33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A vision-free brain-computer interface (BCI) paradigm based on auditory selective attention

Do-Won Kim, Jae-Hyun Cho, Han-Jeong Hwang, Jeong-Hwan Lim, and Chang-Hwan Im

***Abstract*— Majority of the recently developed brain computer interface (BCI) systems have been using visual stimuli or visual feedbacks. However, the BCI paradigms based on visual perception might not be applicable to severe locked-in patients who have lost their ability to control their eye movement or even their vision. In the present study, we investigated the feasibility of a vision-free BCI paradigm based on auditory selective attention. We used the power difference of auditory steady-state responses (ASSRs) when the participant modulates his/her attention to the target auditory stimulus. The auditory stimuli were constructed as two pure-tone burst trains with different beat frequencies (37 and 43 Hz) which were generated simultaneously from two speakers located at different positions (left and right). Our experimental results showed high classification accuracies (64.67%, 30 commands/min, information transfer rate (ITR) = 1.89 bits/min; 74.00%, 12 commands/min, ITR = 2.08 bits/min; 82.00%, 6 commands/min, ITR = 1.92 bits/min; 84.33%, 3 commands/min, ITR = 1.12 bits/min; without any artifact rejection, inter-trial interval = 6 sec), enough to be used for a binary decision. Based on the suggested paradigm, we implemented a first online ASSR-based BCI system that demonstrated the possibility of materializing a totally vision-free BCI system.**

I. Introduction

Brain-computer interface (BCI, sometimes referred to as brain-machine interface) is a technology that translates brain signals into simple commands that can control external devices or into messages with which one can communicate [1]. The major targets of BCI systems have been disabled individuals who cannot freely move or control specific parts of their body because of serious neurological disease or injury, such as amyotrophic lateral sclerosis (ALS, also referred to as Lou Gehrig's disease) or brainstem stroke.

Manuscript received March 23, 2011. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0015957).

D.-W. Kim is with the Department of Biomedical Engineering, Yonsei University, Wonju, Kangwon-do, Republic of Korea (e-mail: dowon.kim@yonsei.ac.kr).

J.-H. Cho is with the Department of Biomedical Engineering, Yonsei University, Wonju, Kangwon-do, Republic of Korea (e-mail: jaehyun.cho@yonsei.ac.kr).

H.-J. Hwang is with the Department of Biomedical Engineering, Yonsei University, Wonju, Kangwon-do, Republic of Korea (e-mail: han-jeong@yonsei.ac.kr).

J.-H. Lim is with the Department of Biomedical Engineering, Yonsei University, Wonju, Kangwon-do, Republic of Korea (e-mail: gobbili1@hanmail.net).

C.-H. Im is with the Department of Biomedical Engineering, Hanyang University, Seoul, Korea (corresponding author; Tel.: +82-2-2220-2322; Fax.: +82-2-2296-5943; e-mail: ich@hanyang.ac.kr).

To translate the neural signals acquired from the patients into appropriate commands, various experimental paradigms and tasks have been introduced, such as P300 speller [2]-[4]; steady state neural responses elicited while one is gazing a certain visual stimulus flickering with a specific frequency [5]-[7]; mental tasks associated with motor imagery [8]-[10]; or mental calculation [11], [12]. Most of the paradigms listed above use visual stimuli, visual feedback, or both, and are thereby applicable only to patients whose visual function is not impaired.

In practice, however, some patients with severe neurological disorders, such as ALS and completely locked-in state (CLIS), often have difficulty controlling their voluntary extraocular movements or fixing their gaze on specific visual stimuli. Even for those who have normal visual function, gazing at stimuli for a long time can easily cause fatigue or loss of concentration. A recent experimental study demonstrated that the performance of the P300-based speller paradigm can be substantially influenced by eye gaze [13], which strongly suggests that the use of visual stimuli or cues might not be appropriate for those who have difficulty in gazing at specific target stimuli. Therefore, developing new BCI paradigms that are not dependent on visual stimuli remains one of the challenging issues in modern BCI research [14].

One of the alternative paradigms to build a vision-free BCI system has been auditory based ones. One of the previous research conducted by Lopez et al. [15] investigated whether the auditory steady-state response (ASSR) is modulated by auditory selective attention (ASA) to a specific sound stream and discussed the possibility of using the ASSR as a new BCI paradigm. In six out of eight participants, the spectral density of alpha rhythm was inversely proportional to that of the modulation frequency for the left ear (38 Hz), providing evidence that selective attention can modulate ASSR. They also showed, using the self-organizing map (SOM) method, that the attended and ignored conditions could be clearly classified into two clusters, demonstrating the possibility of using ASSR modulated by auditory selective attention as a new BCI paradigm.

In the present study, inspired by the pilot study of Lopez et al. [15], we further investigated whether ASSR can be a feasible feature for a practical BCI system by implementing a modified BCI paradigm to classify one's auditory selective attention and by evaluating the classification accuracy of the BCI system. Indeed, to the best of our knowledge, our paradigm is one of the first auditory BCI paradigms that did not use any visual information during the entire experiment. Furthermore, we used the proposed paradigm and analysis methods to implement an online ASSR-based BCI system to further demonstrate whether our paradigm could be used as a successful BCI paradigm.

II. Methods

A. Participants

Six healthy volunteers (one female and five male, mean age 25.0 ± 5.0 years) with no neurological or psychiatric disorders or previous head injury were recruited among the graduate and undergraduate students in the Department of Biomedical Engineering of Yonsei University. Before the experiment, all participants were given a detailed, written summary of the experimental procedures. Participants signed a written consent and received adequate reimbursement for their participation. The study protocol was approved by the Institutional Review Board (IRB) of Yonsei University, Korea.

B. Auditory Stimuli

We chose two frequencies, 37 Hz and 43 Hz, as the modulation frequencies (beat frequencies in the present study). The carrier frequencies of the two auditory stimuli were set to 2.5 kHz and 1 kHz, respectively, so that the subjects could easily distinguish each sound stream [15]. We used pure tone burst trains; each generated using MATLAB (The MathWorks, Natick, MA, USA, Version 7.7.0) at a sampling rate of 44,100 Hz. The pulse widths of the 37 Hz and 43 Hz pure tone pulses were 13.5 ms and 11.6 ms, respectively. The duration of each trial was 20 seconds.

C. Experimental Protocols

Participants sat in a comfortable armchair in front of a pair of speakers and were asked to adjust the position of the chair to a comfortable location while maintaining equal distance (less than 60 cm from the speakers) from the two speakers (see Fig. 1). In each trial, the participants were presented with 2.5 kHz tone burst trains with 37 Hz beat frequency for their left sound field and 1 kHz tone burst trains with 43 Hz beat frequency for their right sound field. Subjects were asked to close their eyes and remain as still as possible, particularly during the acquisition intervals.

One segment of the auditory stimulus lasted for 20 seconds and a random interval of 6–10 seconds was inserted between each trial. Two seconds before the stimulus onset, five pulses of pure tone sounds were generated randomly from either the left or right side, to indicate which sound source they were to concentrate on. The five pulses of pure tone sounds had the same carrier frequency and beat frequency as the main auditory stimulus (2.5 kHz carrier frequency and a 37 Hz beat frequency for the left sound field; 1 kHz carrier frequency and a 43 Hz beat frequency for the right sound field) to help the participants recognize the direction of the stimulus more accurately. Our paradigm was implemented with TeleScan 2.2 for Windows (Laxtha, Inc., Daejoen, Korea), which was also used for the EEG data acquisition.

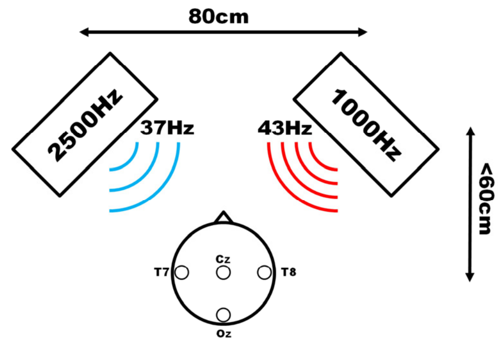

Fig. 1. Overall experimental environment; A schematic diagram to elucidate the experimental environment. Two speakers were placed 80 cm apart. The participants were asked to adjust the position of the chair to a comfortable location while maintaining equal distance (less than 60 cm from the speakers) from the two speakers. The participants were presented with 2.5 kHz tone burst trains with a 37 Hz beat frequency for their left ear and 1 kHz tone burst trains with a 43 Hz beat frequency for their right ear.

Each session consisted of 25 trials and lasted for approximately 10 minutes. Before the recording, one training session was performed to familiarize participants with the paradigm. The main experiment was performed in two sessions with a 10-minute inter-session rest. In total, we acquired EEG data sets for 50 trials: 25 for selective attention to the left-sided stimulus and the other 25 trials for selective attention to the right-sided stimulus.

D. Data Acquisition and Processing

Electrodes were attached on the participants' scalp according to the international 10-20 system. The EEG signals were acquired at four electrodes (Cz, Oz, T7, T8), which represent the motor, visual, and auditory cortical areas, using a multi-channel EEG acquisition system (WEEG-32, Laxtha Inc., Daejeon, Korea) in a dimly lit, soundproof room. The sampling rate was set at 512 Hz in all experiments. The ground electrode was placed behind the left ear with the reference electrode on the opposite side.

The raw EEG data were segmented into 20-second epochs from the beginning of the main auditory stream. No preprocessing methods, such as re-referencing, band-pass filtering, or artifact rejection, were applied to the present analysis. The frequency spectrums of each epoch were calculated using the fast Fourier transform (FFT) algorithm with a 1 second long sliding window with a 50% overlap. The estimated frequency spectrums were accumulated and averaged over time for each epoch.

E. Feature Selection and Classification

As candidates of feature vectors, we first evaluated the EEG spectral densities of each electrode averaged over 37 ± 1 Hz (denoted as Cz_{37}, Oz_{37}, $T7_{37}$, $T8_{37}$) and 43 ± 1 Hz (Cz_{43}, Oz_{43}, $T7_{43}$, $T8_{43}$). We also evaluated the ratios between all possible pairs of spectral densities evaluated at the same modulation frequency ($Cz_{37}/T7_{37}$, $Cz_{37}/T8_{37}$, Cz_{37}/Oz_{37}, $T7_{37}/T8_{37}$, $T7_{37}/Oz_{37}$, $T8_{37}/Oz_{37}$, $Cz_{43}/T7_{43}$, $Cz_{43}/T8_{43}$, Cz_{43}/Oz_{43}, $T7_{43}/T8_{43}$, $T7_{43}/Oz_{43}$, $T8_{43}/Oz_{43}$) as well as the ratios between the spectral powers of each electrode evaluated at different modulation frequencies (Cz_{37}/Cz_{43}, $T7_{37}/T7_{43}$, $T8_{37}/T8_{43}$, Oz_{37}/Oz_{43}).

To investigate the changes in classification accuracy with respect to the number of feature vectors, we calculated the classification accuracy for all possible combinations of the 24 feature candidates listed above, assuming the number of selected features to be one, two, or three. To show the influence of the analysis window sizes (or analysis interval sizes) on the classification accuracy, we also tested different analysis window sizes (2 - 20 seconds from the main auditory stimulus onset with a step size of 1 second).

For the classification, we used a 10-fold cross-validation method considering the small number of trials. We first divided the 50 trials into 10 equal-size folds, and for each validation 45 trials were used as a reference data set and the other 5 trials were used as a test set. For each trial of the test set, Euclidean distances from the average feature vectors (each averaged to the left and right stimuli) computed on the reference data set were compared, and the trial was assigned to a class based on whichever had the shorter distance. The cross-validation was done separately for each of all possible feature sets.

F. Online Experiment Procedure

We also implemented a pilot online ASSR-based BCI system and tested it to one of the participants (JP, female, 24 years old). Right before the online experiment, we selected an optimal feature set from a preliminary offline experiment. The experimental paradigm and analysis methods used for the feature selection were identical to those of previous offline experimental studies, except that the location of the reference electrode was moved from left ear to the participants' forehead. This change was made to avoid the potential influence of the reference electrode on the laterality between the electrodes T7 and T8. Since the participant was asked to close her eyes during the entire offline and online experiments, we confirmed that EOG artifact did not affect the recorded signals.

In our online experiment, the participant was instructed verbally to attend one of the auditory stimuli, left stimulus or right stimulus, in a random order. After the instruction was made, the experimenter manually turned on a switch that starts generating two different tone burst trains from speakers located on the left and right sides of the participant. Then, the main computer system started recording the EEG signals, and at the same time calculated the values of 3 feature vectors. After 10 seconds from the beginning of the recording, our BCI system classified the participant's selective attention in real time and displayed the decision on the monitor screen so that the instructor can evaluate the result. All the analysis methods were identical to those used in the offline analyses.

III. Results

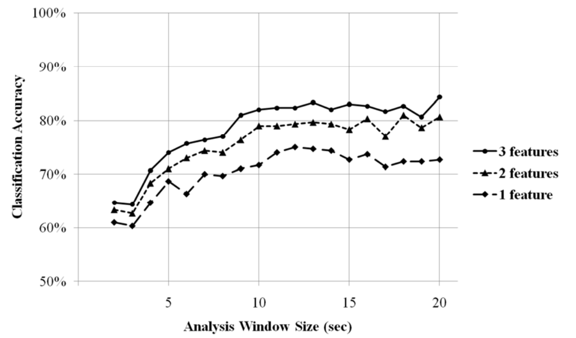

Fig. 3. Classification accuracy averaged over six participants with respect to different analysis window sizes and different numbers of feature vectors (1, 2, and 3).

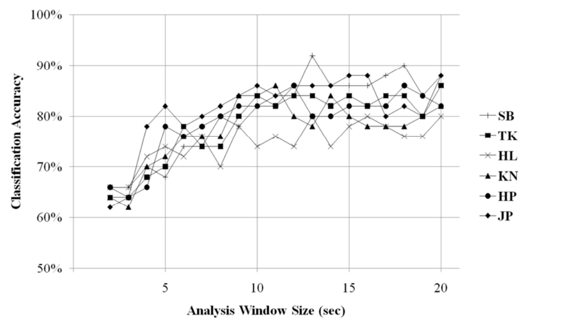

Fig. 4. Classification accuracy for each participant with respect to the analysis window sizes when three feature vectors were selected.

Fig. 3 shows the variations in classification accuracy averaged over the six participants with respect to the analysis window sizes and the number of feature vectors. We observed that higher classification accuracy could be obtained when larger numbers of feature vectors were used for the classification. The classification accuracy nearly monotonically increased with respect to the analysis window sizes, but after approximately 10 seconds the accuracy no longer increased. Since short analysis window size guarantees more possible commands per minute, the analysis window size of approximately 10 seconds was the most appropriate.

Fig. 4 shows variations in the classification accuracy evaluated for each participant with respect to the analysis window size when three feature vectors were selected. The six individual graphs show similar and consistent shape with those of Fig. 3 and show very small differences in the overall averaged classification accuracies, with a standard deviation of only 2.11%. The maximum classification accuracy of each participant was found at different analysis window sizes and varied from 80% to 92%. The average of the maximum classification accuracy of each subject was 86.33 ± 3.54%, and the analysis window size that resulted in the highest accuracy was 14.00 ± 2.94 seconds. Subject SB showed the highest classification accuracy value (92%) among all of the participants, while subject JP showed the highest overall classification accuracy (81.26%). Although our cross-validation results using small number of trials might be somewhat biased for specific feature sets, the high classification accuracy consistently exceeding the chance level (50%) demonstrates the possibility of using ASSR for the binary decision of BCI.

The online experiment consisted of 14 continuous trials (7 for right stimulus and 7 for left stimulus) and showed a fair classification accuracy of 71.4%. Since the participant was asked to close her eyes during the experiment, she could not have any information on whether the previous decision was right or wrong. In our pilot online experiment, we did not provide the participant with any feedbacks as they might affect her attention. The readers can watch the full video of our online experiment at http://cone.hanyang.ac.kr/BioEST/Kor/research/update/ASSR_BCI.avi.

IV. Discussions

In the present study, we investigated whether ASSR modulated by selective attention to a specific sound stream can be used to create a practical auditory BCI system, with the goal of classifying the intentions of individuals who have difficulty in controlling their vision. Inspired by the conventional SSVEP-based BCI paradigms that use multiple spatially separated visual stimuli with different flickering frequencies, we presented the participants with multiple spatially separated auditory stimuli with different tones and modulation frequencies. In our experiments performed to six healthy volunteers, we were able to discriminate which sound source the participants were selectively attending to with high classification accuracy fairly exceeding the chance level of a binary decision (50%), demonstrating the feasibility of using ASSR modulated by selective attention as one of the promising BCI features.

Our paradigm has several advantages that are suitable for use in practical BCI systems. First, we did not apply any complex preprocessing procedures. In fact, we did not even use basic filtering or artifact rejection processes, which would be advantageous in realizing an efficient real-time BCI system. Second, since the paradigm was simple and intuitive, the participants could easily understand and get accustomed to the target tasks, for which they were only asked to concentrate their attention on either the left or right sound source. Therefore, the proposed paradigm overcomes one of the drawbacks of mental task-based BCI paradigms that require complex and time-consuming training processes. Moreover, we did not use any visual information during the whole experiment, considering that the main targets of auditory BCI systems would be patients with advanced ALS or CLIS, who have difficulty controlling visual fixation.

References

[1] J.R. Wolpaw, N. Birbaumer , D.J. McFarland and G. Pfurtscheller, T.M. Vaughan, “Brain-computer interfaces for communication and control”, *Clin. Neurophysiol.*, vol. 113, pp. 767-91, 2002

[2] L. A. Farwell and E. Donchin, “Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials”, *Electroen. Clin. Neuro.*, vol. 70, pp. 510-523, 1988

[3] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan and J. R. Wolpaw, “Toward enhanced P300 speller performance”, *J. Neurosci. Meth.*, vol. 167, pp. 15-21, 2008

[4] E. W. Sellers and E. Donchin, “A P300-based brain-computer interface: Initial tests by ALS patients”, *Clin. Neurophysiol.*, vol. 117, pp. 538-48, 2006

[5] E.C. Lalor, S.P. Kelly, C. Finucane, R. Burke, R. Smith, R.B. Reilly and G. McDarby, “Steady-state VEP-based brain-computer interface control in an immersive 3D gaming environment”, *Eurasip. J. Appl. Sig. P.*, pp. 3156-3164, 2005

[6] Z.L. Lin, C.S. Zhang, W. Wu and X.R. Gao, “Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs”, *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 2610-2614, 2006

[7] M. Middendorf, G. McMillan, G. Calhoun and K.S. Jones, “Brain-computer interfaces based on the steady-state visual-evoked response”, *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 211-214, 2000

[8] J. Decety and D.H. Ingvar, “Brain structures participating in mental simulation of motor behavior: a neuropsychological interpretation”, *Acta. Psychol. (Amst)*, vol. 73, pp. 13-34, 1990

[9] M. Jeannerod and V. Frak, “Mental imaging of motor activity in humans”, *Curr. Opin. Neurobiol.*, vol. 9, pp. 735-739, 1999

[10] G. Pfurtscheller and C. Neuper, “Motor imagery activates primary sensorimotor area in humans”, *Neurosci. Lett.,* vol. 239, pp. 65-68, 1997

[11] Z. A. Keirn and J. I. Aunon, “Man-machine communications through brain-wave processing”, *IEEE Eng. Med. Biol. Mag.*, vol. 9, pp. 55-57, 1990

[12] W. D. Penny, S. J. Roberts, E. A. Curran and M. J. Stokes, “EEG-Based communication: A pattern recognition approach”, *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 214-215, 2000

[13] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof and G. Schalk, “Does the P300 Speller Depnd on Eye-Gaze?”, *J. Neural. Eng.*, vol. 7, 056013 (9pp), 2010

[14] F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N. Birbaumer and A. Kubler, “An auditory brain-computer interface (BCI)”, *J. Neurosci. Meth.*, vol. 167, pp. 43-50, 2008

[15] M. A. Lopez, H. Pomares, F. Pelayo, J. Urquiza, J. Perez, “Evidences of cognitive effects over auditory steady-state responses by means of artificial neural networks and its use in brain-computer interfaces”, *Neurocomputing*, vol. 72, pp. 3617-3623, 2009