

PAPER

Task-relevant stimulus design improves P300-based brain–computer interfaces

To cite this article: Jongsu Kim *et al* 2024 *J. Neural Eng.* **21** 066046

View the [article online](#) for updates and enhancements.

You may also like

- [A novel task-oriented optimal design for P300-based brain–computer interfaces](#)
Zongtan Zhou, Erwei Yin, Yang Liu *et al.*
- [Optimizing the P300-based brain–computer interface: current status, limitations and future directions](#)
J N Mak, Y Arbel, J W Minett *et al.*
- [An adaptive P300-based control system](#)
Jing Jin, Brendan Z Allison, Eric W Sellers *et al.*



PAPER

Task-relevant stimulus design improves P300-based brain–computer interfaces

Jongsu Kim¹ , Yang Seok Cho^{2,*} and Sung-Phil Kim^{1,*}¹ Department of Biomedical Engineering, UNIST, Ulsan, Republic of Korea² School of Psychology, Korea University, Seoul, Republic of Korea

* Authors to whom any correspondence should be addressed.

E-mail: yscho_psych@korea.ac.kr, kjstn737@unist.ac.kr and spkim@unist.ac.kr**Keywords:** P300-based BCI, stimulus design, task relevance, selective attention, single-trial BCI, calibration-free, user variabilitySupplementary material for this article is available [online](#)RECEIVED
30 April 2024REVISED
20 November 2024ACCEPTED FOR PUBLICATION
18 December 2024PUBLISHED
30 December 2024**Abstract**

Objective. In the pursuit of refining P300-based brain–computer interfaces (BCIs), our research aims to propose a novel stimulus design focused on selective attention and task relevance to address the challenges of P300-based BCIs, including the necessity of repetitive stimulus presentations, accuracy improvement, user variability, and calibration demands. *Approach.* In the oddball task for P300-based BCIs, we develop a stimulus design involving task-relevant dynamic stimuli implemented as finger-tapping to enhance the elicitation and consistency of event-related potentials (ERPs). We further improve the performance of P300-based BCIs by optimizing ERP feature extraction and classification in offline analyses. *Main results.* With the proposed stimulus design, online P300-based BCIs in 37 healthy participants achieve an accuracy of 91.2% and an information transfer rate (ITR) of 28.37 bits/min with two stimulus repetitions. With optimized computational modeling in BCIs, our offline analyses reveal the possibility of single-trial execution, showcasing an accuracy of 91.7% and an ITR of 59.92 bits/min. Furthermore, our exploration into the feasibility of across-subject zero-calibration BCIs through offline analyses, where a BCI built on a dataset of 36 participants is directly applied to a left-out participant with no calibration, yields an accuracy of 94.23% and the ITR of 31.56 bits/min with two stimulus repetitions and the accuracy of 87.75% and the ITR of 52.61 bits/min with single-trial execution. When using the finger-tapping stimulus, the variability in performance among participants is the lowest, and a greater increase in performance is observed especially for those showing lower performance using the conventional color-changing stimulus. *Significance.* Using a novel task-relevant dynamic stimulus design, this study achieves one of the highest levels of P300-based BCI performance to date. This underscores the importance of coupling stimulus paradigms with computational methods for improving P300-based BCIs.

1. Introduction

Brain–computer interfaces (BCIs) are rapidly evolving at the intersection of neuroscience and engineering, offering revolutionary communication and control channels independent of peripheral neural and muscular activity [1]. A P300-based BCI is one of the non-invasive BCIs that leverages the P300 component in event-related potentials (ERPs) of electroencephalography (EEG), which is elicited by

recognizing a target stimulus in the oddball paradigm [2]. P300-based BCIs have been widely adopted in various applications from assistive technologies to gaming and virtual reality (VR) [3, 4]. However, P300-based BCIs still face challenges, including repeated stimulus presentations, accuracy enhancement, individual variation, and frequent calibration requirements, which need to be addressed through innovative approaches for the enhancement of usability and applicability.

1.1. Challenges in P300-based BCIs

First, P300-based BCIs generally require repeated stimulus presentations for accurate ERP detection, limiting real-world application efficiency [5–9]. Despite advancements in single-trial P300-based BCIs, their performance falls short compared to multi-trial approaches [10, 11]. Second, P300-based BCIs necessitate personalized feature extraction and decoding algorithms as well as separate calibration sessions for each use, thus hindering daily practicality [12]. Transfer learning strategies have been developed to minimize the extensive calibration requirements; yet, attaining the robustness necessary for consistent applications across diverse datasets has proven to be challenging [13–16]. Third, the P300 component depends on cognitive states such as attention and working memory. Thus, strategies such as increasing stimulus saliency are necessary to enhance the quality of ERPs [4, 17–19]. These challenges emphasize the need for balanced P300-based BCI designs that can effectively integrate technical as well as user-centered aspects.

Addressing these three challenges could significantly enhance the usability of P300-based BCIs. In this study, we focus on the common characteristic underlying these challenges: the need for consistent and effective ERP elicitation. The requirement for multiple stimulus presentations arises because single-trial ERPs are often not sufficiently distinguishable. Similarly, the difficulty of applying transfer learning is largely due to variability in ERP patterns across users. Furthermore, the dependency on user attention is directly linked to ERP elicitations. There are users whose ERPs are consistently difficult to elicit, often resulting in poor performance (a phenomenon referred to as ‘BCI illiteracy’) [19], while there are others whose ERPs are reliably elicited, achieving high performance across conditions. This study aims to address this variability by ensuring consistent ERP elicitation across all participants, thereby addressing the aforementioned challenges and improving overall BCI performance.

1.2. Paradigm considerations in P300-based BCI design

P300-based BCIs utilize ERPs elicited by the oddball paradigm, where the key components are the unequal presentation frequency of stimuli, target probability, inter-stimulus interval (ISI), inter-target interval, and stimulus repetition. These factors influence P300 amplitude and waveform characteristics, shaping how effectively ERPs can be elicited [4, 8, 20, 21]. Central to these design elements is the selective attention of the participant toward the target stimulus, a critical aspect in enhancing cognitive processing during the oddball task [4].

Selective attention within the oddball paradigm can be further delineated into bottom-up attention, which is related to the characteristics of the stimulus itself, and top-down attention, which is associated with the user’s intention. Although these processes have distinct characteristics in visual cognition, studies have demonstrated that they interact during cognitive processing [22]. Therefore, our focus is on optimizing this interaction as it occurs within the oddball paradigm.

In the context of bottom-up attention, previous research has explored various stimulus designs to elicit stronger ERP responses (color-changing [3, 12], dots [23], famous faces [18], emotional faces [24, 25], moving stimuli [25–27], etc). Notably, studies on motion-onset visual evoked potentials (mVEPs) have shown that dynamic stimuli can reduce inter- and intra-subject variability in ERP responses [27], thereby stabilizing ERP elicitation under consistent conditions. Top-down attention has been studied in scenarios ranging from passive viewing of stimuli to active mental tasks [28], such as focusing attention on target stimuli and counting their occurrences. Among these approaches, the task of mentally counting target stimuli has been found to produce superior ERP responses compared to other methods [27, 28].

To further optimize the interaction between these two attention processes, we propose incorporating the concept of task relevance [29]. In previous studies, task relevance in P300-based BCIs has been defined within the framework of the oddball paradigm, focusing on the most appropriate stimulus presentation frequency and the intention to direct attention solely to the target stimulus, which is typically presented less frequently than non-targets [28, 30, 31].

However, we aim to expand this concept by considering the essential function of P300-based BCIs: enabling users to perform intentional selections. Since BCIs function as a user interface (UI), the act of selection in everyday contexts often involves physical actions, such as tapping a touchscreen or clicking with a mouse. Inspired by this observation, we designed a finger-tapping stimulus, where an animated finger appears to press a button, and paired it with a mental task that involves motor imagery (MI) of finger movements when the target stimulus is presented.

We hypothesize that this approach will lead to three key improvements: enhanced classification accuracy through more effective ERP elicitation, a reduction in the number of stimulus repetitions required for accurate classification due to improved ERP quality, and decreased variability in ERP responses due to the use of dynamic stimuli, which could, in turn, enhance the performance of transfer learning models. These anticipated outcomes address several of the current limitations associated with P300-based BCIs.

1.3. Computation considerations in P300-based BCI design

P300-based BCI challenges, as discussed in section 1.1, can be addressed through both computational methods and the development of paradigms. For instance, the number of stimulus presentations can be reduced by adaptively determining an optimal point to halt the stimulation process, thereby enhancing information transfer rates (ITRs) while maintaining accuracy [32]. Enlarging training datasets with transfer learning or generative artificial intelligence methods can also improve decoding models in P300-based BCIs [33].

While adopting advanced computational methods has the potential to provide a promising solution to the challenges in P300-based BCIs, improving decoding models alone may require extensive exploration of optimal algorithms and enormous efforts to validate their efficacy across diverse BCI applications. For instance, variations in the P300-based BCI performance across users partially due to ‘BCI illiteracy’ remain a problem that has not been completely resolved by decoding improvement [19]. Given that BCIs fundamentally depend on ERP quality, high-quality ERP elicitation through innovative task paradigms can complement computational advancements. In line with the principle of ‘garbage in, garbage out’, high-quality ERPs would lead to robust BCI performance. Therefore, we propose addressing P300 BCI challenges not solely through computational methods but by combining these with effective paradigm design.

1.4. Optimization of P300-based BCI design

In this study, we aimed to optimize P300-based BCI design through a two-stage approach that combined paradigm development with computational refinements. Our first goal was to improve the traditional paradigm, which typically uses color-changing stimuli and a counting task, by incorporating task relevance into both stimuli and tasks. The second goal was to identify the most effective computational method to decode ERPs generated by the optimized paradigm. To achieve these objectives, we first validated the paradigm’s effectiveness through an online experiment, using simple preprocessing and decoding techniques to focus on the paradigm’s impact. We then conducted offline analyses, applying various computational methods to further enhance the system’s accuracy, efficiency, and adaptability, addressing key challenges in P300-based BCIs.

1.4.1. Paradigm optimization

Our primary innovation involved introducing a finger-tapping stimulus and MI task to P300-based

BCIs, both designed to add task relevance and potentially enhance selective attention, aimed at improving the quality of ERPs. The finger-tapping animation was chosen as it not only incorporated task relevance but also induced mVEPs that are known to reduce variability in ERP responses. Similarly, the MI task was selected for its potential to engage participants more deeply in the task by requiring a mental simulation of physical actions.

To assess the effect of the proposed approach on P300-based BCIs, we conducted an online BCI experiment comparing six different paradigm designs: three types of visual stimuli (color-changing, icon rotation, and finger-tapping), each combined with two mental tasks (counting and MI). In designing these paradigms, we specifically aimed to manipulate both bottom-up and top-down aspects of attention. For bottom-up attention, we compared dynamic stimuli (the finger-tapping animation and icon rotation) with static stimuli (the traditional color-changing stimulus). For top-down attention, we evaluated how different tasks (MI vs. counting) influenced participants’ engagement with the task.

We examined task relevance by comparing the task-relevant finger-tapping animation stimulus with the task-irrelevant icon-rotating stimulus. The icon-rotating stimulus was specifically used as a control condition for task relevance as it would induce mVEPs similar to the finger-tapping stimulus but lack direct relevance to the selection process.

1.4.2. Computational methods optimization

In the online BCI experiment, we used a conventional decoding approach based on ERP features and a linear SVM classifier to examine the effect of paradigm designs on BCI performance exclusively. From the result of the online experiment, we chose the paradigm design that produced the highest performance. Then, we conducted further offline analyses to explore the full potential of the chosen design. In these offline analyses, we applied various signal processing and classification techniques to identify the most effective computational method for decoding ERPs elicited by the chosen design. Specifically, we examined different feature extraction methods and a variety of classifiers to optimize the decoding performance in each offline analysis.

A series of offline analyses were conducted with different objectives: (1) to maximize decoding performance by identifying the computational method that provides the highest accuracy for the P300-based BCI system tested in the online experiment, where each stimulus was presented twice; (2) to optimize single-trial BCI performance by determining the best computational method that could provide high accuracy even with the single presentation of stimuli; (3) to explore the potential for zero-calibration using

transfer learning, which involved assessing the performance of classifiers trained on other participants' data to decode the ERPs of a new participant, whether using double or single stimulus presentations; and (4) to minimize individual variation by identifying the computational method that resulted in the least variability in performance across different individuals.

2. Methods

2.1. Participants

Based on the statistical power analysis using G*Power 3 [34], a sample size of 33 participants was determined for the experiment (effect size $f = 0.25$, power = 0.95, nonsphericity correction $\epsilon = 0.8$). To meet this minimum requirement, we recruited 37 healthy adults (21.62 ± 3.53 years old, range: 18–30 years old, 8 females). All participants were right-handed and reported no history of mental illness or neurological disorders. Informed consent was obtained from participants in compliance with the Ulsan National Institutes of Science and Technology, Institutional Review Board (UNIST-IRB-18-08-A).

2.2. Stimulus design

We crafted visual stimuli that were tailored by our previous work for the integration of BCIs into different UIs across monitors, augmented reality, and VR environments [12]. Positioned at the four quadrants of the display (top-left, bottom-left, top-right, bottom-right), these stimuli featured icons indicative of on/off, play, stop, and pause functions, set against a blue rectangular background, mirroring the potential control of an external device (i.e. Bluetooth speaker in this study) [12].

We created three types of stimuli. First, a static stimulus was crafted that changed the color of the rectangular background from blue to green (figure 1(a)). This control stimulus was supposed to be the least salient in bottom-up attention among other stimuli used in this study and task-irrelevant. Second, a dynamic stimulus was built by incorporating animations on top of the color-changing stimulation (figure 1(b)). The stimulus was animated by rotating the icons (i.e. rotating the icons by 90° clockwise), similar to other studies [35]. This second stimulus was supposed to be more salient than the first one but still task-irrelevant. Note that we added animation to color-changing effects when designing the second stimulus since we intended to ensure that its bottom-up saliency was greater than the first one. Third, we created the finger-tapping stimulus, designed to be task-relevant and dynamic, by animating the right index finger to press the icons and return to an initial position (figure 1(c)). Again, the finger-tapping animation was added to the color-changing stimulation to increase bottom-up saliency. The third

stimulus was distinct from the second one by making the animation task-relevant.

Each stimulus was presented for 400 ms. This stimulus duration (SD) was determined to render the appearance of finger-tapping as natural as possible within the shortest time frame. Given the importance of the total stimulus presentation duration for practical use in P300-based BCIs, we fixed the ISI to 0 ms.

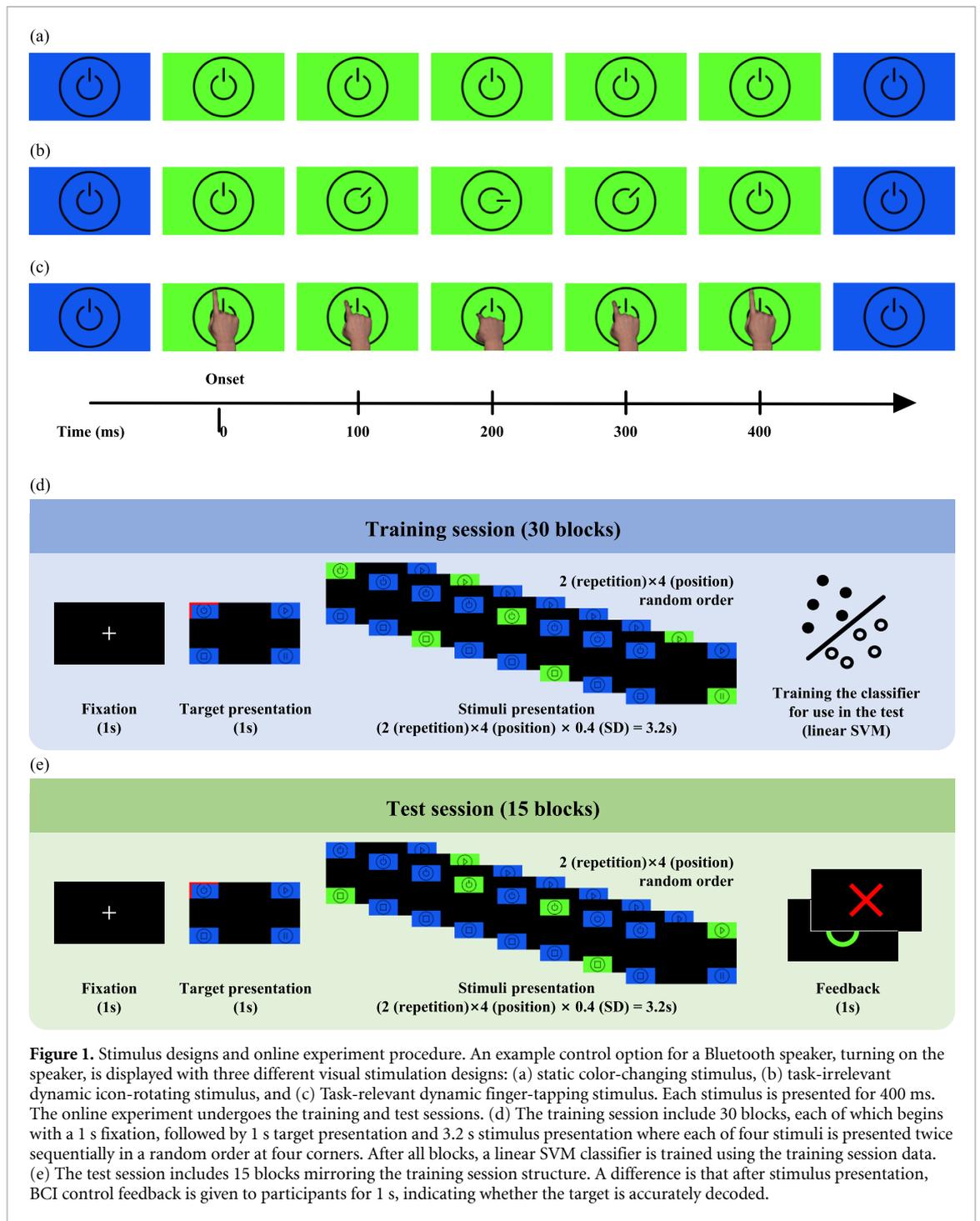
2.3. Task design

In the experimental setup, participants performed two specific mental tasks when they saw a target stimulus: counting and MI. In the counting task, participants silently counted the number of times the target stimulus appeared. This task was designed to help participants focus more intently on the target, making it easier to elicit a P300 response. In the MI task, participants were instructed to imagine pressing the target button with their right index finger. Similar to the counting task, the purpose of MI was to enhance focus on the target and facilitate P300 elicitation. However, in addition to this, the MI task aligns with the intuitive purpose of selection in a BCI context. By imagining the action of pressing the target, MI introduces an element of task relevance, making it a more intuitive and goal-oriented task that corresponds to the actual function of selection. To account for differences in individual performance of MI, a guideline was provided using visual examples of finger-tapping that matched the stimulus design. Participants were then asked to mentally replicate these finger movements.

2.4. Online experimental procedure

Throughout the online experiment of P300-based BCIs, participants were comfortably seated in front of a 27 inch monitor. Before the start of the experiment, participants read a manual detailing the experimental procedure, received thorough explanations, and had a question-and-answer session. To ensure the successful performance of MI during the experiment, sufficient training was provided with the guidelines described above until participants felt confident to proceed with their MI. They were instructed to self-control the progression of the experiment by pressing the space bar (see below for details). An electromyography (EMG) electrode was placed over each of the metacarpophalangeal and distal interphalangeal joints of the participant's right index finger to monitor potential muscle activity during the MI. Before the experiment began, we confirmed that deviant EMG was observed when participants flexed or extended their right index finger.

A block of the P300-based BCI operation underwent four phases as follows (figures 1(d) and (e)): (1) fixation: participants gazed at a central cross for 1000 ms; (2) target indication: the target location



among four corners was indicated by displaying a red border around the corresponding icon for 1000 ms; (3) stimulus presentation: each of the four corners of the monitor screen displayed a blue box, with each box occupying a visual angle of $10.4^\circ \times 5.7^\circ$ on the screen. Inside each box, icons representing individual functions of controlling a Bluetooth speaker were shown. During the stimulus presentation period, one of the boxes was highlighted by turning its color to green and, optionally, showing additional animations such as icon rotation or finger tapping. This stimulus presentation lasted for 0.4 s per box. The presentation

of the four stimuli occurred sequentially in a random order, and this sequence was repeated twice, resulting in a total of eight stimulus presentations that summed up to 3.2 s.; and (4) feedback presentation (only for test blocks): participants received immediate feedback on whether the BCI system correctly identified the target stimulus, where the BCI output was displayed on the screen for 1000 ms. Participants were instructed to pay attention to the presentation of a target stimulus while performing a given mental task. They were given 1 target and 3 non-target stimuli twice for a total of 3200 ms (8×400 ms). During

training, the feedback presentation phase was omitted. Each stimulus presentation will be referred to as a trial hereafter.

The experiment consisted of six sessions based on the combination of the three types of stimuli (section 2.2) and the two mental tasks (section 2.3): color change with counting, color change with MI, icon rotation with counting, icon rotation with MI, finger tapping with counting, and finger tapping with MI. The order of these six sessions was randomized for each participant to control for order effects. Each session included 45 blocks, with 30 blocks for training and 15 blocks for testing.

Before the main experiment, participants underwent a two-stage practice process. First, they completed MI training. In this stage, participants fixed their right wrist on the table and repeatedly performed dorsiflexion and plantar flexion of their right index finger, replicating the finger-tapping motion depicted in the stimulus. They gradually reduced the range of movement until they were instructed to imagine the movement without physically moving their finger. Once participants reported that they could vividly imagine their index finger moving, they proceeded to the next stage. In the second stage, participants engaged in six practice sessions, each consisting of five blocks, to become acquainted with the tasks and stimuli. The practice sessions were designed in the following sequence: (1) color changing with counting, to introduce participants to a basic P300-based BCI task; (2) finger-tapping with counting, and (3) finger-tapping with MI, to help participants learn to synchronize their MI with the finger-tapping stimulus; (4) color changing with MI, to continue practicing MI with a different stimulus, reinforcing what was learned in the previous sessions (1); and finally, (5) icon-rotating with counting and (6) icon-rotating with MI, to expose participants to novel stimuli and practice MI, thereby increasing their adaptability to various BCI tasks. In this stage, participants had to execute MI in synchrony with the presented stimulus. For MI proficiency, they were interrogated on their ability to execute MI upon presentation of the stimuli; any snag in execution by any participant extended practice beyond the prescribed amount of training. Training continued until each subject demonstrated sufficient competence in performing the MI task as instructed. With this two-stage practice, we intended participants to be prepared for the MI task in the main experiment.

2.5. Data acquisition and preprocessing

EEG data were recorded using 31 active wet electrodes (FP1, FPz, FP2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, and O2) following the international 10–20 system (American Clinical

Neurophysiology Society Guideline 2). Signals were transmitted to an EEG amplifier (actiCHamp, Brain Product GmbH, Gilching, Germany) at a 500 Hz sampling rate. The left and right mastoids were used as reference and ground, respectively. Electrode impedance was kept below 10 k Ω throughout the experiment.

EEG preprocessing underwent in the following order: (1) 1 Hz high-pass filtering, (2) 50 Hz low-pass filtering, (3) bad-channel removal and interpolation, (4) re-referencing using common average reference, (5) artifact removal using artifact subspace reconstruction (ASR) (cutoff 10), and (6) 12 Hz low-pass filtering. Filters were designed as finite impulse response filters using a Hamming window to attenuate unwanted frequencies. Bad channels were detected by using a modified version of the 'clean_channels' function from the EEGLAB toolbox. A channel's correlation with its neighbors was evaluated over 5 s, with those showing a correlation coefficient below 0.8 deemed suspect. Channels were classified as 'bad' if their abnormal state persisted over 40% of the recordings, ensuring the exclusion of consistently noisy data from subsequent analyses.

EMG signals were recorded with 2 passive wet electrodes at the metacarpophalangeal and distal interphalangeal joints of the participant's right index finger. EMG signals were transmitted to an EEG amplifier through a BIP2AUX adapter (Brain Products GmbH, Gilching, Germany), allowing for integration and recording at the same sampling rate as the EEG data. EMG preprocessing was conducted as follows: (1) 1 Hz high-pass filtering to remove low-frequency drifts, (2) 150 Hz low-pass filtering to eliminate high-frequency noise, and (3) 60 Hz notch filtering to attenuate power line interference. Subsequently, the signals were rectified and smoothed using a 50 ms moving average.

In the online experiment, preprocessing was conducted after the completion of all 30 blocks of the training part. For the test part, which consisted of 15 blocks, preprocessing was performed immediately after each block. Due to the real-time operation of BCIs during testing, re-applying bad-channel removal, interpolation, and ASR could be problematic in the preprocessing of the test block data. Thus, we adopted the bad-channel information and ASR parameters obtained from the preprocessing of the training data, instead of recalculating them for each test block.

2.6. Online BCI decoding

EEG signals at each channel were epoched from –200 ms to 600 ms relative to the stimulus onset. The epoched EEG signals were baseline-corrected by subtracting the mean amplitude during the pre-stimulus

period. Averaging these signals over the two trials resulted in an ERP waveform for each stimulus.

ERP features were extracted simply by concatenating the ERP waveforms within a post-stimulus period (0–600 ms) from all channels into a 1D vector that contained 9300 features (31 channels \times 300 time points) for each stimulus. A linear support vector machine (SVM) classifier was trained using 120 feature vectors from the 30 training blocks; 30 target stimuli labeled as +1 and 90 non-target stimuli labeled as –1. Then, in each testing block, we created feature vectors in the same way as in training, yielding four vectors corresponding to each stimulus. The trained linear SVM classifier produced the classification score for the j th stimulus as follows:

$$f_j = w \cdot x_j + b, j = 1, 2, 3, 4 \quad (1)$$

where w is the weight vector consisting of support vectors in linear SVM, x_j is the feature vector for the j th stimulus, and b is a bias term. The stimulus with the highest score (f_j) was predicted as a target. In the online experiment, a linear SVM model was employed to decode the ERP signals. This model was selected for its simplicity and efficiency, allowing us to observe performance differences arising from the stimulus design alone, without the help of advanced decoding techniques.

2.7. BCI performance evaluation

The performance of the BCI system was quantified using three metrics: accuracy, ITR, and coefficient of variation (CV).

Accuracy was defined as the ratio of the number of successful test blocks to the total number of test blocks. A test block was successful if the BCI system correctly identified a test stimulus.

ITR was calculated in bits per minute (bits/min) using the following formula:

$$\text{ITR} = \frac{60 \left(P \log_2 P + (1 - P) \log_2 \frac{P-1}{N-1} + \log_2 N \right)}{T} \quad (2)$$

where P is accuracy, N is the number of selectable options ($N = 4$ here), and T is the time taken for one selection in seconds ($T = (\text{SD} + \text{ISI}) \times N \times \text{repetition of stimuli}$).

CV was used to assess the consistency of accuracy and ITR among participants, calculated as the standard deviation divided by the mean of accuracy (or ITR) across participants.

2.8. Offline optimization of computational models

In the post-hoc offline analysis of the online BCI experimental data, we optimized the design of computational models to further improve P300-based BCIs. The optimization was conducted to tackle the aforementioned three challenges in P300-based BCIs:

the need for multiple stimulus repetitions, individual calibration requirements, and individual variations in performance. The optimization of computational models was undertaken in two parts: feature extraction and classification.

Among many models to extract ERP features, we opted for spatial filtering and manifold transformation. First, we chose to use xDAWN for spatial filtering as it is known to extract ERP features appropriately in a supervised manner, particularly the P300 component [36]. The ability of xDAWN to enhance ERP features was expected to mitigate the limited quality of ERPs associated with reduced stimulus repetitions. Second, we chose to use Riemannian geometry (RG) to transform ERP features onto manifolds as it is adept at representing complex data structures and well-suited for addressing EEG signal complexity. Especially, using xDAWN and RG together has proven to enhance P300-based BCI performance [14, 37]. For feature extraction, we first generated ERPs by utilizing the post-stimulus period of the baseline-corrected signals, as described in section 2.6. We then applied xDAWN spatial filtering to these ERPs and calculated the symmetric positive definite (SPD) covariance matrices of the filtered signals. These SPD matrices were represented on the Riemannian manifold and projected onto the tangent space to transform ERP features onto manifolds [37]. We utilized the pyRiemann Python package [38] for xDAWN filtering and RG computation.

As for classification models, we investigated both conventional machine learning models and deep learning models that have been used for BCIs. Linear SVM and logistic regression (LR) were adopted as conventional machine learning models. For deep learning, we used EEGNet [39], shallow ConvNet [40], and deep ConvNet [40], which have been particularly shown to be suitable for classifying P300-based BCI data.

We optimized computational models for P300-based BCIs in a greedy manner by examining the best combination of feature extraction and classification models: three feature extraction models (none, xDAWN (XD), xDAWN, and RG (XDRG)) and five classification models (SVM, LR, EEGNet, shallow ConvNet, and deep ConvNet). We sought the best combination for individual problems as described below.

The offline analysis employed computational techniques, such as xDAWN and RG-based feature extraction. Unlike the online phase, which prioritized simplicity and speed, the offline analysis explored decoding methods optimized for high accuracy, as they provided insights into the developed paradigm's full potential.

2.8.1. Minimization of stimulus repetitions

Although we already reduced the number of stimulation presentations to 2 in our online experiment, we

investigated if we could further reduce it to 1 to realize a single-trial P300-based BCI (i.e. no repetition). To this end, we selected EEG data in the first round of stimulus presentation from the online experiment and explored a combination of feature extraction and classification models that produced the highest BCI performance when stimuli were presented once. For comparison, we also optimized feature extraction and classification models offline when stimuli were presented twice.

2.8.2. Across-subject zero-calibration

In our pursuit of a zero-calibration, plug-and-play P300-based BCI, we applied transfer learning to the online experiment data with the leave-one-subject-out cross-validation (LOSO CV) scheme. In this method, we trained a decoder using data from 36 participants and then tested it on the remaining participants' data to assess whether the BCI could be used without individual calibration. For each iteration of LOSO, the training set included all 45 blocks from each of the 36 participants (30 training blocks and 15 testing blocks). We used the testing blocks because they also begin with the presentation of the target stimulus, making them suitable for training the decoder. This resulted in a total of 1620 blocks used for training in each iteration. The trained decoder was then tested on the excluded participant's 15 test blocks. Furthermore, as in section 2.8.1, by analyzing first-trial data, we also explore the possibility of achieving zero-calibration at the single-trial level. This CV demonstrates the feasibility of zero-calibration that can realize a plug-and-play P300-based BCI.

2.8.3. Individual variation of BCI performance

We explored whether the proposed BCI design could reduce individual variations of BCI performance. To this end, we analyzed the CV values across six different BCI paradigms. Furthermore, we assessed whether the proposed design could improve the performance of participants who showed relatively lower performance using conventional BCI designs—i.e. the color-changing stimulus with counting in our case. We first evaluated changes in BCI performance from using the color-changing stimulus with counting to using the finger-tapping stimulus with counting. Then, we calculated Pearson's correlation coefficient between the performance using the color-changing design and the performance change induced by the finger-tapping design.

2.9. Statistical analysis of BCI performance

A two-way repeated-measures ANOVA (rmANOVA) was employed to assess the BCI performance across different paradigms. The analysis focused on two

dependent variables: accuracy and ITR. The independent variables under examination were the stimulus type (with three levels: static as in color change, and dynamic as in icon-rotating and finger tapping) and the mental task (with two levels: counting and MI of finger tapping). To assess the statistical differences in performance across various combinations of sessions and analysis methods, a one-way repeated-measures ANOVA (one-way rmANOVA) was employed. This statistical approach was particularly utilized to analyze differences among distinct scenarios combining different sessions and computational strategies. When the assumption of sphericity was violated, the Greenhouse–Geisser correction was applied to ensure the validity of the rmANOVA results. Post-hoc analyses, using Tukey's honestly significant difference procedure (HSD), further dissected the effects of stimulus types, mental tasks, and their interaction on BCI performance. For the offline analysis, we utilized a paired *t*-test to compare the efficacy between classification methods.

2.10. ERP and SVM weight vector analysis

In addition to the standard ERP analysis, we conducted a detailed analysis of the N200 and P300 components in response to target and non-target stimuli across different stimulus presentation paradigms (color-changing, icon-rotating, and finger-tapping). We focused on the key EEG channels (Cz, Pz, and Oz) that are known to exhibit prominent ERP responses in the oddball task. The N200 component was analyzed within the 150–240 ms window and the P300 component within the 240–350 ms window. For each component, we extracted the peak amplitude and latency.

Furthermore, to understand the impact of specific EEG channels on classification accuracy, we analyzed the weight vectors from SVM models trained on individual participants. The absolute values of these weight vectors were used to infer the importance of different channels [41]. We selected the top 10% of weight vectors for each participant to evaluate which channels were most influential. Additionally, we performed a comprehensive analysis across all channels by calculating the absolute difference between target and non-target ERPs and summing these differences. This approach allowed us to assess the significance of the stimulus presentation method across all channels.

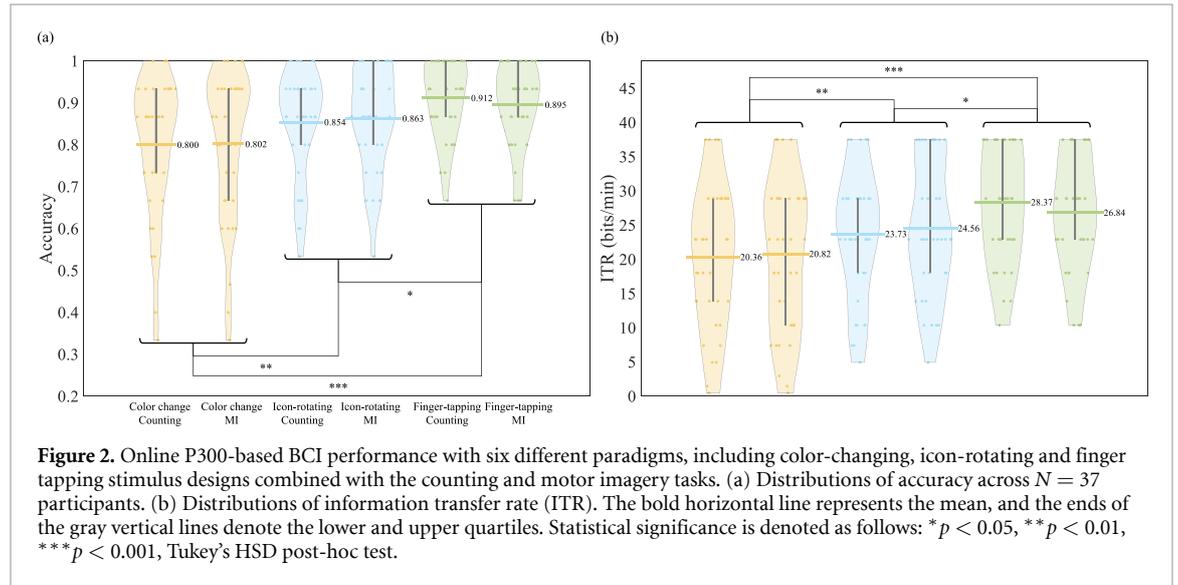
3. Results

3.1. Online BCI control

We assessed the performance of online P300-based BCIs with six different designs in terms of accuracy and ITR. We found the highest accuracy with the finger-tapping stimulus and counting, followed by the finger-tapping stimulus and MI, the icon-rotating

Table 1. Online P300-based BCI performance assessed by accuracy and ITR for three stimulus designs combined with two mental tasks. Mean and standard deviation across $N = 37$ participants. Bold fonts represent the highest performance. # stimulus repetitions = 2.

Stimulus designs	Mental task	Accuracy		ITR (bits/min)	
		Mean	Std.	Mean	Std.
Color-changing	Counting	0.8000	0.1663	20.3615	10.2798
	MI	0.8018	0.1760	20.8218	11.0490
Icon-rotating	Counting	0.8541	0.1285	23.7372	9.4733
	MI	0.8631	0.1227	24.5625	9.8569
Finger-tapping	Counting	0.9117	0.0917	28.3704	8.4638
	MI	0.8955	0.0988	26.8445	8.4952



and MI, the icon-rotating and counting, the color-changing and MI, and the color-changing and counting (see table 1 for details). The highest accuracy with the finger-tapping and counting was 91.17% on average whereas the lowest with the color-changing and counting was 80.00%, resulting in the improvement of 11.17% by adopting a new BCI paradigm (figure 2(a)). Similar trends were observed in ITR, where using the finger-tapping stimuli outperformed other paradigms (figure 2(b)).

A two-way rmANOVA on accuracy revealed the significant main effect of the stimulus type ($p = 3.4834 \times 10^{-9}$), while neither the main effect of the mental task nor the interaction effect was found. A post-hoc analysis showed significant differences between color-changing and finger-tapping ($p = 7.8790 \times 10^{-6}$), color-changing and icon-rotating ($p = 0.0037$), and finger-tapping and icon-rotating ($p = 0.0279$), affirming the superior performance by using the finger-tapping stimulus. Similarly, a two-way rmANOVA on ITR showed the significant main effect of the stimulus type ($p = 5.3242 \times 10^{-7}$). A post-hoc analysis also showed significant differences between the stimulus types: color-changing < icon-rotating < finger-tapping ($ps < 0.05$). We verified that the EMG amplitude

was not different between the MI and counting tasks for each of the three stimulus designs (paired t -test, $ps > 0.05$), which indicates that MI did not modulate the EMG signals in the experiment.

3.2. ERP component and SVM weight vector analysis

The ERP analysis revealed significant differences in the N200 and P300 components across different stimulus presentation paradigms. Although the N200 component was not observed at the Pz channel, it was visible in channels such as P3 and P4 (figure A1). While overall waveforms remained consistent (figure 3), the finger-tapping stimulus resulted in significantly shorter N200 latencies at Oz channels ($p = 6.699 \times 10^{-8}$) compared to other stimuli (figure 4). Post-hoc analysis confirmed that the finger-tapping stimulus had a shorter latency than both the color-changing ($p = 1.527 \times 10^{-5}$) and icon-rotating stimuli ($p = 1.4195 \times 10^{-5}$). N200 peak amplitudes showed no significant differences across stimulus types or tasks at Oz.

For the P300 component, the Cz channel exhibited significant differences in peak amplitude based on stimulus type ($p = 0.0096$) and task ($p = 0.0098$). The finger-tapping stimulus produced a smaller

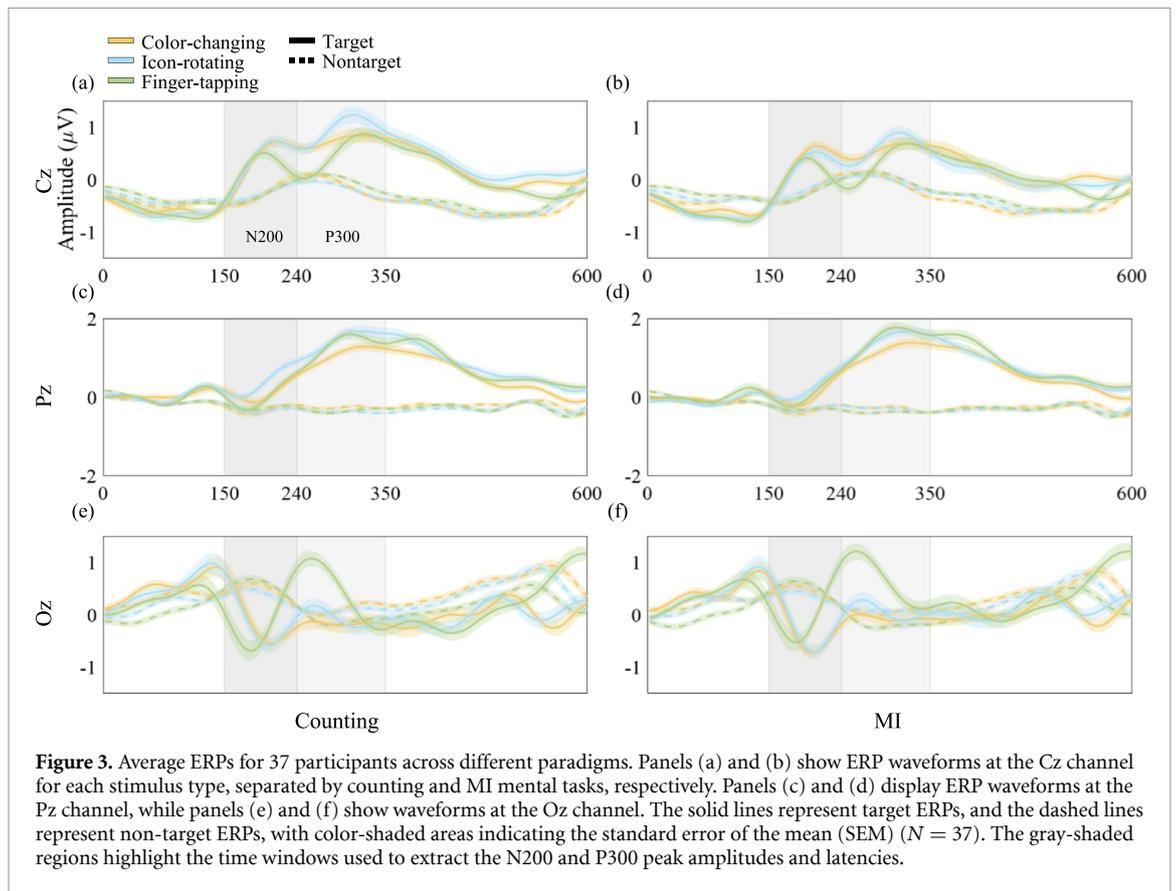


Figure 3. Average ERPs for 37 participants across different paradigms. Panels (a) and (b) show ERP waveforms at the Cz channel for each stimulus type, separated by counting and MI mental tasks, respectively. Panels (c) and (d) display ERP waveforms at the Pz channel, while panels (e) and (f) show waveforms at the Oz channel. The solid lines represent target ERPs, and the dashed lines represent non-target ERPs, with color-shaded areas indicating the standard error of the mean (SEM) ($N = 37$). The gray-shaded regions highlight the time windows used to extract the N200 and P300 peak amplitudes and latencies.

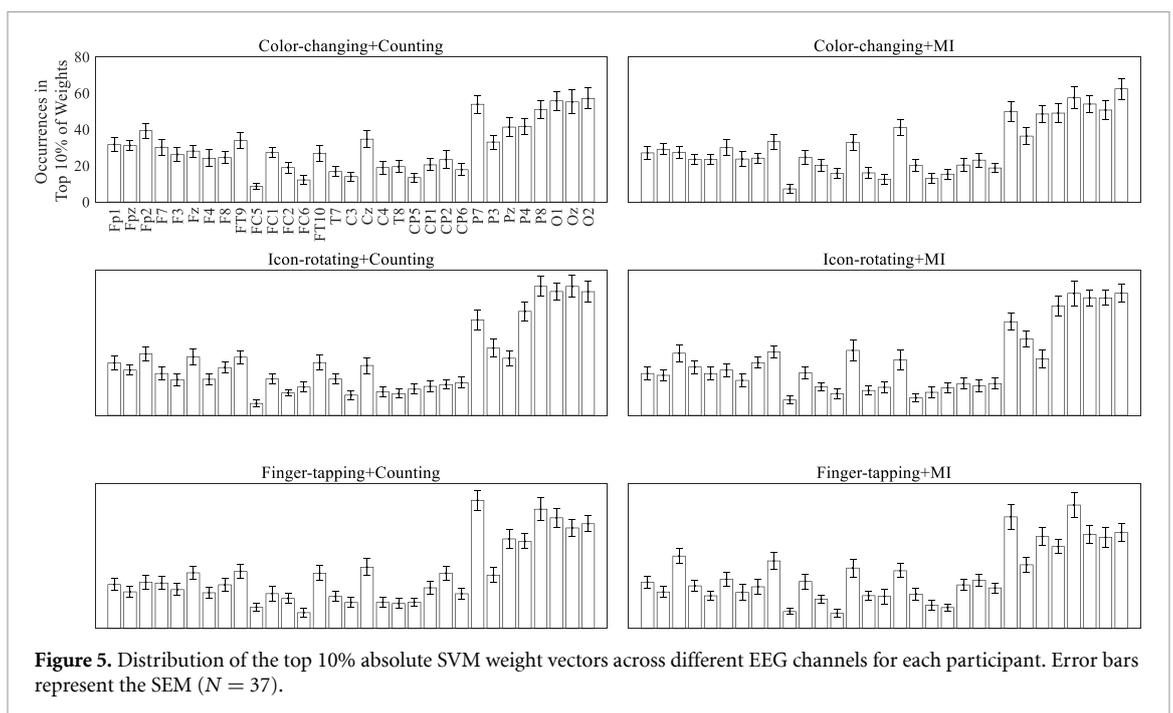
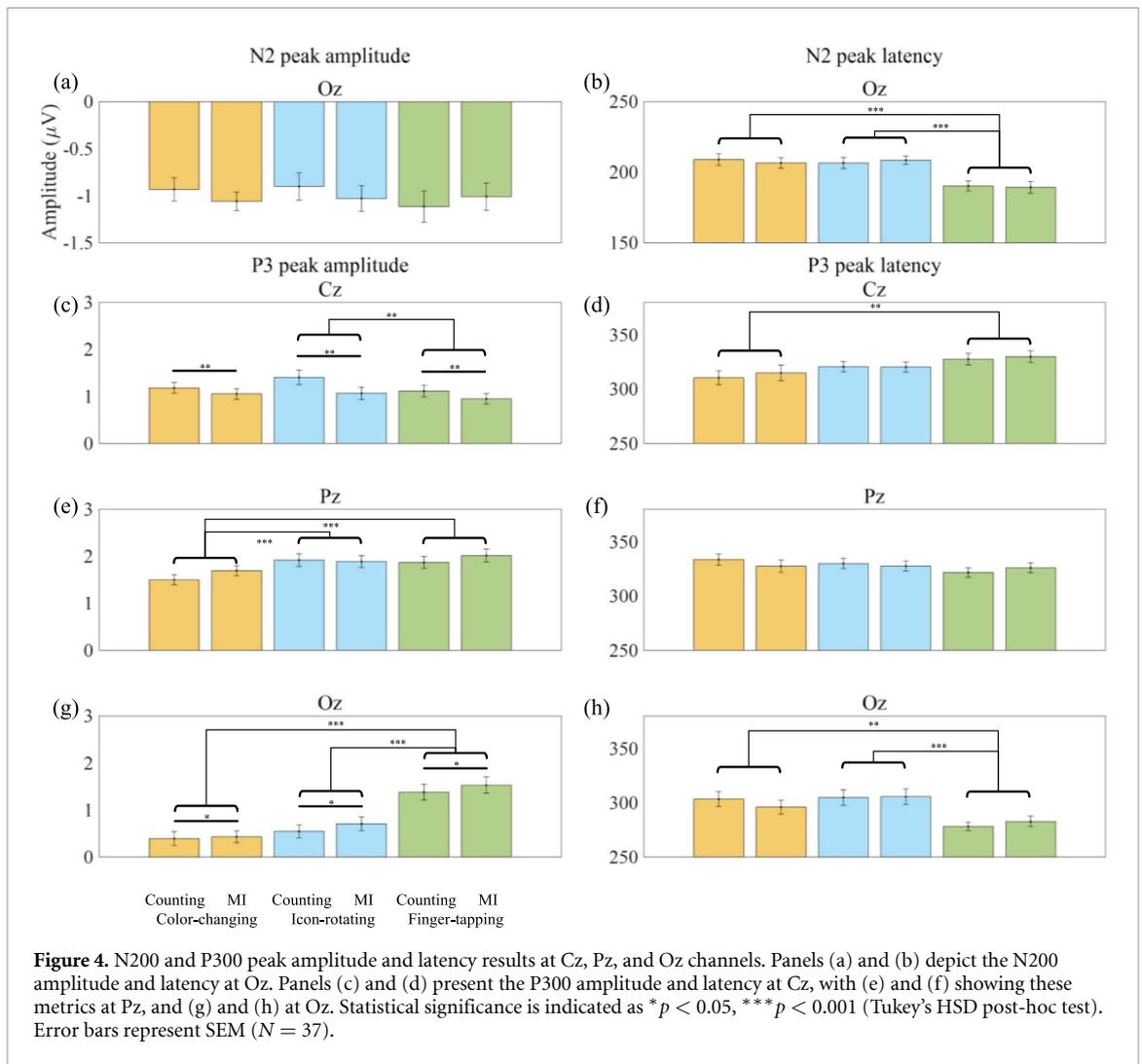
amplitude compared to the icon-rotating stimulus ($p = 0.0036$), and the MI task resulted in a smaller amplitude than counting ($p = 0.0098$). Latency differences at Cz were significant across stimuli ($p = 0.0029$), with the finger-tapping stimulus showing a longer latency than the color-changing stimulus ($p = 0.0046$). At Pz, P300 peak amplitude differed significantly by stimulus type ($p = 1.2381 \times 10^{-6}$), with the finger-tapping stimulus producing a larger amplitude than the color-changing ($p = 6.0886 \times 10^{-5}$). The icon-rotating stimulus also had a larger amplitude than the color-changing stimulus ($p = 0.0002$). There was no statistically significant difference in latency at Pz. At Oz, P300 peak amplitude was significantly affected by both stimulus type ($p = 6.8048 \times 10^{-11}$) and task ($p = 0.0212$). The finger-tapping stimulus produced a larger amplitude than the color-changing ($p = 5.8796 \times 10^{-8}$) and icon-rotating stimuli ($p = 5.18 \times 10^{-8}$), and the MI task yielded a larger amplitude than counting ($p = 0.0212$). Oz latency differences were significant across stimuli ($p = 0.0002$), with the finger-tapping stimulus exhibiting a shorter latency than both the color-changing ($p = 0.0075$) and icon-rotating stimuli ($p = 0.0002$). These findings indicate that while the finger-tapping stimulus might reduce the amplitude of the P300 at Cz, it enhances the response in the posterior channels, particularly at Pz and Oz,

which are crucial for P300 detection. This suggests that the finger-tapping stimulus may be particularly effective in elevating visual attention to stimuli, leading to larger and faster P300 responses in the posterior brain regions.

In the analysis of the SVM weight vectors, we found that the Parietal and Occipital channels consistently had the highest weight values among participants, indicating their importance in classification (figure 5). This finding aligns with the ERP analysis, where the finger-tapping stimulus enhanced the ERP responses in the Oz and Pz regions, correlating with higher classification accuracy.

3.3. Optimization toward single-trial BCI

In the offline analysis of BCI performance depending on the number of stimulus presentation repetitions, among all combinations of three feature extraction methods (none, XD, XDRG) and five classifiers (SVM, LR, EEGNet, shallow ConvNet, and deep ConvNet), using XDRG and LR produced the highest accuracy and ITR when we used single-trial ERPs as well as when we used the average ERPs from 2 repetitions (see table A1 for the full results of all combinations). The highest performance was achieved with the finger-tapping stimulus and counting, echoing the findings from section 3.1.



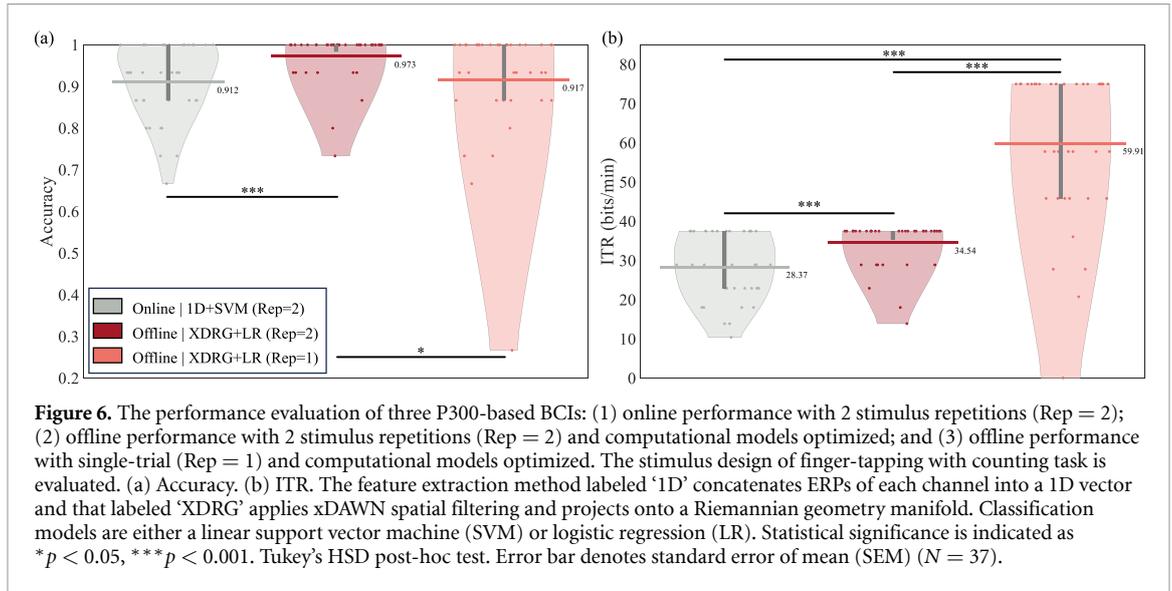


Table 2. The performance of three P300-based BCIs, including online performance with 2 stimulus repetitions (rep = 2), offline performance with 2 stimulus repetitions (rep = 2) and computational models optimized, and offline performance with single-trial presentation and computational models optimized. Refer to figure 6.

	Computational models	Stimuli	Mental task	Accuracy		ITR (bits/min)	
				Mean	Std.	Mean	Std.
Online performance (rep = 2)	1D vector + linear SVM			0.9117	0.0917	28.3704	8.4638
Offline performance (rep = 2)	XDRG + LR	Finger-tapping	Counting	0.9730	0.0598	34.5455	5.9422
Offline performance (single-trial)						0.9171	0.1409

Using the paradigm of the finger-tapping stimulus with counting and the optimized models (XDRG and LR), one-way rmANOVA revealed a significance difference in accuracy among three cases: online 2 repetitions without optimization, offline 2 repetitions with optimization, and offline no repetition with optimization ($p = 0.0059$). Offline accuracy from 2 repetitions of stimulus presentation with optimization was higher than offline accuracy (figure 6(a)) from single-trial presentation with optimization ($p < 0.05$) and online accuracy from 2 repetitions of stimulus presentation without optimization ($p < 0.001$). However, there was no significant difference in accuracy between offline single-trial presentation with optimization and online 2 repetitions of presentation without optimization. One-way rmANOVA also revealed a significance difference in ITR among three cases. ITR was the highest from offline single-trial presentation compared to offline and online ITRs from 2 repetitions of presentation (figure 6(b)), due to reduced presentation duration ($p < 0.001$). Therefore, the offline performance of single-trial P300-based BCIs with optimized computational models yielded a similar level of accuracy and

doubled improvement of ITR compared to the online performance of multi-trial (2 repetitions) P300-based BCIs without optimized computational models (see table 2).

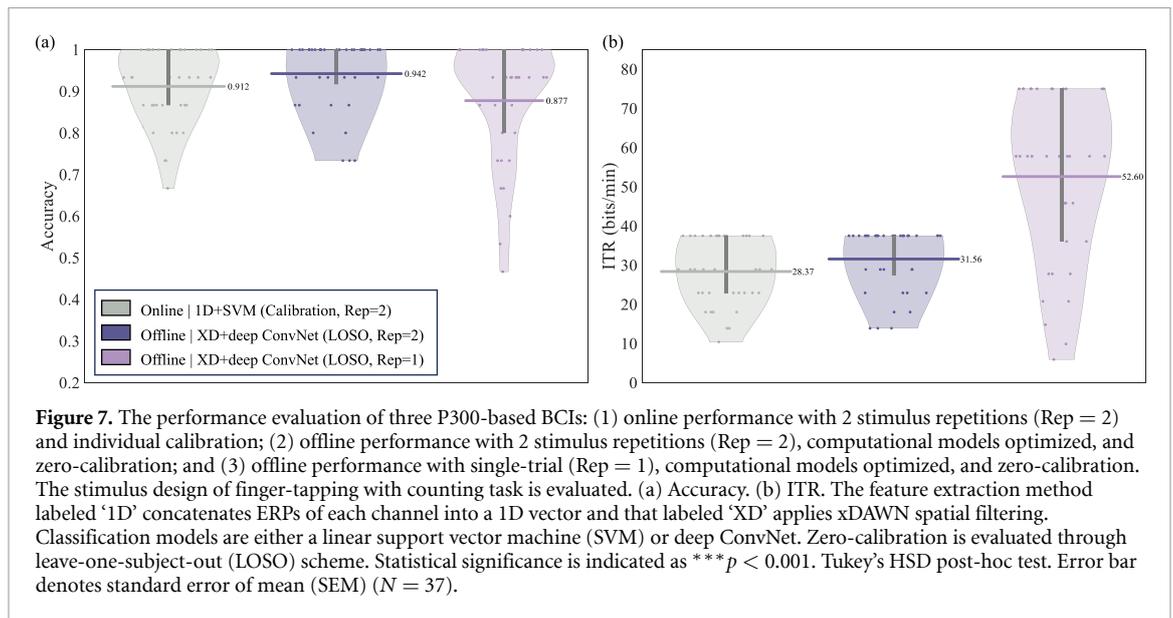
3.4. Optimization toward zero-calibration

In the offline analysis of zero-calibration BCIs via transfer learning, among all combinations of three feature extraction methods and five classifiers, using XD and deepConvNet produced the highest accuracy and ITR from LOSO CV (see table A2 for full results of all combinations).

Again, the highest zero-calibration BCI performance was achieved using the finger-tapping stimulus with counting. Using the finger-tapping stimulus with counting and optimized models (XD and deepConvNet), one-way rmANOVA showed a significant difference in accuracy among three cases: online individual calibration of 2 repetitions without optimization, offline zero-calibration of 2 repetitions with optimization, and offline zero-calibration of no repetition with optimization ($p = 0.0038$) (figure 7(a) and table 3). A post-hoc analysis showed that the accuracy in offline zero-calibration with two

Table 3. Performance of three P300-based BCIs: (1) online performance with 2 stimulus repetitions (rep = 2) and individual calibration; (2) offline performance with 2 stimulus repetitions (rep = 2), computational models optimized, and zero-calibration; and (3) offline performance with single-trial, computational models optimized, and zero-calibration. Refer to figure 7.

	Computational models	Stimuli	Mental task	Accuracy		ITR (bits/min)	
				Mean	Std.	Mean	Std.
Online individual calibration (rep = 2)	1D vector + linear SVM			0.9117	0.0917	28.3704	8.4638
Offline zero-calibration (rep = 2)	XD + deep ConvNet	Finger-tapping	Counting	0.9423	0.0863	31.5602	8.2012
Offline zero-calibration (single-trial)				0.8775	0.1436	52.6084	21.316



stimulus repetitions surpassed that of no calibration ($p < 0.001$). There was no difference in accuracy between online individual calibration without optimization and offline zero-calibration with optimization in 2 repetitions ($p = 0.2122$) and no repetition ($p = 2183$). Also, one-way rmANOVA showed a significant difference in ITR among three cases ($p = 1.6628 \times 10^{-11}$) (figure 7(b) and table 3). The highest ITR was achieved in offline single-trial presentations, outperforming both offline and online ITRs with two repetitions ($p < 0.001$). There was no significant difference in ITR between online individual calibration without optimization and offline zero-calibration with optimization for two repetitions of stimulus presentation. These results, consistent with those in section 3.2, indicate the potential for both single-trial and zero-calibration P300-based BCIs.

3.5. Individual variation of BCI performance

The analysis of CV revealed a trend that using dynamic stimuli exhibited lower CV values compared to using static stimuli (figure 8). Moreover, using the finger-tapping stimuli reduced CV more than using

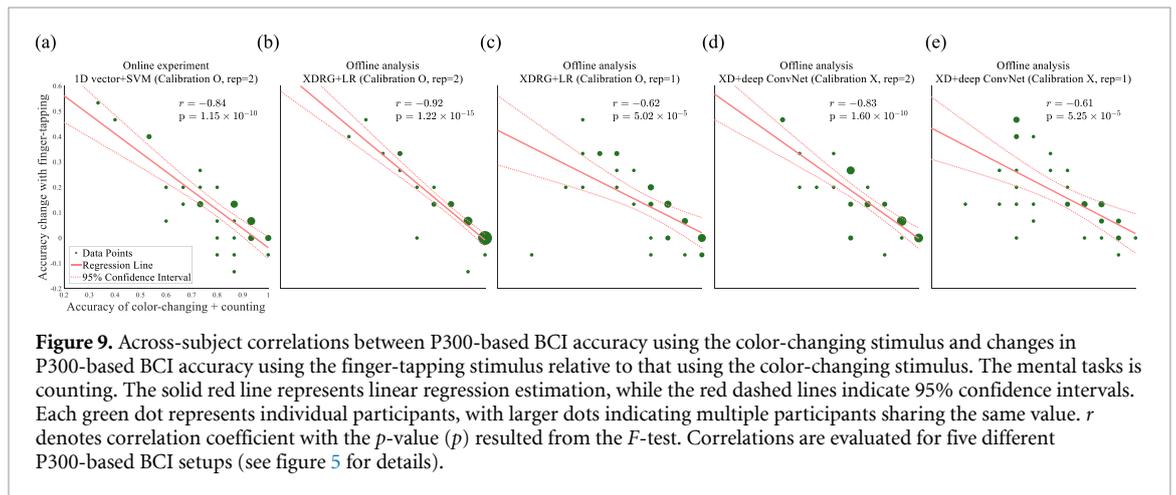
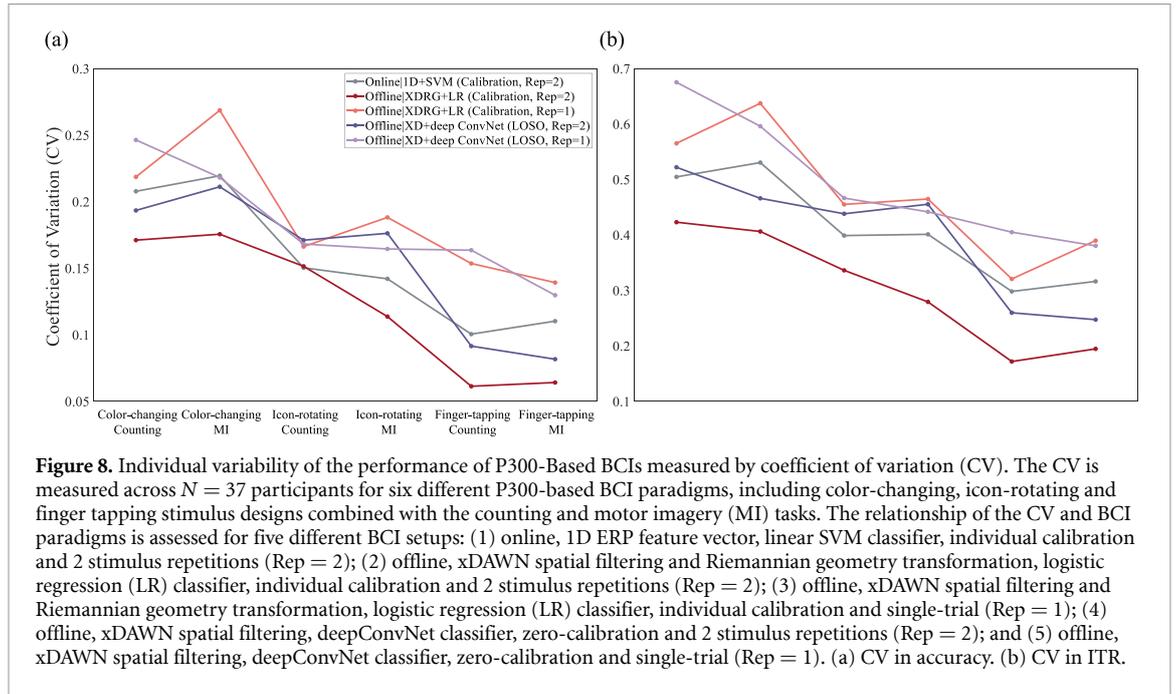
other stimuli for all online and offline analyses (see sections 3.2 and 3.3 for details of each analysis). The lowest CV was observed in the XDRG with LR during the offline analysis of BCIs with 2 repetitions of finger-tapping stimuli. In this case, the accuracy ranged from 0.7333 to 1 and the ITR ranged from 13.8882 to 37.5 bits/min across 37 participants.

Additionally, significant correlations were observed across all online and offline analyses between the accuracy in the color-changing with counting sessions and the change in accuracy when using finger-tapping stimuli ($ps < 0.05$) (figure 9). Negative correlations showed that participants who had lower accuracy in the color-changing with counting tended to achieve more improvement in accuracy by finger-tapping with counting.

4. Discussion

4.1. Overview of study and key findings

In this study, we proposed a novel stimulus design employing the finger-tapping animation relevant to target selection in the oddball task for P300-based BCIs. By eliciting more vivid ERPs using the



proposed stimulus, aimed to address key challenges in P300-based BCIs, including repeated stimulus presentations, individual calibration, and variations in individual performance. P300-based BCIs with the finger-tapping stimulus showed superior performance to those with conventional stimuli, reaching an online accuracy of 91.17% across 37 participants in selecting one of the four commands for controlling an external device. Further offline optimization of computational models improved the accuracy of P300-based BCIs with the finger-tapping stimulus up to 97.3%. An offline test of single-trial P300-based BCIs revealed an ITR of 59.91 bits/min while maintaining accuracy above 90%. Another offline across-subject evaluation demonstrated the plausibility of zero-calibration showing no difference in accuracy from individual calibration. Finally, using the finger-tapping stimulus further reduced variations in individual BCI performance compared to using conventional stimuli. Notably, greater performance

enhancement by the finger-tapping stimulus in participants who showed relatively lower performance using conventional stimuli indicates that the proposed stimulus design was particularly effective for poorer BCI performers. Our results demonstrated the potential of our novel stimulus design to enhance BCI performance and realize plug-and-play BCI systems.

4.2. Impact of stimulus design on BCI performance

A key hypothesis driving our research was that elevated attention by an intuitive stimulus design could enhance ERPs leading to the improvement of P300-based BCIs. This was substantiated in our online experiment, where a mere alteration in the stimulus paradigm resulted in remarkable performance gains. Our experiment results demonstrate the crucial role of stimulus design in the development of P300-based BCIs, showcasing that eliciting reliable ERPs by well-designed stimuli would be as important

Table 4. Comparative analysis of current study with previous research in P300-based BCIs.

Studies	Accuracy (%)	ITR (bits/min)	Number of participants	Repetition of stimuli
Present study with individual calibration	91.71	59.92	37	Single-trial
	97.30	34.54	37	2
Present study with zero calibration	87.75	52.61	37	Single-trial
Kshirsagar <i>et al</i> [43]	92.64	55.45	9	15
Kundu <i>et al</i> [44]	99	—	2	15
Ayğün <i>et al</i> [45]	94.56	7.73	30	15
Blanco-Díaz <i>et al</i> [46]	76	11.38	10	Single-trial
Du <i>et al</i> [47]	72.32	—	10	Single-trial

as applying advanced computational algorithms to decode ERPs. This significant performance improvement, achieved with minimal stimulus repetitions, may pave the way for the development of high-performance BCIs more efficiently.

4.3. Optimization of computational models

The optimization of computational models for ERPs elicited by the finger-tapping stimuli during the off-line analysis further enhanced P300-based BCIs. Even with single-trial configurations, accuracy improved to 0.92 with an ITR nearing 60 bits/min. The accuracy achieved through individual calibration (97.3%) ranks the second-highest (99% being the highest) among benchmarks in previous studies (see table 4). However, note that the study reporting the highest accuracy at 99% involved only two participants [42], potentially limiting its reliability. Moreover, the performance of single-trial BCIs built in this study surpasses existing benchmarks (table 4), notably recording the highest ITR, to the best of our knowledge. These outcomes collectively suggest that our innovative stimulus paradigm enables us to build one of the most proficient P300-based BCIs to date.

This study systematically explored the impact of different computational models on the performance of P300-based BCIs, particularly focusing on deep learning models like deep ConvNet and EEGNet. As highlighted in supplementary tables A1 and A2, deep neural networks generally outperformed traditional machine learning models such as SVM and LG. Among the DNNs, while EEGNet showed high performance, deep ConvNet yielded slightly better results. The difference in performance between these models may stem from their architectural differences. Specifically, deep ConvNet has a straightforward architecture with a series of stacked convolutional layers that directly capture both temporal and spatial features. In contrast, EEGNet employs a more complex architecture involving depthwise and separable convolutions designed to reduce the parameter count and enhance generalizability across different EEG paradigms. This simpler, yet deeper structure of deep ConvNet may allow it to more effectively leverage the discriminative ERP responses elicited by

the proposed stimulus presentation paradigm, leading to slightly better classification performance. We speculate that the more discriminative ERP responses generated by our paradigm could be classified more effectively by a less complex model like deep ConvNet, while also improving generalizability.

The outcomes of our optimization processes reveal some insights into the design of computational models for P300-based BCIs. First of all, spatial filtering such as xDAWN appears to be a key process for P300-based BCIs as shown in different analyses for single-trial BCIs or zero-calibration. Employing more sophisticated methods such as RG transformation and deep ConvNet also contributed to improving performance. Yet, using both RG and deep neural networks did not improve performance further. It may imply that simply mixing different sophisticated algorithms would not help much for P300-based BCIs but a combination of different algorithms optimized for given ERP data would be more important.

The high performance of zero-calibration P300-based BCIs demonstrated in this study may suggest that transfer learning without extensive data augmentation or domain adaptation is plausible for P300-based BCIs. It also points to the effectiveness of the xDAWN process and RG approach in transfer learning as shown by previous reports [14]. But we also suspect that ERPs elicited by the finger-tapping stimuli would be more common among participants than those by conventional stimuli, supported by performance differences between stimulus designs in table A2. Considering the universality of P300 components among people, elevating attention to a target stimulus by the proposed design would effectively mitigate variations of ERPs across participants.

4.4. Influence of task relevance and stimulus characteristics

Our results indicating better performance with finger-tapping than icon-rotating suggest that perceived task relevance likely plays a significant role in eliciting robust neural responses [42, 48]. Yet, there can be alternative explanations for the performance enhancement observed with the finger-tapping stimulus. One perspective is that the finger-tapping

stimulus, composed of a more colorful image, could simply increase stimulus saliency. This increased saliency might lead to heightened bottom-up attention, consequently improving BCI performance, with minimal engagement of the task-relevance aspect of attention. Alternatively, the difference in performance between icon-rotating and finger-tapping stimuli could be attributed to the representation of body-related stimuli. The animation of finger-tapping might draw more attention due to its relevance to bodily movements. Given that human perception tends to be more sensitive to stimuli related to human movements compared to other moving stimuli [49, 50], dynamic stimuli involving finger movements might attract greater attention than dynamic stimuli featuring rotating icons.

4.5. Interpretation of ERP and SVM weight vector analyses

Our analysis of ERP components and SVM weight vectors provides insights into why the finger-tapping stimulus outperformed other stimuli in P300-based BCIs. The finger-tapping stimulus resulted in significantly shorter N200 latencies and larger P300 amplitudes at the Pz and Oz channels, suggesting an enhanced attentional focus on these regions. Additionally, the analysis of SVM weight vectors revealed that the Parietal and Occipital channels contributed most significantly to classification accuracy, aligning with the stronger ERP responses observed at these sites. Furthermore, the greater differentiation between target and non-target ERPs across all channels with the finger-tapping stimulus supports its effectiveness in improving BCI performance. These findings suggest that the finger-tapping stimulus enhances the neural signals critical for BCI performance by focusing attention on key brain regions, leading to more accurate and reliable classification.

4.6. Considerations regarding mental tasks

Our findings indicated no significant differences in BCI performance between the counting and MI tasks, which may be due to several factors. First, participants were required to count the target stimulus only twice per block, reducing the task's cognitive load and its role in enhancing attention within the oddball paradigm. In contrast, MI relies on participants' ability to produce vivid mental imagery, which can be challenging within a fast-paced oddball paradigm that demands sustained attention and rapid task shifts.

The limited effectiveness of MI in this study may also reflect the brief training provided to participants. Prior research on MI-based BCIs suggests that MI signals are often highly variable and typically require extensive training to achieve consistent and reliable

responses [51]. In our study, participants underwent only brief training, likely contributing to the absence of a clear performance advantage for MI over counting. Additionally, post-experiment feedback from some participants indicated difficulties in performing MI while following the finger-tapping animation, which could further dilute MI's attentional impact.

Despite the lack of a significant performance difference, MI was included in this study to investigate its potential as a task-relevant mental process that aligns with the selection goal of the BCI, aiming to increase attentional focus on the target through task relevance. However, the absence of extended training and the complexity of combining MI with the oddball paradigm may have limited this effect. Furthermore, potential interference from movement-related cortical potentials in the sensorimotor area [52] could potentially interfere with ERPs elicited by target stimuli. Nonetheless, the consistently high performance observed with the finger-tapping stimulus across different mental tasks suggests the feasibility of building P300-based BCIs without requiring a specific accompanying mental task, thereby simplifying BCI use with minimal cognitive effort.

4.7. Interaction between MI and target recognition in event-related desynchronization (ERD) responses

We investigated the interaction between visual P300 and MI by analyzing ERD components. Contrary to our hypothesis, MI did not enhance P300-based BCI performance, as no significant differences were found between counting and MI tasks. Further offline analysis showed no improvement in BCI performance by adding spectral sensorimotor rhythm (SMR) features alongside ERP features (figure A2).

To understand this, we examined SMRs [53] during target selection, focusing on the mu (8–12 Hz) and beta (18–26 Hz) bands, and calculated the band power after the Hilbert transform of the filtered signals. Using the band power for the mu and beta rhythms, we generated and compared topographical maps between target and non-target stimuli (see figures A3 and A4). We observed ERD of both mu and beta rhythms in response to the target stimulus while no ERD in response to the non-target stimulus, regardless of the task (i.e. counting or MI). The ERD of the mu rhythm was more pronounced than that of the beta rhythm. This observation was consistently made for all color-changing, icon-rotating, and finger-tapping stimuli. ERD responses in the central lobe are often associated with motor execution, imagery, and observation, regions that share overlapping activation [54]. Given this, we considered it essential to assess whether dynamic stimuli like finger-tapping could influence these ERD responses.

This reasoning led us to conduct a two-way repeated measures ANOVA on ERD responses at C3, Cz, and C4, using stimulus type and mental task as factors. The analysis revealed no significant main effects of either stimulus type or mental task ($p > 0.05$) and no significant interaction between them. The absence of differences between the mental tasks suggests that the observed ERD responses are likely tied to cognitive processes involved in target recognition within the oddball paradigm rather than specifically to MI.

Since the mu band overlaps with the traditional alpha band, the observed alpha ERD may be related to cognitive processes associated with brain-wide alpha modulation rather than solely SMR modulation. Previous studies have reported alpha and beta modulation during the oddball task. For instance, one study observed an ERS followed by ERD of the alpha rhythm when recognizing the target during the oddball task, along with the elicitation of the P300 component [55]. Other studies have subdivided the alpha band and found immediate ERD responses to the target without preceding ERS [56], which aligns with our findings. Similarly, the beta band also exhibited an ERD response, though less pronounced, consistent with previous studies observing a minor ERD response in the beta band during the oddball task along with the generation of the P300 component [56].

These findings align with previous research, suggesting that ERD in the mu and beta bands during the oddball task is more closely tied to cognitive processes like attention and target recognition, rather than MI alone. Further studies are needed to explore these interactions in greater detail, especially with a broader range of MI tasks.

4.8. Regional ERP differences for finger-tapping vs. icon-rotating stimuli

Our results revealed distinct regional ERP patterns between the task-relevant finger-tapping and task-irrelevant icon-rotating stimuli (figure 3). The icon-rotating stimulus elicited a stronger P300 response in the central (Cz) area, likely reflecting a P3a component associated with attention to novel or infrequent stimuli. P3a is typically observed in response to non-target, unexpected events that capture general attentional resources [4]. Although both stimuli were infrequent, the task-irrelevant nature of the icon-rotating stimulus likely contributed to a more pronounced P3a response in the central region.

In contrast, the finger-tapping stimulus, designed to be task-relevant by mimicking the physical action of selection, showed enhanced ERP responses in the posterior and occipital (Pz and Oz) regions, indicative of a P3b component. P3b is commonly associated with task relevance and intentional evaluation processes [4], aligning well with the action-oriented mental imagery engaged by the finger-tapping

stimulus. These findings suggest that task relevance can modulate ERP responses regionally, with task-relevant stimuli preferentially enhancing P3b in posterior regions where task-related processing is more likely to occur. This interpretation aligns with our design intent, which posited that task relevance could amplify selective attention and contribute to higher-quality ERP signals.

These results are further reinforced by examining the feature weights of the classifier, provided in figure 5: the highest-weighted features—one-dimensional spatial feature vectors—along the absolute value axis were concentrated in the parietal and occipital regions. This shows that while the finger-tapping stimulus did not yield the highest overall peak amplitude of the ERP, it did elicit discriminative ERP components in posterior regions which the classifier was able to make effective use of. It also agrees with the nature of the finger-tapping task, given that parietal and occipital areas are involved in visual processing and attention to task-relevant stimuli. On the other hand, although the icon-rotating stimulus did elicit the larger P300 peak in the central region, the classifier could not benefit that much from that signal due to its spatial distribution and also task irrelevance. Therefore, the better performance of the finger-tapping with counting task may be explained because, in this task, the posterior regions contributed more strongly to the ERP features, underlining both the spatial and temporal features of the ERP to improve the classification accuracy.

4.9. Validation of the proposed decoding model in an online setting

To validate the optimized decoding model from the offline analysis (see section 3.3) in an online environment, we conducted an additional online experiment. This experiment used the finger-tapping with counting paradigm, along with the decoding model composed of the xDAWN filter and deep convNet classifier. Transfer learning was also applied by training the decoding model on the dataset collected in the main experiment and used for the offline analysis and testing it in new participants in the additional online experiment without individual calibration.

The additional online experiment involved 10 new participants who completed 30 test blocks (note that no training blocks were used because of transfer learning). The results showed a mean accuracy of 82.67% (SD = 12.45%) and a mean ITR of 43.31 bits/min (SD = 20.89 bits/min). While being slightly lower than the offline result (accuracy = 87.75%, ITR = 52.61 bits/min, see section 3.3), the online performance closely aligned with the offline analysis prediction, demonstrating the feasibility of the proposed stimulus paradigm and decoding model in online environments (table 5).

Table 5. Performance metrics of the proposed model in an additional online experiment.

Subject #	Accuracy	ITR (bits/min)
1	1.0000	75.0000
2	0.9667	65.1122
3	0.8667	45.8311
4	0.7667	31.7399
5	0.7000	24.1208
6	0.8000	36.0405
7	0.7333	27.7763
8	0.7667	31.7399
9	0.6667	20.7519
10	1.0000	75.0000
Mean	0.8267	43.3113
STD.	0.1245	20.8944

Although the small sample size limits broad generalizations, the additional online experimental result suggests the potential of establishing single-trial, calibration-free P300-based BCI systems. Based on this result, future studies will aim at further improving and validating the proposed BCI system.

4.10. Effectiveness of the two-stage training process

The two-stage training process in this study was designed to prepare participants for performing MI tasks effectively within a P300-based BCI paradigm. The first stage focused on developing a MI ability through the repetitive MI of finger-tapping movements, gradually transitioning from physical to purely mental tasks. The second stage introduced participants to various paradigms to help generalize their MI skills in different stimulus contexts.

Although the absence of a control group without MI training prevents direct evaluation of its impact, previous MI-BCI studies have shown that structured MI training enhances performance by improving participants' ability to generate reliable MI signals [57, 58]. Feedback from participants indicated that this process helped familiarize themselves with the task and synchronize their MI with the stimulus dynamics, suggesting that the training was effective in reducing variability in MI performance.

Regarding familiarity, the fixed sequence of tasks during the second practice stage was designed to gradually introduce participants to task complexities. Participants began with 'color-changing with counting', a simple paradigm widely used in P300-based BCI research, before progressing to more complex paradigms such as 'finger-tapping with MI'. This fixed order aimed to ensure that participants understood each task before advancing. However, the potential for familiarity effects due to this fixed order could arise. To mitigate familiarity effects, the main experiment featured randomized paradigm sequences, with each paradigm presented in 15 blocks following 30 training blocks. This design ensured that participants completed nine times more trials in a randomized

order compared to the fixed sequence in the practice phase. Additionally, if familiarity had significantly influenced performance, tasks practiced later in the fixed sequence, such as 'icon-rotating with MI', would have exhibited better performance. However, the results did not show this pattern, as 'finger-tapping with counting', practiced earlier, yielded better performance than these later tasks. These findings suggest that any familiarity effects introduced during the practice phase were likely minimal and would not significantly affect the overall results.

4.11. Comparison of counting and MI tasks

Counting and MI exhibited distinct strengths and limitations as mental tasks for P300-based BCIs. Counting is simple, requires minimal pre-training, and is less prone to inter-subject variability, making it practical for general BCI applications [27, 28]. However, its simplicity may fail to sustain user engagement in single-trial paradigms, where counting each stimulus just once lacks meaningful context.

In contrast, MI offers greater cognitive engagement and task relevance, particularly when aligned with motion stimuli such as 'finger-tapping with MI'. Post-hoc feedback from a number of participants in this study indicated that MI felt intuitive when paired with task-relevant stimuli, as it encouraged mental focus on the action being simulated. However, MI basically requires extensive training and exhibits higher inter-subject variability due to differences in individuals' mental imagery abilities. Moreover, some participants reported difficulty maintaining synchronization between their imagery and the stimulus, highlighting the challenge of integrating MI into practical BCI applications.

As such, future research should explore new mental strategies that combine the simplicity of counting with the engagement benefits of MI. Such strategies could enhance the usability and performance in P300-based BCIs, particularly in single-trial settings where consistent user engagement and task relevance are critical.

4.12. Study limitations and future research directions

Despite the remarkable improvement of P300-based BCIs with innovative stimulus design, the present study has several limitations that need to be addressed further. As mentioned earlier, it is challenging to precisely determine why finger-tapping stimuli led to high performance. Understanding the relationship between stimulus design and BCI performance in light of cognitive processing will be critical to advance the development of P300-based BCIs. Furthermore, the evaluation of BCI performance was limited to 15 blocks per test session, which may seem insufficient. However, this limitation was imposed by the experimental time constraints that require participants to

maintain focus. Lastly, MI generally requires training and is subject to significant variability among participants, which may have limited its effectiveness in our study. Due to the rapid and repetitive nature of our paradigm, we were unable to develop a variety of MI tasks. Future research that includes a broader range of MI tasks may potentially lead to better performance.

4.13. Conclusion and future directions

In conclusion, our research marks a significant advancement in enhancing the performance of P300-based BCIs. Moving beyond the conventional focus on computational models, this study emphasizes the importance of designing stimulus paradigms that align closely with human cognitive processes. By designing a user-centric stimulus based on attention mechanisms and cognitive engagement, we have laid the groundwork for more intuitive and efficient P300-based BCIs. These advancements may pave the way for the development of BCIs that are more accessible and user-friendly for everyday use. Particularly noteworthy is the potential demonstrated in this research for single-trial, zero-calibration P300-based BCIs, making a pivotal advancement. The next research step naturally involves translating these results into highly usable plug-and-play BCI systems, aiming to broaden the spectrum of practical applications and make P300-based BCIs more readily available and convenient for diverse users. Our follow-up study will delve into the feasibility of such plug-and-play P300-based BCIs.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2023-00302489).

We acknowledge the use of ChatGPT for language assistance in this manuscript. ChatGPT was employed solely to enhance the clarity and readability of the text through grammar and language corrections.

ORCID iD

Jongsu Kim  <https://orcid.org/0009-0004-1613-8706>

References

- [1] Värbu K, Muhammad N and Muhammad Y 2022 Past, present, and future of EEG-based BCI applications *Sensors* **22** 3331
- [2] Mak J, Arbel Y, Minett J W, McCane L M, Yuksel B, Ryan D, Thompson D, Bianchi L and Erdogmus D 2011 Optimizing the P300-based brain-computer interface: current status, limitations and future directions *J. Neural Eng.* **8** 025003
- [3] Krusienski D J, Sellers E W, McFarland D J, Vaughan T M and Wolpaw J R 2008 Toward enhanced P300 speller performance *J. Neurosci. Methods* **167** 15–21
- [4] Polich J 2007 Updating p300: an integrative theory of P3a and P3b *Clin. Neurophysiol.* **118** 2128–48
- [5] Wolpaw J R and Wolpaw E W 2012 Brain-computer interfaces: something new under the sun *Brain-Computer Interfaces: Principles and Practice* (Oxford University Press) p 14
- [6] Won K, Kwon M, Ahn M and Jun S C 2022 EEG dataset for RSVP and P300 speller brain-computer interfaces *Sci. Data* **9** 388
- [7] Cohen J and Polich J 1997 On the number of trials needed for P300 *Int. J. Psychophysiol.* **25** 249–55
- [8] Artzi N S and Shriki O 2018 An analysis of the accuracy of the P300 BCI *Brain-Comput. Interfaces* **5** 112–20
- [9] Townsend G, LaPallo B K, Boulay C B, Krusienski D J, Frye G E, Hauser C K, Schwartz N E, Vaughan T M, Wolpaw J R and Sellers E W 2010 A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns *Clin. Neurophysiol.* **121** 1109–20
- [10] Blankertz B, Lemm S, Treder M, Haufe S and Müller K-R 2011 Single-trial analysis and classification of ERP components—a tutorial *NeuroImage* **56** 814–25
- [11] Serby H, Yom-Tov E and Inbar G F 2005 An improved P300-based brain-computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **13** 89–98
- [12] Kim M, Kim J, Heo D, Choi Y, Lee T and Kim S-P 2021 Effects of emotional stimulations on the online operation of a P300-based brain-computer interface *Front. Hum. Neurosci.* **15** 612777
- [13] Lotte F et al 2018 A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update *J. Neural Eng.* **15** 031005
- [14] Li F, Xia Y, Wang F, Zhang D, Li X and He F 2020 Transfer learning algorithm of P300-EEG signal based on xDAWN spatial filter and Riemannian geometry classifier *Appl. Sci.* **10** 1804
- [15] Kundu S and Ari S 2019 MsCNN: a deep learning framework for P300-based brain-computer interface speller *IEEE Trans. Med. Robot. Bionics* **2** 86–93
- [16] Kindermans P J, Schreuder M, Schrauwen B, Müller K-R and Tangermann M 2014 True zero-training brain-computer interfacing—an online study *PLoS One* **9** e102504
- [17] Zhang Y, Zhao Q, Jing J, Wang X and Cichocki A 2012 A novel BCI based on ERP components sensitive to configural processing of human faces *J. Neural Eng.* **9** 026018
- [18] Kaufmann T, Schulz S M, Grünzinger C and Kübler A 2011 Flashing characters with famous faces improves ERP-based brain-computer interface performance *J. Neural Eng.* **8** 056016
- [19] Allison B Z and Neuper C 2010 *Could Anyone Use a BCI?, in Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction* ed D S Tan and A Nijholt (Springer) pp 35–54
- [20] Polich J 1987 Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli *Electroencephalogr. Clin. Neurophysiol.* **68** 311–20
- [21] Gonsalvez C J and Polich J 2002 P300 amplitude is determined by target-to-target interval *Psychophysiology* **39** 388–96
- [22] Connor C E, Egeth H E and Yantis S 2004 Visual attention: bottom-up versus top-down *Curr. Biol.* **14** R850–R2

- [23] Jin J, Zhang H, Daly I, Wang X and Cichocki A 2017 An improved P300 pattern in BCI to catch user's attention *J. Neural Eng.* **14** 036001
- [24] Eimer M and Holmes A 2007 Event-related brain potential correlates of emotional face processing *Neuropsychologia* **45** 15–31
- [25] Jin J, Allison B Z, Kaufmann T, Kübler A, Zhang Y, Wang X and Cichocki A 2012 The changing face of P300 BCIs: a comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement *PLoS One* **7** e49688
- [26] Jin J, Allison B Z, Wang X and Neuper C 2012 A combined brain–computer interface based on P300 potentials and motion-onset visual evoked potentials *J. Neurosci. Methods* **205** 265–76
- [27] Guo F, Hong B, Gao X and Gao S 2008 A brain–computer interface using motion-onset visual evoked potential *J. Neural Eng.* **5** 477–85
- [28] Potts G F 2004 An ERP index of task relevance evaluation of visual stimuli *Brain Cogn.* **56** 5–13
- [29] Navalpakkam V and Itti L 2005 Modeling the influence of task on attention *Vis. Res.* **45** 205–31
- [30] Gamer M and Berti S 2010 Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials *Psychophysiology* **47** 355–64
- [31] Potts G F, Patel S H and Azzam P N 2004 Impact of instructed relevance on the visual ERP *Int. J. Psychophysiol.* **52** 197–209
- [32] Bianchi L, Liti C and Piccialli V 2019 A new early stopping method for P300 spellers *IEEE Trans. Neural Syst. Rehabil. Eng.* **27** 1635–43
- [33] Gao W, Huang W, Li M, Gu Z, Pan J, Yu T, Yu Z L and Li Y 2023 Eliminating or shortening the calibration for a P300 brain–computer interface based on a convolutional neural network and big electroencephalography data: an online study *IEEE Trans. Neural Syst. Rehabil. Eng.* **31** 1754–63
- [34] Faul F, Erdfelder E, Lang A-G and Buchner A 2007 G* power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences *Behav. Res. Methods* **39** 175–91
- [35] Kaufmann T and Kübler A 2014 Beyond maximum speed—a novel two-stimulus paradigm for brain–computer interfaces based on event-related potentials (P300-BCI) *J. Neural Eng.* **11** 056004
- [36] Rivet B, Souloumiac A, Attina V and Gibert G 2009 xDAWN algorithm to enhance evoked potentials: application to brain–computer interface *IEEE Trans. Biomed. Eng.* **56** 2035–43
- [37] Zanini P, Congedo M, Jutten C, Said S and Berthoumieu Y 2017 Transfer learning: a Riemannian geometry framework with applications to brain–computer interfaces *IEEE Trans. Biomed. Eng.* **65** 1107–16
- [38] Barachant A et al 2023 pyRiemann/pyRiemann: v0.5. *Zenodo* (<https://doi.org/10.5281/zenodo.8059038>)
- [39] Lawhern V J, Solon A J, Waytowich N R, Gordon S M, Hung C P and Lance B J 2018 EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces *J. Neural Eng.* **15** 056013
- [40] Schirrneister R T, Springenberg J T, Fiederer L D J, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W and Ball T 2017 Deep learning with convolutional neural networks for EEG decoding and visualization *Hum. Brain Mapp.* **38** 5391–420
- [41] Li Y, Pan J, Wang F and Yu Z 2013 A hybrid BCI system combining P300 and SSVEP and its application to wheelchair control *IEEE Trans. Biomed. Eng.* **60** 3156–66
- [42] Kok A 2001 On the utility of P3 amplitude as a measure of processing capacity *Psychophysiology* **38** 557–77
- [43] Kshirsagar G B and Londhe N D 2019 Weighted ensemble of deep convolution neural networks for single-trial character detection in Devanagari-script-based P300 speller *IEEE Trans. Cogn. Dev. Syst.* **12** 551–60
- [44] Kundu S and Ari S 2020 P300 based character recognition using convolutional neural network and support vector machine *Biomed. Signal Process. Control* **55** 101645
- [45] Aygün A B and Kavsaoğlu A R 2022 An innovative P300 speller brain–computer interface design: easy screen *Biomed. Signal Process. Control* **75** 103593
- [46] Blanco-Díaz C, Guerrero-Méndez C and Ruiz-Olaya A 2023 Enhancing P300 detection using a band-selective filter bank for a visual P300 speller *IRBM* **44** 100751
- [47] Du P, Li P, Cheng L, Li X and Su J 2023 Single-trial P300 classification algorithm based on centralized multi-person data fusion CNN *Front. Neurosci.* **17** 1132290
- [48] Squires K C, Donchin E, Herning R I and McCarthy G 1977 On the influence of task relevance and stimulus probability on event-related-potential components *Electroencephalogr. Clin. Neurophysiol.* **42** 1–14
- [49] Harnad S 1990 The symbol grounding problem *Physica D* **42** 335–46
- [50] Barsalou L W 2008 Grounded cognition *Annu. Rev. Psychol.* **59** 617–45
- [51] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain–computer communication *Proc. IEEE* **89** 1123–34
- [52] Shibasaki H, Barrett G, Halliday E and Halliday A M 1980 Components of the movement-related cortical potential and their scalp topography *Electroencephalogr. Clin. Neurophysiol.* **49** 213–26
- [53] McFarland D J et al 2000 Mu and beta rhythm topographies during motor imagery and actual movements *Brain Topogr.* **12** 177–86
- [54] Hardwick R M, Caspers S, Eickhoff S B and Swinnen S P 2018 Neural correlates of action: comparing meta-analyses of imagery, observation, and execution *Neurosci. Biobehav. Rev.* **94** 31–44
- [55] Klimesch W 1999 EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis *Brain Res. Rev.* **29** 169–95
- [56] Sochurková D, Brázdil M, Jurák P and Rektor I 2006 P3 and ERD/ERS in a visual oddball paradigm: a depth EEG study from the mesial temporal structures *J. Psychophysiol.* **20** 32–39
- [57] Park S, Ha J, Kim D-H and Kim L 2021 Improving motor imagery-based brain–computer interface performance based on sensory stimulation training: an approach focused on poorly performing users *Front. Neurosci.* **15** 732545
- [58] Zhong Y, Yao L, Wang J and Wang Y 2022 Tactile sensation assisted motor imagery training for enhanced BCI performance: a randomized controlled study *IEEE Trans. Biomed. Eng.* **70** 694–702