A Plug-and-Play P300-Based BCI With Zero-Training Application

Jongsu Kim and Sung-Phil Kim

***Abstract*— The practical deployment of P300-based brain–computer interfaces (BCIs) has long been hindered by the need for user-specific calibration and multiple stimulus repetitions. In this study, we build and validate a plug-and-play, zero-training P300 BCI system that operates in a single-trial setting using a pre-trained xDAWN spatial filter and a deep convolutional neural network. Without any subject-specific adaptation, participants could control an IoT device via the BCI system in real time, with decoding accuracy reaching 85.2% comparable to the offline benchmark of 87.8%, demonstrating the feasibility of realizing a plug-and-play BCI. Offline analyses revealed that a small set of parietal and occipital electrodes contributed most to decoding performance, supporting the viability of low-density, high-accuracy BCI configurations. A data sufficiency simulation provided quantitative guidelines for pre-training dataset size, and an error trial analysis showed that both stimulus timing and preparatory attentional state influenced real-time decoding performance. Together, these results demonstrate the real-time validation of a fully pre-trained, zero-training P300 BCI operating on a single-trial basis, without stimulus repetition or user-specific calibration, and offer practical insights for developing scalable, robust, and user-friendly BCI systems.**

***Index Terms*— BCI, zero-training, EEG, P300, plug-and-play.**

I. Introduction

BRAIN-COMPUTER interfaces (BCIs) are a rapidly growing technology that enable users to interact with computers, machines, or digital environments directly through brain activity, without the need for any physical movement or muscular control [1]. BCIs interpret neurophysiological signals into actionable commands, promising applications in assistive technology, smart devices, rehabilitation, and gaming [2], [3], [4], [5]. Among the three main paradigms of BCIs—active, reactive, and passive—reactive BCIs have received significant attention due to their high reliability and relatively low training burden [6], [7]. In reactive BCI paradigms, users respond passively to external stimuli, and their brain's involuntary responses are decoded to infer intent.

Two of the most established paradigms in reactive BCI research utilize steady-state visual evoked potentials (SSVEP) [8] and P300-based event-related potentials (ERPs) [9]. SSVEP-based BCIs detect periodic EEG responses that are time-locked to the frequency of a flickering visual stimulus. These responses are strong, consistent, and relatively easy to classify, which enables SSVEP systems to support high-speed control. However, SSVEP performance can degrade when stimuli are placed too closely ($<4°$ visual angle) due to mutual interference, or too far apart ($>13°$) due to reduced amplitude [10], [11], [12], and responses also drop sharply beyond $\sim4°$ eccentricity [13]. While Nakanishi et al. demonstrated a 40-target speller with densely packed stimuli [14], the visual angles were not reported, and later studies have continued to confirm spatial and visual-angle constraints in SSVEP paradigms. At a 60 cm viewing distance, 4° corresponds to $\sim$4 cm spacing, which limits scalability on small screens or narrow fields of view (e.g., AR/VR headsets) [15]. These constraints, along with visual fatigue from prolonged flicker exposure [16], make SSVEP less practical in mobile or wearable contexts. In contrast, P300-based BCIs have shown robust performance even with very narrow inter-stimulus spacing, reflecting their reduced dependence on spatial separation (e.g., Speller). On the other hand, P300-based BCIs rely on detecting the P300 ERP component, a positive EEG deflection that typically occurs about 300 milliseconds after the user perceives a rare but meaningful stimulus in an "oddball" paradigm [17], [18]. Because P300 does not rely on frequency-tagged flicker or precise spatial separation, it affords substantially greater flexibility in stimulus design. Visual cues can vary in shape, color, size, animation, or symbolic form without degrading performance, and stimuli can be distributed across central or peripheral regions—even at irregular distances—provided they remain attentionally salient. Moreover, P300-based BCIs are inherently supramodal, operating with visual, auditory, or tactile stimuli as long as the oddball paradigm is employed. This versatility stands in contrast to SSVEP-based BCIs,

Received 20 May 2025; revised 19 August 2025; accepted 25 August 2025. Date of publication 29 August 2025; date of current version 4 September 2025. This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by Korean Government [Ministry of Science and ICT (MSIT)] under Grant RS-2025-00522357; and in part by the 'Alchemist Project' (Fully Implantable Closed-Loop Brain to X for Voice Communication) funded by the Ministry of Trade, Industry and Energy (MOTIE), South Korea, under Grant 20012355 and Grant NTIS 1415181023. *(Corresponding author: Sung-Phil Kim.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Ulsan National Institute of Science and Technology under Application No. UNIST-IRB-18-08-A, and performed in line with the Declaration of Helsinki.

The authors are with the Department of Biomedical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea (e-mail: kjstn737@gmail.com; spkim@unist.ac.kr).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2025.3603979, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2025.3603979

where stimulus properties such as color, size, and layout must be carefully optimized to maintain robust frequency tagging [19], [20], [21], [22], and prolonged high-frequency flicker can induce visual fatigue [23]. Such design flexibility makes P300 paradigms particularly well-suited for deployment in real-world and visually diverse contexts, including smart homes, mobile interfaces, AR/VR systems, and large ambient displays.

Despite their advantages, P300-based BCIs face several technical challenges that have limited their transition from laboratory settings to real-world applications. One of the primary challenges is the need for multiple trials of repeated stimulus presentations to obtain reliable ERPs [24], [25], [26], [27], [28]. Since single-trial P300 responses are often contaminated by noise, repeated presentations are used to average out this noise and enhance the signal-to-noise ratio [29], [30]. However, repetition significantly slows down interaction and induces user fatigue [27], [31], reducing the overall efficiency of the system. Another major challenge is the reliance on user-specific calibration sessions, which are time-consuming and require users to provide labeled data before the system can be used effectively [32]. This undermines the concept of "plug-and-play" BCIs and is particularly problematic in clinical or large-scale deployment scenarios. Furthermore, there is substantial inter-subject variability in P300 responses due to differences in attention, fatigue, and cognitive engagement, which complicates the design of generalized models [33], [34], [35], [36]. Lastly, the stimulus design itself plays a crucial role in the quality of the elicited P300. Poorly designed or low-relevance stimuli often fail to engage the user's attention, leading to inconsistent ERP generation [37], [38], [39].

Beyond algorithmic and paradigm-specific limitations, translational deployment further hinges on hardware form factors and robustness to motion artifacts in naturalistic settings. Recent wearable EEG advances—such as around-ear cEEGrids, headphones-integrated sensors, and mobile P300 during whole-body navigation—demonstrate promising feasibility outside the lab, while simultaneously highlighting challenges in signal quality under movement and constrained electrode layouts [40], [41], [42]. These developments emphasize the need for a compact set of electrodes over P300-relevant regions and preprocessing that mitigates motion-related noise, providing a concrete path toward plug-and-play BCIs that maintain reliability during everyday use.

To address these issues, prior studies have explored improvements in both stimulus design and signal processing. Our previous work proposed a novel stimulus paradigm using task-relevant finger-tapping animations to simulate physical interaction, thereby enhancing attention and improving ERP responses [43]. Using this design, we demonstrated enhanced decoding performance at the single-trial level without individual calibration in an offline analysis. On the signal processing side, recent advances such as xDAWN spatial filtering [44], Riemannian geometry-based features [34], and deep learning classifiers [45] have also shown promise in improving performance across users. However, most of these findings were derived from offline evaluation in which training and testing were performed on fixed datasets with cross-validation, rather than being validated on entirely new users in real time. As such, the real-world generalizability and robustness of these approaches remain uncertain. Therefore, it remains unclear whether P300 BCI systems equipped with the novel stimulus design and advanced signal processing techniques can function robustly in real-time, user-independent settings.

In the present study, we validate a zero-training, single-trial P300 BCI system in an online experiment. Thirty-four participants controlled a smart IoT lighting device by focusing on visually presented stimuli. The system used a pre-trained xDAWN with a deep convolution network pipeline, optimized from prior offline datasets, and processed EEG signals in real time without user-specific training. Participants issued commands such as turning the light on/off, changing color, or adjusting brightness, and the system responded in real time.

The results demonstrated high classification accuracy and information transfer rate (ITR) under a fully plug-and-play setting using only single-trial ERPs. To complement the real-time experiment, we conducted offline analyses: (1) channel ablation to quantify EEG channel contributions, (2) data sufficiency simulation to estimate training data requirements, and (3) error trial analysis examining behavioral and neural correlates of misclassifications, including alpha-band power and stimulus sequence effects.

We also explored whether predictable (fixed-order) presentation could reduce attentional lapses in single-trial settings. However, as discussed later, results indicated that maintaining stimulus unpredictability is essential. These findings underscore the importance of attentional engagement and expectancy in P300 elicitation. The online results highlight the potential of single-trial, zero-training plug-and-play BCIs deployable beyond controlled laboratory settings.

II. Methods

A. Participants

A total of fifty healthy adults participated in the online BCI experiment. Among them, thirty-four participants (mean age = 23.41 years old, SD = 4.72) completed the main experiment using a random-order stimulus presentation. Sixteen additional participants mean age = 22.44 years old, SD = 3.30) were tested under a fixed-order condition to examine whether predictable stimulus sequencing could reduce missed responses in single-trial settings. Results from the fixed-order condition are reported in the Discussion. All participants had normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Written informed consent was obtained from participants in accordance with the UNIST Institutional Review Board (UNIST-IRB-18-08-A).

B. Experiment Design

The experiment was designed with a four-option P300 oddball task simulating smart home light control. Participants selected one of the four visual icons on a screen to issue commands to a smart light through brain signals. The four commands included turning the light on, turning it off, adjusting brightness, and changing color. Each command was represented by a distinct icon displayed at a fixed location in one of the screen's four corners (Fig. 1A). The task began with

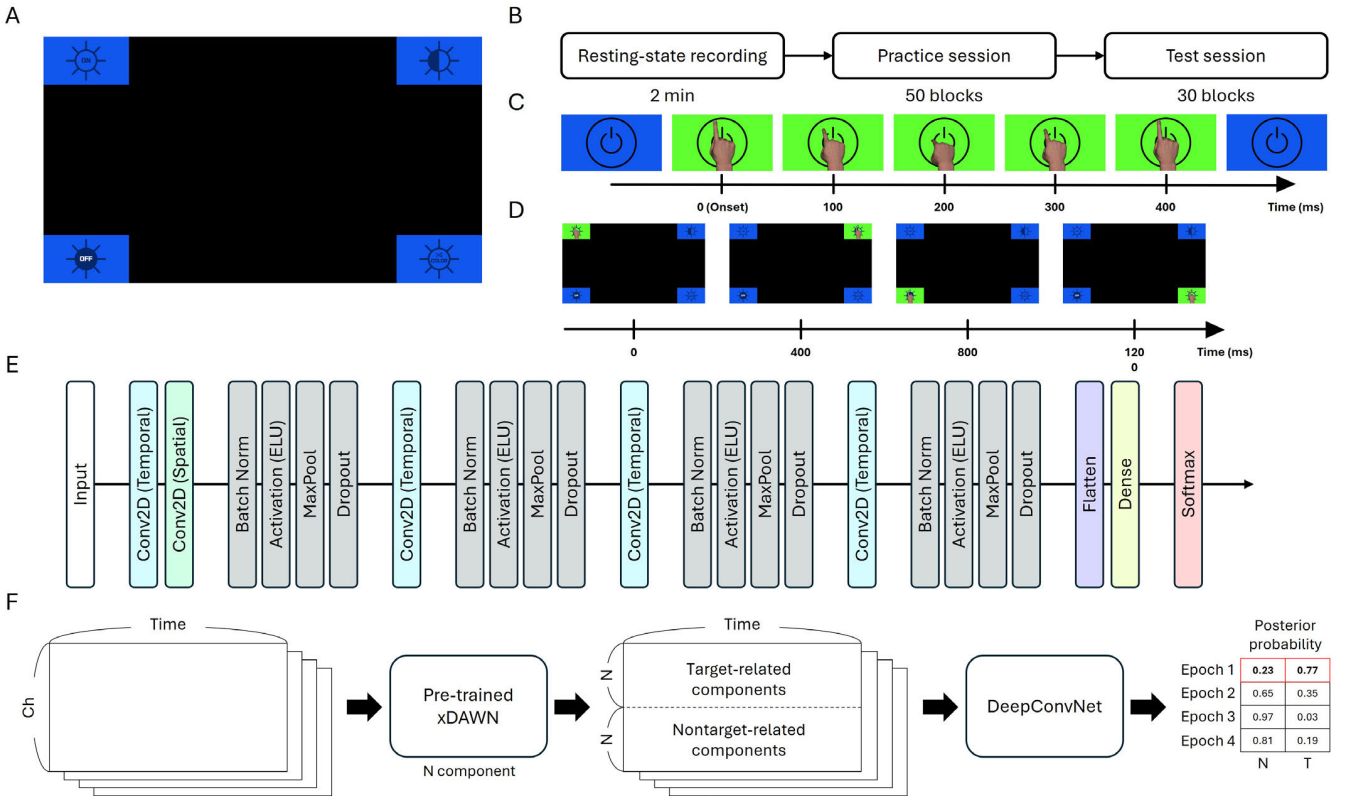

Fig. 1. Overview of the experimental paradigm, visual stimuli, and signal processing pipeline. A) The user interface (UI) presented to participants during the P300-based IoT smart lighting control task. Four icons representing distinct light commands (e.g., power on/off, brightness adjustment, color change) were positioned at each corner of the screen. B) The overall experimental procedure, consisting of a 2-minute resting-state EEG recording, a 50-block practice session, and a 30-block test session with real-time BCI operation. C) Example of the finger-tapping animation used as visual stimuli. Each stimulus was shown for 400 ms, visually mimicking the act of pressing a virtual button. D) Illustration of stimulus presentation order in the fixed-order condition, where the four stimuli are shown sequentially in a predetermined spatial order. In the random-order condition, the same stimuli were presented in a randomized sequence (not depicted here for brevity). E) The architecture of DeepConvNet used for classifying P300 responses. The model included multiple temporal and spatial convolution layers, batch normalization, max pooling, dropout regularization, and a final softmax layer. F) The signal processing pipeline for real-time classification. EEG data were first transformed using a pre-trained xDAWN spatial filter, separating target- and non-target-related components. These components were then passed into the DeepConvNet, which computed the posterior probability for each class based on the four stimulus presentations in each trial.

the light off; users first turned it on. Subsequent commands (brightness, color, off) were randomized, ensuring a logical and variable interaction cycle.

Before the main experiment, participants completed a 2-minute baseline EEG recording while fixating on a central cross and minimizing movement (Fig. 1B). This baseline was recorded only once for each participant immediately after EEG setup, and the resulting data were used to calibrate artifact subspace reconstruction (ASR) parameters and to detect bad channels. These parameters were then fixed and consistently applied to all online signal processing throughout the experiment without any additional calibration.

Each trial consisted of the following sequence. First, a fixation cross was presented in the center of the screen for 1 second, during which participants were asked to maintain a steady gaze and minimize movement. Next, one of the four icons was highlighted as the target, displayed for 1 second to indicate the command participants should focus on. This cue served as an instruction to mentally prepare for the upcoming selection process. Following the cue, the four icons were each sequentially highlighted with a 400-millisecond finger-tapping animation that visually simulated button pressing (Fig. 1C). No inter-stimulus interval (ISI) was used: each animation was presented for 400 ms, and the next animation began immediately after the previous one ended, resulting in a stimulus-onset asynchrony (SOA) of 400 ms. There was no temporal overlap between animations. The four icons were highlighted once per block in a randomized order. Participants were instructed to focus on the target and ignore other non-targets. This constituted a single-trial oddball paradigm, where each option appeared once per block without repetition. A block consisted of four trials corresponding to each icon. No inter-stimulus interval was given in the block.

The experiment consisted of a practice session followed by a test session. In both sessions, participants interacted with a smart lighting device (Philips Hue; Signify N.V., Eindhoven, Netherlands) through a P300-based interface. During the practice session, system control was not based on EEG signals. Instead, light responses were automatically executed using the aforementioned scenario. Participants were informed that the system was learning from their brain signals and were instructed to concentrate as if their EEG data were actively being processed. The purpose of this approach was to encourage engagement and stabilize attentional focus, both of which

are critical for eliciting reliable P300 responses. Participants were also told that automatic control of the light was intended to help them concentrate by minimizing frustration.

The 50-block practice session familiarized participants with the task and stabilized attentional readiness for reliable P300 responses. Participants were encouraged to choose their mental strategy—such as mental counting or imagery—rather than following a prescribed mental task.

The test session consisted of 30 blocks. In this session, EEG signals were processed and classified in real time using a pre-trained decoder (see Section II-D-F). The number of blocks was chosen to balance adequate performance assessment with minimizing fatigue, which could negatively affect attention in single-trial tasks. Based on the classification results, the system executed the corresponding light control commands, and participants received immediate visual feedback through changes in the lighting device (supplementary video 1). This allowed us to evaluate the BCI system's real-time performance under conditions approximating practical use.

C. EEG Data Aquisition

EEG signals were recorded using a 31-channel actiCHamp system (Brain Products GmbH, Germany) with active wet electrodes positioned according to the international 10–20 system. Electrode sites included frontal, central, parietal, and occipital regions to ensure comprehensive spatial coverage of ERP components, particularly those associated with visual attention and P300 generation. The left mastoid was used as the reference electrode, and the right mastoid served as the ground. Signals were sampled at 500 Hz, and electrode impedance was kept below 10 kΩ throughout the recording. Data quality was continuously monitored to ensure minimal contamination from muscle, eye, or movement artifacts.

D. Preprocessing and Feature Extraction

All EEG data underwent a preprocessing pipeline prior to feature extraction and classification. In both the online and offline settings, raw EEG signals were first bandpass filtered using a 1-Hz high-pass filter to remove slow drifts and a 30-Hz low-pass filter to suppress high-frequency noise and muscle artifacts. Filters were implemented as finite impulse response (FIR) filters using a Hamming window design, following the same principles used in EEGLAB's pop_eegfiltnew function. The 1 Hz high-pass filter employed a fixed filter order of 1500 taps, while the 30 Hz low-pass filter used a variable order determined by the cutoff frequency, sampling rate (500 Hz), and a transition bandwidth ratio of 0.25—for example, resulting in an order of 826 taps for the 30 Hz cutoff.

Bad channels were identified from a two-minute resting-state EEG recording before the task. Using a modified 'clean_channels' function in EEGLAB, channel–neighbor correlations were assessed in 5-s windows, and channels with correlations below 0.8 in more than 40% of the baseline were flagged as bad. These channels were interpolated throughout the online experiment using spherical interpolation. ASR was applied to suppress transient noise such as eye blinks and muscle activity, with parameters estimated from the same baseline. For offline analyses, the same preprocessing (filtering, bad channel interpolation, ASR) was applied for consistency.

After preprocessing, the EEG was segmented into epochs from –200 to 600 ms. Baseline correction was performed using the –200 to 0 ms interval. These corrected epochs were then passed through a pre-trained xDAWN spatial filter, which enhances discriminability between target and non-target ERPs by maximizing the signal-to-signal-plus-noise ratio [44], [46]. In this study, xDAWN filters were trained offline using pre-training data (see Section II-F) and applied without further adaptation during online classification.

E. Pre-Training Dataset

The xDAWN spatial filters and the classifier used in this study were pre-trained on a dataset collected from thirty-seven healthy participants in our previous study [43]. In that experiment, participants performed a four-class P300 BCI task over six sessions, which involved three different stimulus types and two types of mental strategies. Each session consisted of 45 blocks, with each of the four stimuli presented twice per block, resulting in eight sequential presentations per block. This yielded a total of 9,990 target-class samples (45 blocks × 6 sessions × 37 participants) and 29,970 non-target samples, which were three times as numerous as the target class. The protocol and visual paradigm were consistent with those of the current study, making the dataset well-suited for training models in zero-training online BCI systems.

F. Classification

The spatially filtered EEG features were input to a deep convolutional neural network (DeepConvNet) [45], which had been trained offline using the pre-training dataset. The DeepConvNet architecture (Fig. 1E) consisted of an initial temporal convolution layer followed by a spatial convolution layer, and then three additional convolutional blocks, each comprising batch normalization, a non-linear activation function (ELU), max pooling, and dropout regularization. This architecture was designed to effectively capture both the temporal dynamics of P300 responses and the discriminative spatial features derived from target and non-target xDAWN components (Fig. 1F). The model produced a binary prediction (target vs. non-target) for each stimulus presentation. A final decision for each block was made by selecting the stimulus with the highest posterior probability of being the target among the four presented options. The time to run the pre-trained xDAWN spatial filter per trial was 2.03 milliseconds, and the pre-trained DeepConvNet classifier per trial was 668.43 milliseconds on average (Intel Core i9-12900K CPU, 32 GB DDR4 RAM, NVIDIA GeForce RTX 3060 Ti GPU), supporting real-time BCI system operation.

G. Offline Analyses

Following the online experiment, several offline analyses were conducted to evaluate system robustness, identify informative EEG channels, and provide empirical guidelines for dataset construction in zero-training P300 BCIs. First, a channel ablation analysis was performed to assess the

contribution of each EEG channel to classification performance. For each channel, the corresponding ERP signal was replaced with zeros, and the classification process—pre-trained xDAWN spatial filters and the DeepConvNet—was re-applied to the entire test dataset. The resulting drop in accuracy was used to quantify the importance of each channel. In addition to zeroing-based ablation under a fixed pre-trained decoder, we conducted a leave-one-channel-out retraining (LOCO-R) analysis. For each electrode, we retrained the xDAWN spatial filters and the DeepConvNet with that channel excluded from both training and inference, and evaluated the resulting model on the same test data. This procedure estimates the performance impact when the pipeline is allowed to re-optimize in the absence of a given channel. Second, a data sufficiency simulation was conducted to determine the amount of training data required to build effective pre-trained models. To emulate real-world conditions where BCI datasets are gradually accumulated, we incrementally increased the number of participants (1–37) and blocks per participant (10–270), following the actual data collection order. The lower bound of 10 blocks per participant was chosen to ensure sufficient data for partitioning into training and validation sets, which is necessary for stable DeepConvNet optimization. Data were accumulated in steps of five blocks per participant to balance temporal resolution with computational feasibility. At each iteration, a new xDAWN + deep ConvNet model was trained on the incrementally accumulated data and evaluated using the EEG recordings from the online experiment. To model the trend in decoding accuracy, we fitted an inverse-saturation function of the form:

$$\hat{y} = 1 - \frac{1}{a \cdot S + b \cdot B + c} \tag{1}$$

where $\hat{y}$ denotes the predicted classification accuracy, S is the number of training subjects, and B is the number of blocks per subject. The parameters a, b, and c were estimated from the data, capturing the respective contributions of subject count, block count, and an intercept term. This inverse saturation function reflects the empirical observation that BCI accuracy asymptotically approaches a ceiling as the amount of training data increases.

Finally, an error trial analysis was conducted to investigate the behavioral and neural correlations of misclassifications. Post-experiment interviews collected subjective reports such as loss of attention or target uncertainty. Quantitatively, baseline alpha-band power (8–14 Hz) during error and correct trials was compared to assess attentional readiness [47], [48], [49]. Baseline activity was defined as the pre-stimulus interval (−200 ms to stimulus onset) within each block. In addition, we examined whether the temporal position of the target stimulus within a block influenced error likelihood. To this end, we used a binomial generalized linear mixed model (GLMM), where each trial was treated as an observation. The model included a subject-specific random effect and an intercept-only fixed effect. The intercept's deviation from the theoretical chance level of 0.25 (equivalent to –1.099 on the logit scale) was assessed using a Z-test. Using GLMM allowed us to assess whether the target position introduced a systematic bias in classification accuracy beyond what would be expected by chance. In parallel, we performed a permutation-based analysis to evaluate whether the observed error rates at each of the four target positions differed from chance. We pooled all misclassified trials across subjects and computed the empirical proportion of errors at each position. To establish a null distribution, we randomly permuted the target positions in the error trials 100,000 times, preserving the overall number of errors. For each permutation, the error rate for each position was calculated, and the mean and standard deviation were obtained across permutations. One-sided p-values were computed by counting the proportion of permutation samples that exceeded (or fell below) the empirical value, depending on the direction of the effect. This allowed us to quantify statistical deviations from chance without assuming a specific distribution.

H. Performance Evaluation

BCI performance was assessed using two primary metrics: classification accuracy and ITR. Accuracy was defined as the proportion of test blocks in which the target command was correctly identified by the BCI system. ITR was computed in bits per minute according to the standard BCI formula:

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

where P denotes the classification accuracy, N represents the number of selectable options (with N=4 in the present study), and T refers to the time required for a single selection in seconds, calculated as T=(Stimulus duration)×N×(Number of repetitions per stimulus). For our system, this corresponds to $T = 2.5$s (1.6s stimulus presentation + 0.2s post-stimulus + 0.67s classification). The additional 0.2 s post-stimulus window reflects our ERP epoch setting of –200 to 600 ms relative to stimulus onset, ensuring that the full P300 time window is captured before classification.

I. Statistical Analysis

To evaluate the relationship between attentional state and classification performance, we conducted two analyses: a bootstrap test for alpha-band power and a permutation test for temporal error bias. For prestimulus readiness, alpha power (8–14 Hz) was compared between correct and error trials at five posterior and occipital electrodes (P4, P8, O1, Oz, O2) using 100,000 bootstrap resamplings, with significance determined by whether the 95% confidence interval excluded zero.

To test temporal error bias, we conducted 100,000 permutations by shuffling target positions within error trials while preserving error counts, and compared empirical error proportions with null distributions. In addition, a binomial GLMM was fitted for each position with subject as a random effect, and intercepts were compared against the theoretical logit of 0.25 using a Z-test.

III. Results

A. Online Decoding Performance and Comparison With Zero-Training Offline Analysis

We compared real-time decoding with a benchmark offline analysis from our previous zero-training P300 BCI study using leave-one-subject-out (LOSO) cross-validation [43]. In the offline analysis, xDAWN filters and a deep ConvNet were trained on thirty-six participants and tested on the remaining one, yielding an average accuracy of 0.878 (SD = 0.144) and an ITR of 52.6 bits/min (SD = 21.3). In the present online experiment, the same pre-trained models achieved an average accuracy of 0.852 (SD = 0.106) and an ITR of 46.4 bits/min (SD = 17.8) (Fig. 2). A Wilcoxon rank-sum test showed no significant difference between offline and online accuracies ($p = 0.132$) or ITRs ($p = 0.132$), confirming that real-time decoding performance was statistically equivalent to the offline benchmark.

B. Channel Ablation Analysis for Identifying Channel Contributions

Channel ablation analysis revealed that removing five electrodes (P3, P4, P8, O1, Oz) significantly reduced accuracy (paired t-test, Bonferroni corrected, $p < 0.05$) (Fig. 3A). These channels were located in the parietal and occipital regions, which are well known for generating P300 responses. Posterior electrodes generally reduced accuracy when ablated, underscoring their importance. Fig. 3B shows the topographical distribution of all 31 electrodes, with the five statistically significant channels highlighted. These findings indicate that the proposed P300 BCI system relies primarily on ERPs recorded over parietal and occipital areas and provide data-driven guidance for future optimization of low-density BCI configurations. While the zeroing-based analysis indicated larger drops over parieto–occipital electrodes (reflecting the fixed decoder's immediate sensitivity), the LOCO-R analysis showed that no single electrode yielded a statistically significant decrease in accuracy after multiple-comparison correction (Fig. S1), consistent with spatial redundancy among neighboring EEG sensors. Together, these findings support the practical conclusion that a compact posterior cluster is sufficient for high performance, rather than any single electrode being indispensable.

C. Data Sufficiency Simulation for an Effective Zero-Training System

To determine how much training data is required for reliable calibration-free P300 BCIs, we performed a data sufficiency simulation using a dataset of 37 participants. We varied the number of participants (1–37) and blocks per participant (10–270). Fig. 4A shows that classification accuracy increased nonlinearly with dataset size, with diminishing returns. The fitted inverse-saturation function (Section II-G) yielded coefficients $a = 0.1713$, $b = 0.0063$, and $c = 1.5525$, all significant (ps < 0.001), indicating contributions from both subject and block count. Fig. 4B shows iso-accuracy contours (0.70–0.85) and the trade-off between participants and blocks. With 30 or more participants, 0.85 accuracy was reached with only 10 blocks per subject. For example, the point on the 85% accuracy contour closest to the origin lies at 30 participants with 12 blocks. Along the same trajectory, 20 participants with 270 blocks also reach 85%. This suggests that the same target performance can be achieved by either increasing the number of participants or the number of blocks per participant, allowing researchers to select the strategy that best matches their experimental goals and resource constraints.

D. Error Trial Analysis

1) Prestimulus Attentional State: Alpha-Band Power: We examined whether prestimulus attention influenced errors by analyzing alpha-band power (8–14 Hz) during the baseline period (–200 ms to onset). Using a bootstrap approach (100,000 iterations) due to class imbalance, we found significantly lower alpha power in error trials than in correct trials at P4 ($p = 0.0094$), P8 ($p = 0.0116$), O1 ($p = 0.0094$), Oz ($p = 0.0226$), and O2 ($p = 0.0267$) (Fig. 5). This indicates that reduced alpha activity, reflecting diminished attentional readiness, increased error probability.

2) Target Presentation Order and Temporal Error Bias: We assessed whether target position (1st–4th) influenced error probability. Error rates were highest at the 4th position and lowest at the 2nd (Fig. 6A). Permutation tests confirmed significantly lower error rates at the 2nd and 3rd positions and a higher rate at the 4th (ps smaller than 0.001). A binomial GLMM with subject as a random effect showed that the 4th position had a significantly higher intercept ($p = 0.005$) and the 2nd a significantly lower one ($p = 0.002$) compared to chance (0.25) (Fig. 6B).

IV. Discussion

In this study, we developed and validated a single-trial, zero-training P300-based BCI system that combines a pre-trained xDAWN spatial filter with DeepConvNet. Unlike previous studies that relied solely on offline validation or required subject-specific calibration, our system demonstrated robust decoding performance in real-time, achieving 85.2% accuracy—comparable to the offline benchmark of 87.8%. The entire processing pipeline was designed to be computationally efficient and robust to noise, enabling low-latency operation without the need for individualized training. Together with a cognitively engaging stimulus paradigm and real-time interaction with IoT devices, these results mark a significant step toward practical, plug-and-play BCI systems that are scalable, user-independent, and ready for real-world deployment. Specifically, we believe this study is the first to validate a fully pre-trained P300-based BCI system that operates in a true single-trial setting under real-time conditions, without requiring stimulus repetition or subject-specific adaptation. This advances previous work that has either relied on repeated trials, offline validation, or limited online demonstrations.

A. Comparison to Existing Studies

Compared to prior work on zero-training P300-based BCIs, our study demonstrates one of the most robust systems to date in terms of decoding performance and participant scale.

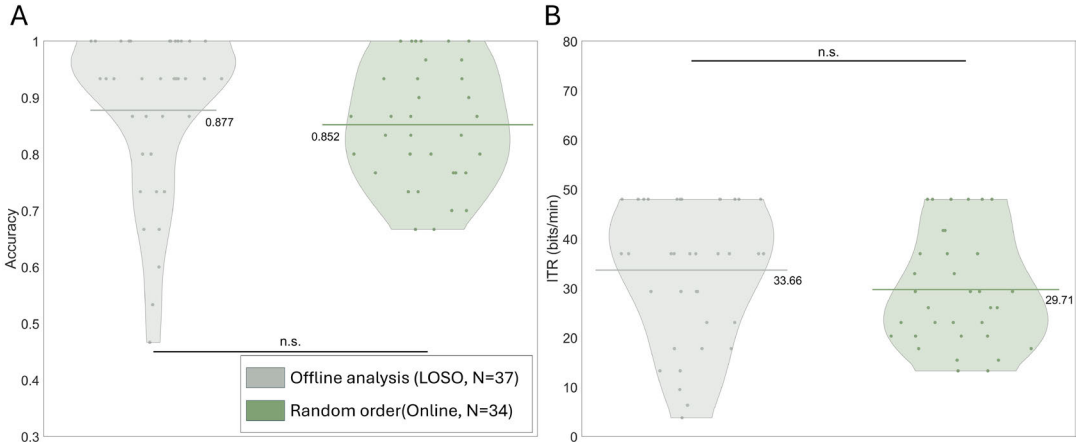

Fig. 2. Comparison of classification performance between offline benchmark and online random-order condition. A) Classification accuracy and B) information transfer rate (ITR) are presented for two conditions: zero-training offline analysis and online decoding under random-order stimulus presentation.

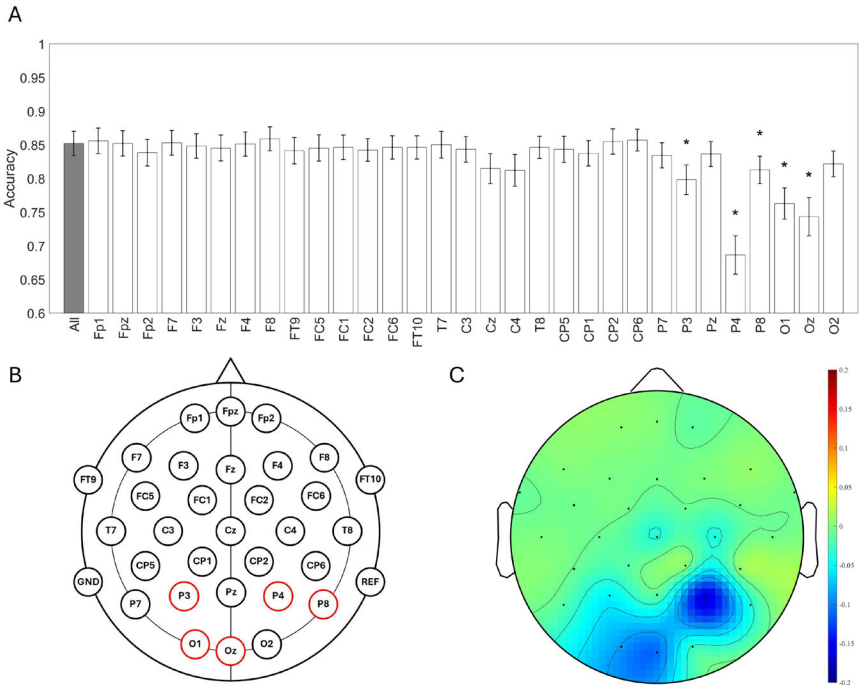

Fig. 3. Channel ablation analysis to identify electrode importance. A) Decoding accuracy using the full 31-channel setup (leftmost bar) is compared with accuracy from ablation conditions, where each individual channel was removed (set to zero) one at a time. Error bars indicate standard error. Asterisks mark electrodes whose removal caused a statistically significant performance drop after Bonferroni correction ($p < 0.0016$). B) Scalp map of the 31 EEG channels used in the study. Channels showing significant impact on accuracy (P3, P4, P8, O1, Oz) are highlighted in color. Results indicate that parietal and occipital electrodes contribute most critically to P300 classification. C) Topographic representation of the accuracy decrease across all channels. Color intensity indicates the magnitude of performance drop when each channel was ablated. This provides a spatial overview of each channel's contribution to classification accuracy, with larger drops (in blue) observed over parietal and occipital regions.

While several recent studies have pursued subject-independent approaches, few have validated them under a true plug-and-play scenario—i.e., applying a pre-trained model to entirely new users without any subject-specific calibration. Among the previous studies summarized in Table I, only Hu et al. [50] and our work performed online experiments using a fully pre-trained decoder. In contrast, Gao et al. incorporated a few calibration trials for adaptation [51], and Huang et al. conducted only an offline analysis with subject-specific ERP normalization [52]. These differences emphasize that true real-time plug-and-play validation remains rare.

In terms of methodology, all comparative studies employed CNN-based classifiers, but their network structures and training schemes differ substantially from ours. Hu et al. utilized temporal attention within a CNN to enhance generalization, but still relied on light calibration [50]. Gao et al. leveraged a CNN trained on a large-scale EEG dataset with transfer learning, but required brief subject-specific data [51]. Huang et al. adopted a CNN followed by a cosine similarity classifier for generalization across subjects; however, their work focused on wearable systems and lacked online validation [52]. In contrast, we employed a standard DeepConvNet architecture

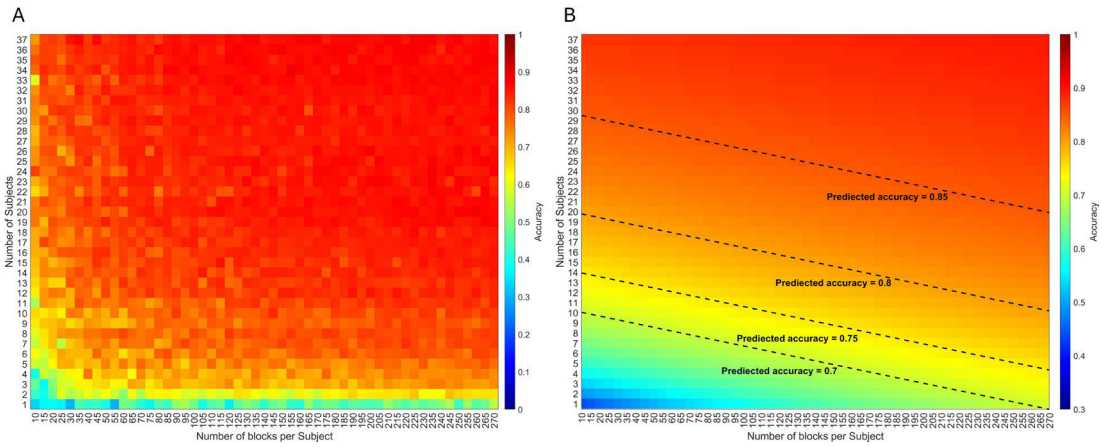

Fig. 4. Data sufficiency analysis for calibration-free model training. A) Classification accuracy as a function of the number of training participants and blocks per participant. Accuracy improves with more data, following the actual collection order. B) Predicted accuracy surface from an inverse regression model with overlaid contours indicating target accuracy levels (e.g., 0.70–0.85). Results show that using data from 30 or more subjects requires only ∼10 blocks per subject to achieve ≥0.85 accuracy, providing practical guidelines for dataset construction.

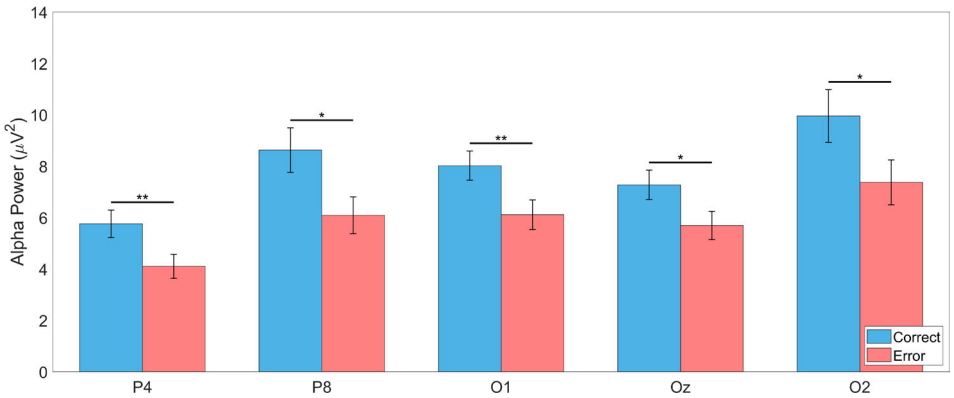

Fig. 5. Prestimulus alpha-band power in correct and error trials. Bootstrap-estimated mean and SEM of prestimulus alpha power (-200 to 0 ms) across five posterior electrodes: P4, P8, O1, Oz, and O2. Each pair of bars represents the average alpha power for correct and error trials based on 100,000 bootstrap samples, with error bars indicating standard error of the mean. Across all channels, alpha power was consistently higher in correct trials, suggesting that increased prestimulus alpha activity is associated with enhanced attentional readiness and improved BCI performance. Asterisks indicate channels with statistically significant differences ($p < 0.05$) where the 95% bootstrap confidence interval of the group difference did not include zero. Asterisks denote significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

integrated with pre-trained xDAWN spatial filters, used as-is during real-time testing without any user-specific adaptation or tuning. This simplicity in design reinforces the practicality of our pipeline while achieving competitive performance.

In our system, single-trial decoding was performed without repetition or user-specific tuning, yet achieving high accuracy (85.2%)—second among the compared works—and the highest ITR, owing to the single-trial stimulus design. We expect that the accuracy could be further improved if we increase the number of stimulus presentations, as demonstrated in our previous work Kim et al. Furthermore, our online BCI system was tested in 34 participants, exceeding the scale of other previous real-time studies. Collectively, our results underscore the practical feasibility and scalability of plug-and-play P300 BCIs.

TABLE I
COMPARISON OF CLASSIFICATION PERFORMANCE ACROSS ZERO-TRAINING BCI CONDITIONS

	Channel	Subject	Repetition	Online/ Offline	Accuracy (%)	ITR (bits/min)
Present study	31	34	1	Online	85.20	46.41
[50]	3	10	10	Online	73.23	19.85
[51]	30	20	10	Online	94.00	43.99
[52]	32	20	6	Offline	80.40	36.4

B. Implications of The Channel Ablation Analysis

The ablation analysis showed that parietal and occipital electrodes (e.g., P3, P4, P8, O1, Oz) contributed most to classification, consistent with prior findings that P300 responses are strongest in these regions [37], [53], [54]. This suggests that targeting posterior electrodes can maintain decoding performance while simplifying system design.

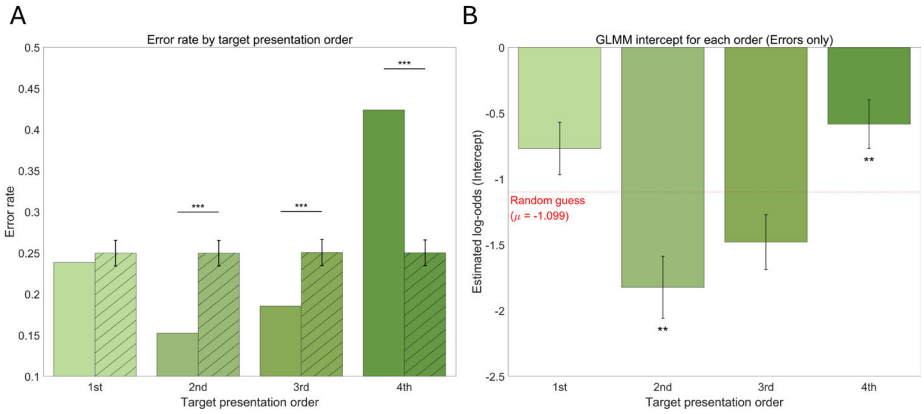

Fig. 6. Temporal position effects on classification errors in single-trial BCI. A) Error rates for each target presentation position (1st to 4th) are shown using solid bars for the observed data and hatched bars for the null distribution derived from 100,000 permutations assuming random error assignment. Observed bars represent fixed proportions from the full dataset and thus do not include error bars. For the null distribution, the mean and standard deviation across permutations are shown as bars and error bars, respectively. One-sided p-values were calculated by comparing the observed error rate to the permutation distribution: the proportion of permutations with greater values when the observed rate was higher, or with smaller values when it was lower. B) Intercepts from a binomial generalized linear mixed model (GLMM) fitted separately for each target position. The red dashed line indicates the theoretical log-odds under random guessing ($\mu = -1.099$). Intercepts for the 2nd and 3rd positions were significantly below chance, indicating improved detection accuracy, while the 4th position approached chance level. Asterisks denote significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Unlike previous studies that emphasized subject-specific optimization of EEG channel selection [55], [56], [57], our approach evaluates channel importance in a cross-subject, pre-trained computational framework. This generalizable approach aims to identify a consistent set of informative channels applicable across users, providing a foundation for constructing zero-training BCI systems with fixed, low-density electrode layouts. Our dual ablation results suggest that posterior channels are most informative to the deployed decoder (zeroing), yet individually replaceable after model re-optimization (LOCO-R). Hence, electrode selection should prioritize cluster-level coverage over single-sensor indispensability, aligning with low-density, wearable plug-and-play designs. Such configurations are critical for improving usability and enabling plug-and-play deployment in daily life.

C. Evaluation of Dataset Sufficiency and Model Generalization

The data sufficiency simulation showed the relationship between the training dataset size and expected classification performance. The inverse regression model predicted that with 30 or more participants, as few as 10 blocks per subject are sufficient to achieve an average accuracy above 0.85. Notably, our result indicates that collecting data from a large number of people would be more effective than collecting many data samples from a small number of people. Large-scale datasets are challenging to collect, and public EEG datasets were incompatible with our paradigm [58].

To validate the predictive model, we performed a bi-directional evaluation: the online dataset from 34 participants (random-order condition) was used to pre-train the model, which was then tested on the original 37-participant dataset. This yielded an average accuracy of 0.8198 (SD = 0.1490), compared with a regression-predicted value of 0.8547. Although the model slightly overestimated accuracy, the deviation remained moderate, supporting the validity of the predictive approach.

D. Target Presentation Order and Error Bias

The analysis of error trials revealed that target stimuli presented later—particularly in the fourth position—were more likely to result in classification errors. Post-experiment interviews with participants provided qualitative support, noting that the fixed temporal rhythm made the last stimulus predictable and less salient, reducing attentional engagement. Interestingly, the error rate for the second position was significantly lower than expected by chance, possibly because attention was still sustained and the stimulus retained novelty without being as abrupt as the first. These behavioral findings align with our neural analysis of prestimulus alpha-band activity. While previous studies have associated alpha desynchronization (i.e., reduced alpha power) with increased visual attention during stimulus processing [37], [59], [60], [61], [62], our study focused on the preparatory period before stimulus onset. Higher alpha power has been interpreted as a marker of top-down inhibition of irrelevant neural processing [47], [48], [49], and we found that alpha power preceding error trials was significantly lower, particularly at posterior sites. Taken together, these findings suggest that both randomness of target location and temporal unpredictability are critical for attentional readiness and robust P300 responses, underscoring the importance of optimizing stimulus timing and incorporating attentional state monitoring into future adaptive BCI systems.

E. ERP and Feature Representation Differences Between Fixed and Random Conditions

In addition to the main random-order condition, we performed an online experiment with a fixed-order stimulus presentation to address a practical concern of single-trial

BCIs: missing a single target leads to classification failure. We hypothesized that predictable sequencing might reduce lapses and tested 16 participants (mean age = 22.44, SD = 3.30). Although both conditions followed a four-class oddball paradigm, performance differed. In the fixed-order condition (consistent spatial sequence; Fig. 1D), accuracy was 0.7146 (SD = 0.1398) and ITR 28.40 bits/min (SD = 16.82). In contrast, the random-order condition yielded higher accuracy (0.8520, SD = 0.1055) and ITR (46.41 bits/min, SD = 17.77) with lower variability (Supplementary Fig. S2). A Kruskal–Wallis test with post hoc Tukey–Kramer confirmed significantly lower fixed-order performance ($p = 0.0124$). These findings suggest that temporal unpredictability preserves attention and BCI performance. To further examine neural differences, we analyzed grand-averaged ERPs at Fz, Cz, and Pz (Supplementary Fig. S3). Fixed-order ERPs were attenuated compared to random-order, and xDAWN-enhanced signals also showed weaker peaks and reduced target–nontarget differentiation (Supplementary Fig. S4). These results indicate that stimulus predictability may weaken attentional engagement and reduce target salience, impairing neural responses and classification.

It is important to note that the fixed-order condition was designed as a supplementary experiment, conducted after the main study, to test a practical hypothesis. The group sizes between the random-order ($n = 34$) and fixed-order ($n = 16$) conditions were intentionally imbalanced, as the fixed-order condition was not intended for primary performance evaluation. Accordingly, we employed non-parametric statistical comparisons (Wilcoxon rank-sum test) to account for this imbalance. This addition does not affect the interpretation of our main findings, which are based on the more robust random-order dataset.

F. Comparing Plug-and-Play Potential Across BCI Types

The present study demonstrates a plug-and-play implementation of a P300-based BCI system, leveraging spatial filtering and deep learning to achieve subject-independent performance. This general framework naturally raises the question of whether similar approaches could be extended to other major BCI paradigms, particularly SSVEP- and MI-based BCIs. SSVEP-based BCIs, like P300, fall under the category of reactive BCIs. SSVEP signals are relatively stable across individuals and typically require fewer occipital channels, relying on frequency-domain features that can be robustly extracted. This makes SSVEP-based systems good candidates for plug-and-play implementation, especially when a large pre-training dataset is available. In particular, our analysis revealed that having more subjects' data in the pre-training set appears to be more important than having more trials per subject (Fig. 4). In addition, the spatial filters were trained using diverse subjects' data, allowing the algorithms to generalize across individuals. During the DeepConvNet training, we introduced a validation set to ensure robustness to subject differences and prevent overfitting. These methodological components collectively contributed to a generalizable, subject-independent system. We believe that a similar approach can be extended to SSVEP-based BCIs.

G. Generalizability Beyond the Present Paradigm

Our zero-training model was trained and tested with the same finger-tapping animation paradigm, so its current generalizability is confined to that visual layout. While this confirms cross-subject robustness, it does not guarantee equal performance with markedly different stimuli or screen arrangements. Extending the approach is conceptually straightforward: by merging data from varied stimulus designs into the pre-training pool and retuning model parameters, a broader, paradigm-agnostic decoder could be built. Investigating such multi-paradigm training remains an important direction for future work.

H. User Fatigue and Cognitive Workload

Although the current study did not incorporate direct assessments of subjective fatigue or cognitive workload (e.g., NASA-TLX or attention probes), several factors suggest that participants were unlikely to experience substantial burden. The entire online session—including EEG cap setup and BCI operation—lasted approximately 10 minutes per participant. Moreover, the paradigm design was consistent with established P300-based BCI protocols and did not impose unusually high attentional demands. The stimulus presentation rate (400 ms per stimulus without an inter-stimulus interval) was comparable to or slower than that of other recent studies using single-trial paradigms. Nevertheless, we agree that incorporating standardized measures of user experience, such as subjective workload ratings or objective attention probes, would strengthen the interpretability and user-centered evaluation of plug-and-play BCI systems. We consider this an important direction for future work aimed at ensuring practical usability in real-world environments.

I. General Implications and Future Directions

Overall, our results demonstrate that zero-training, single-trial P300 BCI systems can operate effectively in real-time. The ability to train decoders on pre-existing data and decode new users' intentions without subject-specific calibration represents a meaningful advance in BCI usability. In addition, our error trial analysis indicated that prestimulus alpha-band power—reflecting preparatory attentional state—was significantly lower before incorrect classifications. This suggests that real-time monitoring of alpha power may enable online system adjustments (e.g., prompting the user to refocus or delaying stimulus onset) to mitigate attentional lapses. Furthermore, our comparison of fixed and random stimulus sequences revealed that temporal unpredictability enhances attentional engagement and neural discriminability. Together, these insights point to the future work of implementing adaptive timing mechanisms, such as randomized inter-stimulus intervals, to prevent user fatigue and maintain optimal attentional states throughout BCI operation. Future work should prioritize the scaling of pre-training datasets, refinement of low-channel configurations, and incorporation of real-time system adjustments. Our results contribute to the growing body of evidence that practical, generalizable BCIs are within reach and lay a foundation for their integration into everyday applications.

V. Conclusion

This study presents the real-time validation of a zero-training, single-trial P300-based BCI system under ecologically realistic conditions. Using a pre-trained xDAWN spatial filter and a deep convolutional neural network trained on independent users, we enabled plug-and-play BCI without subject-specific calibration. The online system achieved performance comparable to offline benchmarks, while also demonstrating practical real-time interaction with IoT devices. Unlike prior zero-training approaches that relied on repeated trials, offline validation, or limited online demonstrations, our system enables immediate, single-trial decoding without stimulus repetition. Key elements included randomized stimulus sequences, minimal but effective preprocessing, and a compact deep learning architecture for real-time deployment. Offline analyses further provided guidelines for training data sufficiency and showed that no single electrode was indispensable, with posterior clusters being sufficient for robust decoding. Although real-time feasibility was established, future work should address challenges such as refining end-to-end latency estimation, mitigating visual fatigue, and adapting the system to wearable and mobile EEG configurations. Together, these findings demonstrate the feasibility of real-time, zero-training P300 BCIs and offer strategies for robust, scalable, and deployable systems beyond laboratory settings.

References

[1] J. R. Wolpaw et al., "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] A. Finke, A. Lenhardt, and H. Ritter, "The MindGame: A P300-based brain–computer interface game," *Neural Netw.*, vol. 22, no. 9, pp. 1329–1333, Nov. 2009.

[3] K. M. Pitt and J. B. Boster, "Identifying P300 brain–computer interface training strategies for AAC in children: A focus group study," *Augmentative Alternative Commun.*, vol. 2025, pp. 1–10, Apr. 2025.

[4] M. A. Khan, R. Das, H. K. Iversen, and S. Puthusserypady, "Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103843.

[5] M. Kim, M.-K. Kim, M. Hwang, H.-Y. Kim, J. Cho, and S.-P. Kim, "Online home appliance control using EEG-based brain–computer interfaces," *Electronics*, vol. 8, no. 10, p. 1101, Sep. 2019.

[6] T. O. Zander, C. Kothe, S. Jatzev, and M. Gaertner, "Enhancing human–computer interaction with input from active and passive brain–computer interfaces," in *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*. Cham, Switzerland: Springer, 2010, pp. 181–199.

[7] T. O. Zander, "Enhancing human-machine systems with secondary input from passive brain-computer interfaces," in *Proc. 4th Int. BCI Workshop*, 2008, pp. 44–49.

[8] D. Zhu, J. Bieger, G. G. Molina, and R. M. Aarts, "A survey of stimulation methods used in SSVEP-based BCIs," *Comput. Intell. Neurosci.*, vol. 2010, pp. 1–12, Jan. 2010.

[9] J. N. Mak et al., "Optimizing the P300-based brain–computer interface: Current status, limitations and future directions," *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025003.

[10] Y. Gao, A. Ravi, and N. Jiang, "Does inter-stimulus distance influence the decoding performance of SSVEP and SSMVEP BCI?" in *Proc. 10th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2021, pp. 507–510.

[11] Y. Gao, A. Ravi, and N. Jiang, "Effect of competing stimuli for steady-state visually evoked potential and steady-state motion visually evoked potential," *IEEE Access*, vol. 9, pp. 129820–129829, 2021.

[12] E. P. Zambalde, L. R. Borges, G. Jablonski, M. B. De Almeida, and E. L. M. Naves, "Influence of stimuli spatial proximity on a SSVEP-based BCI performance," *IRBM*, vol. 43, no. 6, pp. 621–627, Dec. 2022.

[13] N. Zhang et al., "Retinotopic and topographic analyses with gaze restriction for steady-state visual evoked potentials," *Sci. Rep.*, vol. 9, no. 1, p. 4472, Mar. 2019.

[14] M. Nakanishi, Y. Wang, X. Chen, Y.-T. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, Jan. 2018.

[15] Y. Ke, P. Liu, X. An, X. Song, and D. Ming, "An online SSVEP-BCI system in an optical see-through augmented reality environment," *J. Neural Eng.*, vol. 17, no. 1, Feb. 2020, Art. no. 016066.

[16] P. Diez, L. Orosco, A. Garcés Correa, and L. Carmona, "Assessment of visual fatigue in SSVEP-based brain–computer interface: A comprehensive study," *Med. Biol. Eng. Comput.*, vol. 62, no. 5, pp. 1475–1490, May 2024.

[17] D. E. J. Linden, "The p300: Where in the brain is it produced and what does it tell us?" *Neuroscientist*, vol. 11, no. 6, pp. 563–576, Dec. 2005.

[18] A. Rezeika, M. Benda, P. Stawicki, F. Gembler, A. Saboor, and I. Volosyak, "Brain–computer interface spellers: A review," *Brain Sci.*, vol. 8, no. 4, p. 57, Mar. 2018.

[19] X. Duart, E. Quiles, F. Suay, N. Chio, E. García, and F. Morant, "Evaluating the effect of stimuli color and frequency on SSVEP," *Sensors*, vol. 21, no. 1, p. 117, Dec. 2020.

[20] R. Tello, S. Muller, T. Bastos, and A. Ferreira, "Evaluation of different stimuli color for an SSVEP-based BCI," in *Proc. XXIV Congr. Brasileiro Eng. Biomed. (CBEB)*, 2014, pp. 25–28.

[21] C. Reitelbach and K. Oyibo, "Optimal stimulus properties for steady-state visually evoked potential brain–computer interfaces: A scoping review," *Multimodal Technol. Interact.*, vol. 8, no. 2, p. 6, Jan. 2024.

[22] Y. B. Eisma, S. T. van Vliet, A. J. Nederveen, and J. C. F. De Winter, "Assessing the influence of visual stimulus properties on steady-state visually evoked potentials and pupil diameter," *Biomed. Phys. Eng. Exp.*, vol. 10, no. 6, Nov. 2024, Art. no. 065044.

[23] E. Schrag, D. Comaduran Marquez, A. Kirton, and E. Kinney-Lang, "An investigation into the comfort and neural response of textured visual stimuli in pediatric SSVEP-based BCI," *Sci. Rep.*, vol. 15, no. 1, p. 26168, Jul. 2025.

[24] J. R. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. London, U.K.: Oxford Univ. Press, 2012.

[25] K. Won, M. Kwon, M. Ahn, and S. C. Jun, "EEG dataset for RSVP and P300 speller brain–computer interfaces," *Sci. Data*, vol. 9, no. 1, p. 388, Jul. 2022.

[26] J. Cohen and J. Polich, "On the number of trials needed for P300," *Int. J. Psychophysiology*, vol. 25, no. 3, pp. 249–255, Apr. 1997.

[27] N. S. Artzi and O. Shriki, "An analysis of the accuracy of the P300 BCI," *Brain-Comput. Interfaces*, vol. 5, no. 4, pp. 112–120, Oct. 2018.

[28] G. Townsend et al., "A novel P300-based brain–computer interface stimulus presentation paradigm: Moving beyond rows and columns," *Clin. Neurophysiol.*, vol. 121, no. 7, pp. 1109–1120, Jul. 2010.

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

[30] H. Serby, E. Yom-Tov, and G. F. Inbar, "An improved P300-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 1, pp. 89–98, Mar. 2005.

[31] J. Polich, P. C. Ellerson, and J. Cohen, "P300, stimulus intensity, modality, and probability," *Int. J. Psychophysiology*, vol. 23, nos. 1–2, pp. 55–62, Aug. 1996.

[32] M. Kim, J. Kim, D. Heo, Y. Choi, T. Lee, and S.-P. Kim, "Effects of emotional stimulations on the online operation of a P300-based brain–computer interface," *Frontiers Hum. Neurosci.*, vol. 15, Feb. 2021, Art. no. 612777.

[33] F. Lotte et al., "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[34] F. Li, Y. Xia, F. Wang, D. Zhang, X. Li, and F. He, "Transfer learning algorithm of P300-EEG signal based on XDAWN spatial filter and Riemannian geometry classifier," *Appl. Sci.*, vol. 10, no. 5, p. 1804, Mar. 2020.

[35] S. Kundu and S. Ari, "MsCNN: A deep learning framework for P300-based brain–computer interface speller," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 1, pp. 86–93, Feb. 2020.

[36] P.-J. Kindermans, M. Schreuder, B. Schrauwen, K.-R. Müller, and M. Tangermann, "True zero-training brain–computer interfacing—An online study," *PLoS ONE*, vol. 9, no. 7, Jul. 2014, Art. no. e102504.

[37] J. Polich, "Updating p300: An integrative theory of P3a and P3b," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007.

[38] K. C. Squires, E. Donchin, R. I. Herning, and G. McCarthy, "On the influence of task relevance and stimulus probability on event-related-potential components," *Electroencephalogr. Clin. Neurophysiology*, vol. 42, no. 1, pp. 1–14, Jan. 1977.

[39] A. Kok, "On the utility of P3 amplitude as a measure of processing capacity," *Psychophysiology*, vol. 38, no. 3, pp. 557–577, May 2001.

[40] M. Eidel, M. Pfeiffer, P. Ziebell, and A. Kübler, "Recording the tactile P300 with the cEEGrid for potential use in a brain–computer interface," *Frontiers Human Neurosci.*, vol. 18, Jun. 2024, Art. no. 1371631.

[41] S. Ladouce, M. Pietzker, D. Manzey, and F. Dehais, "Evaluation of a headphones-fitted EEG system for the recording of auditory evoked potentials and mental workload assessment," *Behavioural Brain Res.*, vol. 460, Mar. 2024, Art. no. 114827.

[42] L. Vařeka and S. Ladouce, "Prediction of navigational decisions in the real-world: A visual P300 event-related potentials brain–computer interface," *Int. J. Hum.-Comput. Interact.*, vol. 37, no. 14, pp. 1375–1389, Aug. 2021.

[43] J. Kim, Y. S. Cho, and S.-P. Kim, "Task-relevant stimulus design improves P300-based brain–computer interfaces," *J. Neural Eng.*, vol. 21, no. 6, 2024, Art. no. 066046.

[44] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "XDAWN algorithm to enhance evoked potentials: Application to brain–computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2035–2043, Aug. 2009.

[45] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[46] B. Rivet, H. Cecotti, A. Souloumiac, E. Maby, and J. Mattout, "Theoretical analysis of xDAWN algorithm: Application to an efficient sensor selection in a p300 BCI," in *Proc. 19th Eur. Signal Process. Conf.*, Aug. 2011, pp. 1382–1386.

[47] O. Jensen and A. Mazaheri, "Shaping functional architecture by oscillatory alpha activity: Gating by inhibition," *Frontiers Human Neurosci.*, vol. 4, p. 186, Mar. 2010.

[48] M. S. Worden, J. J. Foxe, N. Wang, and G. V. Simpson, "Anticipatory biasing of visuospatial attention indexed by retinotopically specific α-bank electroencephalography increases over occipital cortex," *J. Neurosci.*, vol. 20, no. 6, p. RC63, Mar. 2000.

[49] P. Sauseng et al., "A shift of visual spatial attention is selectively associated with human EEG alpha activity," *Eur. J. Neurosci.*, vol. 22, no. 11, pp. 2917–2926, Dec. 2005.

[50] L. Hu et al., "Subject-independent wearable P300 brain–computer interface based on convolutional neural network and metric learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 3543–3553, 2024.

[51] W. Gao et al., "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.*, vol. 31, pp. 1754–1763, 2023.

[52] Z. Huang, J. Guo, W. Zheng, Y. Wu, Z. Lin, and H. Zheng, "A calibration-free approach to implementing P300-based brain–computer interface," *Cognit. Comput.*, vol. 14, no. 2, pp. 887–899, Mar. 2022.

[53] C. Bledowski et al., "Localizing P300 generators in visual target and distractor processing: A combined event-related potential and functional magnetic resonance imaging study," *J. Neurosci.*, vol. 24, no. 42, pp. 9353–9360, Oct. 2004.

[54] O. Bai, P. Lin, D. Huang, D.-Y. Fei, and M. K. Floeter, "Towards a user-friendly brain–computer interface: Initial tests in ALS and PLS patients," *Clin. Neurophysiol.*, vol. 121, no. 8, pp. 1293–1303, Aug. 2010.

[55] M. Xu et al., "Channel selection based on phase measurement in P300-based brain–computer interface," *PLoS ONE*, vol. 8, no. 4, Apr. 2013, Art. no. e60608.

[56] X. Zhao et al., "A regional smoothing block sparse Bayesian learning method with temporal correlation for channel selection in P300 speller," *Frontiers Hum. Neurosci.*, vol. 16, Jun. 2022, Art. no. 875851.

[57] M. T. McCann, D. E. Thompson, Z. H. Syed, and J. E. Huggins, "Electrode subset selection methods for an EEG-based P300 brain–computer interface," *Disab. Rehabil., Assistive Technol.*, vol. 10, no. 3, pp. 216–220, May 2015.

[58] J. Lee et al., "A comprehensive dataset for home appliance control using ERP-based BCIs with the application of inter-subject transfer learning," *Frontiers Hum. Neurosci.*, vol. 18, Feb. 2024, Art. no. 1320457.

[59] K. M. Spencer and J. Polich, "Poststimulus EEG spectral analysis and p300: Attention, task, and probability," *Psychophysiology*, vol. 36, no. 2, pp. 220–232, Mar. 1999.

[60] J. Yordanova, V. Kolev, and J. Polich, "P300 and alpha event-related desynchronization (ERD)," *Psychophysiology*, vol. 38, no. 1, pp. 143–152, Jan. 2001.

[61] P. Jasiukaitis and G. Hakerem, "The effect of prestimulus alpha activity on the P300," *Psychophysiology*, vol. 25, no. 2, pp. 157–165, Mar. 1988.

[62] W. Peng, L. Hu, Z. Zhang, and Y. Hu, "Causality in the association between P300 and alpha event-related desynchronization," *PLoS ONE*, vol. 7, no. 4, Apr. 2012, Art. no. e34163.