Towards an Enhanced ERP Speller based on the Visual Processing of Face Familiarity

Seul-Ki Yeom, Siamac Fazli, Klaus-Robert Müller, and Seong-Whan Lee

***Abstract*— In this study, a novel P300 based brain-computer interface (BCI) system using random set presentation pattern and employing the effect of face familiarity has been proposed and developed. While the effect of face familiarity is widely studied in the cognitive neurosciences, it has so far not been addressed for the purpose of BCI. We compare P300-based BCI performances of a conventional row-column (RC)-based paradigm with our novel approach. Our experimental results indicate stronger deflections of the ERPs in response to face stimuli and thereby improving P300-based spelling performance. This leads to a significant reduction of stimulus sequences required for correct character classification. These findings demonstrate a promising new approach for improving the speed and thus fluency of BCI-enhanced communication with the widely used P300-based BCI setup.**

I. INTRODUCTION

Several types of EEG-based BCIs have been developed and can be categorized by the brain activity patterns, which they employ for BCI control: Event-related potentials (ERPs) [1], steady state visual evoked potentials (SSVEP) [2], [3], sensorimotor rhythms (SMRs) [4], and slow cortical potentials (SCP) [5].

The P300 is the positive ERP component, which occurs over the parietal cortex, approximately 300ms after a rare (surprising) but meaningful stimulus presentation among a series of many irrelevant stimuli (i.e. oddball paradigm) [6] and has been widely investigated over the past few years due to the high information transfer rates (ITRs), simplicity, and the need for small sample training data [7].

Since the first introduction of the P300-based matrix speller (aka P300 Speller) in 1988 [1] many extensions to the original RC paradigm have been proposed in order to improve its performance in terms of speed and accuracy. Some of the various configurations include: (1) electrode montages, (2) stimulus (or matrix) property (i.e. color, size, rate and motion) or type alteration (e.g. face) [3], [8], [9], (3) variations of inter-stimulus intervals (ISIs) (or stimulus onset asynchrony (SOA)) and target-to-target intervals (TTIs) [10], and (4) redesign of visual stimulus representation patterns [11], [12].

* This work was supported by National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. 2012-005741).

S.-K. Yeom, S. Fazli, and S.-W. Lee are with the Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 136-714, Republic of Korea (corresponding authors to provide e-mail: {fazli, sw.lee}@korea.ac.kr).

K.-R. Müller is also with the Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 136-714, Republic of Korea and with Machine Learning Group, Berlin Institute of Technology, Marchstr. 23, 10587 Berlin, Germany.

In this paper, two improvements to the above mentioned issues of P300 spelling are proposed: 1) to minimize adjacency-distraction errors we adopted a random set-based stimulus representation pattern (RASP), 2) the present study is dedicated to further investigate the effects of face familiarity on the performance of BCIs using stimuli of facial images. In a previous study, we found that brain activity responses to one's own face are markedly unique and show stronger responses when compared to familiar or unfamiliar non-self faces and this phenomenon was defined as 'face-specific visual self-representation' in [13] for the neurophysiological basis thereof we refer to [14]. To this end we designed the paradigm, such that self-faces as well as non-self faces were presented in a randomized order.

This paper is organized as follows: We start with the Methods section to introduce the setup and experimental paradigm of our EEG study as well as data analysis techniques we applied. In the Results section we describe the experimental results and finally the Discussion section concludes the work by discussing our findings and putting them into perspective with future work.

II. MATERIALS AND METHODS

A. Participants

Fifteen healthy university students who were between 26 and 32 years (mean 27.7 $\pm$ 1.5, right-handed, all males) took part in our experiments. Participants were seated comfortably in a chair with armrests in a quiet room at a distance of 60 $\pm$ 5 cm from a standard 19 inch LCD monitor (60 Hz refresh rate, 1280 $\times$ 1024 screen resolution).

B. Equipment and Data Acquisition

EEG signals were recorded with a sampling rate of 500 Hz with a BrainAmp multichannel EEG amplifier by Brain Products from the following 29 Ag/AgCl electrodes on a cap (actiCAP, Brain Products, Munich, Germany), according to the international 10-20 system: F3, F4, Fz, FC1, FC2, FC5, FC6, C3, C4, Cz, T7, T8, CP1, CP2, CP5, CP6, P3, P4, Pz, P7, P8, PO3, PO4, POz, PO7, PO8, O1, Oz, and O2. Channels were nasion-referenced and grounded to electrode Fpz.

Face images were acquired using a 3dMD face capture system. All the face images were processed to remove external features such as hair and then cropped into a common oval frame which was placed on a black uniform background. Face images were scaled to an image size of 400 $\times$ 500 pixels.

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

(a) (b) (c)

Fig. 1. The different conditions of the paradigm. (a) The classical row-column (RC) paradigm, (b) The proposed random set presentation (RASP) paradigm, (c) RASP-F paradigm with flashing face in one row of the virtual matrix. Both RASP and RASP-F stimuli were shown semi-transparently to the participants such that the characters were still visible.

C. *Experimental Stimuli and Paradigm*

All three matrix spellers were presented with a 6 × 6 matrix and highlighted characters or faces were flashed consecutively in random order.

In the first proposed variant, called random set-based stimulus representation pattern (RASP), letters are randomly shuffled in a virtual six-by-six matrix, prior to stimulus presentation and then 12 stimulus flashes are presented to the subject. As a result users see a unique combination of letters in each stimulus during a given sequence. The number of stimuli were equal for the RC and RASP paradigms and the temporal distribution of TTIs was the same on average.

As a second variant, also based on RASP, the characters were overlaid with face stimuli semi-transparently. This variant is termed RASP-F. The types of face stimuli, which were used in the experiments can be divided into 2 categories. Self-face (when a row is selected) and non-self-face (when a column is selected) images were used for stimulation. A self-face image consists of the image of the subject, while a non-self-face image consists of a familiar face such as his/her friends or of unfamiliar faces whom he/she has never seen before. Ratio between self-face and non-self-face presentation is 50:50. See Figure 1.

Each experiment consists of 2 phases: a training phase and a test phase. The presentation order of the spellers was randomized across participants. In each session, participants were provided with strings of letters they were supposed to spell. The whole string was displayed at the top left of the monitor and the next item-to-spell (the target letter) was displayed above the letter matrix (see Figure 1). During the initial training phase, subjects had to copy-spell one sentence ‘BRAIN_COMPUTER_INTERFACE’. There was no feedback and EEG was recorded for offline analysis. In the second phase subjects had to copy-spell another sentence ‘KOREA UNIVERSITY’ (without the space). The participant’s task was to attend to (or count) the number of times the target character flashed. Each run started with a 2 s countdown. For all speller conditions, each set of characters flashed for 135 ms, followed by an ISI of 50 ms. When subjects were instructed to copy-spell, the spelling of each letter consisted of 10 sequences without a prolonged inter-sequence interval. One sequence consists of 12 flashes. Note, that for all cases the target will flash twice.

D. *Data Analysis*

We used the BBCI toolbox[1] for our analysis. EEG data was band-pass filtered between 0.1 and 30 Hz with a 5^{th} order Butterworth digital filter. In each experimental session, the data was epoched from -200 ms to 800 ms with respect to stimulus onset. Epoched EEG signals were baseline-corrected by subtracting the mean amplitudes in the -200 to 0 ms pre-stimulus interval from every epoch. Then, averaged features of the ERPs were extracted from 8 selected discriminative intervals, which were selected by a well established heuristic, which depends on signed r-values [15]. After that, these features from the training phase were validated with the data from the test phase with the help of a regularized linear discriminant analysis (RLDA) classifier with analytic shrinkage of the covariance matrix [15], [16]. For the evaluation of the 3 matrix spellers classification accuracies (a 0-1 loss function was used) as well as Information Transfer Rates (ITRs) were computed. ITRs iven as *bits per unit time* [bits $\min^{-1}$] are commonly used as an evaluation measurement for BCIs.

III. RESULTS

A. *Classification Accuracy and ITR*

Figure 3 depicts the classification accuracy for each subject as well as averaged accuracies and ITRs for all subjects. The number of sequences were varied from one to ten sequences (x-axis) for all three different spellers. Classifier accuracy was significantly increased when face stimuli were used as compared to highlighting characters. In the RASP-F condition, on average fewer sequences (M = 1.1 ± 0.3) were necessary for achieving an accuracy level of ≥ 70% as compared to the RC (M = 2.5 ± 1.3) and RASP conditions (M = 1.9 ± 1.0). This threshold has previously been argued to be the minimum accuracy level for meaningful communication [17]. Offline selection accuracies for selecting one symbol out of 36 by using single sequence data were 58.4% ± 1.6% for RC, 61.3% ± 1.6% for RASP and 84.0% ± 1.2% for RASP-F. As expected, performance increases sharply with the number of repetition ($F = 63.79, p < 0.001$). The

[1] http://bbci.de/toolbox

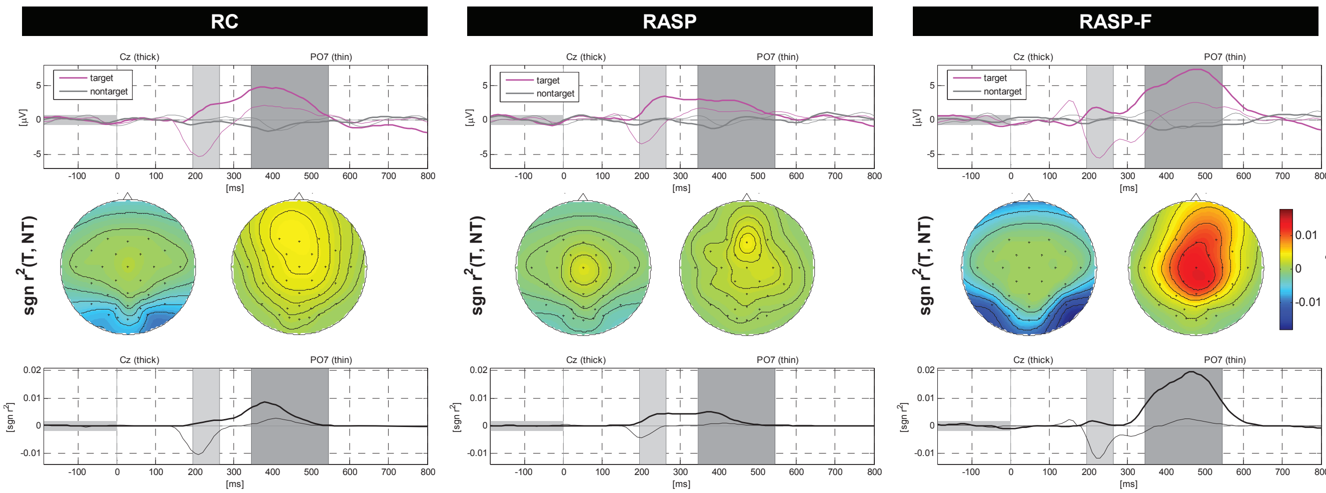

Fig. 2. Grand average ERPs and scalp topographies for the three conditions RC, RASP, and RASP-F. Top row: ERPs for targets and nontargets at two selected electrodes Cz and PO7. The two shaded areas in each ERP plot mark the intervals for which scalp maps are shown underneath. Center: sgn r^2 Scalp plots for ERP responses to the target and nontarget classes. Bottom row: Temporal distribution based on sgn r^2 at two selected electrodes Cz and PO7. N250 and P300 components show a higher discriminability for RASP-F as compared to the two other spellers at the central and parieto-occipital sites.

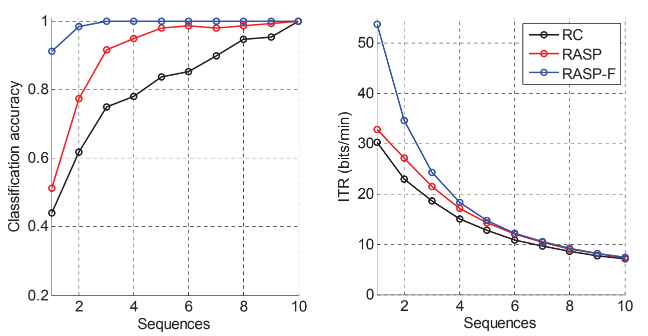

Fig. 3. Averaged classification accuracy and ITR curves on all subjects for three conditions ('RC', 'RASP', and 'RASP-F') using one to ten sequences.

best ITRs of 53.7 ± 11.8 bits/min was achieved by RASP-F as compared to the 30.3 ± 13.3 bits/min for RC and 32.8 ± 13.8 bits/min for RASP. ITRs also increase sharply with the number repeated sequences ($F = 205.22, p < 0.001$).

B. ERP Analysis

Figure 2 shows grand average ERPs and scalp topographies at representative electrodes Cz and PO7. To test whether ERP amplitudes differed across spellers, we examined the data, which was averaged across all subjects, within selected time intervals at N170 (between 190 and 250 ms, PO7), N250 (between 250 and 350 ms, PO7) and P300 (between 350 and 500 ms, Cz), respectively.

N170 amplitudes evoked by the RC paradigm were slightly larger than those evoked by RASP with face (RC > RASP-F: $t = 0.42, p = 0.678$, RASP < RASP-F: $t = 3.65, p = 0.003$) mainly due to larger N250 amplitudes evoked by the face-related stimuli as compared to the the highlighted characters, especially at the parieto-occipital sites (see center of Figure 2). The P300 amplitude analysis also revealed a significant difference among spellers, especially at the central sites (also see center of Figure 2) (RC < RASP-F: $t = 3.78, p = 0.002$, RASP < RASP-F: $t = 6.16, p < 0.001$).

C. Error and Variation on Target-to-Target Interval Analysis

Figure 4 illustrates the topographical distribution of errors in relation to the target item for the RC and RASP-F based paradigms. All target items have been centered in this matrix for display purposes; the numbers in the black cells represent the number of correct selections for each paradigm. The numbers in other cells correspond to the locations of errors relative to the target location. In the RC many errors occurred in the direct neighborhood. These non-targets were flashed simultaneously with the target item in the rows and columns. The upper matrices show results, if only one sequence is considered, the lower matrices consider three sequences. As can be seen, the RASP-F based paradigms successfully reduce the number of errors, because combinations of letters were shuffled within each sequence.

IV. DISCUSSION

Accurate target detection with fewer sequence data is still a challenging problem. For this reason the development of new paradigms with more effective 1) visual stimulus types, and 2) stimulus presentation patterns, which elicit stronger differential ERP responses, is considerably important for improving the performance of such BCI systems.

To further increase the speed of character selection one has to focus on reducing the number of stimulus sequences used for averaging. However, usually several P300 responses must be averaged for the response to be recognized due to the low signal-to-noise ratio. By reshuffling and thus creating unique

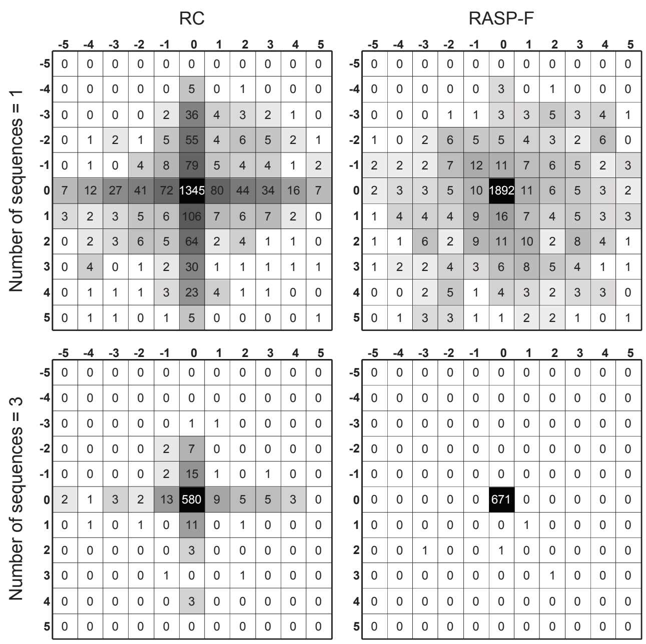

Fig. 4. Error distributions for the RC (left) and RASP-F (right) as increasing the number of sequence. All target items have been centered in each matrix. The number in a black centered cell corresponds the number of correct selections and numbers in other cells represent the number of error corrected selections occurring in each cell relative to the target location for each speller.

combinations of letters for each flash our findings indicate increased performance for the same number of sequences (see Figure 3).

Face stimuli including self- and non-self-faces yielded significantly higher accuracies and ITRs than those of highlighted characters. This implies that stimuli with higher cognitive task requirements such as facial images, are more effective than the intensified stimuli of dull characters for a P300-based BCI system. For achieving a performance level of $\geq 70\%$ RASP-F can reduce the overall time needed to spell a character by a factor of 2.3 on average in comparison to RC and by a factor of 1.7 in comparison to RASP.

While some individual variation is evident, the individual participants' averaged ERPs conform to the grand mean shown in 2, which shows that both the target and non-target ERPs differ in several respects across spellers. N250 amplitudes are significantly enlarged at parieto-occipital sites. Contrary to our expectations no significant differences of N170 amplitudes were found across spellers (see Figure 2). P300 tends to be more pronounced at the central sites for face stimuli, against those evoked by the highlighted character (RC and RASP). This suggests higher level of cognitive components in the central areas through the face perception task. Such cognitive components associated with face perception result in more discriminative features.

We would like to finally remark that our approach as virtually all other work on P300 spellers is gaze dependent. However, as pointed out in their seminal contribution [11], a clear path to gaze independent BCI spellers can be pursued. Future work will therefore extend the present paradigms towards gaze independency.

References

[1] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[2] B. Allison, J. Faller, and C. Neuper, *BCIs that use steady state visual evoked potentials or slow cortical potentials. Brain-Computer Interfaces: Principles and practice. Wolpaw and E. W. Wolpaw, Eds.* Oxford University Press, 2012.

[3] H. Bakardjian, T. Tanaka, and A. Cichocki, "Emotional faces boost up steady–state visual responses for Brain–Computer Interface," *Neuroreport*, vol. 22, no. 3, pp. 121–125, 2011.

[4] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive berlin Brain–Computer Interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.

[5] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralysed," *Nature*, vol. 398, no. 6725, pp. 297–298, 1999.

[6] S. Sutton, M. Braren, J. Zubin, and E. John, "Evoked-potential correlates of stimulus uncertainty," *Science*, vol. 150, no. 3700, pp. 1187–1188, 1965.

[7] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih, and A. Kübler, "P300 Brain Computer Interface: Current challenges and emerging trends," *Frontiers in Neuroengineering*, vol. 5, 2012.

[8] T. Kaufmann, S. Schulz, C. Grünzinger, and A. Kübler, "Flashing characters with famous faces improves ERP-based Brain–Computer Interface performance," *Journal of Neural Engineering*, vol. 8, no. 5, p. 056016, 2011.

[9] Y. Zhang, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "A novel BCI based on ERP components sensitive to configural processing of human faces," *Journal of Neural Engineering*, vol. 9, no. 2, p. 026018, 2012.

[10] C. Polprasert, P. Kukieattikool, T. Demeechai, J. A. Ritcey, and S. Siwamogsatham, "New stimulation pattern design to improve P300-based matrix speller performance at high flash rate," *Journal of Neural Engineering*, vol. 10, no. 3, p. 036012, 2013.

[11] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based Brain–Computer Interface," *Behavioral and Brain Functions*, vol. 6, no. 1, p. 28, 2010.

[12] G. Townsend, B. LaPallo, C. Boulay, D. Krusienski, G. Frye, C. Hauser, N. Schwartz, T. Vaughan, J. Wolpaw, and E. Sellers, "A novel P300-based Brain–Computer Interface stimulus presentation paradigm: Moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109–1120, 2010.

[13] M. Miyakoshi, N. Kanayama, T. Iidaka, and H. Ohira, "EEG evidence of face-specific visual self-representation," *NeuroImage*, vol. 50, no. 4, pp. 1666–1675, 2010.

[14] H. Ninomiya, T. Onitsuka, C.-H. Chen, E. Sato, and N. Tashiro, "P300 in response to the subject's own face," *Psychiatry and Clinical Neurosciences*, vol. 52, no. 5, pp. 519–522, 1998.

[15] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components–A tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, 2011.

[16] 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.

[17] A. Kübler and N. Birbaumer, "Brain–Computer Interfaces and communication in paralysis: Extinction of goal directed thinking in completely paralysed patients?" *Clinical Neurophysiology*, vol. 119, no. 11, pp. 2658–2666, 2008.