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

Patrick Schemrbi[a], Mariusz Pelc[b] and Jixin Ma[c]
Department of Computer and Information Systems, University of Greenwich, Greenwich, London, U.K.

Keywords: Brain-Computer Interface (BCI), Electroencephalography (EEG), Event-Related Potential (ERP), P300 Speller, Distractions, Taxonomy.

Abstract: In this paper, we investigate the effect that an auditory distraction with differing levels of intensity has on the signal of a visual P300 Speller in terms of accuracy, amplitude, latency, user preference, signal morphology, and overall signal quality. This work is based on the P300 speller BCI (oddball) paradigm and the xDAWN algorithm, with ten healthy subjects; while using a non-invasive Brain-Computer Interface (BCI) based on low fidelity electroencephalographic (EEG) equipment. Our results suggest that the accuracy was best for the no music (M0), followed by music at 90% (M90), music at 60% (M60) and last music at 30% (M30), which results were in identical order to the subjects' preferences. In addition, the amplitude did not show any statistical significance in all scenarios while the latency exhibited a minor statistical difference. This work is part of a larger EEG based project where we are introducing different categories of distractions that are being considered alongside the development of a taxonomy. These results should give some insight into the practicability of the current P300 speller to be used for real-world applications.

1 INTRODUCTION

In this paper, we analyze the effect that an auditory distraction, explicitly that of digital music, with different levels of intensity in regards to volume (off, low, mid, high) have on the accuracy, amplitude, latency, user preference, signal morphology, and overall signal quality. Our research makes use of non-invasive Brain-Computer Interface (BCI) on the basis of Electroencephalography (EEG). The work presented here is part of a larger EEG based project and in continuation of our previous papers (Schembri et al., 2017); (Schembri et al., 2018); (Schembri et al., 2018); (Schembri et al., 2019).

Event-related potentials (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 either cognitive, sensor or motor stimuli. The simplest paradigm for eliciting an ERP is by focusing attention on the target stimuli (occurs infrequently) embedded randomly in an array of non-targets (occurs frequently). The methodology used derives from the oddball paradigm; first used in ERPs by Nancy, Kenneth and Steven (Squires et al., 1975), 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 (Sutton et al., 1965). The name is derived from the fact that it is a positive wave that appears around 300ms after the target stimulus. Unless otherwise noted herein, the term P300 (P3) will always refer to a visual P300b (P3b) which is elicited by task-relevant stimuli in the centro-parietal.

BCI research and development has been predominantly focused on 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

[a] https://orcid.org/0000-0002-7808-5871
[b] https://orcid.org/0000-0003-2818-1010
[c] https://orcid.org/0000-0001-7458-7412

Schemrbi, P., Pelc, M. and 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.
DOI: 10.5220/0008065200500058
In *Proceedings of the 3rd International Conference on Computer-Human Interaction Research and Applications (CHIRA 2019)*, pages 50-58
ISBN: 978-989-758-376-6

with unrealistic conditions, where the subject sits in a sound-attenuated room without any distractions (Kam et al., 2019) (Bradford et al., 2018). Only a few research papers such as (Nam et al., 2010) and (Valentin et al., 2019) focus on real-world contexts, however they were either using medical grade equipment (Oliveira et al., 2016) and/or focusing on auditory ERPs (Zink et al., 2016).

The need for this study originated to broaden the utilization of this technology for both healthy subjects and especially to those individuals with severe neuromuscular disabilities (Sellers, et al., 2006), by providing a solution based on low fidelity equipment which is assessed outside lab conditions, and into noisy environments. Our null hypothesis based on preceding related and tantamount medical grade research are that this type of distraction does not show any statistically significant effect on accuracy, task performance, amplitude, or latency.

In this work, we report a study where ten healthy subjects used Farwell and Donchin P300 speller paradigm in conjunction with the xDAWN algorithm (Rivet et al., 2009) while utilizing low fidelity equipment. The subjects were asked to communicate five alphanumeric characters, referred to as symbols, in each of the four separate scenarios i.e. off, low, mid and high volume. The main goal for this study was to methodically investigate the usability of a P300 BCI system, explicitly that of a P300 Speller, in a specific context. Empirical experiments were conducted to assess how environmental factors such as music, with different levels of intensity, affect the task performance and quality of P300 component. This work is part of a larger EEG based project where we are introducing different categories of distractions which are being considered alongside the development of taxonomy as introduced in Figure 1.

This paper is structured as follows: the equipment, participants and experimental procedures are described in Section 2. The offline and online ERP results are presented in Section 3. Conclusions and future work are given in Section 4.

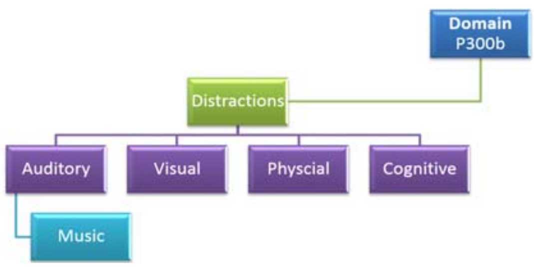

Figure 1: Development of extendable Hierarchal Taxonomy of Distractions for P300b.

2 METHODOLOGY

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

2.1 Low Fidelity Hardware

The work reported herein is based on an OpenBCI 32-bit board (called *Cyton*) connected with an *Electro-Cap* using the international 10/20 system for scalp electrode placement in the context of EEG experiments. The *Cyton* board's microcontroller is the PIC32MX250F128B with a 32-bit processor and a maximum speed of 50MHz; storage of 32KB of memory and is Arduino compatible. The board uses the ADS1299 IC developed by Texas Instruments, which is an 8-Channel, 24-Bit, simultaneous sampling delta-sigma, Analogue-to-Digital Converter used for biopotential measurements. The system comes with a pre-programmed USB dongle for wireless communication which communicates with the low-cost RFDuino RFD22301 microcontroller built on the *Cyton* board. An additional feature which is included in the board is a 3-axis accelerometer from ST with model LIS3DH. This can prove to be quite useful; such as, for sensing a change in orientation of the head or sensing rough motion. A more thorough explanation of the hardware components of the *Cyton* board can be found in our previous paper (Schembri et al., 2017). The *Electro-Cap* being used in our experiments has the fabric which is made from elastic spandex and has recessed pure tin wet electrodes directly attached to the fabric. The term wet electrodes type implies that the use of an electrolyte gel is required to make effective contact with the scalp otherwise it may result in impedance instability.

A pair of *Creative Labs SBS 15* speakers were used to output the three levels of music. The speakers have a nominal (RMS) output power of 5 Watt per speaker, a frequency response of 90Hz – 20,000Hz and a signal to noise ratio of 90dB.

2.2 Participants

We enlisted a total of $N = 10$ healthy subjects, seven males and three females, aged 29-38 ($M = 33.8$) which voluntarily participated in this study. Nine out of the ten subjects' native language was Maltese and the tenth subject's native language was English. All subjects spoke fluent English and were familiar with the symbols displayed on the P300 Speller. All subjects had previous experience using BCI and formerly performed P300 speller experiments.

One other subject assisted in the initial testing and configuration of the equipment; however, he/she did not take part in the official experiment and hence is not included in the results.

2.3 Data Acquisition

The EEG signals were sampled at 250Hz, while the sampling precision was 24-bit. The recordings were stored anonymously as raw data in OpenVIBE .ov format. These were later converted to a comma-separated value (CSV) files for offline analysis. Eight EEG electrodes were used in different regions of the scalp according to the International 10-20 System. The electrode positions C3, Cz, C4, P3, Pz, P4, O1 and O2 were selected as shown in Figure 2. This is because the spatial amplitude dispersal of the P300 component is symmetric around Cz and its electrical potential is maximal in the midline region (Cz, Pz) (Ogura et al., 1995). A referential montage was selected with the reference electrode being placed on the left earlobe A1 given that, in general, a mastoid or earlobe reference will produce a robust P300 response. The right ear lobe A2 was used as ground. The electrodes are referenced to electrode A1 as follows: Ch1: C3; Ch2: Cz; Ch3: C4; Ch4: P3; Ch5: Pz; Ch6: P4; Ch7: O1; Ch8: O2.

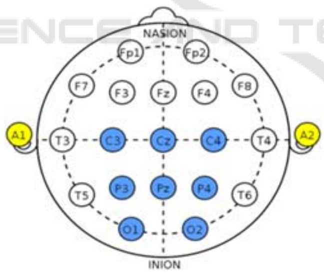

Figure 2: Electrode placement following the 10-20 system.

2.4 P300 Speller and xDAWN

In this paper, we make use of Farwell & Donchin P300 speller, which is based on visual stimuli, in conjunction with the xDAWN algorithm. Figure 3 depicts what is presented to the subject i.e. a six by six grid which is made up of thirty-six alphanumeric characters referred to as symbols. In this methodology, each row and column of the spelling grid is augmented in random order and the subject is asked to distinguish between a common stimulus (nontarget) and a rare stimulus (target). As a result of the (target) stimuli, an exogenous and spontaneous ERP potential known as P300; which is a positive deviation around 300ms after the stimuli; is evoked in the brain. The desired symbol is determined and predicted by the intersection of the (target) row and column. This prediction entails distinguishing between *non-target* i.e. rows/columns stimuli that do not generate a P300 component and *target* i.e. row/column stimuli that generate a P300 component.

In any recorded EEG signal, the P300 component which has a typical peak potential between 5-10µV, is embedded and masked by other brain activities (typical EEG signal +-100µV) such as muscular and/or ocular artefacts (Schembri et al., 2017) leading to a very low Signal-to-Noise Ratio (SNR) of the P300 component. This indicates that it would be very difficult to detect the target stimuli from a single trial, which is denoted by a series of augmentation, in random order, of each of the six rows and six columns in our matrix (i.e. twelve augmentations per trial). A popular way to address the limited SNR of EEG is for each symbol to be spelled numerous consecutive times and the respective column/row epochs are averaged over a number of trials, thus canceling components unrelated to stimulus onset.

The xDAWN process of spatial filtering is (1) a dimensional reduction method that creates a subset of pseudo-channels (referred to as output channels) by a linear combination of the original channels and (2) it promotes the appealing part of the signal, such as ERPs, with respect to the noise. This is applied to the data before performing any classification such as LDA (Linear Discriminant Analysis) which was used in this paper. A more thorough explanation of the xDAWN algorithm can be found in our previous paper (Schembri et al., 2018) or (Rivet et al., 2009).

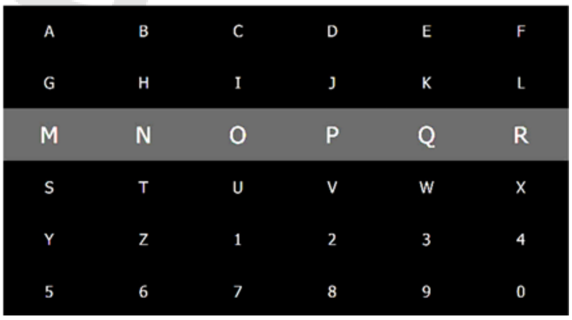

Figure 3: BCI "P300 Speller". The screen as shown to the subjects with the 3rd row highlighted.

2.5 Experimental Design

In this study, there was one independent variable manipulated: (a) digital music (off, low, medium and high) within-subjects variables. In addition, there were several dependent measures used which can be categorized into three types of dependent variables: online performance (accuracy), offline performance (amplitude and latency) and user preference.

2.5.1 Independent Variables

Four-levels of digital music were employed to represent different real-world scenarios: 'off' versus 'low' versus 'medium' versus 'high'. The 'off' level cited as *M0* represents a lab condition scenario, where subjects are seated in a sound-attenuated room. The 'low' level volume cited as *M30* was set at thirty percent i.e. between 20 and 30dB which simulates background music. The 'medium' level volume cited as *M60* was set at sixty percent i.e. between 50 and 60dB which simulated active listening to a movie. The 'high' level volume cited as *M90* was set at ninety percent i.e. between 80 and 90dB which simulated disco level music only i.e. no crowd chatter or noise. This first experiment i.e. *M0* was done as a basis for comparison for *M30*, *M60,* and *M90*.

2.5.2 Dependent Variables

Online Performance (Accuracy): is the number of correctly spelled symbols over the number of planned target symbols to be spelled; in our case 5 symbols which make up the word BRAIN.

Offline Statistics (Amplitude and Latency): 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. *P300 Latency* is considered a measure of cognitive processing time, generally between 300-800ms (Stern et al., 2001) poststimulus i.e. after target stimulus. In simplest terms, it is the time interval between the onset of the target stimulus and the peak of the wave.

User Preference: Throughout a questionnaire, the subjects were asked to rank from one to four, one being the best and four being the worst, their favorite usage condition.

2.6 Experimental Procedure

Each subject was invited and attended an induction session which was aimed to re-educate all subjects on the P300 speller paradigm and the hardware utilized. The subjects' were informed on the following: (1) they would be performing the experiment in five unique conditions, in sequence; (a) in the training phase, in a sound-attenuated room i.e. lab conditions; (b) *M0*, (c) *M30*, (d) *M60*, (e) *M90*, as explained in the independent variable section; (2) the symbols to spell were "BRAIN" for (1b) to (1e) and fifteen random symbols for (1a). Any subjects' query was answered at this stage. Before the start of the experiments, each subject was asked to relax for a few minutes in a seated position. The subject was seated approximately one meter away from the display. The researcher and his equipment were situated on the left side of the subject. The speakers were situated one meter away and facing the subject at a 15-degree angle. The experiment was started when the subject was able to properly perform the task at hand and had no additional questions. Prior to the start of every experiment, the impedance of the electrodes was confirmed to be less than 5KΩ.

The display presented to the subjects is shown in Figure 3 where 36 symbols presented in a 6x6 matrix. The target symbol was preceded by a cue i.e. one of the symbols was highlighted in blue at the beginning of the symbol run. Each row and column in the matrix was augmented randomly for 100ms and the delay between two successive augmentations was 80ms. This led to an interstimulus interval (ISI) of 180ms. For each symbol, six rows and six columns were augmented for fifteen repetitions and there was a 100ms inter-repetition delay and a 300ms inter-trial period between the end of the trials of one symbol and the beginning of trials of the next symbol, which allowed the subject to focus the attention on the next symbol. At the end of each symbol run, the predicted symbol was presented with a green cue, which indicated whether the system predicted the correct target symbol. The subjects were given a short break between experiments.

The training phase (1a) consisted of one session with 15 random symbols by 15 trials each (i.e. 12 flashes of columns/rows per trial * 15 trials = 180 flashes per symbol). The recording of the training phase took approximately 10 minutes. The *M0*, *M30*, *M60* and *M90* task experiments consisted of one session each with the aforementioned conditions and configurations while spelling the symbols "BRAIN" consecutively. Similarly to the training phase, each symbol had fifteen trials each. The recording of each task lasted approximately 6 minutes. In total, there were 15 symbols spelled in the training phase and 5 symbols spelled in each task, per subject. Hence due to the matrix disposition, there were in total 2700 flashes in the training phase, amongst which 450 were targets; and 3600 flashes in each task (900 * 4 tasks), amongst which 600 (150 * 4 tasks) were targets. These values are per subject. The data was stored anonymously by referring to the subjects as subject1-10 respectively.

2.7 Signal Processing

The online system was controlled by OpenViBE 2.0.0 which is a C++ based software platform designed for real-time processing of biosignal data. The *acquisition server* interfaces with the Cyton board and generates a standardized signal stream that is sent to the designer which in turn is used to construct and execute signal processing chains stored inside scenarios. The signal was obtained via the acquisition server which does not communicate directly with the Cyton board. Instead, it provides a specific and dedicated set of drivers that does this task. The signal was obtained at a sampling rate of 250Hz with 8 EEG channels and 3 accelerometer (auxiliary) channels. The experimental paradigm was controlled by the OpenViBE *designer* where a number of scenarios in the "*P300: Basic P300 Speller demo with xDAWN Spatial Filter*" were executed in succession.

In the *offline analysis*, the following procedure is done for each *M0*, *M30, M60,* and *M90*. The captured raw data was converted from the proprietary OpenVIBE .ov extension to a more commonly used .csv format using a particular scenario aimed for this task. The converted data was later imported into MATLAB R2014a and any unnecessary rows and columns such as headers and auxiliary data were removed. Next, we filtered out the data to include the target stimulations with code (33285); non-target stimulations (33286); and visual stimulation stop (32780), which is the start of each flash of row or column. Subsequently, we had to perform a signal inversion due to the hardware and driver implementation. The data (samples and event info) were later imported into EEGLAB for offline processing. The first process was to apply a bandpass filter of 1-20HZ to eliminate the environmental electrical interference (50Hz or 60Hz), to remove any signal harmonics and unnecessary frequencies which are not beneficial in our experiments and to remove the DC offset. Next, the imported data was used in ERPLAB which is an add-on of EEGLAB and is targeted for ERP analysis. We took every event we wanted to average together and assigned that 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 have used the following criteria: "*.{33285}{t<50-150>32780}*" for the target and ".{33286}{t<50-150>32780}" for the non-target. This implies that it is time-locked to the stimuli event 33285 (target) or 33286 (non-target) and must have the event 32780 that happens 50 to 150ms after the target/non-target event. If this criteria is met, it is placed in the appropriate BIN; in our case BIN1 for target and BIN2 for non-target. Next, we extracted the bin-based epochs via ERPLAB (not the EEGLAB version) and set the time period from -0.2s before the stimulus until 0.8s after the stimulus. We have also used baseline correction (pre) since we wanted to subtract the average pre-stimulus voltage from each epoch of data. Next, we passed all channels epoch data through a moving window peak-to-peak threshold artifact detection with the voltage threshold set at 100μV, moving window width at 200ms and window step at 100ms to remove unwanted signals such as blinking and moving artifacts. Subsequently, we averaged our dataset ERPs to produce part of the results shown in Table 5. Lastly, we performed an average across ERPsets (Grand Average) to produce the results in Figure 3 and part of the results in Table 5. The data for Table 5 were generated by the *ERP measurement tool*. A more thorough explanation on segments of the signal processing can be found in our previous paper (Schembri et al., 2018).

3 RESULTS

In this section, we present several results in relation to the dependent variables such as a one-way ANOVA (factorial analysis) to determine the effect that off, low, mid and high level of volume intensity have on the online performance (accuracy), offline statistics (amplitude and latency) and user preference. In the following tables the labels *M0*, *M30*, *M60,* and *M90* represent "no music - lab condition", "music at 30%", "music at 60%" and "music at 90%" volume respectively. Moreover, *M0*, *M30*, *M60,* and *M90* might be interchangeably referred to as *BIN1*, *BIN3*, *BIN5,* and *BIN7* respectively.

3.1 Online Analysis

Following the online experiments, the results achieved per subject are shown in Table 1 which depicts the correct symbols predicted out of five (i.e. symbols BRAIN) 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, subject8 had a success rate of 80% in the M30 scenario, with the symbol R predicted as symbol Q i.e. the row prediction was correct but not the column. However to avoid ambiguity we have decided to assume that both row and column prediction were incorrect when the symbol is predicted incorrectly.

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

Subject	LAB	M30	M60	M90
S1	5 (100%)	5 (100%)	4 (80%)	5 (100%)
S2	5 (100%)	5 (100%)	4 (80%)	4 (80%)
S3	5 (100%)	5 (100%)	5 (100%)	5 (100%)
S4	5 (100%)	5 (100%)	5 (100%)	5 (100%)
S5	5 (100%)	5 (100%)	5 (100%)	5 (100%)
S6	5 (100%)	5 (100%)	5 (100%)	5 (100%)
S7	5 (100%)	5 (100%)	5 (100%)	5 (100%)
S8	5 (100%)	4 (80%)	5 (100%)	5 (100%)
S9	5 (100%)	4 (80%)	5 (100%)	5 (100%)
S10	5 (100%)	5 (100%)	5 (100%)	5 (100%)
Grand Average	100%	96%	96%	98%

We have performed a one-way ANOVA which is based on our independent variable with four levels/groups (M0, M30, M60, and M90) as presented in Table 2, to determine if there is a significant difference between the four means of each group or if they are all the same. We have chosen to use a 5% significance level (0.05) denoted as α (alpha) and rounded all values to the nearest thousandth. Our null hypothesis (H_0) states that the means are all equal i.e. the mean of M0, M30, M60, and M90 are all the same. Our alternate hypothesis (H_1) states that at least two of these means are different.

Table 2: One-way ANOVA test on Accuracy.

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.275	3	0.092	0.805	0.499	2.866
Within Groups	4.1	36	0.114			
Total	4.375	39				

In the first column we have the source of variation, where ANOVA carries out an analysis between groups variation (i.e. M0, M30, M60, M90), and also carry an analysis of the within groups variation i.e. the variation within each of our four groups (refer to Table 1). In the second column, we have the sum of squares (*SS*) of the variation, which is the spread between each individual value and the mean. The third column is the degrees of freedom (*df*) which is the (number of samples – 1). We have four samples of *between groups* which gives three and we have forty samples in *total* which give thirty-nine. That allows us to calculate the *within-group df* which is *total* less *between groups* i.e. a value of thirty-six. The fifth column we have the mean Square Values (*MS*) which is calculated by dividing the *SS* by the corresponding *df*. The sixth column is the *F* statistic which is the key statistic where we divide the *MS between groups* by the *MS within group*. Since our *F* statistic got a result of 0.805 which is smaller than our *F-critical* value (8^{th} column), this implies that we accept the H_0 i.e. that all means are equal and reject H_1. Also, by analyzing that the *P-value* (7^{th} column) which is 0.499 i.e. it is greater than the alpha value of 0.05, so we can also accept H_0 and reject H_1.

3.2 Offline Statistics

In this section, we process and analyze the averaged epoch signal of ten subjects in relation to the independent variable (LAB, M30, M60, and M90).

Figure 3 shows the grand average P300 component for all ten subjects in each scenario which include all eight channels and an average channel (AVG). It is comprised of the grand-averaged raw signals i.e. (5 symbols with 15 trials per symbol); with (12 flashes of columns/rows per trial); with (10 subjects) i.e. 9000 flashes amongst which 1500 were targets. In addition figure 3 shows four overlapping signals, (i) BIN1 - Target for *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, we have omitted BIN2, BIN4, BIN6, and BIN8 which represent the non-target signals.

Table 3 shows the means and standard deviations in parentheses, for the dependent variables (amplitude and latency) according to levels of the independent variable rounded to the nearest hundredth. This data includes the average of all eight recorded electrodes throughout the five symbols and is shown per subject for each BIN1, BIN3, BIN5, and BIN7.

We have performed a one-way ANOVA which is based on our independent variable with four levels/groups (Lab, M30, M60, and M90) for our dependent variables (amplitude and latency) as presented in Table 4 and Table 5 respectively. In Table 4 which represents the amplitude, we can see that the *F statistic* is 0.723 which is smaller than our *F-critical* value of 2.866. In addition, our *P-value* is 0.545 which is greater than the alpha vale. This implies that we can accept the null hypothesis (H_0) and reject the alternate hypothesis (H_1). In Table 5 which represents the latency, we can see that the *F statistic* is 2.982 which is slightly larger than our *F-critical* value of 2.867. In addition, the *P-value* is 0.044 which is slightly smaller than the alpha value. This implies that we reject H_0 and accept H_1. A more thorough explanation of the one-way ANOVA can be found in the previous section.

Table 3: Means and Standard Deviations (in Parentheses) for Two Dependent Measures (Amplitude and Latency).

Subject	LAB (BIN 1)		M30 (BIN 3)		M60 (BIN 5)		M90 (BIN 7)	
	Amplitude (µV)	Latency (ms)	Amplitude (µV)	Latency (ms)	Amplitude (µV)	Latency (ms)	Amplitude (µV)	Latency (ms)
S1	4.90 (0.50)	466.5 (2.97)	2.09 (0.28)	473.5 (12.46)	3.90 (0.40)	470.5 (5.21)	2.88 (0.40)	462.5 (19.00)
S2	2.39 (1.14)	437.0 (82.16)	3.24 (0.35)	434.0 (76.58)	4.26 (0.41)	419.5 (54.15)	3.66 (0.83)	416 (66.69)
S3	3.29 (1.68)	431.0 (83.52)	2.18 (2.24)	417.0 (71.61)	4.03 (0.99)	430.0 (85.68)	3.64 (1.42)	423 (78.49)
S4	3.70 (0.98)	466.0 (107.48)	4.95 (0.54)	427.0 (82.11)	4.14 (0.45)	432.5 (86.76)	3.25 (1.08)	419.5 (75.49)
S5	3.96 (0.78)	444.0 (91.49)	4.11 (0.95)	431.0 (84.58)	4.61 (0.76)	428.5 (80.21)	4.77 (1.17)	436.5 (85.56)
S6	5.39 (1.64)	483.0 (1.85)	3.63 (2.44)	479.0 (17.73)	4.99 (2.40)	484.5 (3.96)	3.70 (2.12)	495 (2.83)
S7	1.70 (1.66)	430.5 (81.75)	4.94 (1.26)	431.5 (84.97)	4.18 (1.24)	427.5 (79.81)	4.58 (1.34)	449.5 (93.64)
S8	2.64 (1.38)	415.0 (71.10)	2.51 (1.57)	440.0 (72.88)	3.52 (1.22)	415.5 (66.91)	2.18 (1.69)	446.0 (90.16)
S9	4.46 (1.94)	439.0 (85.89)	2.50 (1.40)	415.5 (71.66)	2.17 (2.05)	424.0 (76.67)	3.67 (1.89)	411.0 (75.46)
S10	3.58 (0.26)	497.5 (87.40)	4.08 (0.71)	420.5 (80.61)	3.46 (0.61)	420.5 (76.98)	3.51 (0.46)	424.5 (84.43)
Grand Average	3.60 (1.83)	440.0 (85.31)	3.43 (1.66)	429.0 (80.87)	3.93 (1.38)	427.5 (79.98)	3.59 (1.46)	431.0 (81.04)

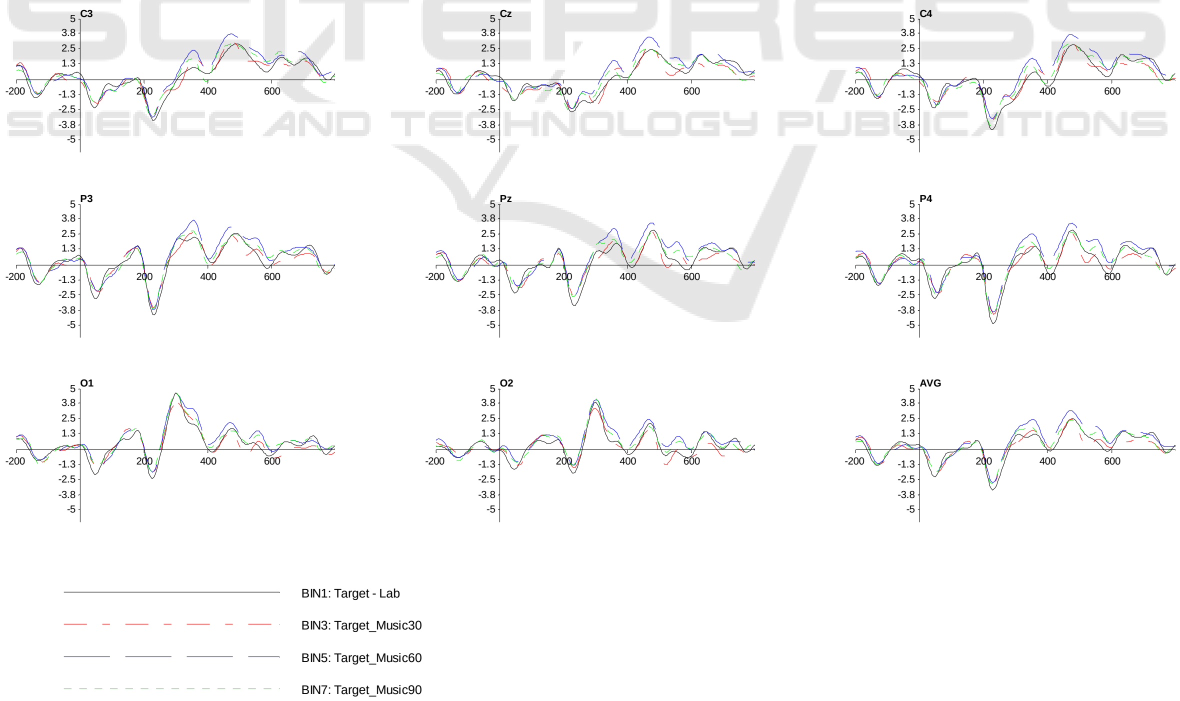

Figure 3: Grand average P300 for all 10 subjects in each scenario with all eight channels and an average channel (AVG).

Table 4: One-way ANOVA on Amplitude.

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.798	3	0.933	0.723	0.545	2.866
Within Groups	46.415	36	1.289			
Total	49.212	39				

Table 5: One-way ANOVA on Latency.

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3236.4	3	1078.8	2.982	0.044	2.867
Within Groups	13025.6	36	361.822			
Total	16262	39				

3.3 User Preference

Exactly after the experiments were finished, each subject was presented with two questionnaires (a) and (b) to specify their preferred BCI usage condition. The ranking consisted of maximum weight value of four (4) as the most desired and minimum weight value of one (1) for the least desired.

In the first questionnaire, the subjects were allowed to give the same ranking to different groups as shown in column (a) of Table 6. Expectedly, the *M0* scenario got the highest ranking, whilst surprisingly *M90* came second, *M60* third and *M30* last. The frequency analysis grouped by the highest value of four (4), shows that the *M0* was given 100%, followed by *M90* with 50% trailed by *M30* and *M60* equally at 40%.

In the second questionnaire, the subjects were asked to give a unique ranking (1-4) to each scenario as shown in column (b) of Table 6. The results are similar to those achieved in the questionnaire (a), where *M0* came first, followed by *M90*, *M60* and *M30* respectively. The frequency analysis shows that *M0* got 40%, followed by *M90* at 23%, *M60* at 22% and lastly by *M30* at 15%.

Our assumption on why *M90* placed second and *M30* placed last in both questionnaires is that, since the experiments were performed in sequence, the subjects were staggered by the difference between *M0* and *M30*, while they gradually got accustomed to the music and hence they were not bothered or distracted as they initially were. In addition, the results indicated that the user preference wasn't affected by the loudness of the music.

Table 6: User preference for two questionnaires with (a) allowing the same ranking and (b) unique ranking.

Subject	LAB		M30		M60		M90	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
S1	4	4	3	2	3	1	4	3
S2	4	4	4	3	3	1	3	2
S3	4	4	3	1	3	3	3	2
S4	4	4	3	1	3	2	4	3
S5	4	4	4	3	4	2	3	1
S6	4	4	4	1	4	3	4	2
S7	4	4	4	1	4	2	4	3
S8	4	4	2	1	3	2	3	3
S9	4	4	2	1	3	3	4	2
S10	4	4	3	1	4	3	3	2
Total	40	40	32	15	34	22	35	23

4 CONCLUSION

In this study, N = 10 healthy subjects were able to perform several experiments using Farwell & Donchin P300 speller in conjunction with the xDAWN algorithm, with a six by six matrix of alphanumeric characters, in *M0*, *M30*, *M60,* and *M90* environments while utilizing low fidelity equipment.

The goal of this study was to investigate the usability of a visual P300 Speller, and assess the extent of effect that an auditory distraction such as digital music, with varying level of intensities (off, low, mid, high), have on the user and overall BCI performance (i.e. the dependent variables: accuracy, amplitude, latency, and user preference). This study is part of a larger-based project where we are introducing different categories of distractions which are being considered alongside the development of a taxonomy and give some insight on the practicability of real-world application of the current P300 speller with our aforementioned low-cost equipment.

Our null hypothesis based on preceding related and tantamount medical grade research was that this type of distraction as elucidated in the independent variable does not show any statistically significant effect on the accuracy, amplitude, and latency dependent variables. The results of a one-way ANOVA factorial analysis accepts our null hypothesis for the accuracy and amplitude dependent variables, however, it was rejected for the latency dependent variable since there was a minor statistical significance as shown in the results section.

Non-statistical results show that the dependent accuracy variable was highest in the *M0* (100%) and surprisingly followed by *M90* (98%) trailed by *M30* and *M60* equally at 96%. Our empirical evidence suggests that the subjects got accustomed to the music in the *M90* environment since these were performed in sequence as explained before. The dependent variable *amplitude* was highest in the *M60* (M=3.93, SD=1.38) followed by *M0* (M=3.60, SD=1.83), *M90* (M=3.59, SD=1.46) and *M30* (M=3.43, SD=1.66). Additionally, the dependent variable *latency* was shortest in *M60*, followed by *M30*, *M90* and finally *M0* as shown in Table 3. It seems that there is no correlation between amplitude and latency. Lastly, the user preference evidently shows that all subjects preferred the *M0*, followed by *M90*, *M60,* and *M30* in both questionnaires as shown in Table 6. This enforces our previous empirical evidence that the subjects seem to get acquainted with the music in the fourth sequential experiment of *M90* while they are staggered by the difference between *M0* and *M30*, which follow each other. These results also indicated that the user preference wasn't affected by the loudness of the music. Moreover, the signals were morphological consistent in all four scenarios, even though they did not yield identical P300 components.

In the future we plan to run the independent variable levels (*M0*, *M30*, *M60*, *M90*) experiments in a randomized order and not sequentially, to avoid the results being affected by subjects accustomization to the distraction. Another important point to take into account in future experiments is the possible impact of mental fatigue with and without the presence of distractions during repetitive exercises.

Our main contribution is the comparative assessment in terms of (a) accuracy, (b) amplitude, (c) latency and (d) user preference, between the levels of the independent variable. Our main goal is to provide insight into the practicability of the current P300 speller to be used in concurrence with several taxonomized distractions.

In this paper, we have introduced our expandable hierarchical taxonomy as depicted in Figure 1. This work is part of a larger EEG based project where we are introducing different categories of distractions which are considered alongside the development of taxonomy while using low fidelity equipment. Our investigation is concerned with the way in which different types of distractions (e.g. audio, visual, with differing intensity/regularity and engagement factor) translate into a reduction of the signal quality and amplitude, or any other change/distortion that occurs.

REFERENCES

Bradford, C. J. et al., 2018. Performance of conformable, dry EEG. *40th Int Conferenc of EMBC,* pp. 4957-4960.

Kam, J. W. et al., 2019. Systematic comparison between a wireless EEG system with dry electrodes and a wired EEG system with wet electrodes. *NeuroImage,* 184(1).

Nam, C. S., Li, Y. & Johnson, S., 2010. Evaluation of P300-Based Brain-Computer Interface in Real-World Contexts. *International Journal of Human-Computer Interaction,* 26(6), pp. 621-637.

Ogura, , Koga , & Shimokochi, , 1995. Recent Advances in Event-related Brain Potential Research: Proceedings of the 11th International Conference on Event-related Potentials (EPIC), Japan, June., 1995. Elsevier.

Oliveira, A. S. et al., 2016. Proposing Metrics for Benchmarking Novel EEG Technologies Towards Real-World Measur. *Frontiers in Human Neuroscience*

Rivet, B., Souloumiac, A., Attina, V. & Gibert, G., 2009. xDAWN Algorithm to Enhance Evoked Potentials: Application to BCI. *IEEE Transactions On Biomedical Engineering,* 56(8), pp. 2035 - 2043.

Schembri, P., Anthony, R. & Pelc, M., 2017. Detection of Electroencephalography Artefacts using Low Fidelity Equipment. *Proceedings of the 4th Int. Conference on Physiological Computing Systems,* pp. 65-75.

Schembri, P., Anthony, R. & Pelc, M., 2018. The Feasibility and Effectiveness of P300 responses using Low Fidelity Equipment in three Distinctive Environments. *5th Int. Conf. Physiological Computing.*

Schembri, P., Anthony, R. & Pelc, M., 2018. The Viability and Performance of P300 responses using Low Fidelity Equipment. *5th Int Conf. on Biomedical Eng. Systems.*

Schembri, P., Pelc, M. & Ma, J., 2019. 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. *41st Annual International Conf of the IEEE Engineering in Medicine and Biology, EMBC 2019.*

Sellers, E. W., Kübler, A. & Donchin, E., 2006. BCI Research at Univ of Florida Cognitive Psychopysiology Laboratory: The P300 Speller. *IEEE Trans On Neural Systems Rehabilitation Engineering,*14(2),pp. 221-224.

Squires, N., Squires, K. & Hillyard, S., 1975. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *EEG Clinical Neurophysiol,*

Stern, , Ray, J. & Quigley, K.S., 2001. *Psycho-physiological Recording*. 2nd ed. Oxford University.

Sutton, S., Braren, M., Zubin, J. & John, E. R., 1965. Evoked-Potential Correlates of Stimulus Uncertainty. *Science,* 150(3700), pp. 1187-1188.

Valentin, O. et al., 2019. Validation and Benchmarking of a Wearable EEG Acquisition Platform for Real-World Applications. *IEEE TRANS ON BIOMEDICAL CIRCUITS AND SYSTEMS,* 13(1), pp. 103-111.

Zink, R., Hunyadi, B., Huffel, S. V. & Vos, M. D., 2016. Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *Journal of Neural Engineering,* 13(4).