Doubling the Speed of N200 Speller via Dual-directional Motion Encoding

Dingkun Liu, Chang Liu, Jingjing Chen, Dan Zhang, *Member, IEEE*, and Bo Hong, *Member, IEEE*

Abstract*—*Objective: **Motion-onset visual evoked potentials (mVEPs)-based spellers, also known as N200 spellers, have been successfully implemented, avoiding flashing stimuli that are common in visual brain-computer interface (BCI). However, their information transfer rates (ITRs), typically below 50 bits/min, are lower than other visual BCI spellers. In this study, we sought to improve the speed of N200 speller to a level above the well-known P300 spellers.** ***Approach:*** **Based on our finding of the spatio-temporal asymmetry of N200 response elicited by leftward and rightward visual motion, a novel dual-directional N200 speller was implemented. By presenting visual stimuli moving in two different directions simultaneously, the new paradigm reduced the stimuli presentation time by half, while ensuring separable N200 features between two visual motion directions. Furthermore, a probability-based dynamic stopping algorithm was also proposed to shorten the decision time for each output further. Both offline and online tests were conducted to evaluate the performance in ten participants.** ***Main results:*** **Offline results revealed contralateral dominant temporal and spatial patterns in N200 responses when subjects attended to stimuli moving leftward or rightward. In online experiments, the dual-directional paradigm achieved an average ITR of 79.8 bits/min, with the highest ITR of 124.8 bits/min. Compared with the traditional uni-directional N200 speller, the median gain on the ITR was 202 %.** ***Significance:*** **The proposed dual-directional paradigm managed to double the speed of the N200 speller. Together with its non-flashing characteristics, this dual-directional N200 speller is promising to be a competent candidate for fast and reliable BCI applications.**

***Index Terms*—Brain-computer interface, N200 speller, dual-directional visual motion, dynamic stopping algorithm, motion-onset visual evoked potentials**

I. Introduction

MOTION-ONSET visual evoked potentials (mVEPs) based brain computer interface (BCI) speller (also known as N200 speller) is a well-established non-flashing visual BCI paradigm [1]–[7]. By attending to the target in which a bar was moving from right to left in an overt [1]–[3] or covert [6] way, participants were able to input characters with a relatively low visual load due to the non-flashing and low contrast characteristics of motion stimuli [2], [8]. N200 spellers have proven its feasibility in both EEG [2], [3] and intracranial EEG (iEEG) [9], and also succeeded in non-control state detection [10]. However, their information transfer rates (ITRs), ranging from 15.9 to 35.6 bits/min [2], [4], were relatively low when compared with the well-known P300 spellers and therefore, greatly limited its possible application.

Efforts have been made to improve the performance of the N200 speller. For instance, high gamma band responses after the motion-onset were utilized as an additional feature in the classification along with the mVEPs features, which resulted in a significant ITR improvement in an iEEG based BCI speller [9]. Smart stopping algorithms, which optimized the number of averaging trials dynamically, have demonstrated their potentials in performance improvement as well [3], [11], [12]. However, even the best ITR ever reported in N200 spellers, which was equivalent to spell about ten correct characters per minute, is not satisfactory for practical applications. New approaches are highly expected to raise its speed further.

One possible solution to improve the ITRs of the N200 BCIs is to reduce the stimuli presentation time for every output decision by increasing the number of targets encoded simultaneously [13]–[16]. Conventional row/column N200 spellers use the stimuli moving in one direction to tag each row/column at one time and therefore need at least $M * ISI$ (M is the number of rows and columns in a speller, and ISI is the necessary inter-stimulus interval) to cover all targets. If multiple directions of visual motion stimuli could be incorporated to encode multiple targets at the same time, it is possible to significantly increase the time efficiency of stimuli presentation, and consequently, boost the typing speed. Indeed, previous studies have observed the directional modulation effects of visual motion stimulus on the cortical responses. Kamitani et al. used fMRI to show that ensemble activity patterns in the human visual cortex contain robust direction-selective information [17]. Meanwhile, Hoffmann et al. demonstrated that at least 28 % of the N200 amplitude reflects the activity of direction-specific elements [18]. Besides, Guo et al.'s study on the N200 BCI speller indicated that there were differences in the spatial patterns of the responses of visual motion stimuli of different directions [1]. These fMRI and EEG studies provided neural evidence which supported the feasibility of employing the directional information of visual motion stimuli as an additional dimension to encode the visual BCI targets.

In this work, we proposed a novel N200 speller based on dual-directional visual motion stimuli. By presenting visual stimuli moving in two different directions at the same time, the new paradigm reduced the stimuli presentation time by half. We also designed a probability-based dynamic stopping algorithm to shorten the decision time further. With the dual-directional visual motion BCI paradigm, an average ITR

*This work was supported by National Science Foundation of China (NSFC grant 61473169 and 61621136008).

D. Liu, C. Liu, and J. Chen, are with the Dept. of Biomedical Engineering, School of Medicine, Tsinghua University.

D. Zhang is with the Dept. of Psychology, School of Social Sciences, Tsinghua University.

B. Hong is with the Dept. of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China. Email: hongbo@tsinghua.edu.cn

of 79.8 ± 24.7 bits/min on ten subjects was achieved in online spelling, with the maximum reached 124.8 bits/min (the theoretical upper bound was approximately 25 characters per minutes). There was a median gain of 202 % on ITR over the conventional uni-directional N200 speller. Besides, the newly proposed dual-directional N200 speller is also superior to the classical row/column P300 speller (typically lower than 50 bits/min [19]–[22]).

II. Methods

A. Stimulation Paradigm

The visual stimuli, with a viewing distance of 50 cm, were displayed on a 23-inch LCD monitor with 60 Hz refresh rate and 1920×1080 resolution. 6×6 rectangle virtual buttons (each with a visual field of $1.64° \times 1.64°$) were placed uniformly on the speller interface (Fig.1a and 1c). In each virtual button, the vertical bar appears (motion-onset) at the right or the left border of the virtual button, moves $1.09°$ (9.5 mm on the screen) towards the opposite side before it disappears (motion offset), forming a 150-ms visual motion stimulus. The color of the vertical bar, which might be red, green, blue, purple, yellow, or brown, was randomly designated in such a protocol that the colors of the six bars in the same row or column were different from each other. The random color was employed to enhance the attention on the desired target as the subject would perform color recognition tasks during spelling silently [2].

For the uni-directional paradigm, an epoch was comprised of six stimuli in one single row/column moved in the same direction. For instance, assume "Z" was the attended target, the 5th row-epoch (R5) and the 2nd column-epoch (C2) would elicit target responses (Fig.1a and b). For the dual-directional paradigm, an epoch was comprised of twelve stimuli in two rows/columns moving in the opposite direction simultaneously (Fig.1c). Note that the two rows/columns of stimuli in one epoch were fixed to be three lines apart (e.g., C1 & C4, R2 & R5). For the 1st to 3rd row/column, the stimuli moved leftward, whereas for the 4th to 6th row/column, the stimuli moved rightward (Fig.1c). Based on the direction combinations of the row and column visual motion stimuli, the dual-directional interface could be divided into four different quadrants (I, II, III, and IV). For instance, the target letter "Z" was located in the quadrant IV (Fig.1c). The 2nd row-epoch (R2, leftward & R5, rightward) and the 2nd column-epoch would elicit responses corresponding to the rightward and leftward target stimuli, respectively. The ISIs were 200 ms for both paradigms.

Here, a trial was defined as a complete series of motion stimuli in all the rows and columns of the keyboard. For the uni-directional paradigm, a trial was comprised of twelve epochs (six row- and six column-epochs, Fig.1b), whereas that of the dual-directional paradigm was comprised of six epochs (Fig.1d). Thus, the stimulus presentation time of each trial in the uni-directional paradigm, which was 2400 ms ($200\ ms \times (6\ rows + 6\ cols)$), was reduced by 50 % to 1200 ms ($200\ ms \times (3\ rows + 3\ cols)$) in the dual-directional paradigm. The interval between two trials was about 1 s.

The software for online stimulus display was developed with Python (Python Software Foundation) and Psychopy [23].

B. Experimental Setup

Subjects participated in two sessions, the offline and the online sessions. During the experiment, subjects were required to gaze at the letters in the center of the target virtual buttons, and distinguish the colors of moving bars that appeared inside the attended target at the same time.

1) Offline training: In the offline session, each target repeated ten times as one block. At the beginning of each block, the border of the to-be-attended virtual button would be drawn with red, serving as a hint to the subjects for 0.6 s. And at the end of each block, the attended virtual button would be filled with black, serving as a hint for the next target for 0.3 s. For the uni-directional paradigm, six virtual buttons on the diagonal ("AHOV29") were sequentially assigned as attended targets [2]. However, for the dual-directional paradigm, twelve virtual buttons on the diagonal ("AHOV29") and counter-diagonal ("FKPUZ4") are required to keep the overall number of epochs the same as that in the uni-directional paradigm. Because the number of epochs in one trial of the dual-directional paradigm is half of that in its uni-directional counterpart. The direction with better offline performances for each subject was used in the online uni-directional paradigm.

2) Online testing: In the online session, subjects were instructed to sequentially type a 36-character sequence from "A" to "9" in both the uni- and dual-directional N200 spellers, which were conducted in random order, with an online classification strategy explained in (Sec.II-E). The stimulation in the online session is the same with the offline session, but the only difference was that the chosen virtual button rather than the assigned one would be filled with black after every selection. When one character was mistyped, the subjects were told not to correct that character but to continue with the next.

C. Subjects and Data Recording

Seven male and three female volunteers aged 20-30 participated in the study. All subjects gave informed consent and were paid for their participation. Each subject showed normal or corrected to normal vision, without any clinical visual diseases. Five of them (S1, S2, S4, S9, and S10) participated in the pilot offline experiment in which a 32-channel wireless EEG amplifier (10-20 system [24], NeusenW, Neuracle Inc.) was employed to record data to investigate the spatial pattern of dual-directional mVEPs. Then, in order to simplify the recording setting, in the formal experiments (including the offline task and online task), eight electrodes (P7, P8, P3, P4, Pz, O1, O2, and Oz) covering occipital and parietal areas [10] were chosen to record the EEG data for all the ten subjects. The electrode CPz and AFz were chosen as REF and GND, respectively. All electrodes' impedance was reduced to $10k\Omega$ before data recording. Signals were sampled at 1000 Hz. Trigger events were acquired simultaneously with stimulus onsets.

D. Data Processing and BCI Classification

First, raw EEG data were digitally filtered forward and backward with a 200 order FIR bandpass filter (passband 1-20 Hz) and then segmented into epochs from 300 ms preceding

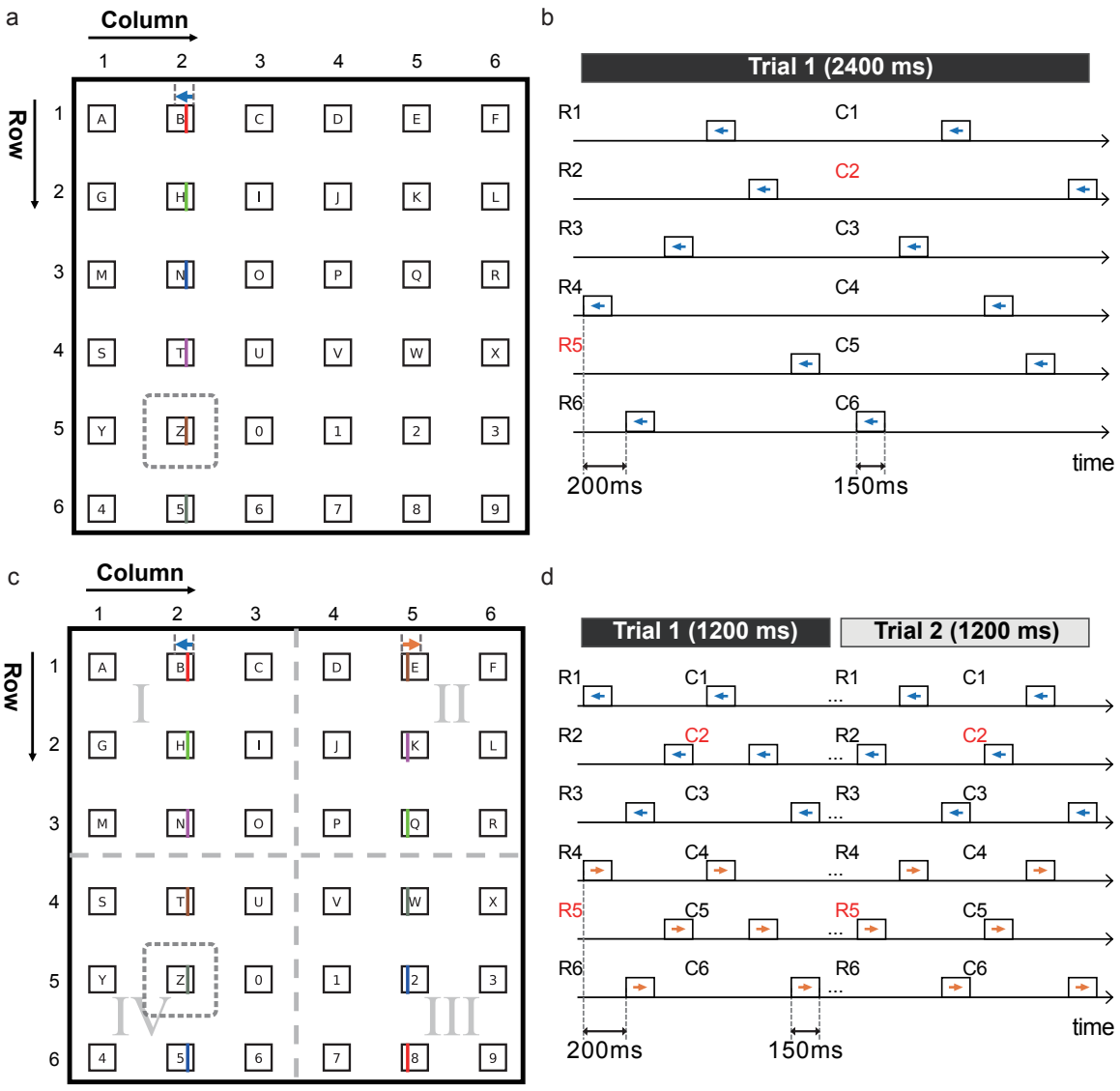

Fig. 1. The stimulus and temporal scheme comparison of the uni- and dual-directional visual motion BCI speller. (a)(c) Interfaces of the uni- and dual-directional N200 spellers, respectively. The leftward and rightward motions are marked by the blue and orange arrows, respectively. The attended targets "Z" are marked by the dashed-line boxes for both paradigms. (b)(d) Temporal schemes of the uni- and dual-directional visual-motion stimulus presentation. Within each trial of stimulus presentation, the stimuli, which were presented in random order, were separated into a column scan and a row scan. Dynamic illustrations for both the paradigms will be available at https://github.com/HongLabTHU/Dual-mVEPs-Speller.

the motion stimulus onset to 600 ms post the onset. The data in -300 to 0 ms to the motion onset was set as the baseline. Afterward, linear detrending and baseline correction was applied to theses 900-ms epochs. For the 32-channel data gathered in the pilot experiment, an independent component analysis (ICA) was then applied in order to remove the EOG artifacts [25]. Given the difficulty in performing ICA online, this step was not applied to the 8-channel data collected in the formal experiments. Next, data from 0 to 500 ms after stimulus onsets, covering N200 and its subsequent components, were employed in the classification. These eight-channel epochs were downsampled to a sampling rate at 40 Hz and normalized by subtracting the mean and dividing the standard deviation for each channel. Finally, the normalized 8-channel data were concatenated together to form a 160-dimension feature vector for classification.

The logistic regression classifier was adopted to discriminate leftward, rightward, and non-target stimuli. One versus rest (OVR) strategy was employed to train a binary classifier for each class. L2 norm regularization was applied to penalize over-fitting. The classifier was evaluated with a fixed two-trial averaging of data trials. Since the ratio of the left, right, and non-target categories in training data was 1:1:4, to deal with the unbalanced data, ROC was chosen as the critical metric as it was robust to unbalanced data [26]. All the offline performances were evaluated with 10-fold cross-validation, whose training and validation sets were uniformly sampled from each class. In online experiments, the ITR [27] was chosen as the metric to evaluate the performance. The ITR is defined in Eq.1, where N is the number of possible selections,

Fig. 2. Probability-based dynamic stopping algorithm. Subject S9 was attending to the target "A" marked by the orange box. The figures showed the likelihood of each button being the real attended one in three consecutive trials.

P is the selection accuracy, and T is the average time required to select a single symbol.

$$ITR = (\log N + P \log P + (1-P) \log \frac{1-P}{N-1})/T \quad (1)$$

All data analyses were carried out using Python (Python Software Foundation) with Numpy [28], Scipy [29], Scikit-learn [30], and MNE [31]. The source code is available at https://github.com/HongLabTHU/dual-mVEPs-speller.

E. Probability-based Dynamic Stopping

In order to find a tradeoff between the decision time and accuracy, we adopted a probability-based dynamic stopping strategy to decide the number of trials required to get a reliable output online. Details of the algorithm could be found in Appendix A. Generally speaking, the dynamic stopping algorithm computes the likelihood of average decision values between trials with softmax function (see Eq.4), which is a widely-used function to map outputs of multiclass classifiers to likelihoods [32]. When the maximum likelihood among the targets or the number of averaging trials reaches a predefined threshold, the algorithm ends the iteration and chooses the most likely target as the predicted one.

As shown in Fig.2, at the first iteration, the likelihood of the attended target "A" was as low as 4 %. At the subsequent iterations, the decision values were averaged between trials, making the likelihood of the target to increase gradually. At the third iteration, the likelihood of the letter "A" to be the attended target was 97 %, which was reliable enough to make the decision. During the iterations, the likelihoods are guaranteed to gradually converge to 1 based on the fact that the standard deviation of average decision values decays at the order of $\sqrt{N}$ (Appendix B). Compared with our previous threshold-based dynamic stopping criteria for mVEPs [3], the newly proposed method is more robust to signal drift [11] since the common part added to the decision values can be reduced by the softmax function.

III. Results

A. Neural Responses of the Dual-directional Visual Motion Stimuli

The grand average ERP waveforms in different motion stimuli conditions are shown in Fig.3. As observed, the negative peaks at 150 to 300 ms after stimulus onset, which corresponded to prominent N200, were elicited on electrodes P3, P7, O1, P4, P8, and O2 when participants attended to the leftward or the rightward stimuli. Different latencies and amplitudes of N200 peaks could be observed between the leftward and the rightward condition. As shown in Fig.4, paired t-tests indicated that N200 peaks of leftward stimuli responses were significantly earlier than that of rightward stimuli responses ($p < 0.001$), and with significantly larger amplitude ($p = 0.017$) on the left hemisphere (electrode P3, P7, O1), while N200 peaks of rightward stimuli responses were significantly earlier than that of leftward stimuli responses ($p = 0.008$), and with significantly larger amplitude ($p = 0.019$, paired t-test) on the right hemisphere (electrode P4, P8, O2), suggesting a contralateral dominant patterns relative to the stimulus onset positions in the visual field under leftward and rightward conditions. Moreover, in the middle line of the brain, N200 peak latencies and amplitudes for the leftward and rightward stimuli showed no significant differences ($p = 0.200$ and $p = 0.059$ for latencies and amplitudes, respectively) (Fig.4).

To obtain the complete topomap of N200 in the dual-directional visual motion paradigm, each of the five subjects (S1, S2, S4, S9, and S10) participated in a pilot offline session with 32-channel EEG recording. The spatial patterns of the N200 responses (average responses between 175 ms and 225 ms) are illustrated in Fig.5. When attending to the stimuli moving from right to left (leftward), the N200 responses could be observed at the left occipital and parietal area. When attending to the stimuli moving from left to right (rightward), the N200 responses could be found at the right occipital and parietal area, showing a contralateral dominance relative to the stimulus onset positions. These neural differences in the

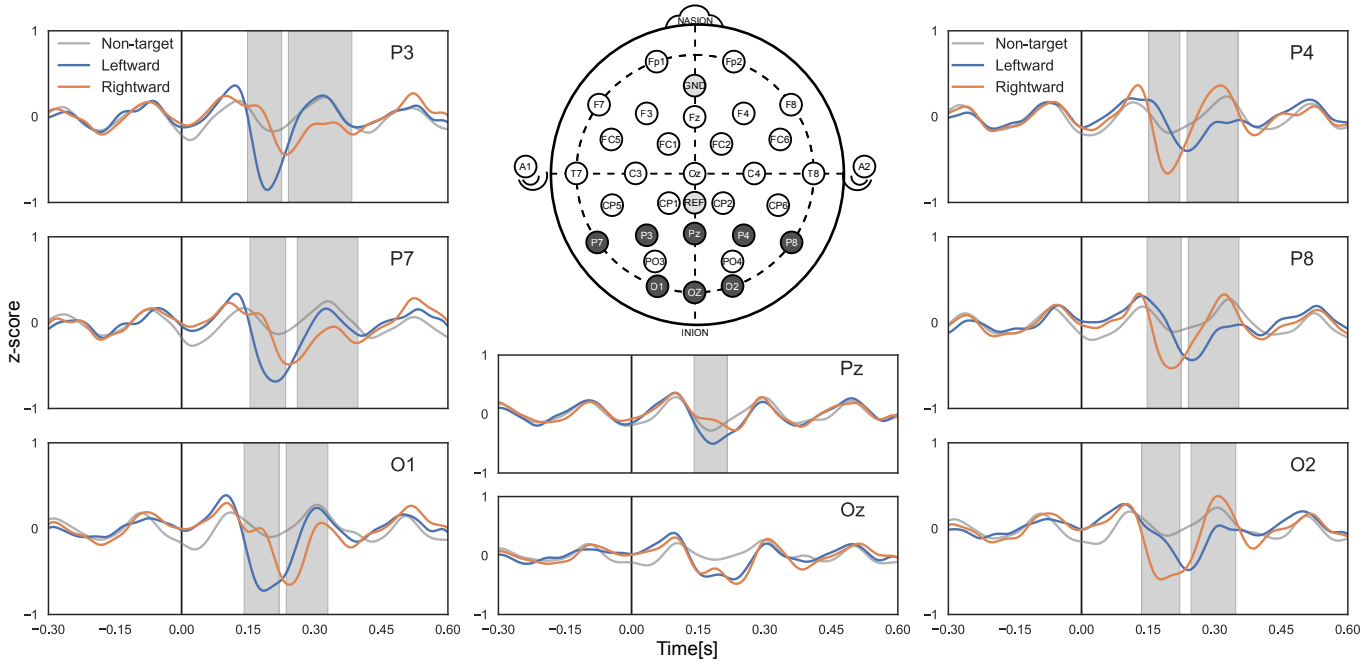

Fig. 3. Grand-average ERP responses for the dual-directional visual motion responses (z-scored). Gray lines represented mean ERP waveforms for unattended targets, and blue lines and orange lines represented leftward and rightward stimulus-related responses, respectively. Gray shadows indicated clusters when responses to leftward stimuli and rightward stimuli were significantly different ($p < 0.001$, clustering permutation test [33]) during 0-500 ms after stimulus onset. The middle upper panel was the electrode layout for the 32-channel EEG used in our pilot experiments. Eight channels used in the formal experiments were marked as dark gray.

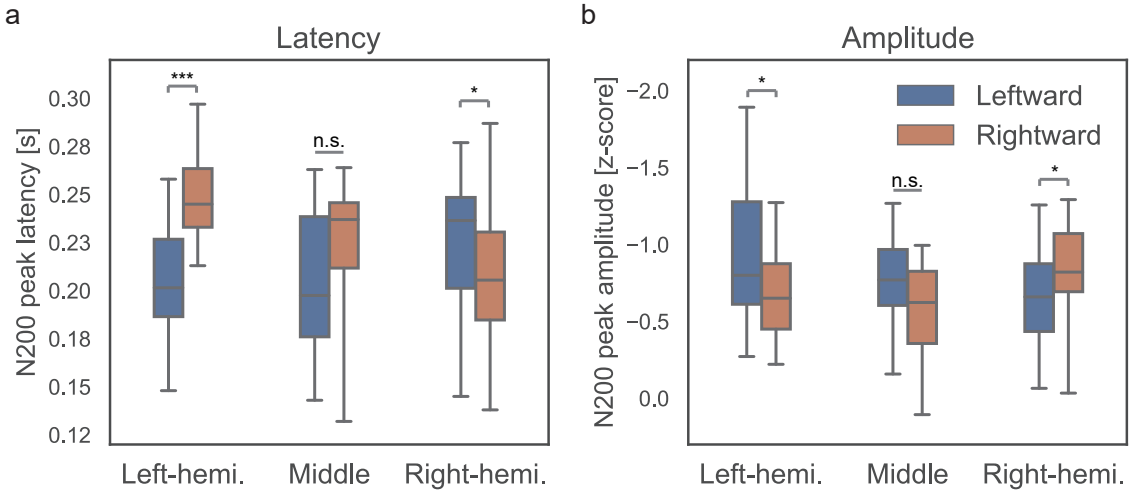

Fig. 4. The comparison of N200 peak latencies (a) and amplitudes (b) between leftward and rightward visual motion stimuli. Both the three boxes in (a) and (b) from left to right represented N200 peak latencies and amplitudes of electrodes on the left hemisphere (P3, P7, O1), in the middle (Pz, Oz), and on the right hemisphere (P4, P8, O2), respectively. *: $p < 0.05$, t-test, ***: $p < 0.001$, t-test, n.s.: no significant differences.

temporal and spatial patterns corresponding to the leftward and rightward stimuli provided separable features to a machine classifier, indicating the feasibility to employ the direction of visual motion stimuli as an additional dimension of encoding BCI targets.

B. Offline Performances

The offline classification performances were reported in Fig.6. With a fixed two-trial averaging, the mean areas under ROCs (AUC) were 0.92 ± 0.05 and 0.93 ± 0.05 for the recognition of the leftward, and rightward stimuli, respectively (Fig.6a), which were far above the chance level (0.5) and showed no significant performance differences between leftward and rightward conditions ($p = 0.592$, paired t-test).

We also test how the numbers of averaging trials influence on the classification performance. As found in Fig.6, the average AUC across all the ten subjects increased gradually as the number of averaging increase, and an average AUC of

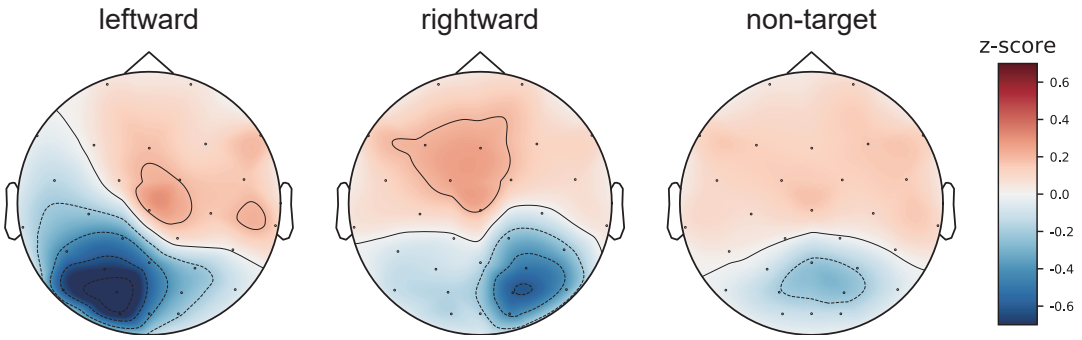

Fig. 5. The average spatial patterns of dual-directional visual motion stimulus responses (z-scored) from five subjects (S1, S2, S4, S9, and S10).

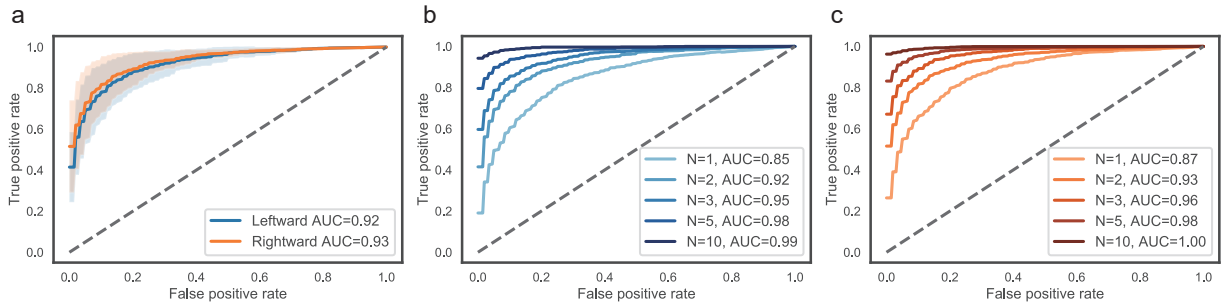

Fig. 6. ROCs of offline classification tasks with a fixed two-trial averaging of ERP data. (a) Average ROCs of different classes. Shadows in different colors indicate the standard deviation for different classes. (b)(c) ROCs of different numbers of averaging trials ($N = 1, 2, 3, 5, 10$) for leftward and rightward responses, respectively.

above 0.95 could be achieved by using only three trials in the offline classification, suggesting a possible real-time and accurate decoding performance (Fig.6b and 6c).

C. Online Performances

Then, online experiments were carried out on ten subjects. The accuracy, decision time, and ITR [27] (see Eq.1) were employed to compare the performance of the uni-directional and dual-directional paradigms.

The online results were listed in (Table I). The average ITR, accuracy, and decision time for the dual-directional paradigm were 79.8 ± 24.7 bits/min, 82.8 ± 8.02 %, and 3.0 ± 0.79 s, respectively, whereas the uni-directional paradigm achieved 42.1 ± 14.9 bits/min, 84.4 ± 8.62 %, and 5.9 ± 1.7 s, respectively. Compared with the uni-directional paradigm, the ITRs in the dual-directional paradigm were largely improved, and the best subject achieved ITR as high as 124.8 bits/min, with averaged 2.2 s per character and accuracy of 94.4 %.

After the experiment, the subject S1 participated in an additional session of copy spelling BCI task to type a sentence, "Not to be known should not grieve you; grieve that ye know not men," which is a famous saying of Confucius, a well-known Chinese philosopher [34]. The subject achieved a remarkable ITR of 162 bits/min in this task (See online video https://youtu.be/eUth16SgnOc).

Then, we took a closer look at what contributed to the improvement of the online performance in the dual-directional paradigm. As shown in Fig.7, the paired t-test found that the new paradigm showed significant improvement in speed ($p < 0.001$, Fig.7a), but no significant difference in accuracy ($p = 0.677$, Fig.7b). The decrease in average decision time and maintenance on accuracy resulted in the significant improvement in the average ITR ($p < 0.001$, Fig.7c). The median gain over the uni-directional paradigm on ten subjects was 202 % (Fig.7d), supporting that the dual-directional paradigm was able to double the speed of the original uni-directional paradigm.

Besides, the proposed dynamic stopping algorithm also contributed to improving the speed of the proposed BCI speller. With this dynamic stopping algorithm, we found a significant negative correlation between offline AUC and online average decision time for the ten subjects ($r = -0.8$, $p = 0.006$, Pearson correlation, Fig.8a). As higher offline AUC often indicated better signal quality, the algorithm saved more time to make decisions for subjects with better signal quality.

Furthermore, we estimated the dynamic stopping decision parameter using the offline data of each subject (see Appendix B) and compared the actual online accuracies with the estimated one. The accuracies theoretically estimated with the offline data were shown in (Fig.8b). For high performance subject S1, medium performance subject S6, and low performance subject S10, number of averaging trials of 1.8, 3.2, and > 5 were required to achieve accuracies of 85 % in the offline estimation, respectively. In the actual online task, better yet comparable performances could be obtained. These results suggested that the threshold of the number of averaging trials estimated with offline data is informative since it is not too

TABLE I
THE COMPARISON BETWEEN UNI- AND DUAL-DIRECTIONAL BCI ONLINE PERFORMANCES

Subject	ITR [bits/min]		Accuracy [%]		Decision time [s]	
	Uni-dir	Dual-dir	Uni-dir	Dual-dir	Uni-dir	Dual-dir
S1	55.7	**124.8**	91.7	94.4	4.7	2.2
S2	63.2	101.1	97.2	86.1	4.6	2.3
S3	37.5	93.9	83.3	91.7	5.9	2.8
S4	61.8	87.7	91.6	77.8	4.2	2.2
S5	51.5	87.2	83.3	86.1	4.3	2.7
S6	38.4	80.4	88.9	83.3	6.4	2.7
S7	40.5	72.6	80.5	77.8	5.1	2.7
S8	32.8	68.7	80.6	83.3	6.3	3.2
S9	22.4	46.8	83.3	83.3	9.8	4.7
S10	17.7	34.7	63.9	63.9	8.1	4.1
Mean	42.1	**79.8**	84.4	82.8	5.9	**3.0**
STD	14.9	24.7	8.62	8.02	1.7	0.79

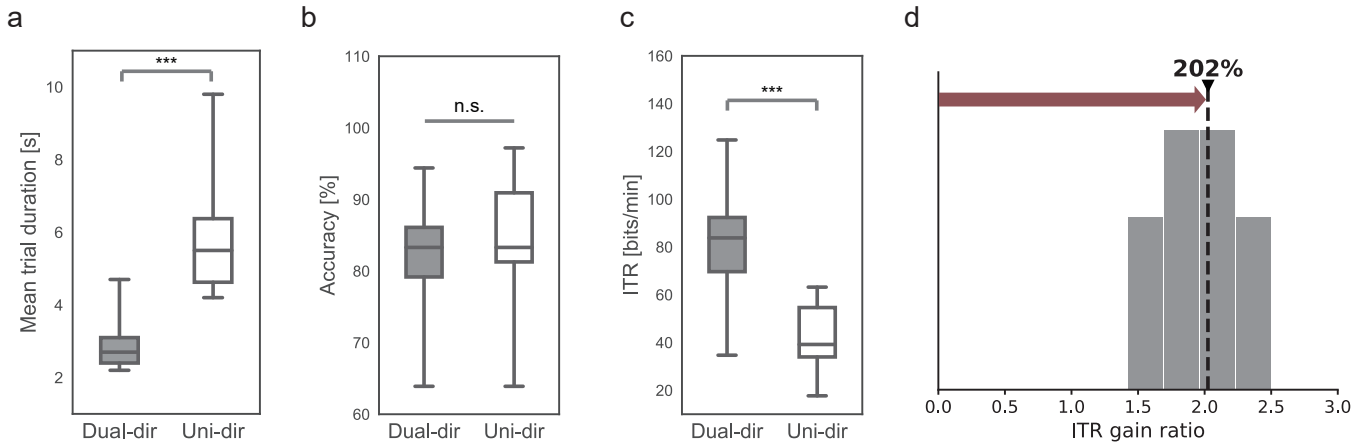

Fig. 7. Comparison of online performances between dual- and uni-directional paradigm. Gray and white boxes in (a)-(c) represented the dual-directional and the uni-directional paradigm, respectively. (a) Comparison of mean decision time. (b) Comparison of spelling accuracies. (c) Comparison of online ITRs. ***: $p < 0.001$, t-test, n.s.: no significant differences. (d) Histogram of dual-directional paradigm ITR gain for results in (Table I). The arrow highlights the median gain with respect to the uni-directional paradigm.

conservative and allows room for faster decisions online.

IV. DISCUSSION

The present study proposed a novel dual-directional N200 BCI speller. By adding the motion direction information in the traditional row/column stimuli paradigm, this speller is able to reduce the stimulation presentation time of one single trial by 50 %. The differences between neural response patterns to visual motion stimuli with different moving directions supported reliable classifications in both the offline and online tests. Together with the dynamic stopping approach, the dual-directional N200 speller achieved an average ITR of 79.8 ± 24.7 bits/min, which was a 2x speed increase compared with the uni-directional N200 speller [2], [3]. The best subject achieved ITR as high as 124.8 bits/min, with a speed of 2.2 s per character and accuracy of 94.4 %. Our results support the feasibility of utilizing dual-directional visual motion stimuli to increase the speed of the N200 speller.

The distinct neural patterns to the visual stimuli with different moving directions lay the neurophysiological basis for the proposed dual-directional N200 paradigm. Consistent with previous EEG and fMRI studies, the directional modulation effect on neural components were found in the present study in spatial patterns as well as temporal dynamics [17], [18], [35]. Furthermore, our offline and online results suggested that the neural patterns encoded by the moving directions could be recognized by the classifiers with a single or a few trials averaged, supporting the feasibility of the real-time direction decoding and its potentials in BCI applications.

More importantly, by utilizing the direction of visual motion stimuli as an encoding dimension, the proposed dual-directional design achieves the best performance in N200 spellers to the best of our knowledge and is also among the fastest ERP based spellers [13], [36]. Kaufmann et al., developed a speller utilizing simultaneously presented "two-stimulus" (a face [22] and a symbol), which achieved twice the speed of the one-stimulus speller, while still maintaining the accuracy of 81 % [13]. This paradigm relied on the different P300 response patterns elicited by different stimuli, e.g., faces and symbols. Townsend et al. replaced the classical row/column paradigm with a checkerboard paradigm and achieved the fastest P300 speller as reported [36]. Note that

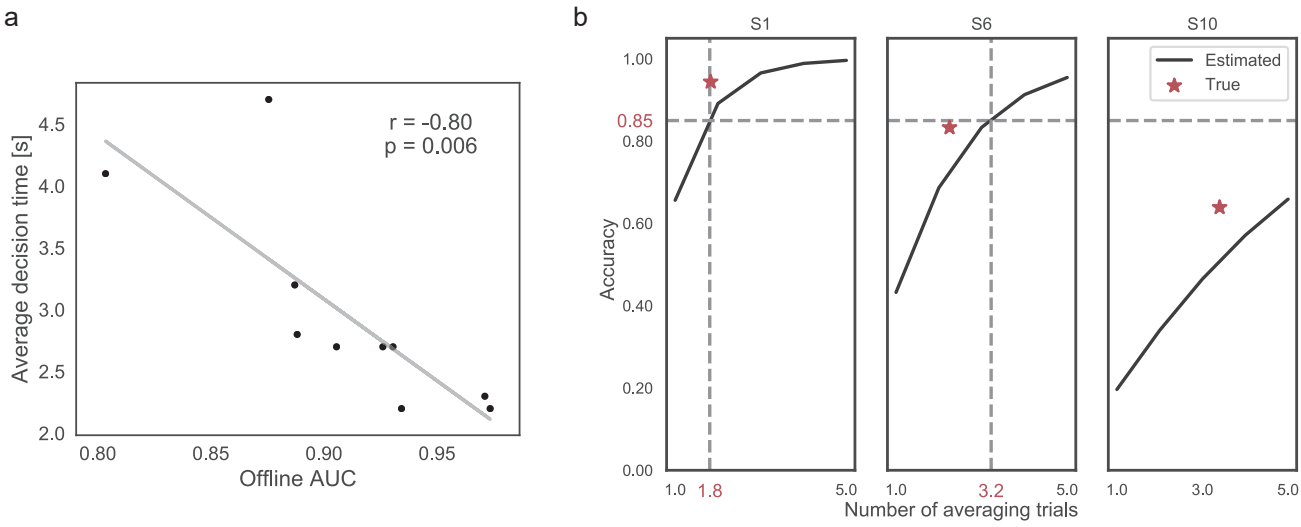

Fig. 8. Results of the probability-based dynamic stopping algorithm. (a) Relationship between offline AUC and average online decision time. (b) Estimated thresholds for the number of averaging for subject S1, S6, and S10, respectively. The black lines indicated the estimated accuracies under different numbers of averaging. The gray dash line indicated the estimated required number of averaging to achieve at least 85 % accuracy for each subject. The red stars marked the true numbers of averaging of each subject in online session.

this work was mainly focused on the optimization of the time courses of the stimulus presentation order while our study added the stimulus categories. Therefore, it is with great feasibility to integrate our dual-directional motion stimuli design and the strategy proposed by Townsend et al. together and push the speed of the ERP speller to a new stage.

As the first N200 speller study to demonstrate the merits of adding direction information in stimuli, there are still lots of room for improvement in future work. Firstly, more parameters should be explored to evaluate their feasibility in reducing the stimuli presentation time, such as more moving directions, like up and down, and more kinds of stimuli, like contraction-expansion and random-dot kinematograms. It is possible to improve the performance of the N200 speller further by encoding targets with richer motion information [20], [37]. Secondly, as the individual differences were found in the present study when using the naive classification method, widely-used machine learning algorithms in ERP BCIs, like the xDAWN [38] and the SIM [39], could be combined in this speller for a wide application in larger population. Thirdly, due to the highly localized cortical visual motion responses, this dual-directional motion design is expected to construct a practical minimally invasive BCI speller, as reported in our previous work [9], [12]. The relatively high signal quality and the high-frequency features of the intracranial responses may help uncover more abundant information to boost the typing speed.

V. Conclusion

In the present study, we proposed a novel dual-directional N200 speller. By using the motion direction as an encoding dimension, and together with a probability-based dynamic stopping algorithm, an average ITR of 79.8 ± 24.7 bits/min was achieved, making the proposed paradigm one of the fastest ERP BCI spellers. Considering its non-flashing and low visual fatigue characteristics, this dual-directional N200 speller is expected to be a competent candidate for wide BCI applications, especially in the scenario of long-time human-computer interaction requiring mild visual stimulus.

Appendix

A. Probability-based Dynamic Stopping Strategy

Each trial consists of three row-epochs and three column-epochs, and we call that group of epochs a scan. Suppose there are N trials in total. Denote the EEG data collected from the $m^{th}, m \in \{1,2,3\}$ epoch in scan $s \in \{row, col\}$, sorted in ascending order by row/column indices, e.g., the epoch covers the 2nd and the 5th row is the 2nd epoch, of trial $i, i \leq N$ as $X^{(i)}_{m,s}$. Denote weights and biases of classifiers for non-target, leftward and rightward classes as $\{\boldsymbol{w}_n, b_n\}, \{\boldsymbol{w}_l, b_l\}, \{\boldsymbol{w}_r, b_r\}$, respectively. At the i^{th} $(i \in [1, N])$ trial, denote decision values in epoch m given by classifiers for d class, $d \in \{l, r\}$, in scan s as $y^{(i)}_{(m,d),s}$:

$$y^{(i)}_{(m,d),s} = \boldsymbol{w}_d X^{(i)}_{m,s} + b_d \tag{2}$$

Denote the mean of decision values across the N trials as:

$$\bar{y}^{(N)}_{(m,d),s} = \frac{1}{N} \sum_{i=1}^{N} y^{(i)}_{(m,d),s} \tag{3}$$

Denote the target at the $a^{th}, a = 1, \ldots, 6$ row and the $b^{th}, b = 1, \ldots, 6$ column as (a, b). Each of the targets has a corresponding row and column decision value, $y_{(m_1,d_1),row}$ and $y_{(m_2,d_2),col}$, respectively.

The row stimuli go leftward in the I and II quadrants and rightward in the III and IV quadrants. The column stimuli go leftward in the I and III quadrants and rightward in the II and IV quadrants. Denote the rule that maps the row-column

index $\begin{bmatrix} a \\ b \end{bmatrix}$ to the epoch index $\begin{bmatrix} (m_1, d_1) \\ (m_2, d_2) \end{bmatrix}$ as $\Phi(\begin{bmatrix} a \\ b \end{bmatrix})$. This rule can be described with Table II. For instance, the target "Z" is located in the 5th row and the 2nd column in the 4th quadrant (Fig.1c). The 2nd row-epoch will elicit a rightward response, and the 2nd column-epoch will elicit a leftward response.

TABLE II
STIMULATION RULE OF THE DUAL-DIRECTIONAL SPELLER

Row \ Col	1-3	4-6
1-3	$\begin{bmatrix} (a,l) \\ (b,l) \end{bmatrix}$	$\begin{bmatrix} (a,l) \\ (b-3,r) \end{bmatrix}$
4-6	$\begin{bmatrix} (a-3,r) \\ (b,l) \end{bmatrix}$	$\begin{bmatrix} (a-3,r) \\ (b-3,r) \end{bmatrix}$

With softmax function (see Eq.4), we can map vector $[\bar{y}^{(N)}_{(m,d)}] \in \mathcal{R}^6, m = 1, \ldots, 3, d \in \{l, r\}$ with six decision values in one trial into a probability distribution with six probabilities, each indicating the likelihood that whether the target is in this row or column.

$$P^{(N)}_{(m,d),s} = \frac{\exp \bar{y}^{(N)}_{(m,d),s}}{\sum_{i=1}^{3} \sum_{j \in \{l,r\}} \exp \bar{y}^{(N)}_{(i,j),s}} \tag{4}$$

The likelihood, denoting as $p_{(a,b)}$, that the target (a, b) is the true attended one, is the product of that the target is in the a^{th} row and b^{th} column:

$$P_{(a,b)} = P_{\Phi_{row}(a),row} \cdot P_{\Phi_{col}(b),col} \tag{5}$$

Theoretically, we could set the number by defining a lower bound of the online accuracy (Appendix B, Fig.8b). The default threshold of the likelihood was set as 0.95, to ensure a high accuracy of classification.

B. Error Rate Analysis for Online Decision

The thresholds mentioned in Appendix A (the maximum number of averaging and the likelihood threshold) need to be set before the online sessions. Hence, akin to the derivation in [40], we derived the relationship between offline training results and online decision error rates to estimate these parameters.

Consider the uni-directional paradigm first. Denote ERP signal as $\boldsymbol{X}$. We can assume that ERPs from different classes are Gaussian, $\boldsymbol{X}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), i \in \{t, n\}$. The non-target responses were considered to obey a mixture Gaussian distribution contingent on the distance from the gazing target [10]. However, the differences between the non-target responses are trivial. For simplicity, we hypothesize that the non-target and the target responses are independent and identical distributed (iid), respectively.

A linear classifier has weight $\boldsymbol{w}$ and bias b, so the decision value y is:

$$y = \boldsymbol{w}\boldsymbol{X} + b \tag{6}$$

So for target and non-target responses:

$$y_i \sim N(\mu_i, \boldsymbol{w}^T \boldsymbol{\Sigma_i} \boldsymbol{w}), i \in \{t, n\} \tag{7}$$

$\mu_i \in \mathcal{R}$, and without loss of generality, let $\mu_t > 0$ and $\mu_n < 0$.

After N times of averaging, the average decision value $\bar{y}$ is Gaussian, with the standard deviation reduced by $\sqrt{N}$ times.

$$\bar{y}^{(N)}_i \sim N(\mu_i, \boldsymbol{w}^T \boldsymbol{\Sigma_i} \boldsymbol{w}/\sqrt{N}), i \in \{t, n\} \tag{8}$$

The sufficient condition (SC) of a correct online decision for the uni-directional paradigm is that the decision values of all non-target classes are less than that of the target class.

$$\forall \bar{y}^{(N)}_n < \bar{y}^{(N)}_t \tag{9}$$

Hence, the correct rate of row/column scan for online decision is:

$$P(c) = \int_{-\infty}^{\infty} p_1(\bar{y}^{(N)}_t)(\int_{-\infty}^{\bar{y}^{(N)}_t} p_2(\bar{y}^{(N)}_n)d\bar{y}^{(N)}_n)^5 d\bar{y}^{(N)}_t \tag{10}$$

A correct online decision requires both the decision for row and column scans to be correct.

$$P(e) = 1 - P(c)^2 \tag{11}$$

Substitute eq.10 into eq.11 and we are able to get eq.12.

Thus, the estimated online error rate is:

$$P(e) = 1 - (\int_{-\infty}^{\infty} p_1(\theta)(\int_{-\infty}^{\theta} p_2(\gamma)d\gamma)^5 d\theta)^2 \tag{12}$$

The θ, γ represent the decision values for target and non-target responses, respectively. $\theta \sim N(\mu_t, \sigma_t/\sqrt{N})$, $\gamma \sim N(\mu_n, \sigma_n/\sqrt{N})$. N denotes the number of averaging trials. The parameters of the distributions $(\mu_t, \sigma_t, \mu_n, \sigma_n)$ can be estimated with maximum likelihood (MLE) algorithm.

Similarly, for the dual-directional paradigm, Eq.9 is also the SC. We assume that each class of N200 obeys the Gaussian distribution. Also, we assume that there must be a target in each trial. Therefore, we can get six different variables for each kind of decision, as shown in Table III.

TABLE III
VARIABLE LIST FOR EACH KIND OF DECISION

Decision \ Class	Non-target	Leftward	Rightward
Leftward	$\gamma_{n,l}$	$\theta_{l,l}$	$\gamma_{r,l}$
Rightward	$\gamma_{n,r}$	$\gamma_{l,r}$	$\theta_{r,r}$

The accuracies for targets at the leftward side or the rightward side are slightly different (see Eq.14-15). The probability that the target at either the leftward side (R/C 1-3) or the rightward side (R/C 4-6) is equal, so the correct rate of a scan is the average of that of the row and the column scans (see Eq.13).

Further, the error rate for the dual-directional paradigm is:

$$P(e) = 1 - P(c)^2 = 1 - (\frac{P_l(c) + P_r(c)}{2})^2 \tag{13}$$

$$\begin{aligned} P_l(c) = &\int_{-\infty}^{\infty} p(\theta_{l,l})(\int_{-\infty}^{\theta_{l,l}} p(\gamma_{l,r})d\gamma_{l,r}) \\ &(\int_{-\infty}^{\theta_{l,l}} p(\gamma_{n,l})d\gamma_{n,l})^2 (\int_{-\infty}^{\theta_{l,l}} p(\gamma_{n,r})d\gamma_{n,r})^2 d\theta_{l,l} \end{aligned} \tag{14}$$

$$P_r(c) = \int_{-\infty}^{\infty} p(\theta_{r,r})(\int_{-\infty}^{\theta_{r,r}} p(\gamma_{r,l})d\gamma_{r,l}) \\ (\int_{-\infty}^{\theta_{r,r}} p(\gamma_{n,l})d\gamma_{n,l})^2(\int_{-\infty}^{\theta_{r,r}} p(\gamma_{n,r})d\gamma_{n,r})^2 d\theta_{r,r} \tag{15}$$

The $\theta_{\cdot,\cdot}$ and $\gamma_{\cdot,\cdot}$ denote the values of the correct and wrong decisions for the targets, respectively. The first subscript indicates the real category to which it belongs, whereas the second indicates the decision made for this target. Each of these variables obeys a Gaussian distribution, whose parameters can be estimated with the MLE algorithm.

With the distributions of decision values in different conditions estimated using the offline data, the relationship between the number of averaging trials and the online error rate could be obtained. Thus, we can pick a suitable value as the threshold for each subject for the system to make a solid decision.

REFERENCES

[1] F. Guo *et al.*, "A brain–computer interface using motion-onset visual evoked potential," *Journal of Neural Engineering*, vol. 5, no. 4, p. 477, 2008.

[2] B. Hong *et al.*, "N200-speller using motion-onset visual response," *Clinical Neurophysiology*, vol. 120, no. 9, pp. 1658–1666, 2009.

[3] T. Liu *et al.*, "An online brain–computer interface using non-flashing visual evoked potentials," *Journal of Neural Engineering*, vol. 7, no. 3, p. 036003, 2010.

[4] J. Jin *et al.*, "A combined brain–computer interface based on P300 potentials and motion-onset visual evoked potentials," *Journal of Neuroscience Methods*, vol. 205, no. 2, pp. 265–276, 2012.

[5] R. Zhang *et al.*, "An adaptive motion-onset VEP-based brain-computer interface," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 4, pp. 349–356, 2015.

[6] S. Schaeff *et al.*, "Exploring motion VEPs for gaze-independent communication," *Journal of Neural Engineering*, vol. 9, no. 4, p. 045006, 2012.

[7] T. Ma *et al.*, "The hybrid BCI system for movement control by combining motor imagery and moving onset visual evoked potential," *Journal of Neural Engineering*, vol. 14, no. 2, p. 026015, 2017.

[8] R. Tootell *et al.*, "Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging," *Journal of Neuroscience*, vol. 15, no. 4, pp. 3215–3230, 1995.

[9] D. Zhang *et al.*, "Toward a minimally invasive brain–computer interface using a single subdural channel: a visual speller study," *Neuroimage*, vol. 71, pp. 30–41, 2013.

[10] ——, "An N200 speller integrating the spatial profile for the detection of the non-control state," *Journal of Neural Engineering*, vol. 9, no. 2, p. 026016, 2012.

[11] M. Schreuder *et al.*, "Optimizing event-related potential based brain–computer interfaces: a systematic evaluation of dynamic stopping methods," *Journal of Neural Engineering*, vol. 10, no. 3, p. 036025, 2013.

[12] D. Li *et al.*, "Minimally invasive brain computer interface for fast typing," in *Neural Engineering (NER), 2017 8th International IEEE/EMBS Conference on*. IEEE, 2017, pp. 477–480.

[13] T. Kaufmann and A. Kübler, "Beyond maximum speed—a novel two-stimulus paradigm for brain–computer interfaces based on event-related potentials (P300-BCI)," *Journal of Neural Engineering*, vol. 11, no. 5, p. 056004, 2014.

[14] J. Chen *et al.*, "A single-stimulus, multitarget BCI based on retinotopic mapping of motion-onset VEPs," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 2, pp. 464–470, 2018.

[15] ——, "Towards a fully spatially coded brain-computer interface: simultaneous decoding of visual eccentricity and direction," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 3091–3094.

[16] D. Liu *et al.*, "Bi-directional visual motion based BCI speller," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2019, pp. 589–592.

[17] Y. Kamitani and F. Tong, "Decoding seen and attended motion directions from activity in the human visual cortex," *Current Biology*, vol. 16, no. 11, pp. 1096–1102, 2006.

[18] M. B. Hoffmann *et al.*, "Directional tuning of human motion adaptation as reflected by the motion VEP," *Vision Research*, vol. 41, no. 17, pp. 2187–2194, 2001.

[19] 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.

[20] Y. Liu *et al.*, "Comparison of stimulus types in visual P300 speller of brain-computer interfaces," in *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*. IEEE, 2010, pp. 273–279.

[21] J.-h. Shi *et al.*, "A submatrix-based P300 brain-computer interface stimulus presentation paradigm," *Journal of Zhejiang University SCIENCE C*, vol. 13, no. 6, pp. 452–459, 2012.

[22] T. Kaufmann *et al.*, "Flashing characters with famous faces improves ERP-based brain–computer interface performance," *Journal of Neural Engineering*, vol. 8, no. 5, p. 056016, 2011.

[23] J. W. Peirce, "Psychopy—psychophysics software in python," *Journal of Neuroscience Methods*, vol. 162, no. 1-2, pp. 8–13, 2007.

[24] R. W. Homan *et al.*, "Cerebral location of international 10–20 system electrode placement," *Electroencephalography and Clinical Neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987.

[25] T.-P. Jung *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[26] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[27] J. R. Wolpaw *et al.*, "Brain-computer interface technology: a review of the first international meeting," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164–173, 2000.

[28] T. E. Oliphant, *Guide to NumPy*, 2nd ed. USA: CreateSpace Independent Publishing Platform, 2015.

[29] P. Virtanen *et al.*, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[30] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[31] A. Gramfort *et al.*, "MNE software for processing MEG and EEG data," *Neuroimage*, vol. 86, pp. 446–460, 2014.

[32] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[33] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG- and MEG-data," *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, 2007.

[34] C. W. Eliot, *The Harvard Classics*. PF Collier & son, 1910, vol. 50.

[35] J. E. Raymond, "Directional anisotropy of motion sensitivity across the visual field," *Vision Research*, vol. 34, no. 8, pp. 1029–1037, 1994.

[36] G. Townsend and V. Platsko, "Pushing the P300-based brain–computer interface beyond 100 bpm: extending performance guided constraints into the temporal domain," *Journal of Neural Engineering*, vol. 13, no. 2, p. 026024, 2016.

[37] J. Jin *et al.*, "The changing face of P300 BCIs: a comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement," *PloS One*, vol. 7, no. 11, p. e49688, 2012.

[38] B. Rivet *et al.*, "xDAWN algorithm to enhance evoked potentials: application to brain–computer interface," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, 2009.

[39] C. Wu *et al.*, "A novel algorithm for learning sparse spatio-spectral patterns for event-related potentials," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 4, pp. 862–872, 2016.

[40] Z. Zhou *et al.*, "A novel task-oriented optimal design for P300-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 11, no. 5, p. 056003, 2014.