From lab to life: assessing the impact of real-world interactions on the operation of rapid serial visual presentation-based brain-computer interfaces

Muhammad Ahsan Awais*, Tomas Ward, Peter Redmond and Graham Healy

Insight SFI Research Centre for Data Analytics, School of Computing, Dublin City University, Dublin, Ireland
* Author to whom any correspondence should be addressed.

E-mail: muhammad.awais2@mail.dcu.ie

Keywords: electroencephalogram (EEG), brain-computer interface (BCI), artefacts, P300, event-related potential (ERP), rapid serial visual presentation (RSVP), noise

Abstract

Objective. Brain-computer interfaces (BCI) have been extensively researched in controlled lab settings where the P300 event-related potential (ERP), elicited in the rapid serial visual presentation (RSVP) paradigm, has shown promising potential. However, deploying BCIs outside of laboratory settings is challenging due to the presence of contaminating artifacts that often occur as a result of activities such as talking, head movements, and body movements. These artifacts can severely contaminate the measured EEG signals and consequently impede detection of the P300 ERP. Our goal is to assess the impact of these real-world noise factors on the performance of a RSVP-BCI, specifically focusing on single-trial P300 detection. *Approach.* In this study, we examine the impact of movement activity on the performance of a P300-based RSVP-BCI application designed to allow users to search images at high speed. Using machine learning, we assessed P300 detection performance using both EEG data captured in optimal recording conditions (e.g. where participants were instructed to refrain from moving) and a variety of conditions where the participant intentionally produced movements to contaminate the EEG recording. *Main results.* The results, presented as area under the receiver operating characteristic curve (ROC-AUC) scores, provide insight into the significant impact of noise on single-trial P300 detection. Notably, there is a reduction in classifier detection accuracy when intentionally contaminated RSVP trials are used for training and testing, when compared to using non-intentionally contaminated RSVP trials. *Significance.* Our findings underscore the necessity of addressing and mitigating noise in EEG recordings to facilitate the use of BCIs in real-world settings, thus extending the reach of EEG technology beyond the confines of the laboratory.

1. Introduction

Electroencephalography (EEG), a method for recording electrical activity from the brain using electrodes placed on the scalp with a conductive gel, is widely employed in brain-computer interfaces (BCIs) and is experiencing rapid growth due to its applications in both clinical and non-clinical settings (Peksa and Mamchur 2023). However, EEG data can be easily contaminated by various noise sources, including environmental factors such as electrical interference, as well as participant-related sources like body movements and oculomotor events. These sources of noise critically impact the quality of recorded EEG data, affecting the performance of BCI applications, and even affecting event-related potential (ERP) analysis. Addressing and mitigating such noise is challenging, yet a crucial step in applications and analysis that use EEG.

Measurement of ERPs can provide valuable insights into how the human brain processes information by capturing the electrical activity generated in response to specific events or stimuli (Luck 2014). ERPs represent the brain's electrical response to

cognitive processes, enabling researchers to investigate neural mechanisms involved in cognitive functions (Bressler and Ding 2006).

One of the most well-known and extensively studied ERPs is the P300 ERP, named for its positive polarity and typical latency of around 300 milliseconds post-stimulus (Pritchard 1981). The P300 component is associated with a wide range of cognitive processes, where its distinct characteristics have made it a crucial response for studying attention, memory, and decision-making. The P300 ERP is typically elicited when an individual detects a rare, unexpected, or salient event within a sequence of standard events (Luck 2014).

The P300 ERP holds significant importance in both cognitive psychology and clinical fields. Clinically, the P300 ERPs is valuable for diagnosing neurological and psychological disorders, such as Alzheimer's disease, schizophrenia, and attention deficit hyperactivity disorder (ADHD) (Li *et al* 2015). Beyond diagnosis, the P300 ERP has practical applications in brain-computer interfacing, allowing individuals with motor impairments to control devices through brain activity (Lee *et al* 2020). Additionally, the P300 component plays a role in lie detection and deception studies, contributing to the development of related technologies (Rosenfeld 2020, Bablani *et al* 2021).

The rapid serial visual presentation (RSVP) in BCI research enables the elicitation of the P300 ERP, whereby the P300's relationship to a user's attention and intent (Lees *et al* 2018, Wang *et al* 2018) can be used to control a BCI. In RSVP, a series of images are displayed at high speed where participants are asked to differentiate between a set of target images and a set of standard images. Due to infrequency of target images, the P300 ERP is evoked by their presentation but not by standard images (Zhang *et al* 2020).

While the RSVP-BCI paradigm can be demonstrated in controlled laboratory environments, translation of this paradigm into ecologically valid contexts requires a better understanding of real-world EEG artifacts that impact EEG signal quality e.g. in online worlds, metaverse, and gaming contexts. Such application scenarios are characterized by less constrained user behavior, where artifact-inducing movements may be frequent during normal interaction with a system. Examples include talking, head, and hand movements, all of which generate artifacts that impede the application of EEG in BCIs in real-world contexts.

The P300-RSVP task is significant in this study, as it provides a clear way to measure accuracy, serving as a valuable dependent variable to assess the effectiveness of noise cleaning and to examine the impact of artifacts when using EEG signals.

1.1. Related work and problem statement

BCIs could enable numerous opportunities in both clinical and non-clinical applications, however, deploying BCIs outside of controlled lab settings introduces challenges due to environmental noise and movement artifacts. Removing unwanted noise from EEG recordings would enable a broader application of BCIs in real-world settings, particularly in clinical and health-focused application spaces, as well as those in human-computer interaction including VR (Virtual Reality).

There is an extensive body of literature that has explored various approaches for noise mitigation in EEG signals within the domain of BCIs, where researchers have investigated challenges and advancements in efficiently filtering unwanted noise to improve the reliability and accuracy of EEG-based BCIs (Rashid *et al* 2020, Rashmi and Shantala 2022). Non-biological sources of noise, such as poor electrode settings, high impedances, and electrical interference, along with non-neural noise sources like eye movements, muscle activity, cardiac signals, and body movements, play a significant role in the contamination of EEG signals (Schmoigl-Tonis *et al* 2023).

Recognizing the importance of artifact removal in EEG-based BCIs, researchers employ a range of techniques, including both fundamental and advanced methods to effectively eliminate noise from EEG signals (Gorjan *et al* 2022).

Researchers have extensively utilized independent component analysis (ICA) to eliminate noise sources from EEG signals recorded in traditional lab environments, with particular emphasis on ocular and muscular artifacts. Dimigen (2020) proposed an artifact removal method based on ICA optimization by systematically varying a range of parameters, including high-pass and low-pass filters, and thresholding for component rejection. They reported a significant improvement in performance with minimal distortion of neural activity. Gao (2023) proposed a combination of canonical correlation analysis (CCA) and ICA to adaptively remove ocular artifacts from the EEG data. Liu *et al* (2018) explored the effect of ocular artifacts on EEG recordings during both rapid eye movement (REM) and non-REM sleeping stages. Shi *et al* (2024) proposed a denosing method by using CCA and second-order blind identification (SOBI) followed by adaptive and strict fixed thresholds to localize and remove artifacts. In addition, Tortora *et al* (2020) and Sburlea *et al* (2015) employed ICAs to eliminate various noise sources, including movement artifacts.

Moreover, the literature showcases other diverse approaches to noise mitigation in EEG signals. Specifically, Cimmino *et al* (2021) and Javed *et al* (2017) leveraged principal component analysis, Bajaj *et al* (2020, Narmada and Shukla (2023) and ter Braack *et al* (2013) applied regression methods and wavelet transforms, whereas Gandhi *et al* (2018, Hwaidi and Chen (2021, Zhang *et al* (2021) and Brophy *et al* (2022) employed GANs and CNNs for addressing various noise sources in EEG signals

recorded in traditional lab settings. However, a significant research gap remains regarding the management of EEG signals within real-world and out of the lab contexts.

In addition to noise removal in BCIs, a notable volume of research has been dedicated to the analysis of ERPs, specifically the P300 component, with a particular emphasis on utilizing algorithms for P300 analysis (Cui *et al* 2023, Sadras *et al* 2023, Yasemin *et al* 2023). Traditional analysis methods have predominantly relied on cleanly recorded EEG data collected within controlled laboratory settings that aims to minimize noise. Nevertheless, a significant research gap persists in the investigation of the P300 ERP within real-world conditions. In real-world conditions, EEG data collection can occur in environments where complex and uncontrolled scenarios can generate significant noise in the EEG recording that hinders detection and use of the P300 in applications such as BCIs. Mitigating this issue to allow effective detection of important ERP phenomena in EEG such as the P300 is not only essential for expanding our understanding of cognitive processes in naturalistic environments but also for the development of applications in fields such as human-computer interaction.

In order to measure and understand the impact of different sources of movement-related noise on P300 detection in a RSVP-BCI image search application, we assessed detection performance using both EEG data captured in clean recording conditions (i.e. where participants were instructed to withhold movements) and a variety of movement conditions where participants intentionally produced movements to contaminate the EEG recordings. We demonstrate variations in classification accuracy based on various systematically induced types of noise.

The primary objective of this research was to predict target versus non-target single-trial responses in a novel dataset, specifically, a RSVP-based P300 dataset (Awais *et al* 2023), which included both clean and intentionally induced noisy recording conditions. We have systematically examined the impact of noise and artifacts on detection performance, examining how this influences the accuracy of predictions. In addition, we also investigated the importance of using clean data for training machine learning models and exploring how utilizing contaminated data affects detection accuracy.

The key contributions of this study are as follows:

(i) Assessing the impact that different noisy recording conditions (i.e. body movement, talking, head movement) have on single-trial P300 detection performance.
(ii) Assessing the impact of noisy data when used to train machine learning models, and measuring this in terms of the performance of an RSVP-BCI using ROC-AUC.

The remainder of the paper is structured as follows. Section 2 describes the methodology, and covers important details of the dataset used in this paper, including preprocessing techniques, the machine learning approach, the evaluation strategy, and a description of the analysis procedure that was carried out. Following this, section 3 presents the experimental results and an in-depth discussion. Finally, section 4 concludes the paper by summarizing key findings and outlining directions for future work.

2. Methodology

In this section, we briefly describe the dataset used, the preprocessing steps carried out, the analysis techniques employed, and the metric used for evaluation and comparison.

Figure 1 illustrates the brain signal acquisition process for two distinct conditions: a controlled laboratory setting and a real-world like environment simulated by intentional EEG data contamination through different artifacts. During EEG data acquisition, participants engaged in an RSVP task presented on a computer screen. Following data acquisition, preprocessing procedures were applied, and subsequent analyses involved target versus non-target prediction to examine the influence of noise on classification performance. Further elaboration on these methodological aspects is provided in the following subsections.

2.1. Dataset

The dataset employed in this work is 'AMBER: Advancing Multimodal BCIs for Enhanced Robustness—A Dataset for Naturalistic Settings' (Awais *et al* 2023). Data collection took place at Dublin City University, Ireland, and the study included ten healthy participants within the age range of 20–35 years. Ethical approval for data acquisition was obtained from the Dublin City University Research Ethics Committee (DCUREC/2021/175). A 32-channel ANT-Neuro eego sports mobile EEG system was used for data collection, with electrodes positioned according to the 10–20 international electrode system. The CPz electrode was used as the online reference channel for data recording and the data was sampled at 1000 Hz.

The dataset was designed to imitate real-world and naturalistic settings, such as where a participant may be using a passive BCI while also performing physical movements. In this dataset, participants engaged in RSVP-BCI image search tasks while also executing various types of instructed physical movements so as to induce artifacts in the EEG. Each image search task block had a duration of 90 s, comprising a total of 360 images, of which 324 were standard trials (non-targets), and 36 were target trials. Figure 1 illustrates an example of target images (i.e. car) and

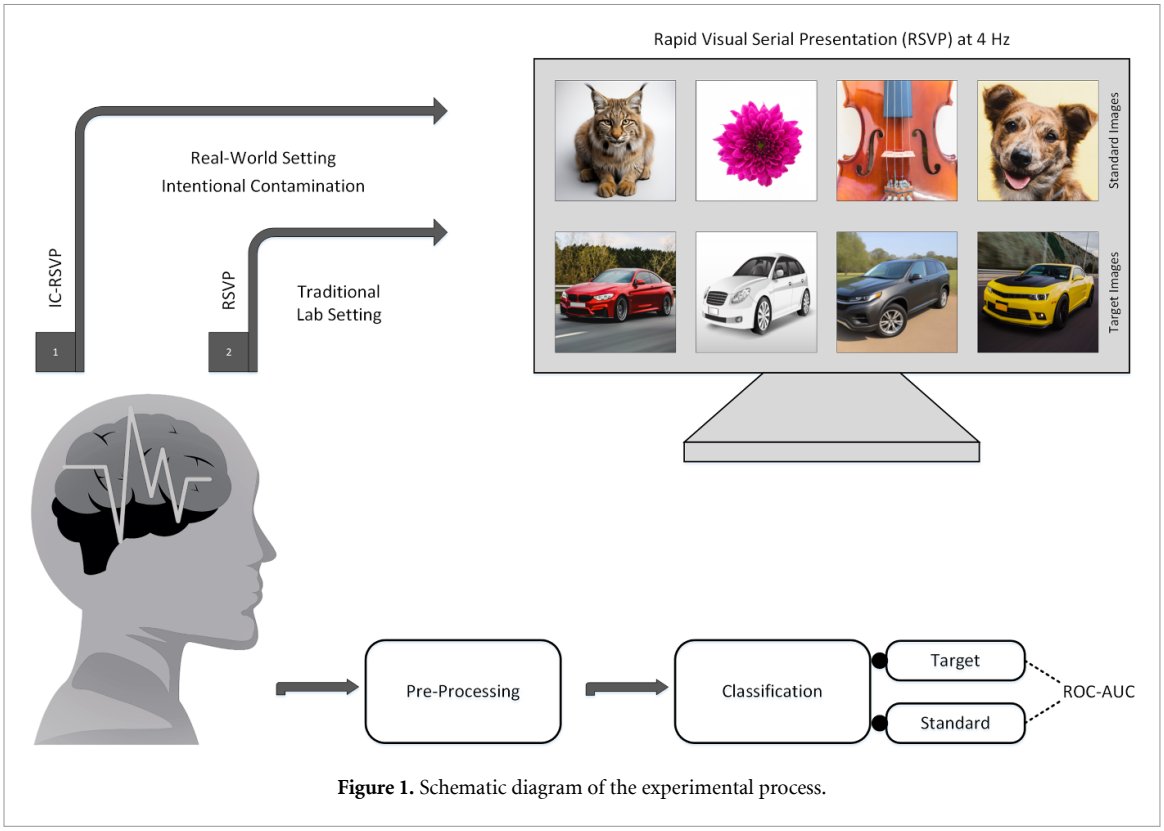

Figure 1. Schematic diagram of the experimental process.

non-target images (i.e. dog, flower, cat, etc) utilized in the RSVP task.

To establish a baseline, clean P300-RSVP tasks were conducted in a controlled manner, devoid of any intentionally induced noise or artifacts, with participants instructed to actively minimize noise contamination by sitting still in the chair, and these tasks are hereafter referred to as 'RSVP' (Traditional—RSVP) throughout the paper.

In contrast, RSVP-BCI image search tasks where participants were instructed to carry out physical activities that would contaminate the EEG with noise are hereafter referred to as 'IC-RSVP' (Intentionally Contaminated—RSVP).

The introduced artifacts in this dataset encompassed the following categories:

(i) Body movement: to induce artifacts from body movements, participants were instructed to repetitively raise and wave their hands, followed by lowering their hands, thereby creating a continuous sequence of movements.

(ii) Talking: to induce artifacts from talking, participants were instructed to audibly count aloud, incorporating a random mixture of numbers and letters within the sequences. The participants had the flexibility to choose the order and sequence of their counting.

(iii) Head movement: to induce artifacts from head movements, participants were instructed to move their heads in an up-down (nodding) or left-right motion at a natural pace throughout the entire block while maintaining gaze with the screen. Participants were instructed to alternate between up-down and left-right head movements.

The purpose of introducing these three distinct artifacts was to imitate real-world scenarios, where individuals engage in activities involving hand and arm movements, verbal communication, and spontaneous head gestures, such as nodding in agreement during conversations or shaking their heads in response to questions, such as in a virtual reality environment.

Each participant completed four sessions, with each session consisting of two 90 s traditional RSVP blocks i.e. without noise, along with one 90 s IC-RSVP block for each of the aforementioned artifact categories. Consequently, a total of eight RSVP blocks and four IC-RSVP blocks corresponding to each artifact type were obtained. The dataset for each participant comprises 2592 standard and 288 target trials for the RSVP type task, and 1296 standard and 144 target trials for each artifact category. This setup reflects a class imbalance, with different numbers of standard and target trials offering a comprehensive and dynamic resource for investigating neural processes in realistic and challenging scenarios.

2.2. Preprocessing

The preprocessing of EEG data is an essential step as it helps enhance the quality and interpretability of the

recorded signals (Awais *et al* 2021). It involves a series of procedures aimed at removing noise and preparing the data for subsequent analysis. In our study, Python along with the MNE library (version 1.6.0) was utilized for data preprocessing. The preprocessing steps in our study comprise the following steps:

(i) Digital filtering: to mitigate high-frequency noise, an Infinite Impulse Response (IIR) band-pass filter with cut-off frequencies of 0.1 Hz and 30 Hz was applied to the EEG data.
(ii) Re-sampling: the original data, initially sampled at 1000 Hz, was downsampled to 100 Hz. This reduction in sampling frequency allows for more manageable data processing while preserving critical temporal information.
(iii) Re-referencing: the EEG data was re-referenced from the initial CPz electrode to a common average reference, a technique that aids in reducing common-mode noise and enhancing signal quality (Ludwig *et al* 2009).
(iv) Epoching: the continuous EEG recordings were segmented into epochs by extracting time series data from −0.2 to 0.8 s relative to the onset of the visual stimulus. This step is crucial for aligning the data with specific events of interest, facilitating subsequent analysis focused on ERPs and other cognitive processes.

2.3. Classification and evaluation

This study did not intend to surpass any existing single-trial P300 detection models. Instead, the classifier was merely a tool to enable systematic evaluation, where the primary objective of this research was to comprehensively examine the impact of noise and artifacts on classification performance for a RSVP-P300 dataset.

2.3.1. Bayesian ridge regression

Bayesian machine learning approaches have gained widespread acceptance among researchers in the field of BCIs due to their effectiveness in addressing classification challenges (Cecotti and Ries 2017, Katyal and Singla 2021, Pan *et al* 2022, Chang *et al* 2023).

Bayesian Ridge Regression is a machine learning approach that combines Bayesian principles with ridge regression shrinks the coefficients to achieve greater numerical stability, resulting in improved computational accuracy (Han *et al* 2022). It is particularly useful when dealing with datasets that have properties such as multicollinearity, i.e. where the input features are highly correlated (Liu *et al* 2020).

In traditional regression analysis, we estimate model parameters based solely on the observed data. However, Bayesian ridge regression incorporates prior information about the parameters into the model, allowing more informed estimates. It assigns prior distributions to coefficients and combines them with a likelihood function to yield a posterior distribution through Bayes' theorem. It then estimates parameters based on this posterior distribution, and introduce regularization to prevent overfitting (Shi *et al* 2016, Ren and Zhou 2023).

In this paper, we used the Bayesian Ridge algorithm (MacKay 1992, Tipping 2001) for single-trial P300 detection as similar to Kindermans *et al* (2014, Wang *et al* (2021, Rahman *et al* (2022), for both IC-RSVP (noisy) and RSVP (clean) conditions using the AMBER dataset. Scikit-learn (Pedregosa 2011) version 1.3.2 was used for model training, hyperparameter tuning and evaluation.

The default configuration for Bayesian ridge was employed. RandomizedSearchCV, a random search approach, was utilized to tune two alpha and two lambda parameters, optimizing the model's performance.

2.3.2. Receiver operating characteristic—area under curve

Model performance was assessed using ROC-AUC (Receiver Operating Characteristic—Area Under Curve) (Fawcett 2006). This metric enables measurement of a model's discriminative ability.

ROC-AUC is insensitive to class imbalances, ensuring that the performance evaluation is not skewed by the dominance of one class over the other, making it particularly useful in scenarios where ranking or prioritization is a key objective. For RSVP EEG research, ROC-AUC is a commonly employed evaluation metric (Wang *et al* 2018, Lees *et al* 2019, Liu *et al* 2020, Cui *et al* 2022).

2.4. Analysis procedure

In this study, P300-based RSVP trials were categorized into two types: RSVP (i.e. traditional clean RSVP) and IC-RSVP (i.e. RSVP trials subjected to intentional artifact production using body movements, talking, and head movements).

The data (both traditional RSVP and intentionally contaminated IC-RSVP) were each further categorized into 'clean' and 'bad' on the basis of peak-to-peak thresholding, following the recommendation by Luck (2014), where a threshold between 50 μV and 200 μV is commonly employed for ERP analysis. This categorization was made to investigate the impact of using clean, bad, or a combination of both types of data (clean + bad) on training and testing, ultimately influencing the classification performance. We applied a 100 μV threshold, consistent with various studies (Moretti *et al* 2003, Barbey 2022, Yip *et al* 2022, Kabbara *et al* 2023).

As our study explores P300 detection to understand how noise affects detection performance, a classifier was trained using clean trials from traditional RSVP, serving as a baseline for understanding the influence of noise on prediction outcomes. After training the model on clean trials, we used the

classifier to test on both bad trials and a combination of clean and bad trials from the same RSVP dataset. In the case of artifact-laden RSVPs (i.e. IC-RSVP), predictions were conducted on all trials (clean and bad), only bad trials, and only clean trials for each of the three different artifact-producing behaviors (body movement, talking, and head movement).

An overview of the total number of trials, the percentage of retained trials (hereafter, referred to as Rt%), and the percentage of dropped trials (hereafter, referred to as Dr%) resulting from the rejection procedure using the peak–peak threshold (as presented in table 1) is crucial in understanding the size of the dataset used for training and testing the model. For instance, in subject 1, out of a total of 2880 trials, 73.30% (2111 trials) were retained as clean trials, while 26.70% (769 trials) were identified as bad trials. With a sampling rate of 100 and an epoch size of 1 s across 32 channels, the model was trained on the clean trials, equivalent to a dataset size of 2111×3200, and tested on the bad trials, amounting to 769×3200.

Next, to highlight the significance of training models on clean trials from traditional RSVP and to assess how detection accuracies are affected when intentionally contaminated RSVPs are used for training, we conducted a comparative analysis. This involved employing different strategies to train and test models on both RSVP and IC-RSVP datasets, aiming to evaluate detection performance under various conditions.

(i) **Condition 1:** training and testing on clean trials from the RSVP dataset i.e. **train** (*clean* RSVP) and **test** (*clean* RSVP). Additionally, for comparative analysis, training on clean trials from the IC-RSVP dataset was performed, and the testing was executed using the same subset from RSVP i.e. **train** (*clean* IC-RSVP) and **test** (*clean* RSVP).
In IC-RSVP, 'clean trials' refer to those that are below the peak-to-peak rejection threshold, and not necessarily completely free from contamination.

(ii) **Condition 2:** training and testing on the combination of clean+bad trials from the RSVP dataset i.e. **train** (*clean+bad* RSVP) and **test** (*clean+bad* RSVP)[2]. Additionally, for comparative analysis, training on the combination of clean+bad trials from the IC-RSVP dataset was performed, and the testing was executed using the same subset from RSVP i.e. **train** (*clean+bad* IC-RSVP) and **test** (*clean+bad* RSVP).

(iii) **Condition 3:** training on clean trials from the IC-RSVP dataset and testing it on the subset within the same category i.e. **train** (*clean* IC-RSVP) and **test** (*clean* IC-RSVP). Additionally, to explore the impact of bad trials within intentionally contaminated RSVP, training on a combination of clean and bad trials from the IC-RSVP dataset was performed and the testing phase was then executed on the subset within the same category i.e. **train** (*clean+bad* IC-RSVP) and **test** (*clean+bad* IC-RSVP).

[2] Although participants were instructed to minimize behaviors that would induce noise in the EEG recording during the RSVP condition, some trials containing artifacts were still present.

The objective was to assess the impact concerning choices when training models on various trial combinations from intentionally contaminated RSVPs and traditional RSVPs. We maintained consistency across all training models by using an equal number of trials for training (1080 trials) and the remaining trials for testing purposes. Models were trained and evaluated on a participant-per-participant basis.

Given the utilization of a 100 μV threshold for trial rejection, we further analyzed it by systematically varying the threshold from 50 μV to 500 μV. This analysis serves as our concluding exploration where the objective was to observe the impact of threshold variations on the percentage of rejected trials, as this factor significantly influences the training and testing phases of the classification.

3. Results and discussion

Figure 2 presents target versus standard averaged epochs for two subjects (i.e. 4 and 8) at channel Pz. The first row shows examples from subject 4, while the second row shows data from subject 8. The first column shows averaged epochs in a traditional lab setting, while the subsequent three columns shows averaged epochs from RSVP contaminated with talking, body movements, and head movements, respectively.

Figure 3 shows grand averages across all subjects at channel Pz and compares the traditional RSVP with intentionally contaminated RSVP, illustrating all artifacts combined, as well as each artifact separately, all within a single figure. In this representation, dotted lines represent the grand average of standard epochs, while the solid lines represents the grand average of target epochs. The high amplitude of noise caused by head movement made it impractical to display the comparison in the combined figure. Therefore, the IC-RSVP (head movements) condition was scaled down to fit within the comparison.

In order to visually analyze it further in depth, figure 4 presents butterfly plots using all the electrode channels instead of only Pz as used in figures 2 and 3. The butterfly plots show the difference between the grand averages of target epochs and standard epochs across all blocks. In addition, topographic analysis has also been made part of the butterfly plot for better understanding. Characteristic P3b activity can be seen at posterior scalp sites approximately between

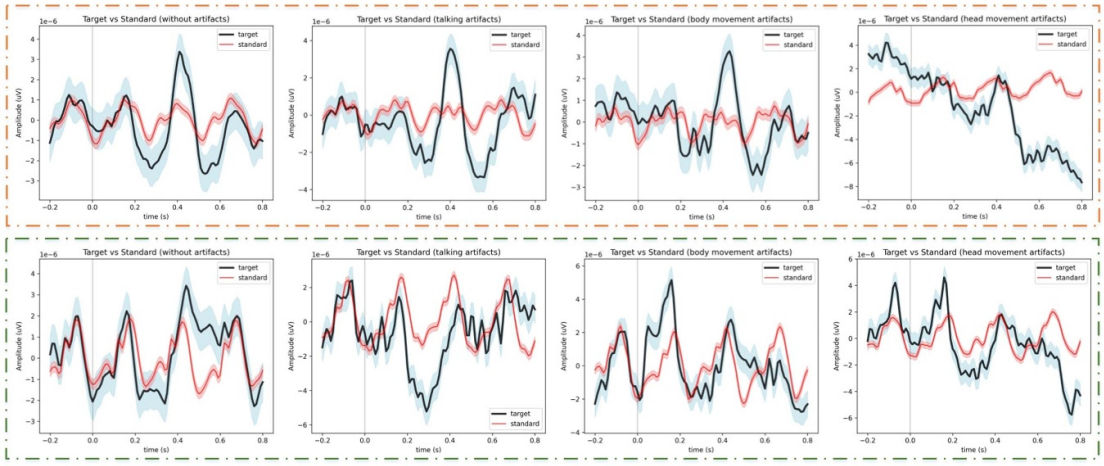

Figure 2. Visualization of averaged epochs at channel Pz for RSVP, talking, body movement and head movement conditions (left to right). The first row shows examples from subject 4, while the second row shows data from subject 8.

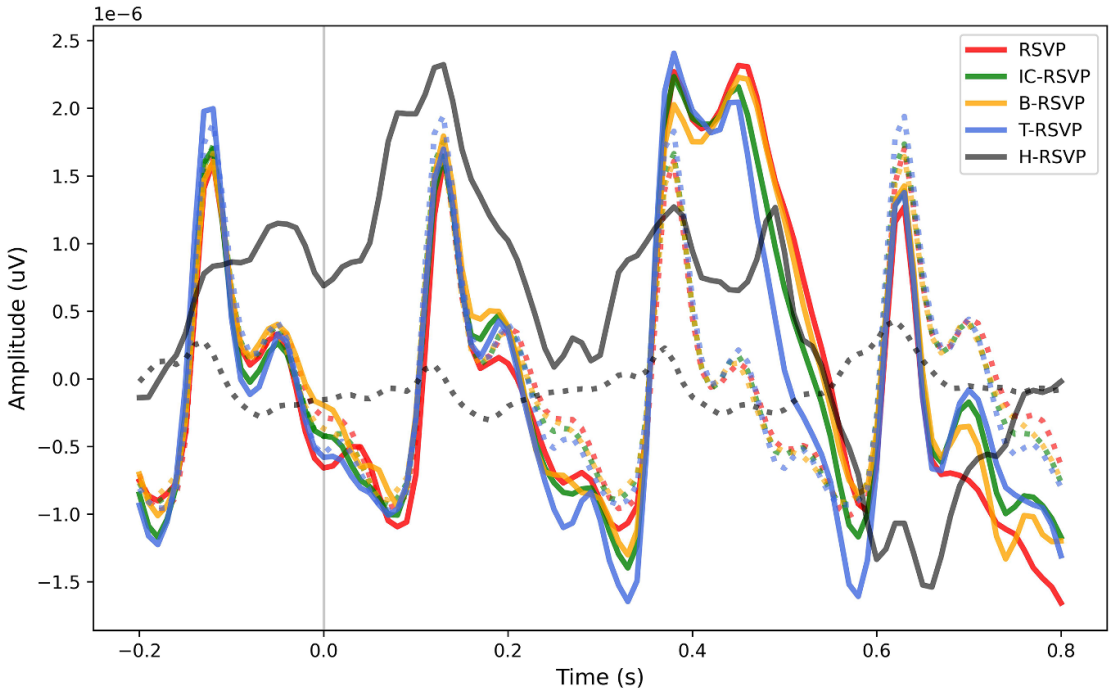

Figure 3. Target vs Standard grand average comparisons under different conditions at channel Pz. The solid lines on the plot represents target epochs whereas the dotted refers to the standard epochs. *IC—intentionally contaminated, T—talking, B—body movement, H—head movement.*

300 ms and 600 ms following target detection, however, it cannot be as readily identified in the topographic maps of intentionally contaminated RSVPs.

From these figures, it becomes clear that studying the effect of artifacts on P300 detection solely through visual inspection of the averaged epochs is challenging. Nevertheless, the impact is clearly visible in topographic plots related to the intentionally contaminated RSVPs specifically of the head movement artifacts i.e. figure 4.

3.1. Bad trials rejection

The initial step in our analysis involved rejecting bad trials using a 100 μV peak-to-peak threshold. Table 1 details the outcomes of the bad trials rejection process, providing information on total trials, the percentage of retained trials after rejection (i.e. clean trials), and the percentage of dropped trials (i.e. bad trials). In table 1, the percentage of retained trials is referred to as Rt% and the percentage of dropped trials is referred to as Dr%.

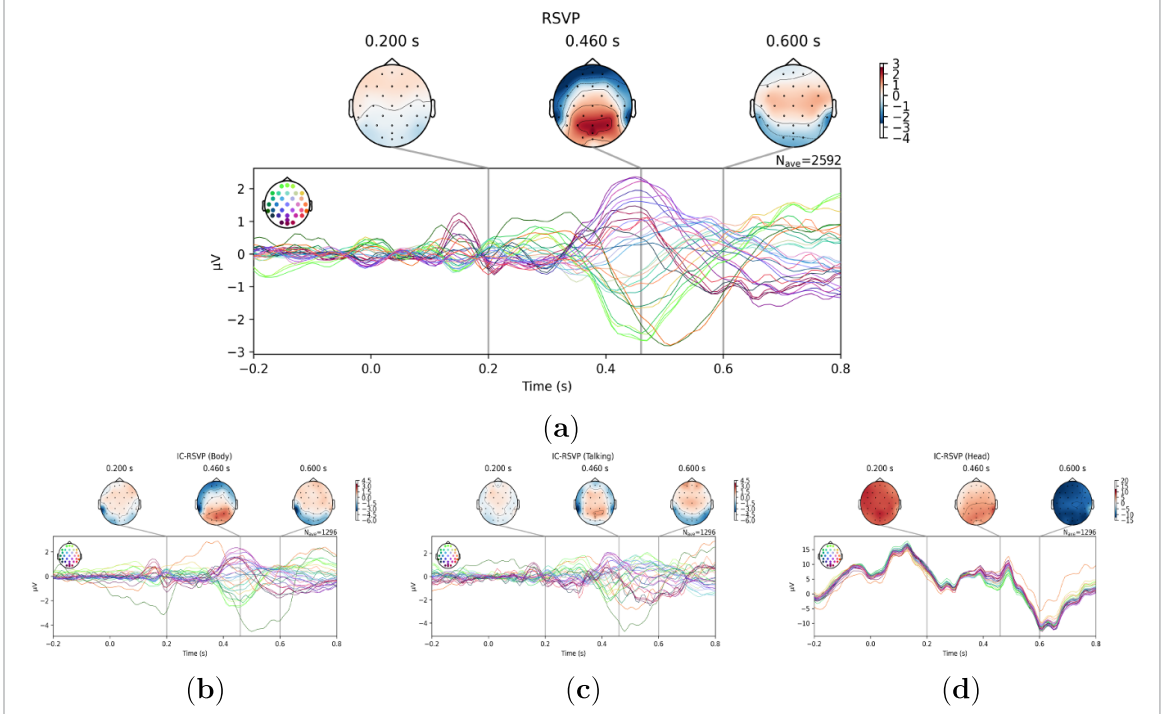

Figure 4. Butterfly plot (ERP averages) of target epochs minus average standard epochs across all blocks. The colors on time-series plots refers to the electrode locations on scalp. (a) Traditional RSVP (b) IC-RSVP (body movement) (c) IC-RSVP (talking) (d) IC-RSVP (head movement).

For the RSVP condition, a total of 2880 trials were recorded. In contrast for IC-RSVP, each artifact category for each participant had 1440 trials. Table 1 displays subject-wise trial rejections, revealing that subject 1 had the highest number of trials rejected for RSVP and IC-RSVP with body movement. However, for talking, subject 6 exhibited the most rejections, while for the head movement category, subjects 4 and 10 had all trials rejected.

IC-RSVP with head movement showed the highest trial rejections, averaging 90.04% across all 10 subjects. Only one subject retained 61%, and another retained approximately 28% of the trials, while the remaining subjects had over 90% of trials rejected. In contrast, the RSVP condition, owing to its non-artefactual nature, had the fewest trials rejected, averaging 91% retention. For IC-RSVP with body movement and talking, an average of 72.7% and 71.3% of trials were retained, respectively.

These results highlight the substantial impact of artifact-producing movement, particularly head movement, on the quality of trials. The observed trial rejection rates highlight the challenges posed by such artifacts and emphasize the need for robust preprocessing techniques in P300-based studies involving real-world scenarios.

3.2. P300 detection

After performing trial rejection to categorize trials into clean and bad, single-trial P300 detection was carried out by training only on clean trials from RSVP, as it serves as a baseline for understanding the influence of noise on prediction outcomes. The results, presented in table 2 as ROC-AUC scores, provide an overview of detection performance under various conditions.

Considering the percentage of dropped trials outlined in table 1, it is noteworthy that for some subjects, prediction was not feasible using clean trials due to insufficient data for testing after trial rejection. Specifically, clean trial predictions for body movement artifacts were unavailable for Subject 1, and sufficient testing data for talking artifacts was not available (i.e. dropped trials $>$ 80%) for Subject 6. Notably, head movement IC-RSVP tasks posed additional challenges, limiting predictions to only two subjects (i.e. 2 and 9). This highlights the intricacies of conducting accurate predictions in real-world scenarios, where noise and artifacts can significantly impact data availability and model performance.

When tested on a combination of clean and bad trials from RSVP, the model exhibited an average ROC-AUC score of 0.85, reflecting its ability to distinguish target and non-target stimuli effectively. However, when subjected to predictions exclusively on bad trials within the RSVP, the average ROC-AUC score decreased to 0.69, highlighting the impact of noise on detection accuracy, even in controlled settings.

Expanding our analysis to trials from artifact-laden RSVP tasks (i.e. IC-RSVP), predictions on clean trials from body movement artifacts demonstrated an

Table 1. Trial rejection using peak-to-peak threshold of 100 μV.

				IC-RSVP						
	RSVP				Body Movement		Talking		Head Movement	
Subject	Total Trials	Trials Rt (%)	Trials Dr (%)	Total Trials	Trials Rt (%)	Trials Dr (%)	Trials Rt (%)	Trials Dr (%)	Trials Rt (%)	Trials Dr (%)
1	2880	73.30	26.70	1440	13.61	86.39	53.96	46.04	1.25	98.75
2	2880	95.66	4.34	1440	91.25	8.75	94.93	5.07	60.97	39.03
3	2880	77.26	22.74	1440	55.69	44.31	81.53	18.47	0.14	99.86
4	2880	87.81	12.19	1440	90.76	9.24	95.90	4.10	0	100
5	2880	99.31	0.69	1440	98.75	1.25	95.35	4.65	0.42	99.58
6	2880	97.05	2.95	1440	71.81	28.19	15.07	84.93	0.07	99.93
7	2880	90.90	9.10	1440	44.65	55.35	86.60	13.40	5.97	94.03
8	2880	77.85	22.15	1440	68.19	31.81	58.06	41.94	2.92	97.08
9	2880	97.95	2.05	1440	97.15	2.85	98.33	1.67	27.85	72.15
10	2880	93.58	6.42	1440	95.07	4.93	33.26	66.74	0	100
Average	2880	89.07	10.93	1440	72.69	27.31	71.30	28.70	9.95	90.04

Trials Rt (%)—Percentage of trials retained, and Trials Dr (%)—Percentage of trials dropped.
RSVP—traditional RSVP, and IC-RSVP—intentionally contaminated RSVP.

Table 2. ROC-AUC scores. The model was trained on traditional RSVP (clean trials) and tested across different conditions from RSVP and IC-RSVP[a].

			IC-RSVP								
	RSVP		Body Movement			Talking			Head Movement		
Subject	Clean+Bad	Bad	Clean+Bad	Clean	Bad	Clean+Bad	Clean	Bad	Clean+Bad	Clean	Bad
1	0.66	0.61	0.62	[b]	0.62	0.68	0.74	0.64	0.53	[b]	0.53
2	0.94	0.68	0.90	0.92	0.69	0.88	0.89	0.69	0.85	0.87	0.81
3	0.77	0.69	0.79	0.84	0.75	0.83	0.87	0.67	0.57	[b]	0.57
4	0.83	0.77	0.87	0.89	0.72	0.87	0.89	0.46	0.62	[b]	0.62
5	0.91	0.92	0.91	0.92	0.71	0.88	0.88	0.83	0.57	[b]	0.57
6	0.89	0.69	0.77	0.79	0.73	0.76	[b]	0.76	0.63	[b]	0.63
7	0.85	0.62	0.73	0.85	0.65	0.79	0.80	0.73	0.58	[b]	0.58
8	0.77	0.56	0.74	0.79	0.66	0.71	0.75	0.66	0.66	[b]	0.65
9	0.89	0.62	0.90	0.90	0.71	0.90	0.90	0.57	0.74	0.76	0.74
10	0.95	0.69	0.85	0.86	0.66	0.75	0.89	0.70	0.56	[b]	0.56
Average	0.85	0.69	0.81	0.86	0.69	0.81	0.85	0.67	0.63	0.82	0.63

[a] Given that the model was exclusively trained on clean trials from traditional RSVP, the table does not include results for the testing scenario using clean trials from RSVP.

[b] There were insufficient clean trials available to perform the prediction.

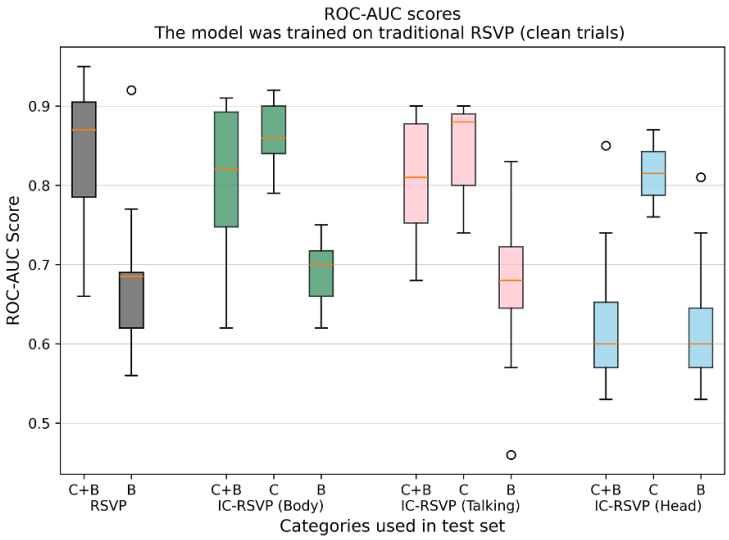

Figure 5. Box plots of ROC-AUC scores across subjects obtained when the model was trained on the clean trials from traditional RSVP. *C—Clean trials, B—Bad trials and C+B—Clean+Bad trials.*

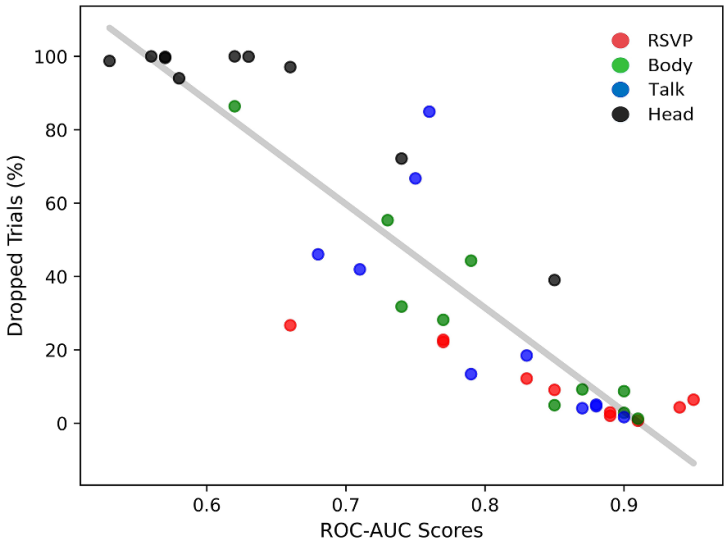

Figure 6. Correlation between the percentage of dropped trials and ROC-AUC scores. for each subject, model was trained on clean trials for traditional RSVP and tested on the combination of clean and bad trials, from both traditional and intentionally contaminated RSVPs.

average ROC-AUC score of 0.86. However, introducing bad trials alongside clean trials in the same category slightly reduced the average ROC-AUC score to 0.81, indicating the challenges associated with mitigating noise effects during predictions in real-world scenarios. Similarly, for talking artifacts, the model achieved an average ROC-AUC score of 0.85 when tested on clean trials. Yet, introducing bad trials alongside clean trials in the same category led to a marginal decrease in the average ROC-AUC score to 0.81, which highlights the importance of addressing verbal interactions as potential noise sources in P300 detection.

The analysis of head movement artifact IC-RSVPs revealed distinct challenges. Predictions on clean trials achieved an average ROC-AUC score of 0.82 for the limited subset of subjects with sufficient data. However, when considering bad trials from all subjects, the average ROC-AUC score dropped to 0.63, emphasizing the challenges in handling intentional head movements as noise sources.

In figure 5, boxplots are use to visually compare ROC-AUC scores obtained when the model trained on clean trials from traditional RSVP is tested across various conditions.

The percentage of dropped trials, as outlined in table 1, significantly impacts detection performance, as reflected in the ROC-AUC scores in table 2. Figure 6 illustrates the inverse relationship between the percentage of dropped trials and the ROC-AUC scores

achieved when tested with a combination of clean and bad trials from each category. A higher percentage of dropped trials results in lower ROC-AUC scores, and vice versa. The red color represents the ROC-AUC scores for traditional RSVP, while green, blue, and black represent the scores for body movement, talking, and head movement RSVPs, respectively.

To statistically assess this correlative relationship, we calculated the Pearson correlation coefficient and p-value between the percentage of dropped trials and ROC-AUC scores showing a significant negative correlation resulting in ($r = -0.907, p = 7.924 \times 10^{-16}$).

From table 1, we know that the percentage of dropped trials varied for each subject, leading to different training sizes. To account for this, we standardized the training size to 1595 trials, corresponding to the maximum number of dropped trials for one subject. We then performed predictions again, and using these predictions and the percentage of dropped trials, we re-calculated the Pearson correlation coefficient and p-value ($r = -0.911, p = 3.738 \times 10^{-16}$).

3.3. Impact of bad trials on training and testing evaluation

To emphasize the importance of training models on clean trials from RSVP, we conducted a comparative analysis. The outcomes of the investigation into various conditions, encompassing both RSVP and IC-RSVP used for training and testing, are presented in table 3.

The observations for each training and testing setup are detailed below:

(i) **Condition 1** involved training the model with clean trials from RSVP and testing it on a subset from the same category, resulting in an average ROC-AUC score of 0.84. To investigate the impact of noise in training data, we trained the model on clean trials from IC-RSVP [3] and tested it on the clean trials from RSVP used in Condition 1. However, due to the insufficient number of trials for training in all subjects except subjects 2 and 9, the analysis could not be extended. For these two subjects, the average ROC-AUC score was 0.89. This analysis highlights the significance of having a sufficient amount of clean data for successful model training.

(ii) **Condition 2** encompassed training the model with a combination of clean and bad trials from RSVP and testing it on the same category, resulting in an average ROC-AUC score of 0.82. To explore the influence of noise in the training dataset, the model was then trained on clean and bad trials from IC-RSVP and tested on the clean trials from RSVP used in Condition 2. The average ROC-AUC score dropped to 0.76, showing a decline in performance.

(iii) **Condition 3**, where only clean trials from IC-RSVP were used for training, resulted in an average ROC-AUC score of 0.90, derived from results obtained from only two subjects. However, the analysis could not be extended to the remaining eight subjects due to insufficient clean trials, the majority of which were rejected during the bad trials removal process with a 100 μV peak-to-peak threshold.
However, the utilization of combined clean and bad trials from IC-RSVP for training, and testing on the subset from the same category, obtained an average ROC-AUC score of 0.74. This highlights the challenges associated with training models on artifactual data.

This finding aligns with our earlier discussions on the challenges posed by intentional artifacts. Connecting this analysis to the previous section where we exclusively trained the model on clean trials from RSVP, the comparison underlines the importance of using non-artifactual data for model training. The performance deterioration when the model was trained on data subjected to intentional artifacts, as compared to accuracy measured by training the model on non-intentionally contaminated data, signifies the importance of having clean and noise-free data for training purposes.

Figure 7 displays a graphical comparison of ROC-AUC scores derived from the three aforementioned conditions as boxplots.

3.4. Trials rejection using different peak-to-peak thresholds

Following the bad trial rejection at the 100 μV threshold and the subsequent prediction analysis performed using this rejection threshold, an extended version of table 1, denoted as table 4, has been included. This table provides insights into the percentage of dropped trials, presenting both average and standard deviation values. We systematically varied the rejection thresholds from 50 μV to 500 μV to cover the range commonly applied in related studies, with particular emphasis on the frequently used 100 μV threshold.

While our study did not set out to determine the optimal threshold for bad trial rejection, our focus was on utilizing a previously established peak-to-peak threshold, namely 100 μV, to conduct our analysis. The primary objective was to highlight the influence of artifacts on prediction accuracies, emphasizing the necessity of addressing these common types of noises that occur in real-world conditions. Our aim is rooted in the broader goal of ensuring EEG data integrity beyond controlled laboratory environments.

[3] In the context of intentionally contaminated RSVP, 'clean' trials refer to those below the peak-to-peak rejection threshold, rather than truly clean trials.

Table 3. ROC-AUC scores. The models were trained in different conditions including only clean trials for training and then training the model using a combination of clean and bad trials.

	Condition 1		Condition 2		Condition 3	
Subject	Train: RSVP (Clean) Test: RSVP (Clean)	Train: IC (Clean) Test: RSVP (Clean)	Train: RSVP (Clean+Bad) Test: RSVP (Clean+Bad)	Train: IC (Clean+Bad) Test: RSVP (Clean+Bad)	Train: IC (Clean) Test: IC (Clean)	Train: IC (Clean+Bad) Test: IC (Clean+Bad)
1	0.68	[a]	0.61	0.52	[a]	0.54
2	0.93	0.91	0.92	0.92	0.92	0.87
3	0.82	[a]	0.78	0.60	[a]	0.55
4	0.86	[a]	0.82	0.80	[a]	0.78
5	0.90	[a]	0.87	0.88	[a]	0.86
6	0.86	[a]	0.88	0.79	[a]	0.76
7	0.81	[a]	0.78	0.73	[a]	0.62
8	0.78	[a]	0.78	0.69	[a]	0.71
9	0.86	0.87	0.88	0.84	0.87	0.86
10	0.92	[a]	0.90	0.86	[a]	0.81
Average	0.84	0.89	0.82	0.76	0.90	0.74

RSVP—traditional RSVP, IC—intentionally contaminated RSVP.

[a] There were insufficient clean trials available to train the model (i.e. more than 80% of the trials were dropped using a 100 μV peak-to-peak threshold).

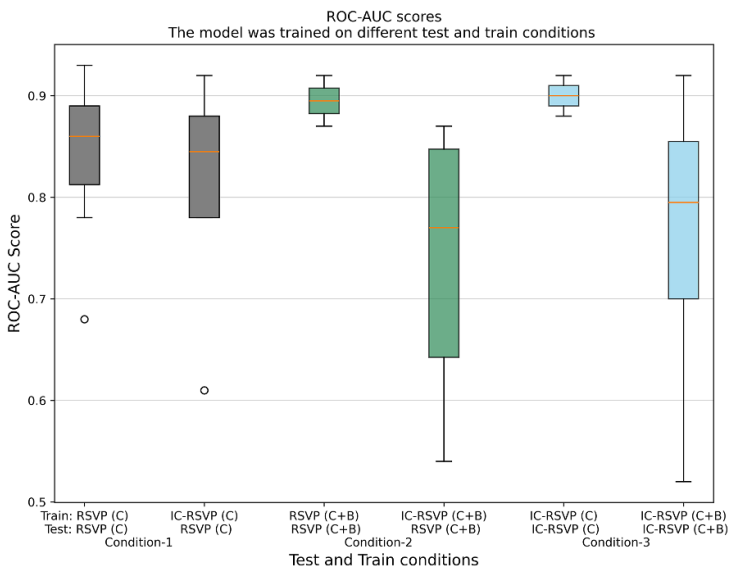

Figure 7. Box plots of ROC-AUC scores obtained when the model was trained and tested on different conditions incorporating both traditional RSVP and IC-RSVP. *C—Clean trials and C+B—Clean+Bad trials.*

Table 4. Trial rejection using different peak-to-peak thresholds, showing percentage of rejected trials.

			IC-RSVP					
	RSVP		Body Movement		Talking		Head Movement	
Threshold (μV)	Avg (%)	Std. Dev (%)	Avg (%)	Std. Dev (%)	Avg (%)	Std. Dev (%)	Avg (%)	Std. Dev (%)
50	33.36	25.09	58.26	35.69	68.37	27.97	99.83	0.46
75	16.46	13.75	37.96	31.45	39.74	33.64	97.06	6.48
100	10.93	9.60	27.31	27.96	28.70	29.62	90.04	19.85
150	4.89	5.28	16.46	19.37	14.81	19.62	77.12	33.70
200	2.39	3.22	11.00	14.52	8.70	13.78	65.79	36.61
250	0.97	1.02	8.30	12.75	5.60	11.24	52.71	36.53
300	0.30	0.29	6.81	11.04	3.98	9.16	42.94	33.67
350	0.16	0.22	5.84	9.62	3.28	8.34	37.13	31.50
400	0.09	0.16	5.20	8.56	2.86	7.43	33.22	29.30
450	0.05	0.14	4.34	7.13	2.43	6.38	30.01	26.92
500	0.05	0.14	3.73	6.08	2.17	5.65	27.34	24.62

Avg—Average, and Std. Dev—Standard deviation.

4. Conclusion

This study extensively explored the impact of ecologically valid interactions on the performance of RSVP-based P300 detection.

The application of a peak-to-peak threshold of 100 μV for identifying clean and bad trials revealed a significant drop in the number of trials, particularly noticeable in instances of head movement artifacts across multiple subjects. The Bayesian Ridge model was then trained exclusively on clean trials from traditional RSVP settings and subsequently tested on diverse subsets of the dataset, examining clean trials, bad trials, and a mix of clean and bad trials, each associated with different artifact condition categories. The use of ROC-AUC scores facilitated the analysis of P300 detection performance, exposing the clear differences between detection performance for clean and bad trials.

Furthermore, our approach of exclusively training the model on clean trials from RSVP settings provided a foundation for later analyses. Training and testing were subsequently conducted under varying clean and noisy conditions for both RSVP and IC-RSVP tasks. The ROC-AUC scores obtained from these analyses reveal the critical importance of noise removal for optimizing P300 detection accuracy, emphasizing its direct association with the quality of data provided for model training. Pearson correlation coefficient was used to statistically access the relationship between ROC-AUC scores and the percentage of dropped trials. Given that these types of noises are ever-present in daily life, addressing this issue becomes paramount for extending the feasibility of EEG technology beyond the confines of the lab and into real-world applications. The outcomes of this study highlight the significance of signal denoising as a crucial step toward realizing the

full potential of EEG technology in noisy real-world scenarios.

As part of our future goals, we plan to extend our research to include additional datasets and explore other BCI paradigms, such as motor imagery (MI), steady-state visually evoked potential (SSVEP), and cognitive state monitoring. This expansion will enable us to evaluate the impact of different data types and paradigms on classification performance and further enhance our understanding of BCI systems, particularly in real-world settings. Moreover, we plan to expand our testing to include scenarios such as sports and exercise, with the goal of taking EEG-based BCIs out of the lab and into true real-world environments. This also aligns with our goal of exploring state-of-the-art denoising algorithms to develop robust methodologies capable of utilizing EEG data for BCI applications that closely correspond with performance outcomes obtained in clean lab settings.

Data availability statement

The datasets used in this study can be found in the GitHub repository. https://github.com/meharahsanawais/AMBER-EEG-Dataset

Acknowledgments

This work was funded by Science Foundation Ireland under Grant Number SFI/12/RC/2289_P2 and CHIST-ERA under Grant Number (CHISTERA IV 2020 - PCI2021-122058-2A).

Ethics statement

The studies involving humans were approved by Dublin City University (DCU) Research Ethics Committee (DCUREC/2021/175). The study was conducted in accordance with local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

References

Awais M A, Redmond P, Ward T E and Healy G 2023 Amber: advancing multimodal brain-computer interfaces for enhanced robustness-a dataset for naturalistic settings *Front. Neuroergonomics* **4** 1216440

Awais M A, Yusoff M Z, Khan D M, Yahya N, Kamel N and Ebrahim M 2021 Effective connectivity for decoding electroencephalographic motor imagery using a probabilistic neural network *Sensors* **21** 6570

Bablani A, Edla D R, Kupilli V and Dharavath R 2021 Lie detection using fuzzy ensemble approach with novel defuzzification method for classification of EEG signals *IEEE Trans. Instrum. Meas.* **70** 1–13

Bajaj N, Carrión J R, Bellotti F, Berta R and De Gloria A 2020 Automatic and tunable algorithm for EEG artifact removal using wavelet decomposition with applications in predictive modeling during auditory tasks *Biomed. Signal Process. Control* **55** 101624

Barbey F M *et al* 2022 Neuroscience from the comfort of your home: repeated, self-administered wireless dry EEG measures brain function with high fidelity *Front. Digit. Health* **4** 944753

Bressler S L and Ding M 2006 Event-related potentials *Wiley Encyclopedia of Biomedical Engineering* (Wiley) (https://doi.org/10.1002/9780471740360.ebs0455)

Brophy E, Redmond P, Fleury A, De Vos M, Boylan G and Ward T 2022 Denoising EEG signals for real-world BCI applications using GANs *Front. Neuroergonomics* **2** 805573

Cecotti H and Ries A J 2017 Best practice for single-trial detection of event-related potentials: application to brain-computer interfaces *Int. J. Psychophysiol.* **111** 156–69

Chang H, Zong Y, Zheng W, Xiao Y, Wang X, Zhu J, Shi M, Lu C and Yang H 2023 EEG-based major depressive disorder recognition by selecting discriminative features via stochastic search *J. Neural Eng.* **20** 026021

Cimmino A, Ciaramella A, Dezio G and Salma P J 2021 Non-linear PCA neural network for EEG noise reduction in brain-computer interface *Progresses in Artificial Intelligence and Neural Systems* (Springer) pp 405–13

Cui Y, Xie S, Xie X, Zhang X and Liu X 2022 Dynamic probability integration for electroencephalography-based rapid serial visual presentation performance enhancement: application in nighttime vehicle detection *Front. Comput. Neurosci.* **16** 1006361

Cui Y, Xie S, Xie X, Zheng D, Tang H, Duan K, Chen X and Jiang Y 2023 LDER: a classification framework based on erp enhancement in RSVP task *J. Neural Eng.* **20** 036029

Dimigen O 2020 Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments *NeuroImage* **207** 116117

Fawcett T 2006 An introduction to ROC analysis *Pattern Recognit. Lett.* **27** 861–74

Gandhi S, Oates T, Mohsenin T and Hairston D 2018 Denoising time series data using asymmetric generative adversarial networks *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conf., PAKDD 2018* (*Melbourne, VIC, Australia June 3–6, 2018 Proc., Part III* vol 22) (Springer) pp 285–96

Gao X *et al* 2023 An adaptive joint CCA-ICA method for ocular artifact removal and its application to emotion classification *J. Neurosci. Methods* **390** 109841

Gorjan D, Gramann K, De Pauw K and Marusic U 2022 Removal of movement-induced EEG artifacts: current state of the art and guidelines *J. Neural Eng.* **19** 011004

Han J, Kim S Y, Lee J and Lee W H 2022 Brain age prediction: a comparison between machine learning models using brain morphometric data *Sensors* **22** 8077
Hwaidi J F and Chen T M 2021 A noise removal approach from EEG recordings based on variational autoencoders *2021 13th Int. Conf. on Computer and Automation Engineering (ICCAE)* (IEEE) pp 19–23
Javed E, Faye I, Malik A S and Abdullah J M 2017 Removal of BCG artefact from concurrent fMRI-EEG recordings based on EMD and PCA *J. Neurosci. Methods* **291** 150–65
Kabbara A, Forde N, Maumet C and Hassan M 2023 Successful reproduction of a large EEG study across software packages *Neuroimage Rep.* **3** 100169
Katyal E A and Singla R 2021 EEG-based hybrid qwerty mental speller with high information transfer rate *Med. Biol. Eng. Comput.* **59** 633–61
Kindermans P-J, Tangermann M, Müller K-R and Schrauwen B 2014 Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller *J. Neural Eng.* **11** 035005
Lee T, Kim M and Kim S-P 2020 Improvement of P300-based brain–computer interfaces for home appliances control by data balancing techniques *Sensors* **20** 5576
Lees S, Dayan N, Cecotti H, McCullagh P, Maguire L, Lotte F and Coyle D 2018 A review of rapid serial visual presentation-based brain–computer interfaces *J. Neural Eng.* **15** 021001
Lees S, McCullagh P, Payne P, Maguire L, Lotte F and Coyle D 2019 Speed of rapid serial visual presentation of pictures, numbers and words affects event-related potential-based detection accuracy *IEEE Trans. Neural Syst. Rehabil. Eng.* **28** 113–22
Li R, Song W-Q, Du J-B, Huo S and Shan G-X 2015 Connecting the P300 to the diagnosis and prognosis of unconscious patients *Neural Regen. Res.* **10** 473
Liu J, Ramakrishnan S, Laxminarayan S, Neal M, Cashmere D J, Germain A and Reifman J 2018 Effects of signal artefacts on electroencephalography spectral power during sleep: quantifying the effectiveness of automated artefact-rejection algorithms *J. Sleep Res.* **27** 98–102
Liu S, Wang W, Sheng Y, Zhang L, Xu M and Ming D 2020 Improving the cross-subject performance of the ERP-based brain–computer interface using rapid serial visual presentation and correlation analysis rank *Front. Hum. Neurosci.* **14** 296
Liu Y, Sun L, Du C and Wang X 2020 Near-infrared prediction of edible oil frying times based on Bayesian ridge regression *Optik* **218** 164950
Luck S J 2014 *An Introduction to the Event-Related Potential Technique* (MIT Press)
Ludwig K A, Miriani R M, Langhals N B, Joseph M D, Anderson D J and Kipke D R 2009 Using a common average reference to improve cortical neuron recordings from microelectrode arrays *J. Neurophysiol.* **101** 1679–89
MacKay D J 1992 Bayesian interpolation *Neural Comput.* **4** 415–47
Moretti D V, Babiloni F, Carducci F, Cincotti F, Remondini E, Rossini P, Salinari S and Babiloni C 2003 Computerized processing of EEG–EOG–EMG artifacts for multi-centric studies in EEG oscillations and event-related potentials *Int. J. Psychophysiol.* **47** 199–216
Narmada A and Shukla M 2023 A novel adaptive artifacts wavelet denoising for EEG artifacts removal using deep learning with meta-heuristic approach *Multimedia Tools Appl.* **82** 1–39
Pan J, Chen X, Ban N, He J, Chen J and Huang H 2022 Advances in P300 brain–computer interface spellers: toward paradigm design and performance evaluation *Front. Hum. Neurosci.* **16** 1077717
Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30
Peksa J and Mamchur D 2023 State-of-the-art on brain-computer interface technology *Sensors* **23** 6001
Pritchard W S 1981 Psychophysiology of P300 *Psychol. Bull.* **89** 506
Rahman S U, O'Connor N, Lemley J and Healy G 2022 Using pre-stimulus EEG to predict driver reaction time to road events *2022 44th Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE) pp 4036–9
Rashid M, Sulaiman N, Abdul Majeed A P P, Musa R M, Nasir A F A, Bari B S and Khatun S 2020 Current status, challenges and possible solutions of EEG-based brain-computer interface: a comprehensive review *Front. Neurorobot.* **14** 25
Rashmi C and Shantala C 2022 EEG artifacts detection and removal techniques for brain computer interface applications: a systematic review *Int. J. Adv. Technol. Eng. Explor.* **9** 354
Ren B and Zhou Q 2023 'Assessing passengers' motion sickness levels based on cerebral blood oxygen signals and simulation of actual ride sensation *Diagnostics* **13** 1403
Rosenfeld J P 2020 P300 in detecting concealed information and deception: a review *Psychophysiology* **57** e13362
Sadras N, Sani O G, Ahmadipour P and Shanechi M M 2023 Post-stimulus encoding of decision confidence in EEG: toward a brain–computer interface for decision making *J. Neural Eng.* **20** 056012
Sburlea A I, Montesano L, de la Cuerda R C, Alguacil Diego I M, Miangolarra-Page J C and Minguez J 2015 Detecting intention to walk in stroke patients from pre-movement EEG correlates *J. Neuroeng. Rehabil.* **12** 1–12
Schmoigl-Tonis M, Schranz C and Müller-Putz G R 2023 Methods for motion artifact reduction in online brain-computer interface experiments: a systematic review *Front. Hum. Neurosci.* **17** 1251690
Shi Q, Abdel-Aty M and Lee J 2016 A Bayesian ridge regression analysis of congestion's impact on urban expressway safety *Accid. Anal. Prev.* **88** 124–37
Shi W, Li Y, Cai N, Chen R, Cao W and Li J 2024 Removal of ocular and muscular artifacts from multi-channel EEG using improved spatial-frequency filtering *IEEE J. Biomed. Health Inf.* **28** 3466–77
ter Braack E M, de Jonge B and Van Putten M J 2013 Reduction of TMS induced artifacts in EEG using principal component analysis *IEEE Trans. Neural Syst. Rehabil. Eng.* **21** 376–82
Tipping M E 2001 Sparse Bayesian learning and the relevance vector machine *J. Mach. Learn. Res.* **1** 211–44
Tortora S, Ghidoni S, Chisari C, Micera S and Artoni F 2020 Deep learning-based BCI for gait decoding from EEG with LSTM recurrent neural network *J. Neural Eng.* **17** 046011
Wang K, Qiu S, Wei W, Zhang C, He H, Xu M and Ming D 2021 Vigilance estimating in SSVEP-based BCI using multimodal signals *2021 43rd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE) pp 5974–8
Wang Z, Healy G, Smeaton A F and Ward T E 2018 Spatial filtering pipeline evaluation of cortically coupled computer vision system for rapid serial visual presentation *Brain Comput. Interfaces* **5** 132–45
Yasemin M, Cruz A, Nunes U J and Pires G 2023 Single trial detection of error-related potentials in brain–machine interfaces: a survey and comparison of methods *J. Neural Eng.* **20** 016015
Yip H M K, Cheung L Y, Ngan V S, Wong Y K and Wong A C-N 2022 The effect of task on object processing revealed by EEG decoding *Eur. J. Neurosci.* **55** 1174–99
Zhang H, Zhao M, Wei C, Mantini D, Li Z and Liu Q 2021 EEGdenoiseNet: a benchmark dataset for deep learning solutions of EEG denoising *J. Neural Eng.* **18** 056057
Zhang S, Wang Y, Zhang L and Gao X 2020 A benchmark dataset for RSVP-based brain–computer interfaces *Front. Neurosci.* **14** 568000