# Motion-Based Rapid Serial Visual Presentation for Gaze-Independent Brain-Computer Interfaces

Dong-Ok Won, Han-Jeong Hwang, Dong-Min Kim, Klaus-Robert Müller, and Seong-Whan Lee, *Fellow, IEEE*

***Abstract*— Most event-related potential (ERP)-based brain–computer interface (BCI) spellers primarily use matrix layouts and generally require moderate eye movement for successful operation. The fundamental objective of this paper is to enhance the perceptibility of target characters by introducing motion stimuli to classical rapid serial visual presentation (RSVP) spellers that do not require any eye movement, thereby applying them to paralyzed patients with oculomotor dysfunctions. To test the feasibility of the proposed motion-based RSVP paradigm, we implemented three RSVP spellers: 1) fixed-direction motion (FM-RSVP); 2) random-direction motion (RM-RSVP); and 3) (the conventional) non-motion stimulation (NM-RSVP), and evaluated the effect of the three different stimulation methods on spelling performance. The two motion-based stimulation methods, FM- and RM-RSVP, showed shorter P300 latency and higher P300 amplitudes (*i.e.*, 360.4–379.6 ms; 5.5867–5.7662 $\mu V$) than the NM-RSVP (*i.e.*, 480.4 ms; 4.7426 $\mu V$). This led to higher and more stable performances for FM- and RM-RSVP spellers than NM-RSVP speller (*i.e.*, 79.06 ± 6.45% for NM-RSVP, 90.60 ± 2.98% for RM-RSVP, and 92.74 ± 2.55% for FM-RSVP). In particular, the proposed motion-based RSVP paradigm was significantly beneficial for about half of the subjects who might not accurately perceive rapidly presented static stimuli. These results indicate that the use of proposed motion-based RSVP paradigm is more beneficial for target recognition when developing BCI applications for severely paralyzed patients with complex ocular dysfunctions.**

***Index Terms*— Brain-computer interface (BCI), gaze-independent, event-related potential (ERP), rapid serial visual presentation (RSVP).**

Manuscript received May 4, 2016; revised November 7, 2016, March 14, 2017, and July 23, 2017; accepted July 24, 2017. Date of publication August 11, 2017; date of current version February 9, 2018. This work was supported in part by the Institute for Information and Communications Technology Promotion (IITP) through the Korea Government (MSIT) under Grant 2017-0-00451, in part by the Development of BCI-based Brain and Cognitive Computing Technology for Recognizing User's Intentions using Deep Learning) and the MSIT, South Korea, under the SW Starlab support program supervised by the IITP under Grant IITP-2015-1107. *(Corresponding author: Seong-Whan Lee.)*

D.-O. Won, D.-M. Kim, and S.-W. Lee are with the Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea (e-mail: wondongok@korea.ac.kr; dm_kim@korea.ac.kr; sw.lee@korea.ac.kr).

H.-J. Hwang is with the Department of Medical IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea.

K.-R. Müller is with the Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea, and also with the Department of Computer Science, Berlin Institute of Technology, Berlin, Germany.

Digital Object Identifier 10.1109/TNSRE.2017.2736600

## I. Introduction

BRAIN-COMPUTER interface (BCI) uses brain signals instead of muscle signals to control external devices such as an exoskeleton, a robotic arm, or a wheelchair [1]–[4]. Electroencephalography (EEG) has been widely used for BCI research [5]–[9] owing to the following advantages: good temporal resolution enabling real-time BCI control, non-invasiveness, and reasonable price. EEG-based BCIs have been categorized according to the type of brain activity used for the development of a BCI system, *e.g.*, event-related potential (ERP) [10], steady-state visual evoked potential [11]–[14], event-related (de)synchronization [15], and slow cortical potential [16].

A BCI speller is a representative application for EEG-based BCIs that enable the user to write characters without muscle movements [17], [18]. Many BCI studies have shown that BCI spelling systems can be implemented using ERPs [10], [19]–[21]. ERP-based spellers (or P300 spellers) use neural activity generated by user attention to input a target character. Conventional ERP spellers based on visual stimuli generally consist of a $6 \times 6$ symbol matrix [10], in which the user concentrates on a target character, and distinct ERPs appear whenever the row and column containing a target symbol is flickered; the user can spell the target character by employing dominant ERPs.

Conventional vision-based ERP spellers have shown excellent performance [21]–[23] when the user has relatively moderate visual functions [21]. However, gaze independence and visual fatigue are remaining challenges in the development of BCI speller systems, as conventional vision-based BCI spellers offer only limited practical applicability for end users with severe oculomotor impairments [24]. The classical vision-based BCI spellers need at least moderate head or eye movement toward target stimuli for achieving reasonable performance. Recently, some studies have considered gaze-independence issues in developing BCI spellers with other sensory modalities such as hearing and somatic sensation [25]–[28]. However, ERPs induced by non-visual stimuli are weaker than those induced by visual stimuli [29], and the number of targets that can be used for implementing BCI systems is limited with non-visual stimuli ($<10$ in general).

Recently, gaze-independent spellers have been introduced using visual ERP paradigms [29]–[32]. For example, a covert attention-based paradigm using peripheral vision was introduced, and its feasibility for developing ERP spellers was proved [29], [30]. However, ERP amplitudes and spelling performance of covert attention-based spellers are weaker and lower than those of overt attention paradigms accompanying eye movements – it is known that central vision has higher visual sensitivity than peripheral vision [33]. In addition, another gaze-independent BCI speller based on motion visually evoked potentials (mVEPs) requires a relatively long detection time; the spelling speed was $1.28 \pm 0.03$ characters/minute [31], because of multiple-step selection strategies to select a single character. There is another gaze-independent BCI speller that uses the rapid serial visual presentation (RSVP) of characters, which has also been applied for rapidly scanning large images [34], [35] by employing real-time cortical feedback for ranking images of interest [32].

To our best knowledge, the classical RSVP paradigm is the fastest gaze-independent paradigm. However, owing to its rapid presentation of stimuli, it is difficult for users to recognize targets. Moreover, visual searching requires a certain amount of mental effort [30]. Thus, novel stimulation method that can be more easily perceived is needed for developing a more practical gaze-independent BCI spellers. It is well known that the visual system is sensitive to moving stimuli, *i. e.*, they are easily perceived [36]–[38]. Moreover, using motion stimuli has an advantage in that they can be perceived with relatively low luminance and contrast [39], [40]. Consequently, these characteristics of motion stimuli can lead to less fatigue and discomfort for users, which again results in enhanced spelling performance by increasing ERP amplitudes and shortening ERP latencies.

In this paper, we propose a novel motion-based RSVP paradigm, which leverages the aforementioned advantages of motion stimulus. To test the feasibility of the proposed motion-based RSVP paradigm, we implemented three different RSVP spellers: i) fixed-direction motion (FM-RSVP), ii) random-direction motion (RM-RSVP), and iii) (the conventional) non- motion stimulation (NM-RSVP). For FM- and RM-RSVP spellers, all symbols are presented one by one in a random sequence and moved to one of the six directions (*i. e.*, 2, 4, 6, 8, 10, and 12 o'clock) within the near-central visual field [30]. A moving direction was randomly set for RM-RSVP and was initially fixed for FM-RSVP. For an NM-RSVP speller, all symbols were randomly presented without any movement, as introduced in [32]. We conducted a series of experiments with the three different RSVP spellers, and evaluated the characteristics of ERP patterns in terms of ERP amplitudes, ERP latencies, and their classification performance.

## II. Methods

### A. Participants

The experiment included 16 participants (12 males and 4 females; $25.38 \pm 3.07$). All participants had corrected-to-normal or normal vision. In addition, no participant had any previous history of visual disorders or neurological disease.

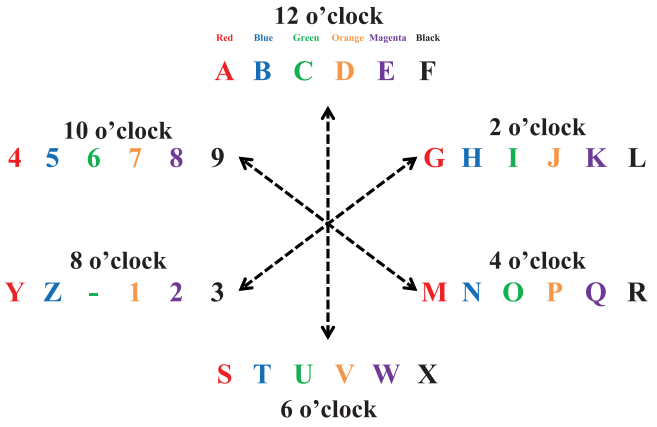


Fig. 1. Characters in each of the color and direction groups in the FM-RSVP speller. Note that characters are randomly moved to one of the six directions in the RM-RSVP speller. Motion directions are given with six clock positions. Each direction contains six symbols with different colors.

The experiments were conducted in accordance with the principles described in the Declaration of Helsinki. This study was reviewed and approved by the Institutional Review Board of Korea University [1040548-KU-IRB-15-163-A-1]. We obtained written informed consent from all participants before the experiment. Five subjects had previous experience with BCI spelling experiments, whereas the other subjects were naïve.

### B. Experimental Stimuli and Paradigm

We used 36 character symbols *i. e.*, 26 letters of the English alphabet (A-Z), nine numbers (1-9), and the hyphen used to separate different words. Because previous studies found that different colors can help discriminate different symbols [31], [32], the 36 characters were divided into six color groups in the spirit of [21] as follows: red: A, G, M, S, Y, and 4; blue: B, H, N, T, Z, and 5; green: C, I, O, U, "1−," and 6; orange: D, J, P, V, 1, and 7; magenta: E, K, Q, W, 2, and 8; and black: F, L, R, X, 3, and 9, as shown in Figure 1. Some character symbols having confusing shapes (*e. g.*, I and 1; O and Q) were divided into different color groups for easier character recognition (*i. e.*, I (green) and 1 (orange); O (green) and Q (magenta)).

*1) NM-RSVP Speller:* In the classic RSVP speller (NM-RSVP), all symbols are randomly shuffled and then presented one by one without any motion [32]. Thus, a target character can be detected for the user by the shape and color of the character. Figure 2a depicts an example of a character presentation sequence of the NM-RSVP speller.

*2) FM- and RM-RSVP Spellers:* There are two motion-based spellers with fixed moving direction (FM-RSVP) and with randomized moving directions (RM-RSVP). Six directions were employed for motion stimulation: 12 o'clock: A, B, C, D, E, and F; 2 o'clock: G, H, I, J, K, and L; 4 o'clock: M, N, O, P, Q, and R; 6 o'clock: S, T, U, V, W, and X; 8 o'clock: Y, Z, "-," 1, 2, and 3; and 10 o'clock: 4, 5, 6, 7, 8, and 9 (Figure 1). Whereas characters randomly move in one of six directions in RM-RSVP, FM-RSVP uses a predetermined direction that was

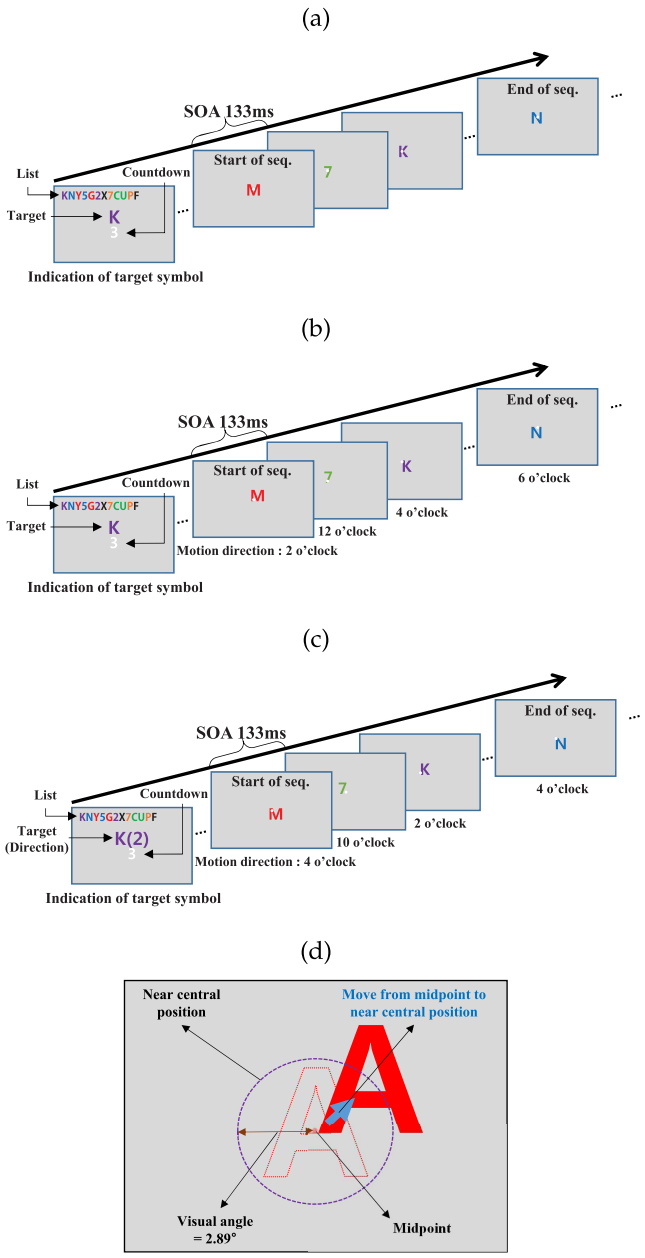


Fig. 2. Experimental paradigms for (a) NM-RSVP, (b) RM-RSVP, and (c) FM-RSVP, and (d) an example of presenting a stimulus with motion. Target character are presented at the top of the display in all conditions (see first picture of (a), (b), and (c)). The interval between each stimulus was set as 133 ms, and 10 sequences were presented for each condition. Please refer to the attached supplementary video files for an example of presenting stimuli.

presented with an indication of the target symbol (see the first picture of Figure 2c). A target character can be recognized for the user by the shape, color, and direction of the character (*e. g.*, the 2 o'clock direction consisted of the G, H, I, J, K, and L characters with red, blue, green, orange, magenta, and black, respectively). All characters were randomly presented in each sequence. Figure 2b and c depict an example of presenting visual stimuli for the RM-RSVP speller and the FM-RSVP speller, respectively. The measured visual angles of the disk areas of the FM- and RM-RSVP spellers are 2.89° for all subjects.

*3) Paradigm:* We designed our experiment to be unbiased with respect to color and direction in the evaluation by using equally distributed letters in each color and each direction through the training ("KNY5G2X7CUPF"; 12 characters) and test sessions ("BSQHDRT-94WJEM36I1"; 18 characters). The remaining letters not used in both training and test sessions ("ALOVZ8"; 6 characters) also have different directions and colors. For all the three RSVP conditions, 36 characters including the other six characters not used as target characters in training and test sessions were presented individually in each sequence (repetition) (see Figure 2), and the participants focused on target symbols to induce target-specific ERPs. In training session, we used the whole 120 target trials (1 target characters × 10 sequences × 12 times corresponding to the number of target characters) and 4200 non-target trials (35 non-target characters × 10 sequences × 12 times corresponding to the number of target characters) for constructing a classifier. In test session, one target trial and thirty-five non-target trials were used for each sequence to evaluate the constructed classifier.

During stimulus presentation, the participants were asked to fix their eyes on the center of a 19-inch monitor with gray background color [32] (model: MY19PS, Samsung), direct their attention toward a target character and silently count whenever they found it. A stimulus interval was set to 133 ms. it means stimuli smoothly move for 133 ms, and the stimulus speed was 2.89° (visual angle)/0.133s = 21.729°/ s for motion-based RSVP. Motion stimuli were presented in the central visual field within the fovea and near-parafovea [41], [42], as shown in Figure 2d, thereby allowing the user to recognize presented characters with minimal eye movement, which did not significantly affect EEG signals (see Figure 6 in advance).

## C. Data Acquisition

The participants seated in a comfortable chair at a distance of approximately 80 cm from the screen. During the experiments, EEG data were recorded at a 1000 Hz sampling rate using 63 electrodes attached based on the extended international 10-20 system. BRAINAMP amplifiers and actiCap active electrodes were used (Brain Products, Germany) and electrode locations were Fp1-2, AF3-4, Fz, F1-10, FCz, FC1-6, FT7-8, Cz, C1-6, T7-8, CPz, CP1-6, P7-8, Pz, P1-10, POz, PO3-4, PO7-10, Oz, and O1-2. Electrooculography (EOG) data were also recorded under the left eye of the subjects [32]. The reference was located on the ridge of the participant's nose, and the ground was located at AFz. The impedance of all electrodes was kept under 10 kΩ during the experiment. The experimental paradigm was implemented with Psychtoolbox (http://psychtoolbox.org).

## D. Data Analysis

*1) Preprocessing:* We used MATLAB (MathWorks, Natick, MA, USA) with the BBCI toolbox (http://bbci.de/toolbox) for data analysis. All EEG data were downsampled to 100 Hz and band-pass filtered at 0.5-30 Hz with a Chebyshev filter for off-line analysis. To remove physiological artifacts contained in the EEG data (*e. g.*, eye and head movements), we applied

an independent component analysis (ICA) by using a temporal decorrelation source separation algorithm [43]. We computed the correlations of the independent components with EOG channels (Fp1, Fp2, F9, F10, and EOG) and determined a conservative threshold (more than two standard deviations) for rejecting independent components as EOG-contaminated components [44]. We also rejected artifact trials based on a min-max criterion (*i.e.*, a min-max voltage difference $>75\ \mu V$) [32]. In case more than 5% of the trials were rejected in a subject, the corresponding data were not used because we assumed that the subject made considerable movements during the experiment. Three subjects showed more than 5% of rejected trials, and therefore data for only 13 participants were used for further analysis.

*2) ERPs and Classification:* For all three stimulus conditions, the data were epoched from −233 to 800 ms on the basis of the stimulus onset. We then selected a pre-stimulus interval (−233 to 0 ms) for baseline correction, and then all epochs for targets and non-targets were independently averaged to see grand-average ERPs. For off-line classification, five most discriminative intervals were selected using a well-established method, signed r-squared values ($sgn\ r^2$) [45], [46]. The window width depended on the signed r-value. We averaged ERP values over all channels in each selected time window. Thus, the features consisted of 315 dimensions (63 channels × 5 average time windows) [45], [46]. The width and location of the selected windows were different for each subject (*e. g.*, subject 1 windows: NM = [225-250 ms, 255-305 ms, 340-355 ms, 360-470 ms, 475-490 ms], RM = [115-130 ms, 135-190 ms, 210-235 ms, 240-325 ms, 330-370 ms], and FM = [100-120 ms, 125-210 ms, 215-220 ms, 225-320 ms, 325-385 ms]). Training and test session data were used for building a classifier, regularized linear discriminant analysis with shrinkage (S-RLDA) of the covariance matrix [32], [45], and determine user-intended characters, respectively. Classification was performed by selecting a maximum classifier output value that was averaged across the sequences [32]. We evaluated the classification accuracy of target characters with respect to the number of sequences, for which we randomly selected the same numbers of trials for each target character from one to ten. Thus, in the results section, N sequence means that N randomly selected trials are simultaneously used for classification. (*i.e.*, three sequences: we randomly selected three trials for the same character). We used 18 characters ("BSQHDRT-94WJEM36I1") as testing characters with a chance level of 2.77 (1/36)%.

*3) Eye Movement Analysis:* In this study, we introduced motion stimuli presented entirely in the central region. The participants were asked to direct their attention toward a target character without any eye movements, but it was inevitable to avoid small eye movements for motion-based stimuli (*i.e.*, RM- and FM-RSVP). To check the impact of the EOGs, we investigated the vertical and horizontal EOGs. The horizontal EOG was calculated by subtracting F9 from F10: left-eye movement was indicated by positive potential, and vice versa [47]. The vertical EOG was calculated by using the EOG channel under the left eye: upward movement was indicated by a negative potential, and downward movement was indicated by a positive potential. For the EOG analysis, we grouped the four directions, 8, 10, 2, and 4 o'clock, into two groups for investigating the horizontal eye movement (left: 8 and 10 o'clock, right: 2 and 4 o'clock), whereas the two directions, 12 and 6 o'clock, were directly used for investigating the vertical eye movement (up: 12 o'clock and down: 6 o'clock). In addition, we evaluated the gamma-band EEG response to more precisely check eye movements, e.g. saccadic movement [48]–[53].

TABLE I
MEAN AND STANDARD ERROR OF THE MEAN OF THE P300 AMPLITUDES ($\mu V$) AND LATENCIES (ms)

| Condition | Ch. | Latency (ms) | Amplitude ($\mu V$) |
|---|---|---|---|
| NM-RSVP | Cz | 480.4 ± 17.0 | 4.7426 ± 0.5057 |
| RM-RSVP | Cz | 379.6 ± 16.7 | 5.7662 ± 0.4980 |
| FM-RSVP | Cz | 360.4 ± 17.0 | 5.5867 ± 0.3778 |

*4) Alpha Power Analysis:* Among the 13 subjects, almost half of them (subjects 3, 4, 6, 8, 11, and 13) showed moderate performance in NM-RSVP condition, but the others showed low performance (see Table 3 and Figure 4 in advance). To investigate the reason of the result, we investigated changes in alpha powers for the three conditions of NM-, RM-, and FM-RSVP, because it has been well documented that alpha power is related attention [30], [54] as well as BCI performance [55]–[57]. For all the subjects, alpha powers ranging from 8 to 12 Hz were estimated using CZ and PO7 channels based on a related study [55] and they were summed.

## III. RESULTS

### A. ERPs and Classification

Figure 3 shows ERP patterns of all the three stimulation conditions for targets and non-targets, respectively, where typical N200 and P300 components are clearly seen for targets regardless of the stimulation condition. We calculated P300 latencies and amplitudes for target stimuli by using the Cz electrode, and N200 ones by using PO7 electrode (Table 1) [29], [31]: Treder *et al.* [29] used Cz and PO7 for P300 and N200 analyses, respectively, and Schaeff *et al.* [31] used channels around Cz (*i.e.*, FCz, Cz, and Pz) and PO7-8 (*i.e.*, P3, P7, and PO7) for P300 and N200 analyses, respectively. The average of the P300 maximum peak amplitudes appeared at 480.4 ± 17.0 ms, 379.6 ± 16.7 ms, and 360.4 ± 17.0 ms with amplitudes of 4.7426 ± 0.5057 $\mu V$, 5.7662 ± 0.4980 $\mu V$, and 5.5867 ± 0.3778 $\mu V$ for NM-RSVP, RM-RSVP, and FM-RSVP conditions, respectively.

The mean P300 latencies of RM-RSVP and FM-RSVP were significantly shorter than those of NM-RSVP (Friedman: $\chi^2$ (chi-square) = 20.46, $p < 0.01$; the Wilcoxon signed-rank test: FM-RSVP = RM-RSVP < NM-RSVP, Bonferroni corrected $p < 0.01$).

In each condition, the average N200 maximum peaks of NM-RSVP, RM RSVP, and FM-RSVP emerged at 310.4 ± 10.5 ms, 216.2 ± 9.8 ms, and 188.1 ± 9.8 ms with amplitudes of −4.1095 ± 0.8272 $\mu V$, −2.8663 ± 0.7481 $\mu V$, and −3.0538 ± 0.7046 $\mu V$, respectively (Table 2). The averaged

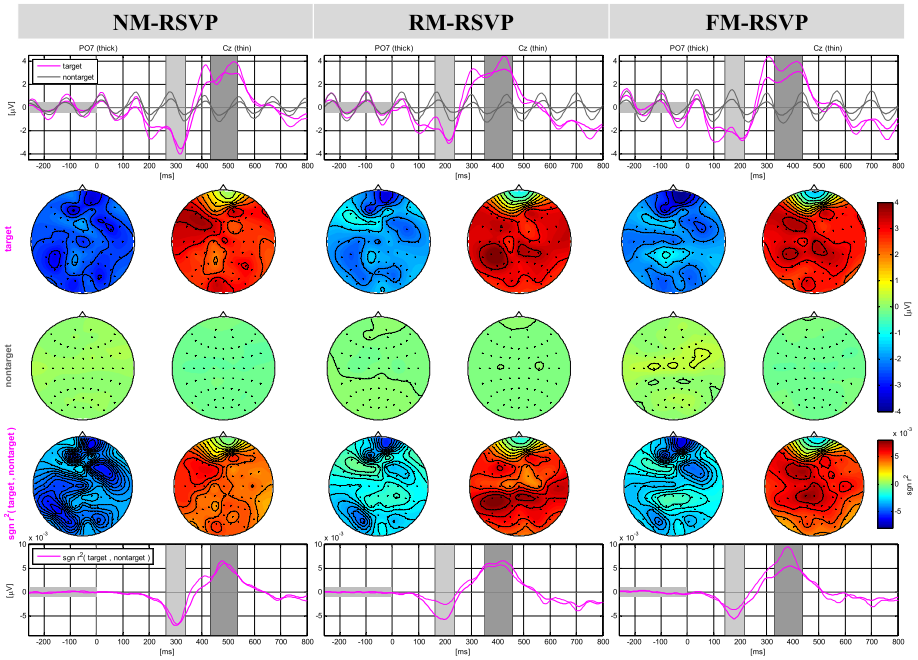


Fig. 3. Grand average ERPs for targets and non-targets in NM-RSVP (first column), RM-RSVP (second column), and FM-RSVP (third column). The light gray and dark gray shadows represent N200 (NM: 265-335 ms, RM: 165-235 ms, FM: 145-215 ms) and P300 (NM: 435-535 ms, RM: 355-455 ms, FM: 335-435 ms) components, respectively (thick magenta line: PO7 and thin magenta line: Cz).

TABLE II
MEAN AND STANDARD ERROR OF THE MEAN OF THE N200 AMPLITUDES ($\mu V$) AND LATENCIES (ms)

| Condition | Ch. | Latency (ms) | Amplitude ($\mu V$) |
|---|---|---|---|
| NM-RSVP | PO7 | 310.4 $\pm$ 10.5 | -4.1095 $\pm$ 0.8272 |
| RM-RSVP | PO7 | 216.2 $\pm$ 9.8 | -2.8663 $\pm$ 0.7481 |
| FM-RSVP | PO7 | 188.1 $\pm$ 9.8 | -3.0538 $\pm$ 0.7046 |

N200 latencies of RM-RSVP and FM-RSVP were significantly shorter than that of NM-RSVP (Friedman: $\chi^2 = 20.49$, $p < 0.01$; the Wilcoxon signed-rank test: FM-RSVP = RM-RSVP < NM-RSVP, Bonferroni corrected $p < 0.01$).

Figure 4 shows the classification accuracies of each subject and their means for each stimulus sequence. On average, the FM-RSVP condition achieved a higher mean accuracy than the other conditions (RM-RSVP and NM-RSVP) on all sequences (Table 3). The classification accuracy of the FM-RSVP condition was significantly higher than that of NM-RSVP for sequences 1, 5, 6, and 10 ($\chi^2 = 6.65 - 10.92$, $p < 0.05$), but no significant differences were found between the accuracies of FM-RSVP and RM-RSVP on all sequences (see the last subplot in Figure 4). And Wilcoxon signed-rank tests showed that the FM-RSVP was significantly higher than NM-RSVP for sequences 1-3, 5, 6, and 8-10.

## B. Eye Movement Analysis

Figure 6 shows the comparison how eye movements affect EEG signals (vertical: UP and DOWN, horizontal: LEFT and RIGHT). No typical EOG patterns are observed for all directions of eye movements. Instead, the patterns looking like ERP ones are observed in EOG channels even though their amplitudes are relatively smaller than the original

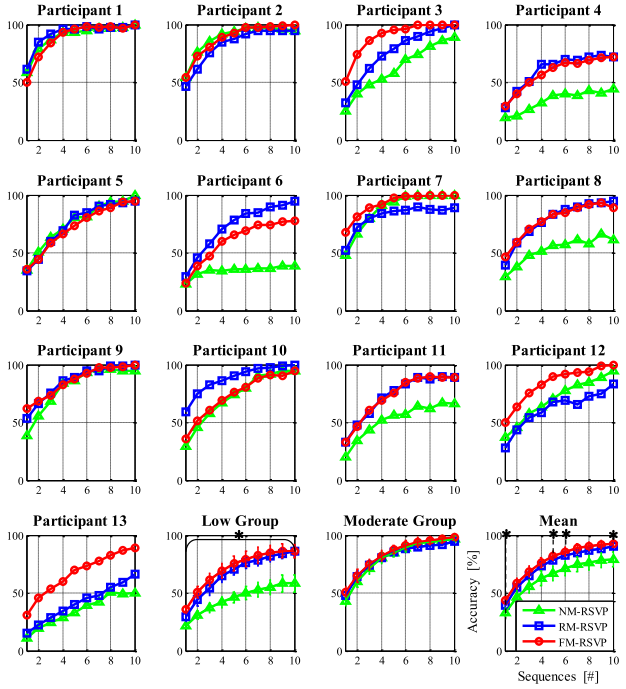


Fig. 4. Relationship between the accuracy and sequences of each subject for all three conditions. In all subjects, a significant difference was noticed between NM-RSVP and FM-RSVP ($\chi^2 = 6.65 - 10.92$, $p < 0.05$, for sequences 1, 5, 6, and 10). In low group (subject 3, 4, 6, 8, 11, and 13; low performing group), a significant difference was noticed between motion-based RSVPs (FM- and RM-RSVP) and NM-RSVP (Friedman: $\chi^2$ (chi-square) = 9-10.17, $p < 0.05$). In moderate group (subject 1, 2, 5, 7, 9, 10, and 12; moderate performing group), no significant difference among all conditions.

ERPs observed in Figure 3. The gamma-band analysis results also verified that no significant saccadic eye movement were contained (see the Supplementary Figures 1 and 2).

TABLE III
AVERAGED CLASSIFICATION ACCURACY FOR 1 TO 10 SEQUENCES IN ALL SUBJECTS

| Seq. | NM-RSVP (%) | RM-RSVP (%) | FM-RSVP (%) | Comparison of RM-RSVP and NM-RSVP (%) | | Comparison of FM-RSVP and NM-RSVP (%) | |
|---|---|---|---|---|---|---|---|
| | | | | Difference | Statistical test | Difference | Statistical test |
| 1 | 33.0 | 39.5 | 43.8 | 6.5 | *p=0.13* | 10.8 | *p=0.01* |
| 2 | 46.5 | 54.7 | 58.3 | 8.2 | *p=0.14* | 11.7 | *p=0.02* |
| 3 | 55.7 | 64.9 | 68.5 | 9.2 | *p=0.06* | 12.8 | *p=0.02* |
| 4 | 62.7 | 73.5 | 76.1 | 10.9 | *p=0.16* | 13.5 | *p=0.10* |
| 5 | 67.0 | 78.7 | 82.1 | 11.8 | *p=0.10* | 15.1 | *p=0.01* |
| 6 | 71.0 | 82.8 | 85.8 | 11.8 | *p=0.14* | 14.8 | *p=0.02* |
| 7 | 74.7 | 84.5 | 88.8 | 9.8 | *p=0.45* | 14.1 | *p=0.08* |
| 8 | 76.2 | 87.0 | 90.4 | 10.8 | *p=0.23* | 14.2 | *p=0.02* |
| 9 | 78.2 | 88.3 | 91.9 | 10.2 | *p=0.32* | 13.7 | *p=0.04* |
| 10 | 79.1 | 90.6 | 92.7 | 11.5 | *p=0.21* | 13.7 | *p=0.03* |

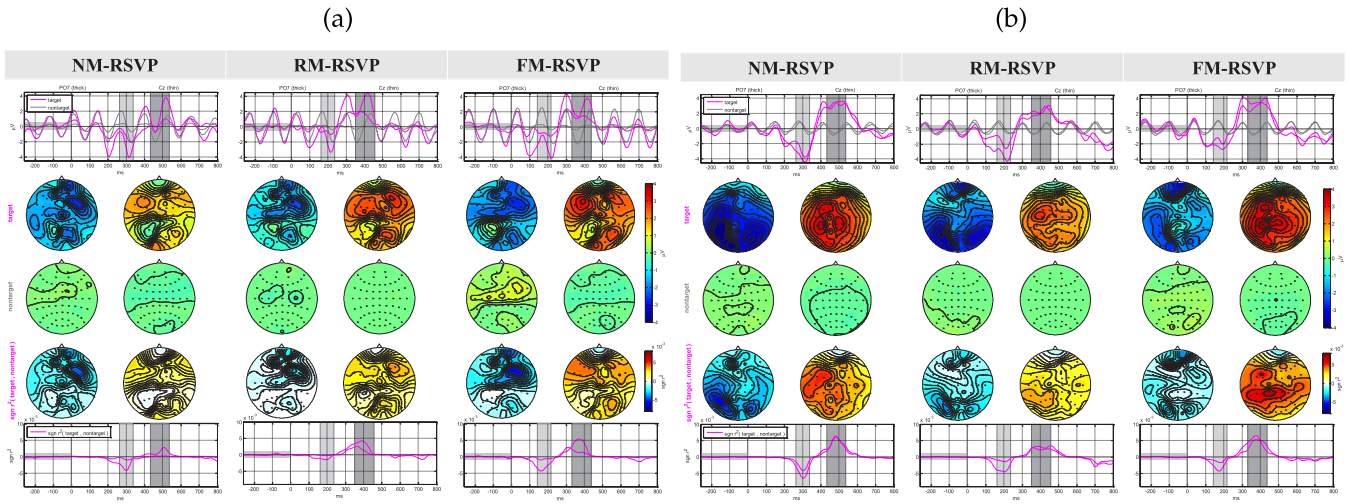


Fig. 5. Grand average ERPs of (a) the low performing group (subject 3, 4, 6, 8, 11, and 13) and (b) the moderate performing group (subject 1, 2, 5, 7, 9, 10, and 12). The light gray and dark gray shadows represent N200 (NM: 265-335 ms, RM: 165-235 ms, FM: 145-215 ms) and P300 (NM: 435-535 ms, RM: 355-455 ms, FM: 335-435 ms) components, respectively (thick magenta line: PO7 and thin magenta line: Cz).

For more details, please see the Supplementary Material about the saccade analysis.

### C. Low and Moderate Performing Group

The performance of the RM- and FM-RSVP spellers for the six subjects (3, 4, 6, 8, 11, and 13) was considerably higher than that of the NM-RSVP, compared to the other subjects (see Figure 4). For convenience, we hereafter refer to the six subjects (3, 4, 6, 8, 11, and 13) and the others as the low performing and moderate performing groups, respectively, with respect to the performance of the NM-RSVP speller. In order to investigate the performance difference between these two groups, we independently analyzed the grand average ERPs for the low performing (Figure 5a) and moderate performing groups (Figure 5b). The moderate performing group showed the standard ERP patterns in all three paradigms. However, the ERP induced by targets of the low performing group showed a relatively faint pattern for NM-RSVP, compared to the other two motion-based conditions, which would lead to classification difference between the two groups.

Figure 7 shows the alpha power analysis results implying that there is no significant difference in alpha power for the moderate performing group between the three stimulation conditions. However, a statistically significant difference is observed between NM-RSVP and motion-based RSVP (RM-RSVP and FM-RSVP) in the low performing group (the Wilcoxon signed-rank test: FM-RSVP = RM-RSVP > NM-RSVP, Bonferroni corrected $p < 0.05$). The results indicate that alpha power is correlated with classification accuracy in the low performing group.

## IV. DISCUSSION

In this study, we implemented three types of RSVP spellers and compared their performance to show the feasibility of the proposed motion-based RSVP paradigm with respect to ERP latency and amplitude as well as classification accuracy. The difference between NM-RSVP and motion-based RSVPs in terms of ERP latencies were statistically significant, *i.e.*, shorter latencies for motion-based RSVPs. Even though ERP amplitudes for targets were not considerably different between the three RSVP conditions, the amplitude difference between target and non-target got larger for FM- and RM-RSVP than NM-RSVP. This result led to enhanced classification accuracy for the motion-based RSVP spellers; $79.06 \pm 6.45\%$, $90.60 \pm 2.98\%$, and $92.74 \pm 2.55\%$ for the NM-RSVP, RM-RSVP, and FM-RSVP speller when employing

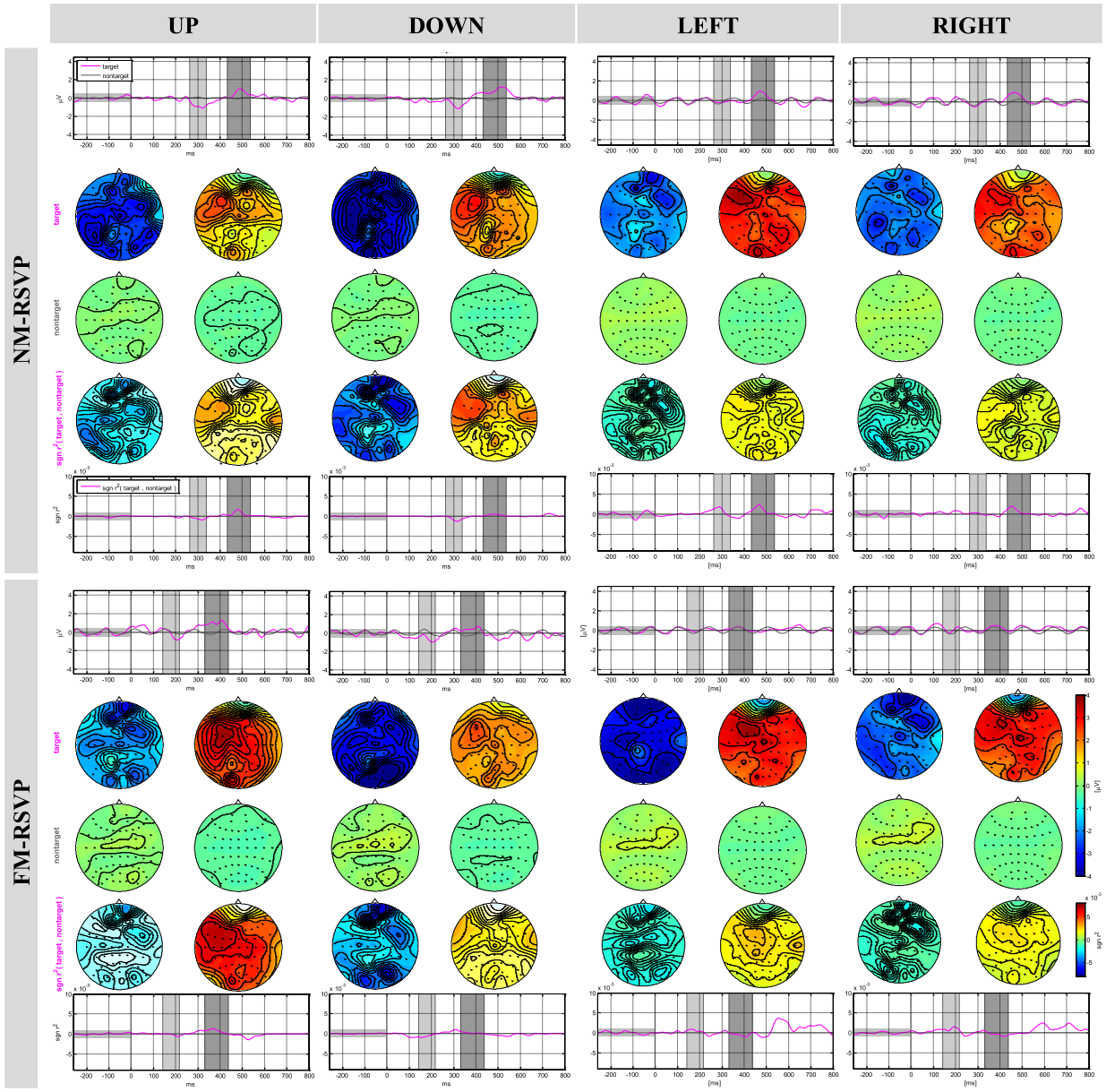


Fig. 6. EEG time series signals measured from EOG channels (the first row), EEG scalp maps (the middle row), and difference between the time series signals induced for targets and non-targets (the last row) for NM-RSVP and FM-RSVP across all subjects. Magenta lines indicate targets, and gray lines indicate non-targets. (Top) NM-RSVP paradigm without motion. (Bottom) FM-RSVP paradigm with motion. The light gray and dark gray shadows represent the time periods showing the most dominant N200 (NM: 265-335 ms, RM: 165-235 ms, FM: 145-215 ms) and P300 (NM: 435-535 ms, RM: 355-455 ms, FM: 335-435 ms) components, respectively.

10 sequence, respectively (see Table 3), demonstrating the feasibility of our proposed motion-based RSVP paradigm.

As shown in Figure 3, the most salient ERP components were N200 and P300. The last row of Figure 3 shows the ERP latencies between the NM-RSVP, RM-RSVP and FM-RSVP conditions, and amplitude differences between targets and non-targets for the three RSVP conditions. The ERPs evoked by the RM- and FM-RSVP conditions were observed earlier than those evoked by the NM-RSVP condition (also see Table 1 and 2). Even though ERP amplitudes looked relatively similar between the three conditions (the first row of Figure 3), the amplitude difference between targets and non-targets was larger for FM-RSVP than NM-RSVP (no statistically significant). The ERP latency and amplitude are affected by the recognition degree of given stimuli. In general, shorter latencies and larger amplitudes correspond with an easier task stimulus evaluation and response production [58]. Thus, it can be thought that motion-based RSVPs are recognized more easily than the NM-RSVP, resulting in significantly shorter ERP latencies and relatively lager ERP amplitudes.

The time series data measured from EOG channels were illustrated in Figure 6, according to the motion direction. In general, EOGs are generally produced right after eye movement with much sharper and higher amplitude than ERPs [59], unlike the patterns looking like ERPs shown in Figure 6. The general eye movement patterns were not elicited by both non-motion stimulation condition and the motion stimulation condition. Therefore, it would be more reasonable to assume

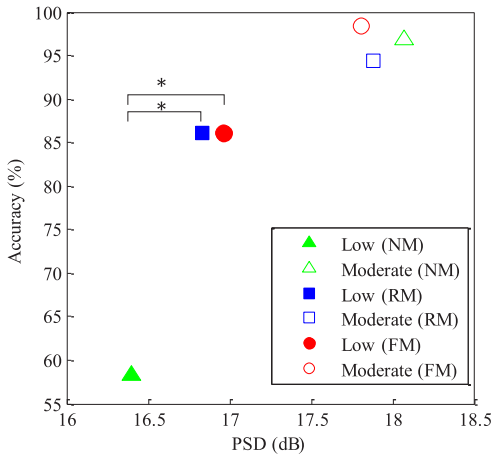


Fig. 7. Relationship between the accuracy and group-wise alpha powers for all three conditions in 10 sequences. In low performing subjects, a significantly different alpha power was noticed between NM-RSVP and motion-based RSVP (the Wilcoxon signed-rank test: FM-RSVP = RM-RSVP > NM-RSVP, Bonferroni corrected $p < 0.05$).

that the patterns recorded from EOG channels are due to the propagation of relatively strong ERPs rather than eye movement. Moreover, we validated no significant saccadic or micro-saccadic eye movement using gamma-band analysis (see the supplementary material). From the result, even though EEG signals would be affected by minimal eye movement, specifically for motion-based RSVP, it seems that these small eye movements do not significantly influence ERPs as well as classification performance. Nevertheless, the accurate impact of eye movements should be further investigated to develop a motions-based RSVP BCI application, such as using an eye-tracking system that directly measures eye movement.

A negative correlation between attention level and alpha activity is shown in general [54] (*e. g.*, higher attention level with lower alpha activity) while a positive correlation between BCI performance and alpha activity (*e. g.*, higher BCI performance with higher alpha activity) [55]–[57]. The low performing group showed a positive correlation between alpha power and BCI performance in the NM-RSVP paradigm, e.g., lower alpha power with lower performance. From the point of view of the attention level, it would be reasonable to interpret the result as low performing subjects would pay more attention to the task, leading to lower alpha activity, because it would be difficult for the low performing group to easily perceive rapidly presented characters the NM-RSVP paradigm, compared to the moderate performing group. Despite their efforts, it seemed that they could not correctly recognize characters presented in NM-RSVP paradigm, which might result in relatively less ERP amplitudes and poor classification performance. However, the low performing group showed larger ERP difference between targets and non-targets and increased classification accuracy when motion-based RSVPs were employed, demonstrating the importance of motions for the RSVP paradigm. On the other hand, the moderate performing group already showed high performance when using the NM-RSVP paradigm, which might be interpreted by the speculation that the moderate performing group has an ability to perceive rapidly presented static stimuli. This speculation should be more precisely investigated to pinpoint the origin of this phenomenon using behavior data, *e. g.*, checking whether the subject correctly counts the number of presented target characters.

Two subjects in the low performing group had experience with BCI experiments and four were naïve, while in the moderate performing group, four were naïve, and three had experience with BCI experiments. Considering the ratios of experienced subjects to non-experienced ones for each group, experience level did not significantly influence the performance. Moreover, the experimental order was randomly designed, which prevented potentially biased results for the stimulation types(*i. e.*, order of experimental stimulation types in the low performing group, subject 3: NM-FM-RM, subject 4: NM-FM-RM, subject 6: FM-NM-RM, subject 8: NM-FM-RM, subject 11: NM-FM-RM, and subject 13: FM-NM-RM, etc.). Note that the FM-RSVP condition showing the best mean results was always tested in either the first or the second session for the low performing group, not in the last session. Therefore, we believe that our experimental design prevented potentially biased results for a particular stimulation condition. Consequently, it can be thought that the improved classification accuracy was not considerably affected by the experimental factors, but the stimulation method itself.

Studying the performance for each sequence revealed remarkable accuracy differences between motion-based RSVP and NM-RSVP for the small numbers of sequences (*i. e.*, 1-5 sequences), ranging from 6.5% to 11.8% in RM-RSVP and from 10.8% to 15.1% in FM-RSVP, respectively (Table 3). Note that the target detection performance with shorter sequences is a challenging issue because it mainly influences overall BCI performance in terms of the operating speed [60]. Thus, the motion-based RSVP condition is beneficial, particularly for a small number of sequences. From the results, it can be suggested that five sequences would be the marginal number of stimulation repetitions for benefiting motion-based RSVP.

## V. Conclusion

In the present study, a novel RSVP paradigm using motion stimuli was proposed and compared with the classic RSVP paradigm with respect to ERP latency and amplitude and spelling performance. The classification accuracy of the RSVP-based speller system was highly improved by using motion stimuli, which was derived from the enhancement of ERP amplitude difference between targets and non-target. We believe that the motion-based RSVP paradigm has important advantages for paralyzed patients with ophthalmoplegia and oculomotor impairments because it does not require moderate eye movement. In future studies, the impact of motion stimuli on eye movement and EOGs should be investigated more accurately with a quantitative way (*e. g.*, eye tracker) for developing a perfectly gaze independent RSVP-based BCI system.

## Acknowledgment

The authors thank for valuable discussions with Benjamin Blankertz and Laura Acqualagna.

## REFERENCES


[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] G. Dornhege, J. D. R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, *Toward Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, 2007.

[3] J. R. Wolpaw, "Brain-computer interfaces as new brain output pathways," *J. Physiol.*, vol. 579, no. 3, pp. 613–619, 2007.

[4] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A lower limb exoskeleton control system based on steady state visual evoked potentials," *J. Neural Eng.*, vol. 12, no. 5, p. 056009, 2015.

[5] M.-H. Lee, S. Fazli, J. Mehnert, and S.-W. Lee, "Subject-dependent classification for robust idle state detection using multi-modal neuroimaging and data-fusion techniques in BCI," *Pattern Recognit.*, vol. 48, no. 8, pp. 2725–2737, 2015.

[6] M. F. Mason, M. I. Norton, J. D. Van Horn, D. M. Wegner, S. T. Grafton, and C. N. Macrae, "Response to comment on 'Wandering minds: The default network and stimulus-independent thought,"' *Science*, vol. 317, no. 5834, pp. 43c–43c, 2007.

[7] I.-H. Kim, J.-W. Kim, S. Haufe, and S.-W. Lee, "Detection of braking intention in diverse situations during simulated driving based on EEG feature combination," *J. Neural Eng.*, vol. 12, no. 1, p. 016001, 2015.

[8] B. Blankertz *et al.*, "The Berlin brain–computer interface: EEG-based communication without subject training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 147–152, Feb. 2006.

[9] B.-K. Min, S. Dähne, M.-H. Ahn, Y.-K. Noh, and K.-R. Müller, "Decoding of top-down cognitive processing for SSVEP-controlled BMI," *Sci. Rep.*, vol. 6, p. 36267, Sep. 2016.

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

[11] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS ONE*, vol. 12, no. 2, p. 0172578, 2017.

[12] B. Herrmann *et al.*, "Occurrence of ventricular late potentials in patients with active acromegaly," *Clin. Endocrinol.*, vol. 55, no. 2, pp. 201–207, 2001.

[13] H.-J. Hwang, J.-H. Lim, Y.-J. Jung, H. Choi, S. W. Lee, and C.-H. Im, "Development of an SSVEP-based BCI spelling system adopting a QWERTY-style LED keyboard," *J. Neurosci. Methods*, vol. 208, no. 1, pp. 59–65, 2012.

[14] D.-O. Won, H.-J. Hwang, S. Dähne, K.-R. Müller, and S.-W. Lee, "Effect of higher frequency on the classification of steady-state visual evoked potentials," *J. Neural Eng.*, vol. 13, no. 1, p. 016014, 2015.

[15] C. Neuper and G. Pfurtscheller, "Event-related dynamics of cortical rhythms: Frequency-specific features and functional correlates," *Int. J. Psychophysiol.*, vol. 43, no. 1, pp. 41–58, 2001.

[16] B. N. Cuthbert, H. T. Schupp, M. M. Bradley, N. Birbaumer, and P. J. Lang, "Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report," *Biol. Psychol.*, vol. 52, no. 2, pp. 95–111, 2000.

[17] N. Birbaumer and L. G. Cohen, "Brain–computer interfaces: Communication and restoration of movement in paralysis," *J. Physiol.*, vol. 579, no. 3, pp. 621–636, 2007.

[18] N. Birbaumer *et al.*, "A spelling device for the paralysed," *Nature*, vol. 398, no. 6725, pp. 297–298, 1999.

[19] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A P300 event-related potential brain–computer interface (BCI): The effects of matrix size and inter stimulus interval on performance," *Biol. Psychol.*, vol. 73, no. 3, pp. 242–252, 2006.

[20] S.-K. Yeom, S. Fazli, K.-R. Müller, and S.-W. Lee, "An efficient ERP-based brain-computer interface using random set presentation and face familiarity," *PLoS ONE*, vol. 9, no. 11, p. e111157, 2014.

[21] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based brain-computer interface," *Behavioral Brain Funct.*, vol. 6, no. 1, pp. 1–28, 2010.

[22] P. Brunner, S. Joshi, S. Briskin, J. Wolpaw, H. Bischof, and G. Schalk, "Does the 'P300' speller depend on eye gaze?" *J. Neural Eng.*, vol. 7, no. 5, p. 056013, 2010.

[23] S. Frenzel, E. Neubert, and C. Bandt, "Two communication lines in a 3×3 matrix speller," *J. Neural Eng.*, vol. 8, no. 3, p. 036021, 2011.

[24] H.-J. Hwang *et al.*, "A gaze independent brain-computer interface based on visual stimulation through closed eyelids," *Sci. Rep.*, vol. 5, p. 15890, Apr. 2015.

[25] J. Höhne, M. Schreuder, B. Blankertz, and M. Tangermann, "A novel 9-class auditory ERP paradigm driving a predictive text entry system," *Frontiers Neurosci.*, vol. 5, p. 99, Aug. 2011.

[26] D. S. Klobassa *et al.*, "Toward a high-throughput auditory P300-based brain–computer interface," *Clin. Neurophysiol.*, vol. 120, no. 7, pp. 1252–1261, 2009.

[27] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, "A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients," *Ann. New York Acad. Sci.*, vol. 1157, no. 1, pp. 90–100, 2009.

[28] A.-M. Brouwer and J. B. Van Erp, "A tactile P300 brain-computer interface," *Frontiers Neurosci.*, vol. 4, p. 19, Feb. 2010.

[29] M. S. Treder, N. M. Schmidt, and B. Blankertz, "Gaze-independent brain-computer interfaces based on covert attention and feature attention," *J. Neural Eng.*, vol. 8, no. 6, p. 066003, 2011.

[30] Y. Liu, Z. Zhou, and D. Hu, "Gaze independent brain–computer speller with covert visual search tasks," *Clin. Neurophysiol.*, vol. 122, no. 6, pp. 1127–1136, 2011.

[31] S. Schaeff, M. S. Treder, B. Venthur, and B. Blankertz, "Exploring motion VEPs for gaze-independent communication," *J. Neural Eng.*, vol. 9, no. 4, p. 045006, 2012.

[32] L. Acqualagna and B. Blankertz, "Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP)," *Clin. Neurophysiol.*, vol. 124, no. 5, pp. 901–908, 2013.

[33] R. W. Rodieck and R. W. Rodieck, *The First Steps in Seeing*. Sunderland, MA, USA: Sinauer Associates, 1998.

[34] P. Sajda *et al.*, "In a blink of an eye and a switch of a transistor: Cortically coupled computer vision," *Proc. IEEE*, vol. 98, no. 3, pp. 462–478, Mar. 2010.

[35] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.

[36] G. Aschersleben and J. Müsseler, "Dissociations in the timing of stationary and moving stimuli," *J. Experim. Psychol., Human Perception Perform.*, vol. 25, no. 6, p. 1709, 1999.

[37] M. Mashhour, *Psychophysical Relations in the Perception of Velocity*. Almquist & Wiksell, 1964.

[38] J. B. Smeets and E. Brenner, "Perception and action are based on the same visual information: Distinction between position and velocity," *J. Experim. Psychol. Human Perception Perform.*, vol. 21, no. 1, pp. 19–31, 1995.

[39] S. P. Heinrich, "A primer on motion visual evoked potentials," *Document. Ophthalmol.*, vol. 114, no. 2, pp. 83–105, 2007.

[40] M. Kuba, Z. Kubová, J. Kremláček, and J. Langrová, "Motion-onset VEPs: Characteristics, methods, and diagnostic use," *Vis. Res.*, vol. 47, no. 2, pp. 189–202, 2007.

[41] R. Engbert, A. Longtin, and R. Kliegl, "A dynamical model of saccade generation in reading based on spatially distributed lexical processing," *Vis. Res.*, vol. 42, no. 5, pp. 621–636, 2002.

[42] M. J. Traxler, *Introduction to Psycholinguistics: Understanding Language Science*. Hoboken, NJ, USA: Wiley, 2011.

[43] A. Ziehe and K.-R. Müller, "TDSEP—An efficient algorithm for blind separation using time structure," in *Proc. ICANN*, 1998, pp. 675–680.

[44] J.-H. Kim, F. Biessmann, and S.-W. Lee, "Decoding three-dimensional trajectory of executed and imagined arm movements from electroencephalogram signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 5, pp. 867–876, Sep. 2015.

[45] 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.

[46] B. Blankertz *et al.*, "The Berlin brain–computer interface: Non-medical uses of BCI technology," *Frontiers Neurosci.*, vol. 4, p. 198, Dec. 2010.

[47] E. Iáñez, J. M. Azorin, and C. Perez-Vidal, "Using eye movement to control a computer: A design for a lightweight electro-oculogram electrode array and computer interface," *PLoS ONE*, vol. 8, no. 7, p. e67099, 2013.

[48] S. Martinez-Conde, S. L. Macknik, X. G. Troncoso, and D. H. Hubel, "Microsaccades: A neurophysiological analysis," *Trends Neurosci.*, vol. 32, no. 9, pp. 463–475, 2009.

[49] D. J. Schwartzman and C. Kranczioch, "In the blink of an eye: The contribution of microsaccadic activity to the induced $\gamma$ band response," *Int. J. Psychophysiol.*, vol. 79, no. 1, pp. 73–82, 2011.

[50] U. Hassler, N. T. Barreto, and T. Gruber, "Induced gamma band responses in human eeg after the control of miniature saccadic artifacts," *Neuroimage*, vol. 57, no. 4, pp. 1411–1421, 2011.

[51] O. Dimigen, M. Valsecchi, W. Sommer, and R. Kliegl, "Human microsaccade-related visual brain responses," *J. Neurosci.*, vol. 29, no. 39, pp. 12321–12331, 2009.

[52] S. Yuval-Greenberg, O. Tomer, A. S. Keren, I. Nelken, and L. Y. Deouell, "Transient induced gamma-band response in EEG as a manifestation of miniature saccades," *Neuron*, vol. 58, no. 3, pp. 429–441, 2008.

[53] A. S. Keren, S. Yuval-Greenberg, and L. Y. Deouell, "Saccadic spike potentials in gamma-band EEG: Characterization, detection and suppression," *Neuroimage*, vol. 49, no. 3, pp. 2248–2263, 2010.

[54] L. Acqualagna *et al.*, "EEG-based classification of video quality perception using steady state visual evoked potentials (SSVEPs)," *J. Neural Eng.*, vol. 12, no. 2, p. 026012, 2015.

[55] M. S. Treder, A. Bahramisharif, N. M. Schmidt, M. A. van Gerven, and B. Blankertz, "Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention," *J. Neuroeng. Rehabil.*, vol. 8, no. 1, pp. 1–24, May 2011.

[56] B. Blankertz *et al.*, "Neurophysiological predictor of SMR-based BCI performance," *Neuroimage*, vol. 51, no. 4, pp. 1303–1309, 2010.

[57] H.-I. Suk, S. Fazli, J. Mehnert, K.-R. Müller, and S.-W. Lee, "Predicting BCI subject performance using probabilistic spatio-temporal filters," *PLoS ONE*, vol. 9, no. 2, p. e87056, 2014.

[58] M. Kutas, G. McCarthy, and E. Donchin, "Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time," *Science*, vol. 197, no. 4305, pp. 792–795, 1977.

[59] Z. Lv, X.-P. Wu, M. Li, and D. Zhang, "A novel eye movement detection algorithm for EOG driven human computer interface," *Pattern Recognit. Lett.*, vol. 31, no. 9, pp. 1041–1047, 2010.

[60] M. A. Conroy and J. Polich, "Normative variation of p3a and p3b from a large sample: Gender, topography, and response time," *J. Psychophysiol.*, vol. 21, no. 1, pp. 22–32, 2007.