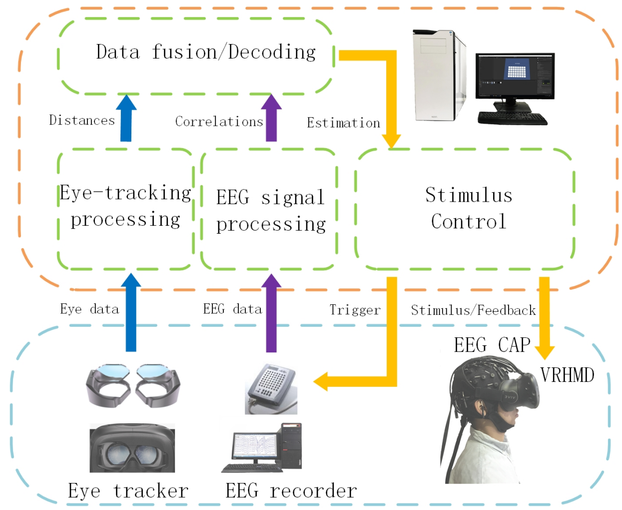

Figure 1. Diagram of the hybrid text entry system. EEG and eye position data are measured and processed simultaneously to detect gaze direction for text entry in VRHMD.

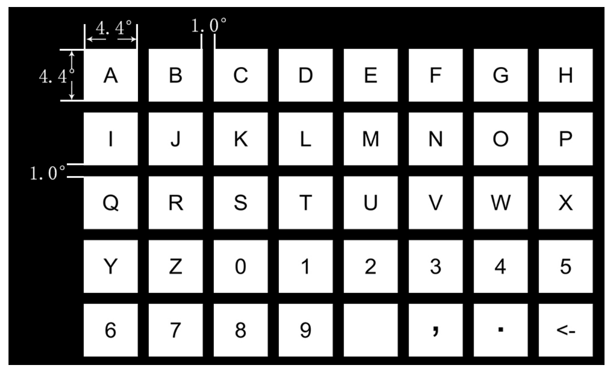

Figure 2. Layout of the virtual keyboard in VR.

As shown in Figure 2, the stimulus program adopted a 40-target virtual keyboard designed for an SSVEP BCI speller [2], which was a 5×8 matrix containing 40 characters (26 English alphabet letters, 10 digits, and 4 other symbols). For multi-target SSVEP coding, 40 characters were tagged differently using a joint frequency-phase modulation (JFPM) method (frequency range: 8-15.8Hz with an interval of 0.2Hz, phase interval between two adjacent frequencies: 0.35π) [2]. As shown in Figure 2, the width and height of each character were 4.4° and the horizontal and vertical distances between two neighbors were 1.0°. The virtual keyboard seen in VRHMD corresponded to the actual situation where a person sat at a distance of 60cm in front of a 22-inch computer monitor.

For data communication, EEG data and eye-tracking data were received and recorded separately by the stimulus program. For offline analysis, data files were saved for further analysis using Matlab (Mathworks, Inc.). For online analysis, an online data processing program was developed in an Anaconda Python environment. EEG data transferred by TCP/IP were directly thrown into the online processing program. Eye-tracking data were saved to files by the stimulus program, and loaded automatically into the online processing program. The online processing program further performed a hybrid data fusion algorithm for target identification, and returned the results to the stimulus program for feedback presentation.

Experiment Design

Eleven healthy subjects (2 females and 9 males, aged 19-29 years) with normal or corrected to normal vision participated in the experiment. The subject first wore EEG cap and then fixed the HMD as shown in Figure 1. Nine electrodes over the parietal and occipital areas (Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, O2) were used to record SSVEPs at a sampling rate of 1000Hz. Eye position data were recorded by the eye-tracking module with a monocular (right eye) mode at a sampling rate of 120Hz. For each subject, eye tracking was calibrated through pupil alignment and a 9-points calibration procedure.

This study designed separate offline and online experiments to evaluate system performance using a cue-guided target selection task [2]. All subjects took part in the offline experiment, and two of them (S1 and S5) joined in the online experiment on a different day. In the offline experiment, all participants were asked to accomplish eight blocks of target selection. In each block, all 40 characters were cued sequentially in a random order. Each trial lasted 1.8 seconds in total. At the beginning of each trial, a red square appeared as the target cue for 0.8 seconds. Subjects were asked to direct their gaze to the target as soon as possible. After the cue, all stimuli started to flicker for 1 second. Subjects were asked to avoid blinks during stimulation. EEG and eye-tracking data were saved for offline analysis. In the online experiment, stimulation duration was reduced to 0.3 seconds towards high information transfer rates (ITRs) [14]. After an 8-blocks training procedure, the subjects performed an online 4-blocks test with algorithms trained from the training data. A blue square at the target location flashed for 0.2 seconds as feedback after online data analysis.

EEG Data Analysis

Data epochs for the nine-channel SSVEP signals were extracted according to event triggers generated by the stimulus program and then down-sampled to 250Hz. A filter bank method [3] with the frequency range of 8-88Hz was used to decompose SSVEP signals into seven sub-band components to cover multiple harmonics of SSVEPs.

For SSVEP detection, this study employed the task-related component analysis (TRCA) algorithm [14] to design spatial filters for SSVEP, and then implemented a template-matching based target identification method. After spatial filtering, the correlation coefficient between the projections of single-trial test data $\boldsymbol{X}$ and an individual template $\overline{\boldsymbol{X}}_{\boldsymbol{k}}$ for the k-th visual stimulus was calculated as:

$$\rho_k = \rho(\boldsymbol{X}^T\boldsymbol{W}, \overline{\boldsymbol{X}}_k^T\boldsymbol{W})$$

where $\boldsymbol{W}$ consisted of spatial filters corresponding to all N_f stimulus frequencies:

$$\boldsymbol{W} = \left[\boldsymbol{w}_1\ \boldsymbol{w}_2\ \dots\ \boldsymbol{w}_{N_f}\right].$$

The stimulus frequency with the maximal correlation coefficient was selected as the target frequency. For each trial, the correlation coefficient vector was used for data fusion with eye-tracking data.

Eye Tracking Data Analysis

Eye positions were tracked in combination with head attitude measured by the HMD. A 9-points calibration was performed to map the gaze points onto the 2D keyboard plane within a normalized coordinate system. The whole virtual keyboard was visible and trackable in the coordinate system. Eye tracking analysis was performed by two methods. The first method was based on the prediction of coordinate values according to calibration. The mean coordinates of the gaze point from a certain period after the visual cue was used to calculate Euclidean distance to all 40 characters. The character with the minimal distance was chosen as the target. The second method was a template-matching method in which the gaze point was estimated by comparing with the training data. The template coordinates for 40 eye-gaze points corresponding to the 40 characters were calculated by averaging across all training trials. The distances between the test gaze point and the template coordinates were calculated. The character with the minimal distance was then selected as the target.

In addition to target detection, this study also evaluated accuracy and precision of the eye-tracking module [18]. Across eight trials for each character, eye-tracking points of each trial were averaged, and then the tracking accuracy and precision were calculated for each character.

Hybrid Data Fusion

To combine information from SSVEP and eye tracking, a hybrid data fusion method was developed to enhance target detection. The hybrid decision model was defined as:

$$R_{hybrid} = R_{EEG} \times ACC_{EEG}^2 + \frac{1}{D_{EYE}} \times ACC_{EYE}^2$$

where R_{EEG} was the correlation coefficients between SSVEP and individual templates, D_{EYE} was the distances between gaze coordinates and trained templates, ACC_{EEG} and ACC_{EYE} were the weights obtained by calculating the averaged classification accuracy of the EEG and eye-tracking methods. Both R_{EEG} and $\frac{1}{D_{EYE}}$ were normalized to zero mean and unit variance. To estimate ACC_{EEG} and ACC_{EYE}, leave-one-out cross-validation was performed with the training data to calculate the averaged classification accuracy corresponding to EEG and eye tracking separately. The hybrid method chose the k-th character with the maximal R_k as the target character.

RESULTS AND DISCUSSION

Performance of SSVEP BCI

This study employed the TRCA-based method to detect SSVEP. The offline classification accuracy and ITR (in bits per minute, bpm) [20] were estimated using a leave-one-out cross-validation method. Table 1 lists the classification accuracy and ITR with a data length of 0.3 seconds. The

Subject	Eye-calibrated		Eye-template		EEG-template		Hybrid	
Metrics	Acc	ITR	Acc	ITR	Acc	ITR	Acc	ITR
S1	60	123	80	194	89	234	**97**	272
S2	65	140	92	247	43	72	**95**	261
S3	47	84	71	160	**96**	268	95	263
S4	77	182	98	277	**99**	287	**99**	287
S5	50	94	87	225	90	235	**96**	270
S6	49	90	84	211	70	156	**91**	243
S7	80	194	99	283	70	157	**99**	287
S8	80	196	97	273	85	217	**99**	285
S9	35	51	91	244	75	176	**95**	260
S10	66	144	83	208	**97**	272	**97**	272
S11	70	156	87	224	89	233	**96**	266
Mean	62	132	88	232	82	210	**96**	270
Std	14	46	8	36	15	60	2	12

Table 1. Accuracy (%) and ITRs (bpm) for all subjects using a data length of 0.3 seconds. Results for the eye tracking, EEG, and hybrid methods were calculated separately for comparison. Values in bold indicate the best accuracy.

averaged accuracy across all subjects is 82±15%, while the averaged ITR is 210 ± 60bpm. Figure 3 illustrates the classification results using different data lengths. The accuracy increases when data length increases. After around 0.3 seconds, the accuracy for EEG becomes comparable to the eye-tracking method. ITR peaks around 0.35 seconds. These results are consistent with previous studies using LCD monitors as stimulus platforms [2, 14], demonstrating the efficiency of SSVEP BCIs in VR environments.

Performance of Eye Tracking

This study calculates two types of accuracy for the eye tracking system. As shown in Table 1, the eye-calibrated method obtained averaged accuracy of 62±14%. The low classification accuracy was caused by the relatively low tracking accuracy of the low-cost eye-tracking module. The supervised eye-template method achieved significantly higher classification accuracy of 88±8%. The tracking accuracy for the experiment setting in this study was roughly estimated as 3.1±2.7 degrees and the tracking precision was 1.8±1.5 degrees. The tracking precision was higher for targets with small viewing angles (i.e., central areas in the keyboard). As shown in Figure 3, eye tracking could reach stable classification accuracy within a very short time window around 50ms. Different from SSVEP, longer data length doesn't improve the classification accuracy for eye tracking. This is due to the fact that accuracy and precision of eye tracking is time invariant.

Performance of the Hybrid System

The hybrid method achieved higher performance than the single-modality methods. As shown in Table 1, the averaged classification accuracy is 96±2% and the ITR is 270±12bpm. One-way repeated-measures analysis of variance (ANOVA) shows that there is significant difference of accuracy between the four different methods ($F(3, 43)=16.45$, $p<10^{-5}$). Bonferroni corrected post-hoc pairwise comparisons using paired t-tests indicate the improvements of accuracy for the hybrid method are significant (vs. eye-calibrated: $p<10^{-5}$, vs. eye-template: $p<0.05$, vs. EEG-template: $p<0.05$). As shown

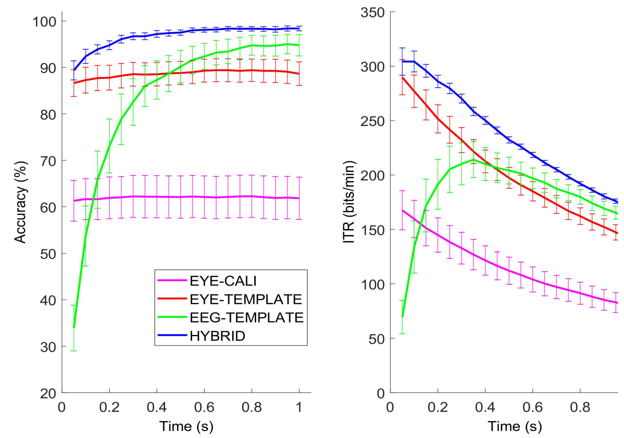

Figure 3. Classification accuracy and ITRs using different data lengths (from 50ms to 1000ms with a step of 50ms). The error bars represent standard errors.

in Figure 3, the hybrid method showed very consistent improvements of accuracy and ITRs at all data lengths compared with the single-modality methods. Furthermore, the hybrid accuracy is significantly higher than the method in which the single modality with higher performance was chosen for each subject (96% vs. 93%, $p<0.01$).

Online Validation

Two subjects (S1, S5) participated in the online experiment. The online processing program took about 16ms to implement the hybrid algorithm. For S1, accuracies for EEG and eye tracking were 88% and 93% respectively, and the hybrid method achieved accuracy of 97%. For S5, accuracies for EEG, eye tracking, and the hybrid method were 89%, 92%, and 98%. Hybrid classification results in the online test were consistent to those in the offline analysis (S1: 97%, S5: 96%), which proved the feasibility of the proposed method for online applications.

CONCLUSION

This study presents a hybrid BCI for high-speed text entry in VRHMD by combining SSVEP and eye tracking. A 40-target virtual keyboard was designed for eliciting SSVEPs and tracking gaze at the same time. A hybrid decision model was proposed for gaze detection. With a cue-guided target selection task, both offline and online experiments indicate that the proposed system can input text at a speed around 10wpm (1.1s per character), leading to an averaged ITR of 270bpm, which is comparable to the highest ITR reported in BCI studies [2]. The results also demonstrate the superiority of the hybrid method over the single-modality methods. Future work will focus on improvement of eye tracking and data fusion methods, as well as the implementation of real text-entry applications in VR environments.

ACKNOWLEDGEMENTS

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), Thousand Youth Talents Plan, and Beijing S&T planning task (Z161100002616019).

RFERENCES

1. Jessica D. Bayliss and Dana H. Ballard. 2000. A virtual reality testbed for brain-computer interface research. *IEEE Transactions on Rehabilitation Engineering* 8, 2: 188-190.
2. Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao. 2015. High-speed spelling with a noninvasive brain-computer interface. *Proceedings of the National Academy of Sciences* 112, 44: E6058-E6067.
3. Xiaogang Chen, Yijun Wang, Shangkai Gao, Tzyy-Ping Jung, and Xiaorong Gao. 2015. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface. *Journal of Neural Engineering* 12, 4, 046008.
4. Shangkai Gao, Yijun Wang, Xiaorong Gao, and Bo Hong. 2014. Visual and Auditory Brain-Computer Interfaces. *IEEE Transactions on Biomedical Engineering* 61, 5: 1436-1447.
5. Ivo Käthner, Andrea Kübler and Sebastian Halder. 2015. Comparison of eye tracking, electrooculography and an auditory brain-computer interface for binary communication: a case study with a participant in the locked-in state. *Journal of Neuroengineering and Rehabilitation* 12, 76.
6. Bonkon Koo, Hwan-Gon Lee, Yunjun Nam, and Seungjin Choi. 2015. Immersive BCI with SSVEP in VR head-mounted display. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC).
7. Anatole Lecuyer, Fabien Lotte, Richard B. Reilly, Robert Leeb, Michitaka Hirose, and Mel Slater. 2008. Brain-Computer Interfaces, Virtual Reality, and Videogames. *Computer* 41, 10: 66-72.
8. Eui Chul Lee, Jin Cheol Woo, Jong Hwa Kim, Mincheol Whang, and Kang Ryoung Park. 2010. A brain-computer interface method combined with eye tracking for 3D interaction. *Journal of Neuroscience Methods* 190, 2: 289-298.
9. Jeong-Hwan Lim, Jun-Hak Lee, Han-Jeong Hwang, Dong Hwan Kim, and Chang-Hwan Im. 2015. Development of a hybrid mental spelling system combining SSVEP-based brain-computer interface and webcam-based eye tracking. *Biomedical Signal Processing and Control* 21: 99-104.
10. Päivi Majaranta and Kari-Jouko Räihä. 2002. Twenty years of eye typing: systems and design issues. *Proceedings of the 2002 symposium on Eye tracking research & applications* (ETRA '02): 15-22.
11. Päivi Majaranta and Kari-Jouko Räihä.2007 Text entry by gaze: Utilizing eye-tracking. In *Text entry systems: Mobility, accessibility, universality*, I. Scott MacKenzie and Kumiko Tanaka-Ishii (eds.). Morgan Kaufmann, San Francisco, 175-187.
12. José del R. Millán, Rüdiger Rupp, Gernot Müller-Putz, Roderick Murray-Smith, Claudio Giugliemma, Michael Tangermann, C. Vidaurre, Febo Cincotti, Andrea Kübler, Robert Leeb, Christa Neuper, Klaus-Robert Müller and Donatella Mattia. 2010. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in Neuroscience* 4, 161.
13. Gernot Müller-Putz, Robert Leeb, Michael Tangermann, Johannes Höhne, Andrea Kübler, Febo Cincotti, Donatella Mattia, Rüdiger Rupp, Klaus-Robert Müller, and José del R. Millán. 2015. Towards Noninvasive Hybrid Brain-Computer Interfaces: Framework, Practice, Clinical Application, and Beyond. *Proceedings of the IEEE* 103, 6: 926-943.
14. Masaki Nakanishi, Yijun Wang, Xiaogang Chen, Yu-Te Wang, Xiaorong Gao, and Tzyy-Ping Jung. 2018. Enhancing Detection of SSVEPs for a High-Speed Brain Speller Using Task-Related Component Analysis. *IEEE Transactions on Biomedical Engineering*, 65, 1: 104-112.
15. Diederick C. Niehorster, Li Li, and Markus Lappe. 2017. The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research. *i-Perception* 8, 3: 1-23.
16. Gert Pfurtscheller, Brendan Z. Allison, Clemens Brunner, Gunther Bauernfeind, Teodoro Solis-Escalante. 2010. The hybrid BCI. *Frontiers in Neuroscience* 4, 30.
17. Thorsten Roth, Martin Weier, Andre Hinkenjann, Yongmin Li, and Philipp Slusallek. 2016. An analysis of eye-tracking data in foveated ray tracing. *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)*: 69-73.
18. Kelly Tai, Stefanie Blain, and Tom Chau. 2008. A Review of Emerging Access Technologies for Individuals With Severe Motor Impairments. *Assistive Technology* 20, 4: 204-221.
19. Tobii Pro Accuracy and precision test method for remote eye trackers: https://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-accuracy-and-precisiontest-method-version-2-1-1.pdf
20. Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. Mcfarland, Gert Pfurtscheller, and Theresa M. Vaughan. 2002. Brain-computer interfaces for communication and control. *Clinical Neurophysiology* 113, 6: 767-791.