Visually Stimulated Brain–Computer Interfaces Compete With Eye Tracking Interfaces When Using Small Targets

Kaori Suefusa[1] and Toshihisa Tanaka[1,2]

***Abstract*— Visually stimulated brain–computer interfacing detects which target on a screen a user is gazing at; however, this is also accomplished by tracking gaze points with a camera. These two approaches have been independently investigated and sometimes doubts about BCI with visual stimuli are raised in terms of usability compared to eye tracking interfaces (ETI). This paper answers this question by investigating information transfer rates (ITR) and recognition accuracies of BCI and ETI having a similar interface design, where subjects were asked to gaze at one of four targets on a screen. Experimental results revealed that BCI is comparable in ITR to ETI and had better performance for relatively small targets on the screen.**

I. INTRODUCTION

Brain–computer interfacing (BCI) is an emerging and potential application of biomedical signal processing and machine learning in human computer interaction. BCI controls a computer or a device by capturing human brain activity [1]. This technology provides another way of communication for those who have difficulty in communicating with the external world, e.g. people suffering from serious movement disorders such as amyotrophic lateral sclerosis (ALS).

To implement command input, BCI captures brain activity and/or response with instruments. A widely-used noninvasive recording of the brain activity is electroencephalography (EEG). Typical responses of the brain measured with EEG are steady-state visual evoked potentials (SSVEP) which are responses of the visual cortex to a periodic visual stimulus such as flickering lights [2], event related potentials (ERP) which are responses to sensory or cognitive event, and so forth. Among them, SSVEP allows BCIs to achieve fast and accurate command input, and various BCIs based on SSVEP have been reported [2]. SSVEP consists of periodic signals with the same and multiples of frequencies of the visual stimulus. A typical SSVEP-based BCI displays multiple visual stimuli whose flickering frequencies differ from one to another, and a user gazes at one of them. Then, with proper signal processing, the BCI recognizes which stimulus a user is gazing at based on the frequency of SSVEP. For the visual stimulus, a checkerboard, which reverses its phase at equal time intervals, is commonly used .

On the other hand, which target on a screen a user gazes at can be also obtained by tracking the point of gaze, i.e. where the user is looking with cameras. This is a so-called eye tracking interface (ETI) which is considered to be fairly robust [3] and a promising component of the user interface in the future [4] because of its inherent advantages such as ease of use and speed [3]. Methods for implementing command selections include the ones to select a target by gazing or dwelling for more than some fixed time [3], making gestures with the gaze [5], blinking or winking, using a physical button [3], or using eye movements with manual pointing which is called MAGIC pointing [6]. Among them, the one to input a command by gazing at a target – known as dwelling is one of the most straightforward and commonly-used ways, also it does not need any physical movement of limbs.

Both BCI and ETI are the user interfaces with sight, their experimental results have been widely reported, and it is known that both of them can achieve high precision of recognition accuracy [3], [2]. However, to our knowledge there have been no attempts to quantitatively compare their performances on the same experimental platform. Therefore, the purpose of this study is to compare their performances, and to clarify their drawbacks and advantages. Specifically, we compare between a SSVEP-based BCI and a dwelling-based ETI. We evaluate their performance by investigating their accuracies and ITR [7] with respect to the target size and the command analysis time, i.e. time window length of EEG analysis or the dwell time.

II. METHODS

A. Subjects and experimental settings

Five males and one female in their twenties took part in our experiment. All subjects were healthy and had normal or corrected-to-normal vision. They were given an informed consent, and the study was approved by the research ethics committee of Tokyo University of Agriculture and Technology.

1) BCI: We used Ag/AgCl active electrodes which are products of Guger Technologies (g.tec) named g.LADYbird, g.LADYbirdGND (for GND), and g.GAMMAearclip (for reference, earclip type) for recording EEG data. These were driven by the power supply unit named g.GAMMAbox (g.tec). The electrodes were located at Pz, Oz, O1 and O2 following the international 10–20 system. The electrodes for GND and reference were AFz and A1, respectively. The signals were amplified by MEG-6116 (Nihon Kohden), that provides lowpass and highpass analog filters for each channel. We set the cutoff frequencies of the lowpass and the highpass filters to 100 Hz and 0.5 Hz, respectively. The EEG signal was sampled by A/D converter (AIO-163202F-PE,

*This work is supported in part by JSPS KAKENHI Grant Number 24360146.

[1]K. Suefusa and T. Tanaka are with Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, 2–24–16 Nakacho, Koganei-shi, Tokyo, 184-8588, Japan suefusa@sip.tuat.ac.jp, tanakat@cc.tuat.ac.jp

[2]T. Tanaka is also affiliated with RIKEN Brain Science Institute, Saitama, Japan

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

TABLE I
FLICKERING FREQUENCIES ASSIGNED TO BCI COMMANDS AND CORRESPONDING TO VISUAL TARGETS ON THE SCREEN.

Target Tk	Location	Frequency f_k
T1	Top	9.2 Hz
T2	Right	10.9 Hz
T3	Bottom	12 Hz
T4	Left	13.3 Hz

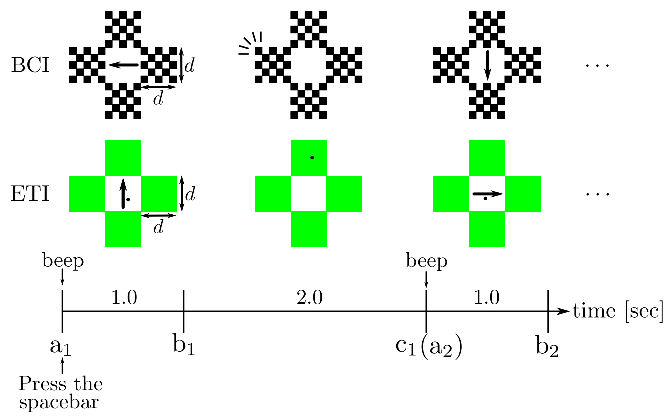


Fig. 1. Time sequence of targets presentation in a trial and displayed targets. A black dot on the targets shown in ETI targets is a gaze point of a user. An arrow in the center of the targets prescribed the user which target to gaze at. In the experiments, three different sizes, d = 20, 40, and 60 mm were used for comparison.

Contec) with a sampling rate of 1200 Hz. The signals were recorded and downsampled to 240 Hz with Data Acquisition Toolbox of the MATLAB (MathWorks).

2) ETI: EyeFrame SceneCamera System (Arrington Research) was used as an eye tracker, which is wearable as frames of glasses and its sampling rate was 60 Hz. Dark pupil technique [8] was used to determine the eye orientation. In this method, an infrared light source was positioned, and it made the iris appear light and the pupil the darkest region in the image. We performed the calibration using Auto-Calibrate of ViewPoint EyeTracker (Arrington Research), which is required to map the eye orientation as the point of gaze on the screen. During the calibration, subjects were asked to fixate on a 3×3 grid of points that are displayed one at a time in random order. Next, we recorded the xy coordinates of the user's gaze point using MATLAB (MathWorks). Throughout the experiment, subjects' heads were fixed using a chin rest. In order to plot a black dot with a diameter of 2 mm as the feedback of the user's gaze point, we used Psychtoolbox of the MATLAB.

B. Design

Targets shown in Fig. 1 were drawn with Psychtoolbox on a display screen. For the BCI experiment, we used a desktop computer connected to a 23 inch display with a resolution of 1920×1080 and a refresh rate of 120 Hz. As illustrated in Fig. 1, the targets were square checkerboards that reverse black and white according to the frequencies shown in Table I. The size of each small square of the checkerboard was 4 mm (visual angle of 0.32 deg). For the ETI experiment, we used a laptop computer connected to a 15.6 inch display with a resolution of 1366×768 and a refresh rate of 60 Hz. During both experiments, subjects sat on a comfortable chair in front of the screen 70 cm away so that they could look at the display straight ahead.

As illustrated in Fig. 1, four square targets were displayed on top, bottom, right and left on the screen. We referred to a target as Tk in clockwise order from the top, thus a target on top is T1 and a target on the left is T4 as listed in Table I. Three different target sizes used in the experiments were 20 mm, 40 mm, and 60 mm on a side, which can be converted to visual angles of 1.6 deg, 3.3 deg, and 4.9 deg. As shown in Fig. 1, an arrow was also displayed in the center of the targets. The arrow prescribed the user which target to gaze at. The distance between the center of the arrow and each target was the same as the target size.

C. Task

Each trial consisted of the fixation for one second and the gaze of the target for two seconds as illustrated in Fig. 1. One command was determined every single trial (e.g., the interval of a1–c1 in Fig. 1 for the first trial). At the beginning of a trial, the arrow was displayed with a beep sound for a duration of 0.1 seconds. After one second, the arrow was hidden for two seconds. Subjects were asked to maintain fixation on the arrow until they started their response, and to start to gaze at the prescribed target as quickly as possible following the disappearance of the arrow. The arrow prescribed each target equally and randomly. In the experiment, three different target sizes (aforementioned 20 mm, 40 mm, and 60 mm) were used for comparison. For each target size, the calibration was performed first, then four sessions were executed. Each session included five successive trials. Therefore, twenty commands were inputted for each size. After each session, the task was stopped in order to reduce the burden of the subjects' eyes. The task was restarted when subjects pressed the space bar.

In addition, as illustrated in Fig. 1, during the BCI experiment, visual stimuli were not flickering while the arrow was displayed, and visual stimuli started flickering soon after the disappearance of the arrow.

Moreover, during the ETI experiment, we performed the calibration every time we changed the target size, since the subjects' heads and eye tracker frames stirred easily, which materially affected mapping the gaze point on the screen.

D. Offline data analysis

We analyzed the recordings (EEG for the BCI and gaze positions for the ETI) offline and detected entered commands. For performance comparison, we calculated the ITR in bits/min defined [1] as:

$$\mathrm{ITR} = \frac{60}{U}\left[\log_2 K + P\log_2 P + (1-P)\log_2 \frac{1-P}{K-1}\right], \quad (1)$$

where U [sec] is the mean time to input one command, K is the number of selectable commands ($K = 4$), and P is the

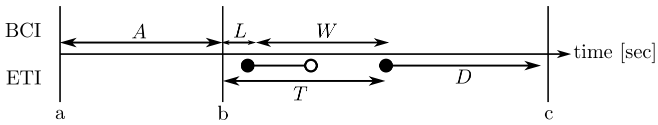


Fig. 2. Data analysis procedure. For the ETI, a closed circle represents a moment when the gaze point enters a target, and the open circle represents a moment when it exits the target. a, b, and c are corresponding to those in Fig. 1.

accuracy, i.e. the probability that the recognized results meet the desired command.

1) BCI: To detect commands, the EEG for W seconds while visual stimuli were flickering was analyzed. We neglected samples just after flickering started due to a delay of SSVEP onset [9] as described later.

We used the method based on canonical correlation analysis (CCA) proposed by Lin *et al.* [10]. Let $\boldsymbol{x}(t) \in \mathbb{R}^M$ be an M-channel EEG signal and $\boldsymbol{y}(t) \in \mathbb{R}^{4 \cdot 4}$ consist of 'Fourier basis functions' of the 1st and 2nd harmonics of simulated stimulus signals, which are ideal SSVEP with frequency f given as

$$\boldsymbol{y}(t) = \left[\{\sin(2\pi f_k t), \cos(2\pi f_k t), \sin(4\pi f_k t), \cos(4\pi f_k t)\}_{k=1}^{4}\right]^{\mathrm{T}}, \tag{2}$$

where the first two components are the sinusoids of the fundamental frequency f and the others are the 2nd harmonics. To detect frequencies of SSVEP components contained in the EEG for the SSVEP-based BCI systems, first, canonical correlation ρ_f corresponding to flickering frequency f is calculated:

$$\rho_f = \max_{\boldsymbol{w}_{\boldsymbol{x}}, \boldsymbol{w}_{\boldsymbol{y}}} \frac{\boldsymbol{w}_{\boldsymbol{x}}^T E[\boldsymbol{x}(t)\boldsymbol{y}^T(t)]\boldsymbol{w}_{\boldsymbol{y}}}{\sqrt{\boldsymbol{w}_{\boldsymbol{x}}^T E[\boldsymbol{x}(t)\boldsymbol{x}^T(t)]\boldsymbol{w}_{\boldsymbol{x}} \boldsymbol{w}_{\boldsymbol{y}}^T E[\boldsymbol{y}(t)\boldsymbol{y}^T(t)]\boldsymbol{w}_{\boldsymbol{y}}}}. \tag{3}$$

Then, frequency $f^\star$ that maximizes the weight vector with respect to $\boldsymbol{y}(t)$ is chosen, such as

$$f^\star = \operatorname*{argmax}_{f_k \in \Omega} \left[\left(w_{f_k}^{\mathrm{s}}\right)^2 + \left(w_{f_k}^{\mathrm{c}}\right)^2 + \left(w_{2f_k}^{\mathrm{s}}\right)^2 + \left(w_{2f_k}^{\mathrm{c}}\right)^2\right], \tag{4}$$

where Ω is the set of frequencies listed in Table I, and $w_{f_k}^{\mathrm{s}}$, $w_{f_k}^{\mathrm{c}}$, $w_{2f_k}^{\mathrm{s}}$ and $w_{2f_k}^{\mathrm{c}}$ are the elements of the weight vector with respect to $\boldsymbol{y}(t)$, such as

$$\boldsymbol{w}_{\boldsymbol{y}} = \left[\left\{w_{f_k}^{\mathrm{s}}, w_{f_k}^{\mathrm{c}}, w_{2f_k}^{\mathrm{s}}, w_{2f_k}^{\mathrm{c}}\right\}_{k=1}^{4}\right]^{\mathrm{T}}. \tag{5}$$

We measured the ITR with respect to two independent variables, the target size (d [mm]), and time window length of EEG signal analysis (W [sec]). To calculate the ITR, we defined U in (1) as follows:

$$U = A + L + W, \tag{6}$$

where as summarized in Fig. 2, A [sec] is the duration the arrow is displayed, L [sec] is the time lag (the delay) of SSVEP onset after the visual stimuli started flickering. We set $A = 1.0$ sec, $L = 0.1$ sec, and $W = 0.5, 0.6, \ldots, 1.8, 1.9$ sec.

2) ETI: A command was determined if the user's gaze has dwelled on the same target for more than the predefined dwell time, D [sec], during the interval from b to c as shown in Fig. 2. If the gaze point got out from the target, we reset the dwell timer. If the gaze did not dwell on any targets for more than the dwell time in the interval between b and c in Fig. 2, we regarded the detection as a failure. We measured the ITR with respect to two independent variables, the target size (d [mm]), and the dwell time (D [sec]). To calculate the ITR, we defined U in (1) as follows:

$$U = A + \overline{T} + D, \tag{7}$$

where

$$\overline{T} = \frac{1}{N}\sum_{i=1}^{N} T_i, \tag{8}$$

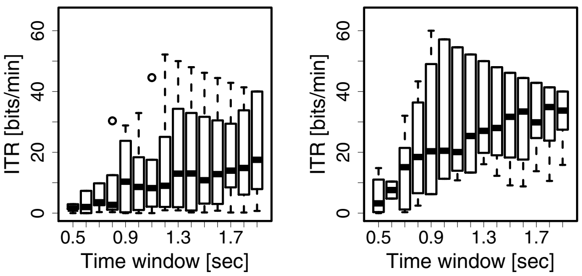
where as summarized in Fig. 2, A [sec] is the duration the arrow is displayed, T_i [sec] is the time until the gaze point enters a target after the arrow was hidden, and N is a number of inputted commands for each size. We set $A = 1.0$ sec, $D = 0.5, 0.6, \ldots, 1.8, 1.9$ sec, and $N = 20$.

3) Remarks on time window and dwell time: Though the shortest time for W and D was set to 0.5, the offline data analysis could allow us to a shorter time window and a dwell time less than 0.5 sec. However, we did not take such a short time, since in practical interfaces, a very short time interval for command input can lead to a unneeded repeat of the same command like a "press-and-hold" keyboard. The duration of 0.5 sec was chosen from the default duration before a press turns into a long press in Android [11].

III. RESULTS AND DISCUSSION

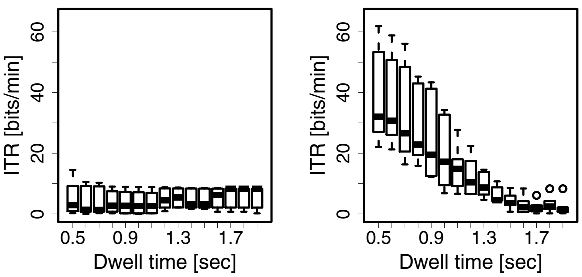
Figs. 3 and 4 show box plots of the distributions of ITR of the BCI and the ETI respectively. The horizontal axis shows the command analysis time (time window, W for the BCI and dwell time, D for the ETI), and the vertical axis shows the ITR. Moreover, (a) and (b) in Figs. 3 and 4 show the results when the target size was 20 mm and 60 mm respectively. The band inside the box represents the median. We omit the results when the target size was 40 mm because of lack of space.

Figs. 3 and 4 illustrate that in terms of the target size, the larger target ($d = 60$ mm) showed higher ITR than the smaller one ($d = 20$ mm) for both the BCI and the ETI. For BCIs, a previous study [12] also indicated that visual evoked potential (VEP) response increased when stimulus field (target size) was enlarged. For ETIs, other studies [3], [4] also indicated that the time to input each command and error rate decreased when the target size was enlarged. On the other hand, in the case of the smaller target ($d = 20$ mm), ITRs of the ETI were consistently low (less than 10 bits/min) regardless of the dwell time, while the BCI showed better performance in ITR as depicted in Fig. 3(a). This result quantitatively suggests that the ETI needs sufficiently large targets, and the BCI has an advantage in the target size over the ETI.



(a) Target size d = 20 mm (b) Target size d = 60 mm

Fig. 3. Box plots of the ITR of the BCI



(a) Target size d = 20 mm (b) Target size d = 60 mm

Fig. 4. Box plots of the ITR of the ETI

TABLE II

Results of recognition accuracy and ITR when the target size was 60 mm. The BCI and the ETI showed similar performance in ITR with the t-test; $p = 0.4249$.

	Accuracy [%]		ITR [bits/min]	
	Mean ± S.D.	Median	Mean ± S.D.	Median
BCI (W = 1.8 sec)	90.0 ± 12.6	95.0	31.2 ± 11.5	34.9
ETI (D = 0.5 sec)	85.8 ± 8.9	82.5	38.1 ± 14.6	32.1

Another aspect is the analysis time (time window for the BCI and dwell time for the ETI). It can be seen from these figures that inappropriate analysis time yields low values of ITR. However, it is worth noting that the behavior of ITR to the analysis time is completely different as seen in Figs. 3(b) and 4(b). The BCI had higher ITR when the time window was longer. This result meets the previous studies [10], [13] Also, longer dwell time was less suitable, as claimed in the previous works [4], [14].

It is important to note that as summarized in Table II, with the larger target (d = 60 mm), the median ITR achieved 34.9 bits/min at W = 1.8 sec, which is slightly higher than the highest median ITR of the ETI (32.1 bits/min at D = 0.5 sec) with the same target size of 60 mm (See Fig. 4(b)). Therefore, both interfaces appeared similar in terms of ITR. Indeed, the result of the t-test showed no significant difference as stated in Table II. Also, it has been suggested [15] that the average time window required by SSVEP-based BCIs using CCA was 1.8 sec, which was the same as the time window length of W = 1.8 sec that gave the best result.

IV. Conclusion

A comparison study on the similar design of brain–computer interfaces (BCI) and eye tracking interfaces (ETI) was conducted in this paper. Through the experiment of visual target selection tasks, we proved that SSVEP-BCI was comparable to ETI in ITR.

We observed that the ETI had higher ITR than BCI when the command analysis time (the dwell time) was short, while BCI had higher ITR than the ETI when the target size was small. From these results, we would recommend the selection of either BCI or ETI based on the size of the screen that displays targets. If it is possible to use a large screen, ETI can achieve fast input of commands. On the other hand, if the size of the screen is highly limited as in the case of smartphones and tablets, BCI would be a proper choice.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, Jun. 2002.

[2] M. Middendorf, G. McMillan, G. Calhoun, and K. S. Jones, “Brain-computer interfaces based on the steady-state visual-evoked response,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 211–4, Jun. 2000.

[3] C. Ware and H. H. Mikaelian, “An evaluation of an eye tracker as a device for computer input,” in *Proc. SGICHI/GI Conf. Human Factors Computing Syst. and Graph. Interface*, May 1987, pp. 183–188.

[4] A. M. Penkar, C. Lutteroth, and G. Weber, “Designing for the eye — design parameters for dwell in gaze interaction,” in *Proc. 24th Australian Comput.-Human Interaction Conf.*, Nov. 2012, pp. 479–488.

[5] A. Hyrskykari, H. Istance, and S. Vickers, “Gaze gestures or dwell-based interaction?” in *Proc. Symp. Eye Tracking Research and Applicat.*, Mar. 2012, pp. 229–232.

[6] S. Zhai, C. Morimoto, and S. Ihde, “Manual and gaze input cascaded (MAGIC) pointing,” in *Proc. SIGCHI Conf. Human Factors Comput-ing Syst.*, ser. CHI ’99, May 1999, pp. 246–253.

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

[8] D. Li, J. Babcock, and D. J. Parkhurst, “Openeyes: A low-cost head-mounted eye-tracking solution,” in *Proc. Symp. Eye Tracking Research & Applicat.*, Mar. 2006, pp. 95–100.

[9] H. Bakardjian, T. Tanaka, and A. Cichocki, “Optimization of SSVEP brain responses with application to eight-command brain–computer interface,” *Neuroscience Letters*, vol. 469, no. 1, pp. 34–38, Jan. 2010.

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

[11] Viewconfiguration — android developers. [Online]. Available: http://developer.android.com/reference/android/view/ViewConfiguration.html

[12] G. Bartl, G. H. Van, Lith, and G. W. Van Marle, “Cortical potentials evoked by a TV pattern reversal stimulus with varying check sizes and stimulus field,” *British J. of Ophthalmology*, vol. 62, no. 4, pp. 216–219, Apr. 1978.

[13] Y. Zhang, G. Zhou, Q. Zhao, A. Onishi, J. Jin, X. Wang, and A. Cichocki, “Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCIs,” in *Neural Inform. Process.*, ser. Lecture Notes in Computer Science, B.-L. Lu, L. Zhang, and J. Kwok, Eds. Springer Berlin Heidelberg, 2011, vol. 7062, pp. 287–295.

[14] R. J. K. Jacob, “The use of eye movements in human-computer interaction techniques: What you look at is what you get,” *ACM Trans. Inf. Syst.*, vol. 9, no. 2, pp. 152–169, Apr. 1991.

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