computers

MDPI

Article

The Effect That Auditory Distractions Have on a Visual P300 Speller While Utilizing Low-Cost Off-the-Shelf Equipment

Patrick Schembri * , **Maruisz Pelc and Jixin Ma**

Department of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK;
m.pelc@greenwich.ac.uk (M.P.); j.ma@greenwich.ac.uk (J.M.)

* Correspondence: p.schembri@greenwich.ac.uk

check for updates

Abstract: This paper investigates the effect that selected auditory distractions have on the signal of a visual P300 Speller in terms of accuracy, amplitude, latency, user preference, signal morphology, and overall signal quality. In addition, it ensues the development of a hierarchical taxonomy aimed at categorizing distractions in the P300b domain and the effect thereof. This work is part of a larger electroencephalography based project and is based on the P300 speller brain–computer interface (oddball) paradigm and the xDAWN algorithm, with eight to ten healthy subjects, using a non-invasive brain–computer interface based on low-fidelity electroencephalographic (EEG) equipment. Our results suggest that the accuracy was best for the lab condition (*LC*) at 100%, followed by music at 90% (*M90*) at 98%, trailed by music at 30% (*M30*) and music at 60% (*M60*) equally at 96%, and shadowed by ambient noise (*AN*) at 92.5%, passive talking (*PT*) at 90%, and finally by active listening (*AL*) at 87.5%. The subjects' preference prodigiously shows that the preferred condition was *LC* as originally expected, followed by *M90*, *M60*, *AN*, *M30*, *AL*, and *PT*. Statistical analysis between all independent variables shows that we accept our null hypothesis for both the amplitude and latency. This work includes data and comparisons from our previous papers. These additional results should give some insight into the practicability of the aforementioned P300 speller methodology and equipment to be used for real-world applications.

1. Introduction

Distractions can be loosely defined as events that take away the attention from what you are supposed to be doing. In the context of a brain–computer interface (BCI), these are unwanted, and the current practice in this field is to avoid them through the careful design of experiments. The work presented here is part of a larger electroencephalography (EEG)-based project and in continuation of our latest papers [1,2], with the goal of building an extendible hierarchical taxonomy aimed at categorizing distractions and the effects thereof. In addition, this paper is an extension of work originally presented in [3], which is also in continuation of the aforementioned papers.

BCI research and development has been predominantly focused on the speed and accuracy of the BCI application but has been wanting in usability, such as the environment in which it is being used. In fact, many BCI applications and experiments were and are still being performed in laboratory settings with unrealistic conditions, where the subject sits in a sound-attenuated room without any distractions [4,5]. A number of research papers, such as [6,7], focus on real-world contexts, however, they were either using medical-grade equipment [8] and/or focusing on auditory event-related potentials (ERPs) [9].

Our goal is to analyze the effect that distractions have on the signal characteristics of a P300 component while using the P300 speller paradigm alongside the continuous development of taxonomy. In this paper, we specifically analyze the effect that auditory distractions, explicitly that of music at 30% (*M30*), music at 60% (*M60*), music at 90% (*M90*), ambient noise (*AN*), passive talking (*PT*), and active listening (*AL*), have on the signal of a visual P300 speller in terms of accuracy, amplitude, latency, user preference, signal morphology, and overall signal quality. This work includes data and comparisons from our previous papers [2,3]. In addition, our research makes use of non-invasive BCI on the basis of EEG. The contribution of this work is to provide researchers with a ground reference taxonomy that enables them to compare and extrapolate EEG signal detection results between different distractions and to be able to predict and assess the effects of a particular distraction and/or an environment containing those distractions.

The requirement for this work was derived from the need to broaden the utilization of this technology for both subjects with neuromuscular disabilities and healthy individuals, by providing a solution which is assessed outside lab conditions, i.e., in noisy environments, and based on low-cost equipment. In the absence of comprehensive studies on the effect that distractions have in relation to the P300 performance, i.e., for accuracy and signal characteristics, an evident necessity for this study was present. Our null hypothesis, based on previous research [2,3], is that this type of distraction does not show any statistically significant effect on accuracy, task performance, amplitude, latency, or signal morphology.

In this work, we report a study where eight to ten healthy subjects used a Farwell and Donchin P300 speller paradigm in conjunction with the xDAWN algorithm [10] while utilizing low-cost off-the-shelf equipment. The subjects were asked to communicate five alphanumeric characters, referred to as symbols, in the different settings mentioned above. The main aim of this study was to systematically examine the usability of a P300 speller BCI, in terms of the effect that the independent variables, i.e., each distraction, has on the dependent variables when compared to the dependent variable lab condition (LC). Empirical experiments were performed to measure how environmental factors, such as the aforementioned settings, affect the signal characteristics and performance of the P300 component. This work forms part of a larger EEG-based project where we have instituted [2] a set of distinctive categories for distractions in conjunction with the continuous development of a hierarchical taxonomy, as portrayed in Figure 1.
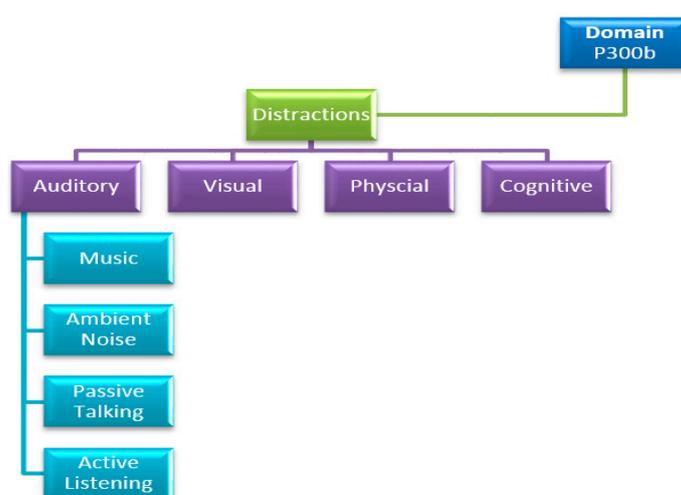
Figure 1. Development of a P300b hierarchical taxonomy for distractions, which is extensible.

This paper is structured as follows: the methodology, which includes the research background, equipment, participants, and experimental procedures, is described in Section 2. The offline and online ERP results and the user preferences are presented in Section 3. The conclusion is given in Section 4.

2. Methodology

The following segment/s of the methodology are from the authors' previous work, as referenced above, and are adopted and outlined in the current paper for readers' convenience.

2.1. Research Background

ERPs are slow voltage fluctuations or electrical potential shifts recorded from the nervous system. These are time locked to perceptual events following a presentation of a stimulus, being cognitive, sensor, or motor stimuli. The simplest paradigm for eliciting an ERP is by focusing attention on the target stimuli (occur infrequently) embedded randomly in an array of non-targets (occur frequently). The methodology used derives from the oddball paradigm first used in ERPs by Squires, Squires, and Hillyard [11], where the subject is asked to distinguish between a common stimulus (non-target) and a rare stimulus (target). The target stimuli elicit one of the most renowned ERP components, known as P300, which is an exogenous and spontaneous component and was first described by Sutton [12]. The name is derived from the fact that it is a positive wave that appears approximately 300 ms after the target stimulus. Unless otherwise noted in this paper, the term P300 (P3) will always refer to the P300b (P3b), which is elicited by task-relevant stimuli in the centro-parietal area of the brain.

There were several studies with the aim of improving the performance of P300-BCI systems. For instance, Qu et al. [13] focused on a new P300 speller paradigm based on three-dimensional (3-D) stereo visual stimuli, while Gu et al.'s [14] primary focus was within the single character P300 paradigm, i.e., they evaluated the posterior probability of each character in the stimuli set to be the target. Even though improving the performance and execution of the application is an important aspect, this should not come at the expense of the environment in which it is used, such as that of a real-world context. In fact, current practice in the field is either to avoid distractions through the careful design of experiments or environments or to filter out these distractions (and the effects thereof). As such, there has not been a rigorous classification of such distractions within a range of intensities, nor is there currently a mapping between distraction type and the effects caused.

For instance, [6] focused on assessing how background noise and interface color contrast affect user performance, while [7] focused on the experimental validation of a readout circuit, which is a wearable low-power design with long-term power autonomy device aimed at the acquisition of extremely weak biopotentials in real-world applications. Additionally, [8] introduced methods for benchmarking the suitability of new EEG technologies by performing an auditory oddball task using three different medical-grade EEG systems, while seated and walking on a treadmill. Moreover, [9] evaluated the ERP and single-trial characteristics of a three-class auditory oddball paradigm recorded in outdoor scenarios while peddling on a fixed bike or freely biking around.

2.2. Hardware

Our work utilized the OpenBCI 32-bit board (dubbed Cyton) in conjunction with the Electro-Cap, as depicted in Figure 2 which, in the context of EEG experiments, is based on the international 10/20 system for electrode placement on the scalp.
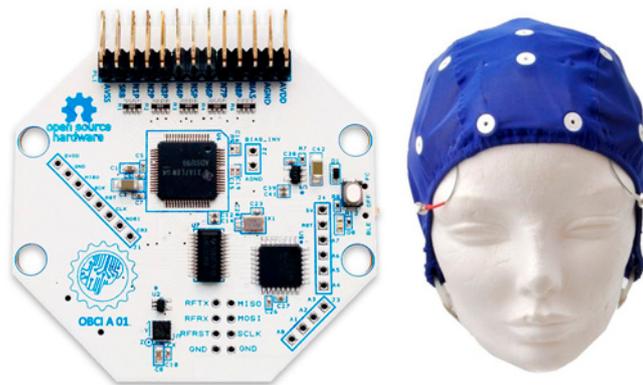
Figure 2. OpenBCI Cyton board and Electro-Cap.

The PIC32MX250F128B microcontroller found on the Cyton board has a 32-bit processor with a 50 MHz ceiling and 32 KB of memory storage, while being Arduino compatible. It also encompasses an ADS1299 IC, which is an eight-channel, 24-bit, simultaneous sampling delta-sigma, analog-to-digital converter used for biopotential measurements, developed by Texas Instruments. In addition, it has a built-in low-cost microcontroller, the RFDuino RFD22301, for wireless communication, which communicates with the provided and pre-programmed USB dongle. In our previous paper [15], the reader can obtain a more in-depth elucidation of the Cyton board and its built-in hardware components, which are depicted in Figure 3 for the reader's convenience.
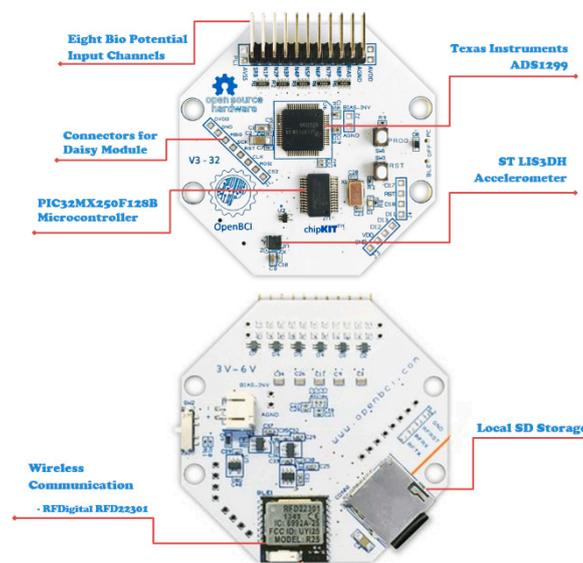
Figure 3. Cyton board (front and back) components.

In addition, our work utilized the Electro-Cap, which is made of fabric, explicitly that of elastic spandex. The electrodes, which are directly attached to the fabric, are made of pure tin and are considered as wet electrodes. This implies that an electrolyte gel must be applied to the electrodes to have effective scalp contact. Otherwise, we could end up with impedance instability.

To output our six diverse distractions, i.e., music at 30%, music at 60%, music at 90%, ambient noise, passive talking, and active listening, we used a pair of Creative Labs SBS 15 speakers, which can output a nominal root mean square (RMS) power of 5 Watts for each speaker, which have a frequency response of 90 Hz–20,000 Hz and a 90 dB signal-to-noise ratio (SNR).

2.3. Participants

For the first session, which included the independent variables *LC*, *M30*, *M60*, and *M90*, we enrolled N = 10 healthy subjects, of which seven were males and three were females. They were aged between 29 and 38, with a mean age of 33.8, and their participation was on a voluntary basis.

For the second session, which included the independent variables *AN*, *PT*, and *AL*, we enrolled N = 8 healthy subjects, of which six were males and two were females. They were aged between 29 and 38, with a mean age of 33.8, and their participation was on a voluntary basis. The native language of seven of the eight subjects was Maltese, while the native language for the final subject was English.

However, all subjects in both sessions were au fait with the English language and were conversant with the alphanumeric symbols portrayed in the P300 speller application. In addition, all subjects had previous experience with P300-BCI experiments.

Another subject aided in the groundwork and initial testing for the configuration of our equipment, and in the development of the methodology, however, he/she did not partake in the official experiments and hence his/her data are not included in the results.

For both sessions, the subjects were instructed to avoid consuming any food or drinks, apart from water, for one hour prior to the start of the experiment. Moreover, the total averaged reported sleep the night before the experiment was 467.5 min (29.05).

2.4. Data Acquisition

We set the sampling rate and sampling precision for the EEG signal at the hardware's ceiling of 250 Hz and 24 bits, respectively. The raw data were stored anonymously in OpenVIBE. ov format, however, for offline analysis, this was converted to a comma-separated value (CSV) format and imported into MATLAB. The stored raw data included the readings of eight EEG electrodes, namely, C3, Cz, C4, P3, Pz, P4, O1, and O2, which were placed in accordance with the International 10–20 System.

Since the spatial amplitude dispersal of the P300 component is symmetric around Cz and its electrical potential is maximal in the midline region (Cz, Pz) [16], we opted to use the aforementioned electrodes. In addition, and in view of the fact that, in general, an earlobe or a mastoid reference generates a robust P300 response, we opted for a referential montage, with the reference electrode placed on the left earlobe (A1) and the ground electrode placed on the right ear lobe (A2).

2.5. P300 Speller and xDAWN

In this work, we utilized the Farewell and Donchin P300 speller in combination with the xDAWN algorithm. This type of methodology is based on visual stimuli which are explained thoroughly below. The subject is presented with thirty-six symbols, i.e., alphanumeric characters, which were positioned in a six by six grid, dubbed the spelling grid. The subject is asked to observe the intensification of each row and column, which for one repetition entails the intensification of six rows and six columns in random order. Then, the subject is asked to differentiate between a rare stimulus (target) which generates a spontaneous and exogenous ERP known as the P300 potential and a common stimulus (non-target) which does not generate this component. This is achieved with the subject focusing his/her attention on the desired symbol (target) while ignoring the other symbols (non-target). This implies that there is one target column and one target row, while there are five non-target columns and five non-target rows, for each repetition. The intersection of the row and column targets predicts what symbol is spelt for that repetition. In the simplest terms, the prediction distinguishes between the target, i.e., a row or column stimulus that produces a P300 component, from the non-target, i.e., a row or column stimulus that does not produce a P300 component. Taking into consideration that the peak potential of the P300 component is between 5–10 µV and that this is entrenched and concealed by artefacts and other brain activities, where the typical EEG signal is ±100 µV, this implies that it would be very hard to predict the correct symbol with one repetition. This also leads to a very low signal-to-noise ratio and the most popular and established way to address this issue is for each symbol to be spelt

numerous consecutive times, i.e., more than one repetition per symbol, and then the corresponding column and/or row epochs are averaged over a number of repetitions, thus canceling out components unrelated to stimulus onset.

The xDAWN algorithm is a process of spatial filtering that (1) is a dimensional reduction method that produces a subset of pseudo-channels, dubbed output channels, by a linear combination of the original channels and (2) promotes the appealing part of the signal, such as ERPs, with respect to the noise. The process of xDAWN is applied to the data prior to any classification, for instance, linear discriminant analysis (LDA), which was used for this work. From an abstract point of view, the xDAWN algorithm can be divided into (1) a least square estimation of the evoked responses and (2) a generalized Rayleigh quotient to estimate a set of spatial filters that maximize the signal plus noise ratio (SSNR).

The following is adapted from [10,17]. Let $X \in \mathbb{R}^{S \times C}$ be the EEG data that contain ERPs and noise, with S samples and C channels. Let $A \in \mathbb{R}^{E \times C}$ be the matrix of ERP signals, while E is the number of temporal samples of the ERP (typically, E is chosen to correspond to 600 ms or 1 s). Let $N \in \mathbb{R}^{S \times C}$ be the noise matrix which contains normally distributed noise. The ERP position in the data is given by a Toeplitx matrix $D \in \mathbb{R}^{E \times S}$. The data model is given by $X = D^T A + N$. A is estimated by a least square estimate using a matrix inverse (pseudoinverse), as shown in formula (1).

$$\hat{A} = \underset{A}{arg\min} = \|X - DA\|_2^2 = \left(D^T D\right)^{-1} D^T X \tag{1}$$

Let $W \in \mathbb{R}^{S \times F}$ be the pseudo-channels while F represents the filters for projection. The result is the filtered data matrix $\tilde{X} = XW$. According to [10], the optimal filters W can be found by maximizing the SSNR, as given by the generalized Rayleigh quotient:

$$\hat{W} = \underset{W}{arg\max} = \frac{Tr\left(W^T \hat{A}^T D^T D \hat{A} W\right)}{Tr(W^T X^T X W)} \tag{2}$$

The optimization problem is solved by combining a QR matrix decomposition (QRD) with a singular value decomposition (SVD). A more thorough explanation is found in [10].

2.6. Experimental Design

In this work, we had seven manipulated independent within-subjects variables: (a) lab condition, (b) music at 30%, (c) music at 60%, (d) music at 90%, (e) ambient noise, (f) passive talk, and (g) active listening,. Moreover, we had several dependent measures which were classified into three types of dependent variables, i.e., online performance, which comprised the *accuracy*, offline performance, which consisted of *amplitude* and *latency*, and *user preference*.

2.6.1. Independent Variables

As previously mentioned, we made use of seven manipulated independent within-subjects variables in two different sessions, which are itemized and elucidated below. It is important to remember that all auditory distractions were performed from the same recording, with the same equipment and at the same level of intensity unless noted otherwise.

1. Lab condition, abbreviated as *LC*, where the experiments were performed in a sound-attenuated room, without any distractions.
2. Music at 30% volume, abbreviated as *M30*, where the volume level was labeled as "low", i.e., between 20 and 30dB, and simulated background music.
3. Music at 60% volume, abbreviated as *M60*, where the volume level was labeled as "medium", i.e., between 50 and 60dB, and simulated active listening to a movie.

4.　Music at 90% volume, abbreviated as *M90*, where the volume level was labeled as "high", i.e., between 80 and 90dB, and simulated disco-level music only, i.e., there was no crowd chatter or other type of noise.

5.　Ambient noise, abbreviated as *AN*, where we introduced an auditory distraction that simulated ambient noise, including city traffic noise.

6.　Passive talk, abbreviated as *PT*, where we simulated persons murmuring.

7.　Active listening, abbreviated as *AL*, where we simulated several persons discussing a particular topic in a clear and audible manner.

The comparison between independent variables was done as follows: *LC* versus *M30*, *M60*, and *M90*, and *LC* versus *AN*, *PT*, and *AL*, respectively, for the dependent variables of amplitude and latency. The setting and configuration for all independent variables were outputted from a recorded simulation by the aforementioned speakers, with the volume set at thirty percent (unless otherwise noted), i.e., between 20 and 30 dB, which simulates a real-world context. We opted to use a recording and the same recording for all subjects since we aimed to make the experiments between subjects as similar as possible. Moreover, this made it possible to increase the level of volume equally in every experiment.

2.6.2. Dependent Variables

In this work, there were three types of dependent measures, which can be sub-categorized into four distinct dependent variables:

1.　Online performance comprised *accuracy*, which is the comparison of the correctly spelled symbols to the planned target symbols, i.e., in our experiments, we had five planned target symbols—BRAIN.

2.　Offline statistics comprised *amplitude* and *latency*, where (a) the P300 amplitude (μV) is related to the distribution of the subject's processing resources assigned to the task. It is defined as the voltage difference between the largest positive peak from the baseline within the P300 latency interval and (b) the P300 latency is considered a measure of cognitive processing time, generally between 300–800 ms [18] poststimulus, i.e., after target stimulus. In the simplest terms, it is the time interval between the onset of the target stimulus and the peak of the wave.

3.　*User preference*, where two questionnaires were provided to the subjects to ask for their favorite usage condition. This was ranked from four to one, four being the highest and one being the lowest. The objective for these questionnaires was to compare if user preference was directly related to an increase in the accuracy—online performance—or to a heightened increase in the amplitude—offline performance. Moreover, the questionnaire results can easily be compared between subjects to assess the overall preference and/or reluctance in using this technology in environments which contain those distractions.

2.7. Experimental Procedure

An induction session was held for each subject which was intended to re-educate the subjects on the P300 speller paradigm and hardware being used. We also notified the subjects that:

1.　They would be doing the P300 speller experiment in five unique settings for the first session, i.e., (i) the training phase which was performed in lab conditions, i.e., in a sound-attenuated room (ii) *LC*, (iii) *M30*, (iv) *M60*, and (v) *M90*, and another five unique settings in the second session, i.e., (vi) training phase (same conditions as the first session), (vii) *LC*, (viii) *AN*, (ix) *PT*, and (x) *AL*, as elucidated in the independent variables.

2.　They would be spelling the symbols "BRAIN" for (1 ii) to (1 v) and for (1 vii) to (1 x), while for (1 i) and (1 vi), i.e., the training phase, they would be spelling fifteen random symbols.

The first experiment was always the training phase in both sessions, i.e., (1 i) and (1 vi) since this was required as training for the system to be able to predict the correct symbols in the subsequent experiments. Then, (1 ii) to (1 v) and (1 vii) to (1 x) were done in a randomized order to circumvent the subjects' accustomization to that particular distraction. Then, the subjects were asked if they had any queries which were promptly answered at this stage. The subjects were then asked to be seated and relax for a few minutes prior to the start of the experiments. The setup was as follows: (a) the subject was seated facing the display, approximately one meter away, (b) the researcher was situated at the left-hand side of the subject and refrained from making any type of movement or noise throughout. We opted to stay to the side of the subject, since from our experience and subjects' feedback, they were not comfortable with someone being behind them, (c) the speakers were placed at a 15-degree angle facing the subject and were situated near the monitor, i.e., one meter away from the subject, and (d) the electrode impedance was verified to be less than 5 KΩ. The experiment only started when the subject had no additional questions, was in a comfortable position, and was able to perform the task at hand.

Subsequently, the spelling grid consisting of 36 symbols in a 6 by 6 matrix was presented to the subjects and the target symbol, i.e., the symbol the subject must focus on, was preceded by a cue, i.e., the target symbol was highlighted in blue, as depicted in Figure 4a. Next, there was a random intensification of 100 ms for each row and column in the matrix and there was an 80 ms delay between two successive intensifications, i.e., after one column and one row was intensified. This implies that we had an interstimulus interval (ISI) of 180 ms. Afterwards, and to predict each symbol, there were fifteen repetitions (i.e., one trial) which consisted of intensifying six rows and six columns for each repetition, and in between each of the groups of six rows and six columns, i.e., one repetition, there was an inter-repetition delay of 100 ms. At the end of trial, i.e., 6 rows by 6 columns by 15 repetitions, the symbol, which was predicted by the system, was presented to the subject by being highlighted with a green cue, as depicted in Figure 4b. The subject would be aware if the system predicted the correct target symbol. Moreover, in between trials, i.e., in between different symbols, there was a 3000 ms inter-trial period. At the end of each experiment, the subjects were given a short break.
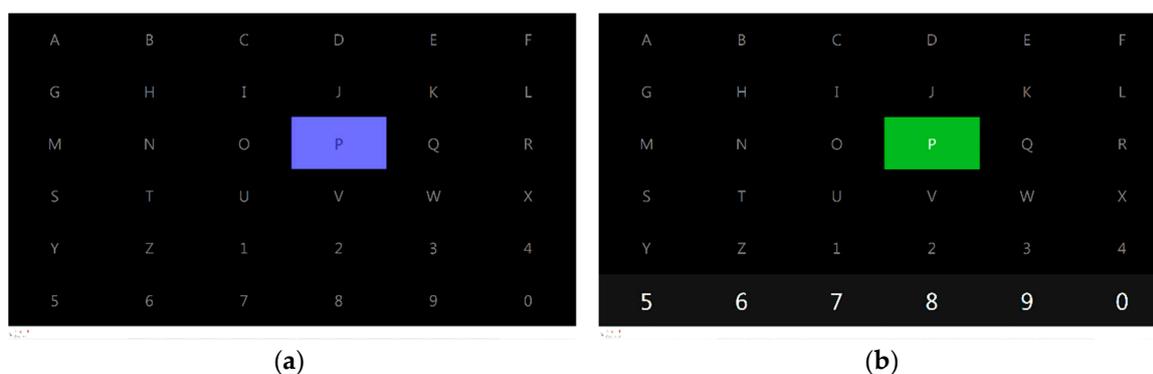
Figure 4. (**a**) Requested symbol highlighted in blue at the beginning of the symbol run and (**b**) the predicted symbol highlighted in green.

As previously mentioned, the training phase ((1 i) and (1 vi)) was made up of one session that consisted of 15 random symbols by 15 trials each (i.e., 6 rows and 6 columns per trial * 15 trials = 180 flashes per symbol), and this took approximately 10 min. The *LC*, *M30*, *M60*, *M90*, *AN*, *PT*, and *AL* were made up of one session each and consisted of five specific symbols that made up the word "BRAIN", and similarly to the training phase, each symbol had 15 trials, with six rows and six columns per trial, i.e., 180 flashes per symbol, and took approximately 6 min for each independent variable.

In total, we had 15 symbols which were spelled in the training phases, and five symbols which were spelled in the other aforementioned independent variables. Hence, due to the matrix disposition, for the training phase, we had 2700 flashes, of which 450 were targets and 2250 were non-targets, and 900 flashes per independent variable, of which 150 were targets and 750 were non-targets.

These values are per subject, the data were stored anonymously, and the subjects were referred to as subject *x*.

2.8. Signal Processing—Online

OpenViBE 2.0.0 was selected as our online system and also took care of the raw signal acquisition. This application is designed for the processing of real-time biosignal data, is a C++-based software platform, and is well known for its graphical language for the design of signal processing chains. It has two main components called the *acquisition server*, which interfaces with the Cyton board, reads the raw data signal, and produces a uniform signal stream, which is transmitted to the *designer*, which in turn is composed of several scenarios where we can structure, construct, and execute signal processing chains.

The acquisition server takes care of obtaining the signal from the Cyton board, however, it does not communicate with the board directly, instead, it provides dedicated drivers to perform this task. The sampling rate of the signal was set at 250 Hz, and consisted of eight EEG channels and three auxiliary channels, with the latter channels being used to broadcast data from the accelerometer. In a single buffer and with accepted values being only powers of two, explicitly from 2^2 to 2^9, we set the *sample count per sent block* to 32, which implies how many samples need to be sent per channel. In addition, the Cyton board reply reading timeout was regulated at 5000 ms, while the flushing timeout was adjusted to 500 ms. Although the version of OpenVibe that we used depends on TCP (Transmission Control Protocol) tagging to align the EEG signal with the simulation markers, we had set the drift tolerance to 20 ms. The issue with this setting is that it can introduce signal artefacts and mask other possible faults, such as driver issues. Even though these issues were not encountered in this work, we have decided to discontinue the use of drift tolerance in future experiments, since when TCP tagging was introduced, it made the drift tolerance mechanism redundant. The P300 speller paradigm was managed by the *designer*, in which there were a number of scenarios that were executed in order, as thoroughly explained below:

The *first scenario* was the acquisition of the signal and stimuli markers for the training phase. The recordings included the raw EEG and stimuli.

The *second scenario* entailed the pre-processing of the signal where it trained the spatial filter using the xDAWN algorithm. The subjects' data recorded in the training session were utilized with the following configuration and modalities. Initially, we chose to eliminate the last three auxiliary channels, which stored the auxiliary data of the accelerometer since the board was firmly placed on the desk and this information was not required. Subsequently, a Butterworth bandpass filter of 1 Hz–20 Hz was applied with an order of 5 and a ripple (dB) of 0.5 to remove the DC offset, the 50 Hz (60 Hz in some countries) electrical interference, and any signal harmonics and unnecessary frequencies which were not beneficial in our experiments. Next, no signal decimation was used since the sampling rate and count per buffer previously used in the acquisition server were not compatible with the actual signal decimation factor due to the Cyton board's sampling rate of 250 Hz (no available value in the sample count per block was factorable with 250 Hz). However, we still passed the signal through *time-based epoching*, which generated "epochs" (signal slices) with a duration of 0.25 s and time offset of 0.25 s between epochs (i.e., we created a temporal buffer to collect the data and forward them into blocks). This implies that there was no overlapping of data and that the inputs for the *xDAWN spatial filter* and the *stimulation-based epoching* were based on epochs of 0.25 s rather than the whole data. In the simplest terms, we had one point for every 0.25 s of data, which made our signal coarser. Subsequently, we passed the time-based epochs and stimulations to the *stimulation-based epoching*, which sliced the signal into chunks of the desired length following a stimulation event. This had an epoch duration of 0.6 s (P300 deviation around 0.3 s after the stimuli) and no offset. Lastly, the stimulations, time-based epochs, and the stimulation-based epochs were passed to the xDAWN trainer, which in the simplest terms, trains spatial filters that best highlight ERPs. The xDAWN expression, utilized in OpenVIBE, which has to be maximized, varies marginally from the original xDAWN (Rivet et al., 2009) formula where the numerator includes only the average of the target signals. In addition, the implemented algorithm

maximizes the quantity via a generalized eigenvalue decomposition method in which the best spatial filters are given by the eigenvectors corresponding to the largest eigenvalues (Clerc et al., 2016). This scenario created twenty-four coefficients values in sequence (i.e., eight input channels by three output channels) that were used in the following scenario.

The *third scenario* carried on the pre-processing of the signal where it trained the classifier, partially with the values from the previous scenario. Once again, the subjects' raw data, which were recorded in the training session, were utilized with the elimination of the last three aux channels, the omission of signal decimation, and the application of a Butterworth bandpass filter of 1 Hz–20 Hz, identical to the previous scenario. Subsequently, the parameters of the xDAWN spatial filter that were generated in the second scenario, which included the 24 spatial filter coefficients, 8 input channels, and output 3 output channels. This spatial filter generated three output channels from the original eight input channels, where each output channel was a linear combination of the input channels. The output channels were computed by performing the "sum on *i* (*Cij* * *Ii*)", as shown in formula (3), where *Ii* represents the input channel (*n* is set to 8), *Oj* represents the output channel, and *Cij* is the coefficient of the *i*th input channel and *j*th output channel in the spatial filter matrix.

$$Oj = \sum_{i=1}^{n} Cij * Ii \tag{3}$$

Subsequently, the outputted signals (i.e., the three output channels) and the stimulations were equivalently and separately passed through *stimulation-based epoching* for the target and the non-target selection. These had an epoch duration of 0.6 s and no offset. The output, i.e., both epoch signals (target and non-target) were again separately computed with block averaging and passed through a feature aggregator that combined the received input features into a feature vector that was used for the classification. This implies that two separate feature vector streams were outputted; the target and non-target selections. Ultimately, both vector streams and the stimulations were passed through our *classifier trainer*. We opted to pass all the data through a single classifier trainer, hence the native multiclass strategy was chosen, which used the classifier training algorithm without a pairwise strategy. The algorithm chosen for our classifier was the regular LDA. The output at this stage was a trained classifier with the settings outputted to a file for use in the next scenario.

The *fourth scenario* consisted of the actual online experiments and was more complex since it was necessary to collect data, pre-process them, classify them, and provide online feedback to the subject. The front-end consisted of displaying the 6 by 6 grid, flashing rows and columns, and give feedback to the subject. The back-end consisted of a number of processes. Primarily, the data were acquired from the subject in real time and, similar to what was done in the previous scenarios, the last three aux channels were eliminated, signal decimation was omitted, a Butterworth bandpass filter of 1 Hz–20 Hz was used, and the parameters of the xDAWN spatial filter that were generated in the *second scenario*, which included the twenty-four spatial filter coefficients, were used. Subsequently, the output and the stimulations were passed through the *stimulation-based epoching*, which had an epoch duration of 0.6 s and no offset. This was then averaged and passed through a *feature aggregator* to produce a feature vector for the classifier. Lastly, the *classifier processor* classified the incoming feature vectors by using the previously learned classifier (*classifier trainer*).

The *fifth scenario* allowed us to replay the experiments by selecting the raw data file and re-process the functions of the fourth scenario.

2.9. Signal Processing—Offline

In the offline analysis, the following procedure was done for each of *LC, M30, M60, M90, AN, PT*, and *AL*. The captured raw data were converted from the proprietary OpenVIBE.*ov* extension to a more commonly used.*csv* format using a particular scenario designed for this task. The outputs were two files in.*csv* format which contained the raw data and stimulations, respectively. These were

later imported into MATLAB R2014a tables called *samples* and *stims* and then converted to arrays. Subsequently, any unnecessary rows and columns in the *samples* array were removed. These consisted of the first rows which contained the time header, channel names, and sampling rate; the first column, which contained the time(s) and the last three columns, stored the auxiliary data of the accelerometer. Next, we filtered out the *stims* array to include the target stimulations with the code (33285); non-target stimulations (33286); visual stimulation stop (32780), which was the start of each flash of a row or column; and segment start (32771), which was the start of each trial (12 flashes, 6 rows, and 6 columns made up one trial). Additional data, such as the *sampleTime*, *samplingFreq*, and *channelNames* variables, were extracted from the data and stored in the workspace. Subsequently, we had to perform a signal inversion due to the hardware and driver implementation.

The *samples* array was later imported into EEGLAB for processing and for offline analysis. The first process was to apply a bandpass filter of 1–20 HZ to eliminate the environmental electrical interference (50 Hz or 60 Hz dependent on the country), to remove any signal harmonics and unnecessary frequencies which were not beneficial in our experiments and to remove the DC offset. Subsequently, we imported the event info (the stimulations—*stim* array) into EEGLAB with the format {*latency*, *type*, *duration}* in milliseconds.

Next, the imported data were used in ERPLAB, which is an add-on of EEGLAB and is targeted for ERP analysis. Although the dataset in EEGLAB already contained information about all the individual events, we created an *eventlist* structure in ERPLAB that consolidated this information and made it easier to access and display; it also allowed ERPLAB to add additional information which was not present in the original EEGLAB list of events. Subsequently, we took every event we wanted to average together and assigned them to a specific bin via the *binlister.* This contained an abstract description of what kinds of event codes go into a particular bin. In our experiments, we used the following criteria: "{33285}{t<50–150>32780}" for the target and "{33286}{t<50–150>32780}" for the non-target. This implied that it was time locked to the stimulus events 33,285 (target) or 33,286 (non-target) and must have had the event 32,780 that happened 50 to 150 ms after the target/non-target event. If this criterion was met, it was placed in the appropriate bin.

Subsequently, we extracted the bin-based epochs via ERPLAB (not the EEGLAB version) and set the time period from −0.2 s before the stimulus until 0.8 s after the stimulus. We also used baseline correction (pre-stimulus) since we wanted to subtract the average pre-stimulus voltage from each epoch of data. Next, we passed all channel epoch data through a moving window peak-to-peak threshold artefact detection with the voltage threshold set at 100 µV, moving window width at 200 ms, and window step at 100 ms to remove unwanted signals, such as blinking and moving artefacts. Lastly, we averaged our dataset ERPs and performed an average across ERP sets (grand average) to produce the results shown in Section 3.2.3, generated by the ERP measurement tool.

3. Results

In this section, we present several results in relation to the dependent variables, such as repeated measures ANOVA, to determine the effect that *LC*, *M30*, *M60*, *M90*, *AN*, *PT*, and *AL* distractions have on the online performance (accuracy), offline statistics (amplitude and latency), and user preference. In the following tables, the labels *LC*, *M30*, *M60*, *M90*, *AN*, *PT*, and *AL* represent "lab condition", "music at 30%", "music at 60%", "music at 90%", "ambient noise", "passive talking", and "active listening" distractions, as thoroughly explained beforehand. As previously mentioned, we included and combined data from our previous works [2,3].

3.1. Online Analysis

Following the online experiments, the results achieved per subject are shown in Table 1, which depicts the predicted symbols out of five, and the percentage in parentheses, rounded to the nearest one, for the accuracy dependent variable. It must be noted that in an incorrect symbol prediction, it might be the case that the column was predicted correctly, whilst the row was predicted

incorrectly or vice versa. For instance, subject6 had a success rate of 80% in the *AN* scenario, with the symbol R predicted as symbol F, i.e., the column prediction was correct but not the row prediction. However, to avoid ambiguity, we decided to assume that both row and column predictions were incorrect when the symbol was predicted incorrectly.

Table 1. Symbols spelled (/5) and percentage (in parentheses) for the dependent variable of accuracy.

Subject	LC	M30	M60	M90	LC	AN	PT	AL
S1	5 (100%)	5 (100%)	4 (80%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)
S2	5 (100%)	5 (100%)	4 (80%)	4 (80%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)
S3	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	4 (80%)	5 (100%)
S4	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	4 (80%)	5 (100%)	5 (100%)
S5	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	3 (60%)
S6	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	4 (80%)	5 (100%)	5 (100%)
S7	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	4 (80%)	3 (60%)
S8	5 (100%)	4 (80%)	5 (100%)	5 (100%)	5 (100%)	4 (80%)	3 (60%)	4 (80%)
S9	5 (100%)	4 (80%)	5 (100%)	5 (100%)				
S10	5 (100%)	5 (100%)	5 (100%)	5 (100%)				
Average	100%	96%	96%	98%	100%	92.5%	90%	87.5%

3.2. Offline Statistics

In the following sub-sections, we process and analyze the averaged epoch signal of eight to ten subjects in relation to the independent variables (*LC*, *M30*, *M60*, *M90*, *AN*, *PT*, and *AL*) as elucidated below:

3.2.1. Session 1

This section processes and analyzes the averaged epoch signal of ten subjects in relation to the independent variables *M0*, *M30*, *M60*, and *M90*.

Amplitude

In Table 2, the descriptive stats for the amplitude dependent variable are presented. The first column denotes the conditions, i.e., *Lab*(*M0*), *M30*, *M60*, and *M90*. The second and third columns show the calculated mean and standard deviation (SD), while the last column *N* represents the number of subjects in the study.

Table 2. Auditory distractions, session 1, amplitude, descriptive analysis.

Descriptives			
Condition	**Mean**	**SD**	**N**
Lab	3.601	1.148	10
M30	3.423	1.086	10
M60	3.926	0.767	10
M90	3.584	0.746	10

Figure 5 depicts a descriptive plot of the conditions for the amplitude dependent variable, where it is visually clear that there is no discriminable difference between the groups.
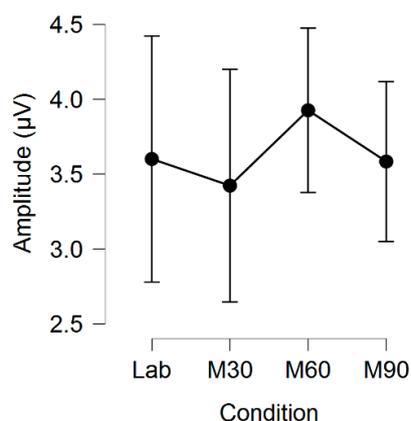
Figure 5. Auditory distractions, session 1, amplitude, descriptive plot.

A repeated measures ANOVA (RMANOVA) was performed, which was based on the independent variable with four levels/groups (*M0*, *M30*, *M60*, and *M90*) for the dependent variables (amplitude and latency), as presented in Table 3. In the first column, the cases are shown, which include the chosen *condition* and the *residuals*. In the second column, the sum of squares (*SS*) of the variation is shown, which is the spread between each individual value and the mean. The third column is the degrees of freedom (*df*), which in the *condition* case is the (number of variables − 1), i.e., there are four variables, which gives us three degrees of freedom. In the *residual* cases, the calculation of (number of variables − 1) * (number of subjects − 1) i.e., (four variables − 1) * (ten subjects − 1) is performed, which gives us twenty-seven (27). The fifth column is the mean square values (*MS*), which are calculated by dividing the *SS* by the corresponding *df*. The sixth column is the *F* statistic, which is the key statistic where the *MS* for the *condition* is divided by the *MS* for the *residual*. The sixth column is the *p*-value, with the alpha (*α*) value being set at 0.05. The seventh and last column shows the estimates of effect size and, since the groups are small, omega squared (*ω2*) is used. This description will only be explained once and will be referred to for future RMANOVA tables.

Table 3. Auditory distractions, session 1, amplitude, repeated measures ANOVA.

		Within-Subjects Effects				
Cases	**Sum of Squares (SS)**	**Degrees of Freedom (df)**	**Mean Square (MS)**	**F**	***p***	**ω²**
Condition	1.334	3	0.445	0.536	0.661	0.000
Residuals	22.374	27	0.829			

Note. Type III sum of squares.

Table 3 reports that our *p*-value result (sixth column) was 0.661, i.e., it was greater than the alpha value of 0.05, so H_0 was accepted, i.e., all means are equal and H_1 should be rejected. In addition, the omega squared (*ω2*) gave us a value of 0.000, which indicates no effect.

Our results indicate that there is no significant difference between all the groups $F_{(3, 27)} = 0.536$, $p < 0.661$, $ω^2 = 0.000$. Therefore, no additional analysis or post hoc tests are required.

Latency

In Table 4, the descriptive stats for the latency dependent variable are presented. The first column denotes the conditions, i.e., *Lab(M0)*, *M30*, *M60*, and *M90*. The second and third columns show the calculated mean and standard deviation (SD), while the last column *N* represents the number of subjects in the study.

Table 4. Auditory distractions, session 1, latency, descriptive analysis.

Descriptives			
Condition	**Mean**	**SD**	**N**
Lab	450.950	26.155	10
M30	436.900	22.148	10
M60	435.300	23.073	10
M90	438.350	25.839	10

Figure 6 depicts a descriptive plot of the conditions for the latency dependent variable, where it is visually clear that there is no discriminable difference between the groups.
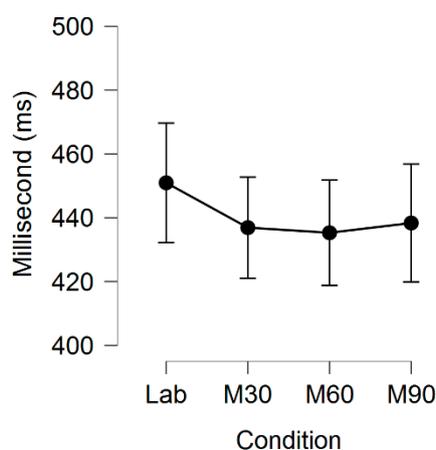
Figure 6. Auditory distractions, session 1, latency, descriptive plot.

Table 5 (refer to explanation for Table 3) reports that our *p*-value result (sixth column) was 0.115, i.e., it was greater than the alpha value of 0.05, so H_0 was accepted, i.e., all means are equal and H_1 should be rejected. In addition, the omega squared (ω2) gave us a value of 0.034, which indicates a small to medium effect. However, under the table, it is noted that the assumption of sphericity was violated.

Table 5. Auditory distractions, session 1, latency, repeated measures ANOVA.

Within-Subjects Effects							
Cases	**Sphericity Correction**	**Sum of Squares (SS)**	**Degrees of Freedom (df)**	**Mean Square (MS)**	**F**	**p**	**ω^2**
Condition	None	1537.625 [a]	3 [a]	512.542 [a]	2.169 [a]	0.115 [a]	0.034
	Greenhouse–Geisser	1537.625	1.468	1047.746	2.169	0.16	0.034
Residuals	None	6379.375	27	236.273			
	Greenhouse–Geisser	6379.375	13.208	482.993			

Note. Type III sum of squares. [a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < 0.05$).

Table 6 gives the results of Mauchly's test of sphericity. It can be seen that there is a significant difference ($p < 0.021$) in the variances of the differences between the groups. It is also important to note that the Greenhouse–Geisser and the Huynh–Feldt epsilon (ε) values are below 0.75. Therefore, the RMANOVA result should be reported based on the Greenhouse–Geisser correction, as reported in Table 5. Since the *p*-value for the Greenhouse–Geisser value is 0.16, and hence greater than the alpha value of 0.05, the H_0 is still accepted and the H_1 is rejected.

Table 6. Auditory distractions, session 1, latency, Mauchly's sphericity test.

	Test of Sphericity						
	Mauchly's W	**Approx. X^2**	**df Sphericity**	***p*-Value**	**Greenhouse–Geisser ε**	**Huynh–Feldt ε**	**Lower Bound ε**
Condition	0.176	13.41	5	0.021	0.489	0.561	0.333

Our results indicate that there is no significant difference between any of the groups F (1.468, 13.208) = 2.169, $p < 0.016$, $\omega^2 = 0.034$. Therefore, no additional analysis or post hoc tests are required.

3.2.2. Session 2

This section processes and analyzes the averaged epoch signal of eight subjects in relation to the independent variable (*LC*, *AN*, *PT*, and *AL*).

Amplitude

In Table 7, the descriptive stats for the amplitude dependent variable are presented. The first column denotes the conditions, i.e., *LC*, *AN*, *PT*, and *AL*. The second and third columns show the calculated mean and standard deviation (SD), while the last column *N* represents the number of subjects in the study.

Table 7. Auditory distractions, session 2, amplitude, descriptive analysis.

	Descriptives		
Condition	**Mean**	**SD**	**N**
LC	3.795	1.168	8
AN	3.695	0.936	8
PT	4.291	0.852	8
AL	2.938	1.054	8

Figure 7 depicts a descriptive plot of the conditions for the amplitude dependent variable, where it is visually clear that there is no discriminable difference between the groups.
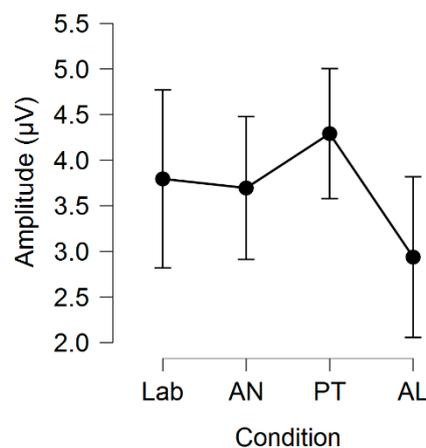
Figure 7. Auditory distractions, session 2, amplitude, descriptive plot.

Table 8 (refer to explanation for Table 3) reports that our *p*-value result (sixth column) was 2.192, i.e., it was greater than the alpha value of 0.05, so H_0 was accepted, i.e., all means are equal and H_1 should be rejected. In addition, the omega squared (ω^2) gave us a value of 0.111, which indicates a medium to large effect.

Table 8. Aauditory distractions, session 2, amplitude, repeated, easures ANOVA.

		Within-Subjects Effects				
Cases	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F	*p*	ω^2
Condition	7.507	3	2.502	2.192	0.119	0.111
Residuals	23.971	21	1.141			

Note. Type III sum of squares.

Our results indicate that there is no significant difference between any of the groups F (3, 21) =2.192, $p < 0.119$, $\omega^2 = 0.111$. Therefore, no additional analysis or post hoc tests are required.

Latency

In Table 9 (refer to explanation for Table 3), the descriptive stats for the amplitude dependent variable are presented. The first column denotes the conditions, i.e., *LC*, *AN*, *PT*, and *AL*. The second and third columns show the calculated mean and standard deviation (SD), while the last column *N* represents the number of subjects in the study.

Table 9. Auditory distractions, session 2, latency, descriptive analysis.

	Descriptives		
Condition	Mean	SD	N
LC	433.063	35.296	8
AN	438.125	52.143	8
PT	403.625	39.473	8
AL	408.063	49.829	8

Figure 8 depicts a descriptive plot of the conditions for the amplitude dependent variable, where it is visually clear that there is no discriminable difference between the groups.
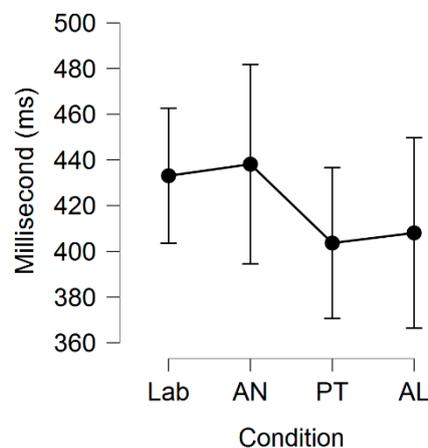
Figure 8. Auditory distractions, session 2, latency, descriptive Plot.

Table 10 (refer to explanation for Table 3) reports that our *p*-value result (sixth column) was 0.162, which was greater than the alpha value of 0.05, so H$_0$ was accepted, i.e., all means are equal and H$_1$ should be rejected. In addition, the omega squared (ω2) gave us a value of 0.051, which indicates a medium effect.

Table 10. Auditory distractions, session 2, latency, repeated measures ANOVA.

	Within-Subjects Effects					
Cases	**Sum of Squares (SS)**	**Degrees of Freedom (df)**	**Mean Square (MS)**	**F**	***p***	**ω^2**
Condition	7261.781	3	2420.594	1.892	0.162	0.051
Residuals	26,866.719	21	1279.368			

Note. Type III sum of squares.

Our results indicate that there is no significant difference between any of the groups $F_{(3, 21)} = 1.892$, $p < 0.162$, $\omega^2 = 0.051$. Therefore, no additional analysis or post hoc tests are required.

3.2.3. Combined Results

Tables 11 and 12 show the means for the dependent variables, amplitude and latency, respectively, according to levels of the independent variable rounded to the nearest hundredth, for both sessions. These data include the average of all eight recorded electrodes for the five symbols and are shown per subject. Descriptive analysis shows that the highest amplitude was for *PT* and the lowest amplitude was for *AL*. On the other hand, the lowest latency was found for *PT*, while the highest latency was found for *AN*, which can be seen in Table 12.

Table 11. Means and standard deviations (in parentheses) for amplitude dependent measure for both sessions.

Subject	LC	M30	M60	M90	LC	AN	PT	AL
S1	4.90 (0.50)	2.09 (0.28)	3.90 (0.40)	2.88 (0.40)	4.85 (0.67)	3.11 (0.96)	3.00 (0.76)	3.78 (0.67)
S2	2.39 (1.14)	3.24 (0.35)	4.26 (0.41)	3.66 (0.83)	2.87 (1.29)	2.57 (0.28)	3.75 (0.92)	2.42 (0.44)
S3	3.29 (1.68)	2.18 (2.24)	4.03 (0.99)	3.64 (1.42)	3.52 (1.00)	2.91 (0.77)	3.94 (0.94)	3.50 (1.69)
S4	3.70 (0.98)	4.95 (0.54)	4.14 (0.45)	3.25 (1.08)	3.95 (0.76)	4.59 (0.74)	3.55 (1.00)	4.22 (0.92)
S5	3.96 (0.78)	4.11 (0.95)	4.61 (0.76)	4.77 (1.17)	5.39 (1.64)	4.62 (2.30)	5.05 (1.88)	0.83 (0.32)
S6	5.39 (1.64)	3.63 (2.44)	4.99 (2.40)	3.70 (2.12)	1.98 (1.52)	4.97 (0.69)	5.04 (1.10)	2.47 (1.46)
S7	1.70 (1.66)	4.94 (1.26)	4.18 (1.24)	4.58 (1.34)	3.00 (1.25)	2.92 (1.02)	5.39 (0.95)	2.92 (0.46)
S8	2.64 (1.38)	2.51 (1.57)	3.52 (1.22)	2.18 (1.69)	4.80 (1.92)	3.87 (1.54)	4.61 (1.56)	3.36 (1.20)
S9	4.46 (1.94)	2.50 (1.40)	2.17 (2.05)	3.67 (1.89)				
S10	3.58 (0.26)	4.08 (0.71)	3.46 (0.61)	3.51 (0.46)				
Average	3.60 (1.83)	3.43 (1.66)	3.93 (1.38)	3.59 (1.46)	3.80 (1.26)	3.70 (1.04)	4.29 (1.14)	2.94 (0.90)

Table 12. Means and standard deviations (in parentheses) for latency dependent measure for both sessions.

S	LC	M30	M60	M90	LC	AN	PT	AL
S1	398.0 ()	473.5 (12.46)	470.5 (5.21)	462.5 (19.00)	428.0 (72.85)	447.0 (88.28)	377.5 (69.74)	456.0 (92.61)
S2	352.0 ()	434.0 (76.58)	419.5 (54.15)	416.0 (66.69)	432.5 (77.37)	499.0 (85.83)	431.0 (83.41)	423.5 (80.04)
S3	343.0 ()	417.0 (71.61)	430.0 (85.68)	423.0 (78.49)	466.0 (107.48)	458.0 (120.29)	385.0 (75.59)	324.0 (23.23)
S4	422.5 ()	427.0 (82.11)	432.5 (86.76)	419.5 (75.49)	445.0 (89.64)	434.5 (85.52)	383.5 (75.97)	438.0 (87.90)
S5	452.0 ()	431.0 (84.58)	428.5 (80.51)	436.5 (85.56)	483.0 (1.85)	492.5 (7.23)	484.5 (8.67)	386.5 (64.32)
S6	489.5 ()	479.0 (17.73)	484.5 (3.96)	495.0 (2.83)	442.5 (101.66)	438.0 (88.60)	418.5 (84.61)	454.0 (101.94)
S7	383.0 ()	431.5 (84.97)	427.5 (79.81)	449.5 (93.64)	387.0 (79.07)	400.0 (98.75)	386.0 (73.85)	434.5 (96.88)
S8	483.0 ()	440.0 (72.88)	415.5 (66.91)	446.0 (90.16)	380.5 (89.91)	336.0 (44.49)	363.0 (60.62)	348.0 (51.31)
S9	371.5 ()	415.5 (71.66)	424.0 (76.67)	411.0 (75.46)				
S10	464.0 ()	420.5 (80.61)	420.5 (76.98)	424.5 (84.43)				
Average	415.00 (69.56)	436.90 (65.52)	435.3 (61.63)	438.4 (67.18)	433.06 (77.48)	438.18 (77.37)	403.63 (66.56)	408.06 (74.78)

Figure 9 shows the grand average P300 component for all ten subjects in session 1, which comprises the grand averaged raw signals, i.e., five symbols with 15 trials per symbol, 12 flashes of columns/rows per trial, and 10 subjects, i.e., 9000 flashes, amongst which 1500 were targets. It also shows four overlapping signals, (i) BIN1—Target for the *M0* scenario shown in black (solid for grayscale), (ii) BIN3—Target for *M30* in red (dash–dot), (iii) BIN5—Target for *M60* in blue (dashed), and (iv) BIN7—Target for *M90* in green (dotted). To avoid ambiguity and for clarity of the illustration, BIN2, BIN4, BIN6, and BIN8, which represent the non-target signals, were omitted. It is important to note that, in general, the highest peak should be of BIN1, followed by BIN3, BIN5, and BIN7.
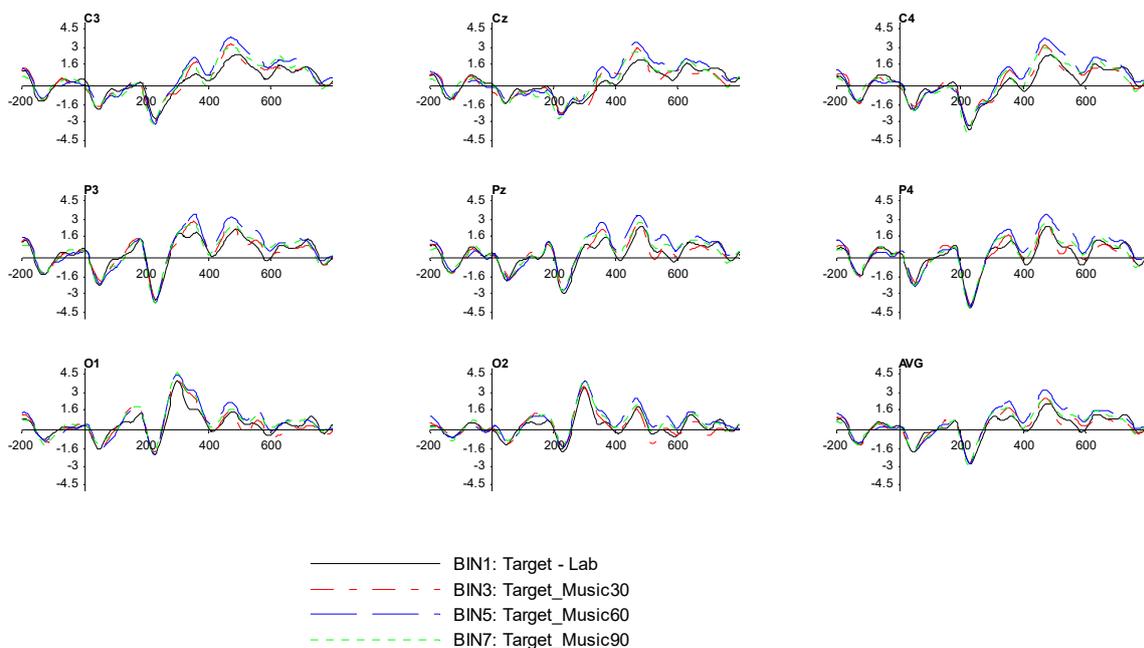
Figure 9. Auditory distractions, session 1, grand average for all eight subjects. The *x*-axis represents time (ms), while the *y*-axis represents the amplitude (μV).

Figure 10 shows the grand average P300 component for all eight subjects in session 2, which comprises the grand averaged raw signals, i.e., five symbols with 15 trials per symbol, 12 flashes of columns/rows per trial, and eight subjects, i.e., 7200 flashes, amongst which 1200 were targets. It also shows show four overlapping signals, (i) BIN1—Target for the *LC* scenario shown in black (solid for grayscale), (ii) BIN3—Target for *AN* in red (dash–dot), (iii) BIN5—Target for *PT* in blue (dashed), and (iv) BIN7—Target for *AL* in green (dotted). To avoid ambiguity and for clarity of the illustration, BIN2, BIN4, BIN6, and BIN8, which represent the non-target signals, were omitted. It is important to note that, in general, the peak of BIN1 should be higher than the peaks of BIN3, BIN5, and BIN7, whilst there was no indication of the expected difference between distractions, i.e., BIN3, BIN5, and BIN7.
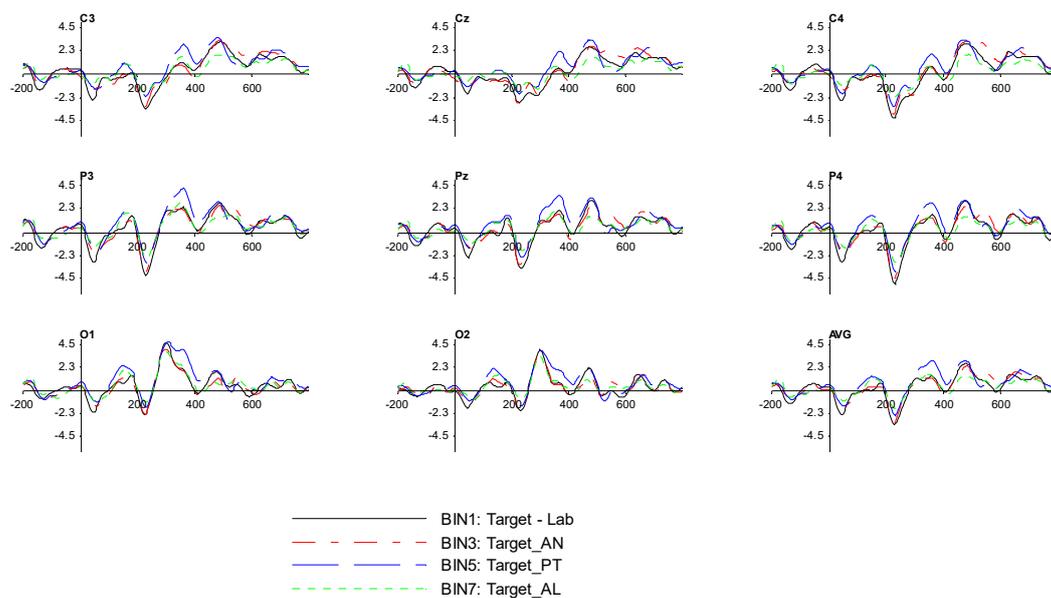
Figure 10. Auditory distractions, session 2, grand average for all eight subjects. The *x*-axis represents the time (ms), while the *y*-axis represents the amplitude (μV).

3.3. User Preference

Immediately after finishing the experiments, each subject was asked to participate in two voluntary questionnaires to specify their preferred usage condition. The ranking consisted of a maximum weight value of four as the most desired and a minimum weight value of one for the least desired. All eight to ten subjects agreed to partake in the questionnaires. In the first questionnaire (a), the subjects were allowed to give the same ranking to each independent variable, while in the second questionnaire (b), the subjects were asked to give a unique ranking to each independent variable.

Table 13 encompasses the user preference analysis of our previous papers [2,3] for both questionnaires (a) and (b). It includes all the user preferences throughout all auditory distractions performed so far.

Table 13. User preference for questionnaire (a), allowing the same ranking, and questionnaire (b), with unique ranking.

S	LC		M30		M60		M90		AN		PT		AL	
	(a)	*(b)*	*(a)*	*(b)*	*(a)*	*(b)*	*(a)*	*(b)*	*(a)*	*(b)*	*(a)*	*(b)*	*(a)*	*(b)*
S1	4	4	3	2	3	1	4	3	4	3	4	2	4	1
S2	4	4	4	3	3	1	3	2	4	3	4	1	4	2
S3	4	4	3	1	3	3	3	2	4	3	3	1	4	2
S4	4	4	3	1	3	2	4	3	2	1	3	2	3	3
S5	4	4	4	3	4	2	3	1	4	3	4	2	2	1
S6	4	4	4	1	4	3	4	2	3	1	3	3	3	2
S7	4	4	4	1	4	2	4	3	4	3	2	2	2	1
S8	4	4	2	1	3	2	3	3	2	3	1	1	3	2
S9	4	4	2	1	3	3	4	2						
S10	4	4	3	1	4	3	3	2						
Total	40/40	40/40	32/40	15/40	34/40	22/40	35/40	23/40	27/32	20/32	24/32	14/32	25/32	14/32
%	100	100	80	37.5	85	55	87.5	57.5	84.38	62.5	75	43.75	78.13	43.75

For questionnaire (a), depicted in Table 13, the results show that, as expected, the *LC* independent variable got the highest ranking, followed by *M90*, *M60*, *AN*, *M30*, *AL*, and *PT*. The frequency analysis shows that *LC* got 100%, followed by *M90* with 87.5%, *M60* with 85%, *AN* with 84.38%, *M30* with 80%, *AL* with 78.13%, and lastly *PT* with 75%. It is noteworthy that the difference between *LC* and *M90*, in second place, amounted to 12.5% and this was the same value found when comparing *M90* to *PT*, in last place.

For questionnaire (b), also depicted in Table 13, the results show that again, as expected, the *LC* independent variable had the maximum ranking, followed by *AN*, *M90*, *M60*, *PT* and *AL* equally, and lastly *M30*. The frequency analysis shows that *LC* got 100%, followed by *AN* at 62.5%, *M90* at 57.5%, *M60* at 55%, *PT* and *AL* at 43.75%, and *lastly* M30 at 37.5%. Once again, it is significant to note that the difference between LC and *AN*, in second place, amounted to 37.5% and was the highest difference, even when comparing *AN* to *M30*, in last place, where the difference was 25%.

These results provide an overwhelming indication that the subjects preferred doing the experiments in a quiet setting, as was originally expected.

4. Discussion

This research area is still in its infancy stages and, even though a great deal of research is currently underway, there are still a large number of areas where research is either at a standstill, moving very slowly, or that can be enriched. The online community is also contributing to this knowledge base; however, their findings are usually ad hoc and not investigated, scrutinized, or documented properly. Our research has performed these experiments thoroughly, methodically, and they are tractable and repeatable.

This work introduced different auditory distractions that were considered alongside the development of a taxonomy. These results should give some insight into the practicability of the current P300 speller and low-cost equipment to be used for real-world applications. Due to the lack of a detailed, scientific, extensible categorization scheme for distractions, and their effects on EEG signals and applications, an evident necessity for this study was present. This work targets this important niche.

The limitations of this study include, but are not limited to:

- Session 1 experiments were done in a sequential order, i.e., *M0*, *M30*, *M60*, and *M90* throughout, for all subjects. Empirical evidence shows that the subjects seemed to become accustomed with the music in the fourth sequential experiment of *M90*, while they were affected by the difference

between *M0* and *M30*, which followed each other. In session 2, we rectified this oversight and performed the experiments in a randomized order.

- The age of the subjects, i.e., the oldest subject was forty and the youngest was eighteen years old. What would the results be with younger or older subjects?
- Owing to the nature of being self-funded research, the study is more of a qualitative nature due to the limited number of subjects for the experimentation sessions.

5. Conclusions

In direct continuity and perpetuation of our previous papers [2,3], this work analyzed the effects of auditory distractions performed throughout, explicitly those listed in our independent variables, i.e., music at 30%, music at 60%, music at 90%, ambient noise, passive talking, and active listening, have on the online performance, i.e., accuracy, the offline performance, i.e., latency and amplitude, and user preference, as elucidated by the dependent variables. This paper combines two sessions, where we had $N = 10$ and $N = 8$ healthy subjects, respectively, and they performed the aforementioned independent variable settings while utilizing low fidelity equipment and using a Farwell and Donchin P300 speller in conjunction with the xDAWN algorithm, with a six by six matrix of alphanumeric characters.

The goal of our study was to develop a taxonomy aimed at categorizing distractions in the P300b domain and the effect that these distractions have on the success rate, signal quality, reduction of amplitude, or any other change/distortion that occurs. This should give some insight into the practicability of the real-world application of the current P300 speller with our aforementioned low-cost equipment. The aim of this paper was to analyze the effects that the aforementioned auditory distractions had on our dependent variables.

Our null hypothesis, based on our previous work [2,3] and preceding related and tantamount medical-grade research, weas that these types of distractions, as elucidated by the independent variables, do not show any statistically significant effect on the amplitude and latency dependent variables. The results of our RMANOVA analysis accepts our null hypothesis for all independent variables and for all dependent variables, as depicted in Table 3, Table 5, Table 8, and Table 10.

Descriptive results for the combined papers [2,3] show that the dependent accuracy variable was highest for *LC* (100%) and surprisingly followed by M90 (98%), trailed by M30 and M60 equally (96%), and shadowed by *AN* (92.5%), *PT* (90%) and *AL* (87.5%), as shown in Table 1. The dependent variable amplitude was highest for *PT* (4.29), followed by *M60* (3.93), *LC* session 2 (3.80), *AN* (3.70), *LC* session 1 (3.60), *M90* (3.59), *M30* (3.43), and *AL* (2.94), as portrayed in Table 11, while the latency was shortest for *PT*, followed by *AL*, *LC* session 1, *LC* session 2, M60, *M30*, *M90*, and *AN*, as depicted in Table 12. The user preference questionnaire (a) showed overwhelmingly that subjects preferred the *LC* condition, as originally expected, followed by *M90*, *M60*, *AN*, *M30*, *AL*, and *PT*, and questionnaire (b) showed that subjects preferred *LC* again, followed by *AN*, *M90*, *M60*, *PT* and *AL* equally, and lastly *M30*, as portrayed in Table 13.

In this paper, we pursued the development of a hierarchical taxonomy aimed at categorizing distractions in the P300b domain, as depicted in Figure 1. Explicitly, we examined the effect that the aforementioned independent variables, categorized under auditory distractions, have on the dependent variables. In the future, we plan to introduce additional types of distractions which are commonly found in a real-world environment and include them within the different categories of our taxonomy.

References

1. Schembri, P.; Pelc, M.; Ma, J. Comparison between a Passive and Active response task and their effect on the Amplitude and Latency of the P300 component for Visual Stimuli while using Low Fidelity Equipment. In Proceedings of the Forty First Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2019, Berlin, Germany, 23–27 July 2019.

2. Schembri, P.; Pelc, M.; Ma, J. The Effect that an Auditory Distraction with differing levels of Intensity have on a Visual P300 Speller while utilizing Low Fidelity Equipment: Alongside the Development of a Taxonomy. In Proceedings of the 7th International Conference on Neurotechnology and Physiological Computing Systems, NEUROPhyCS 2019, Vienna, Austria, 20–21 September 2019.

3. Schembri, P.; Pelc, M.; Ma, J. The Effect that Auxiliary Taxonomized Auditory Distractions have on a P300 Speller while utilising Low Fidelity Equipment. In Proceedings of the 11th Computer Science and Electronic Engineering Conference (CEEC 2019), Colchester, UK, 18–20 September 2019.

4. Kam, J.W.; Griffin, S.; Shen, A.; Patel, S.; Hinrichs, H.; Heinze, H.-J.; Deouell, L.Y.; Knight, R.T. Systematic comparison between a wireless EEG system with dry electrodes and a wired EEG system with wet electrodes. *NeuroImage* **2019**, *184*, 119–129. [CrossRef] [PubMed]

5. Bradford, C.J.; Burke, B.; Nguyen, C.; Slipher, G.A.; Mrozek, R.; Hairston, D.W. Performance of conformable, dry EEG sensors. In Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 4957–4960.

6. Nam, C.S.; Li, Y.; Johnson, S. Evaluation of P300-Based Brain-Computer Interface in Real-World Contexts. *Int. J. Hum.-Comput. Interact.* **2010**, *26*, 621–637. [CrossRef]

7. Valentin, O.; Ducharme, M.; Cretot-Richert, G.; Monsarrat-Chanon, H.; Viallet, G.; Delnavaz, A.; Voix, J. Validation and Benchmarking of a Wearable EEG Acquisition Platform for Real-World Applications. *IEEE Trans. Biomed. Circuits Syst.* **2019**, *13*, 103–111. [CrossRef] [PubMed]

8. Oliveira, A.S.; Schlink, B.R.; Hairston, D.W.; König, P.; Ferris, D.P. Proposing Metrics for Benchmarking Novel EEG Technologies Towards Real-World Measurements. *Front. Hum. Neurosci.* **2016**, *10*, 188. [CrossRef] [PubMed]

9. Zink, R.; Hunyadi, B.; Huffel, S.V.; Vos, M.D. Mobile EEG on the bike: Disentangling attentional and physical contributions to auditory attention tasks. *J. Neural Eng.* **2016**, *13*, 46017. [CrossRef] [PubMed]

10. Rivet, B.; Souloumiac, A.; Attina, V.; Gibert, G. xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain–Computer Interface. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 2035–2043. [CrossRef] [PubMed]

11. Squires, N.; Squires, K.; Hillyard, S. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr. Clin. Neurophysiol.* **1975**, *38*, 387–401. [CrossRef]

12. Sutton, S.; Braren, M.; Zubin, J.; John, E.R. Evoked-Potential Correlates of Stimulus Uncertainty. *Science* **1965**, *150*, 1187–1188. [CrossRef] [PubMed]

13. Qu, J.; Wang, F.; Xia, Z.; Yu, T.; Xiao, J.; Yu, Z.; Gu, Z.; Li, Y. A Novel Three-Dimensional P300 Speller Based on Stereo Visual Stimuli. *IEEE Trans. Hum.-Mach. Syst.* **2018**, *48*, 392–399. [CrossRef]

14. Gu, Z.; Chen, Z.; Zhang, J.; Zhang, X.; Yu, Z.L. An Online Interactive Paradigm for P300 Brain–Computer Interface Speller. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 152–161. [CrossRef] [PubMed]

15. Schembri, P.; Anthony, R.; Pelc, M. Detection of Artifacts Using a Non-invasive BCI on the Basis of Electroencephalography While Utilizing Low-Cost Off-the-Shelf Equipment. *Physiol. Comput. Syst.* **2019**, *10057*, 93–109.

16. Ogura, C.; Koga, Y.; Shimokochi, M. Recent Advances in Event-related Brain Potential Research. In Proceedings of the 11th International Conference on Event-related Potentials (EPIC), Okinawa, Japan, 25–30 June 1995.

17. Woehrle, H.; Krell, M.M.; Straube, S.; Kim, S.K.; Kirchner, E.A.; Kirchner, F. An Adaptive Spatial Filter for User-Independent Single Trial Detection of Event-Related Potentials. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1696–1705. [CrossRef] [PubMed]

18. Stern, R.M.; Ray, W.J.; Quigley, K.S. *Psychophysiological Recording*, 2nd ed.; Oxford University Press: Oxford, UK, 2001.