IMPROVING THE SPEED AND ACCURACY OF THE P300 AUTOMATIC SPELLING SYSTEM THROUGH FACIAL RECOGNITION

By

Thomas Albanese

July 2022

Faculty Advisor: Dr. Sunghan Kim

Major Department: Engineering

Locked-in syndrome is a condition where an individual does not have control of their muscles, including facial muscles and vocal cords which control speech. Instead of oral speech, sign language, or writing, other methods must be used to achieve communication. Brain-computer interfaces relay the intentions of these individuals by using brainwave patterns such as event-related potentials (ERPs) that are collected using electroencephalogram electrodes. An ERP is a time-locked response with many features such as the P300 component. One way to elicit a strong ERP is using the “oddball” paradigm. When a stimulus is presented to the brain, it is classified as a target or non-target event. One of the two events categories is considered “rare” with a lower probability of occurrence than the other. When the “rare” event occurs, an ERP is evoked, with the amplitude of the P300 response being proportional to how small of a probability the event has of occurring. The rarer the stimuli, the stronger the P300 response will be. A P300

spelling system is a brain-computer interface that can allow the P300 component to be converted to text digitally by using signal processing and machine learning techniques. Using EEG measurements, the computer is trained to recognize features in a particular individual's brainwaves. Once these features are identified, they are classified as target and non-target, allowing the computer to decide the desired outcome. However, current P300 spelling systems are still much slower than conventional communication and can be inaccurate. This research aims to improve the speed of the system without compromising accuracy by testing stimulus type and stimulus duration. Specifically, the standard paradigm will be altered to include a familiar face overlay with the intent of generating a stronger P300 component which can be more easily classified as a target or non-target stimulus. Different flash durations will also be tested in an attempt to improve the speed of spelling.

IMPROVING THE SPEED AND ACCURACY OF THE P300 AUTOMATIC SPELLING SYSTEM THROUGH FACIAL RECOGNITION

A Thesis

Presented to the Faculty of the Department of Engineering

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science, Biomedical Engineering

By

Thomas J. Albanese II

July 2022

IMPROVING THE SPEED AND ACCURACY OF THE P300 AUTOMATIC SPELLING SYSTEM THROUGH FACIAL RECOGNITION

by

Thomas J. Albanese II

APPROVED BY:

DIRECTOR OF THESIS: ____

Sunghan Kim, PhD

COMMITTEE MEMBER: ____

Raymond Smith, PhD

COMMITTEE MEMBER: ____

Venkat Gudivada, PhD

CHAIR OF THE DEPARTMENT

OF ENGINEERING: ____

Barbara Muller-Borer, PhD

INTERIM DEAN OF THE

GRADUATE SCHOOL: ____

Kathleen Cox, PhD

TABLE OF CONTENTS

LIST OF FIGURES

I. Introduction

Locked-in syndrome is often caused by damage to a region of the brainstem called the pons. The pons contains numerous neuronal pathways between the cerebellum, cerebrum, and spinal cord. This damage hinders the motor neurons that run from the brain to the spinal cord, blocking the signals that allow for movement and control. This damage is often caused by hemorrhage or infarct caused by conditions such as a stroke or blood clot. Other conditions that can cause locked-in syndrome include infection, tumors, trauma, deteriorations of neuron myelin, amyotrophic lateral sclerosis (ALS), and inflammation [1]. This does not, however, mean that the person's brain is not functioning or that the person is incoherent. Many individuals are still able to see, hear, and comprehend the world around them. Therefore, research has been conducted to create new ways to assist those who suffer from locked-in syndrome.

One way that those with locked-in syndrome can communicate with the world around them is by using a brain-computer interface (BCI). This system acquires brain activity and converts these measurements into a command that can be performed by a computer [2]. The purpose of the BCI is to replace bodily functions that have been lost due to illnesses, disorders, or traumas. Many commonly used BCI measure the neuronal post-synaptic membrane polarity changes, dipoles, that occur due to opened voltage or ion gated channels. To measure these electrical signals, the electroencephalogram (EEG) is used. To create a brain-computer interface, there are four needed components: signal acquisition, feature extractor, feature translation, and device output. Signal acquisition refers to the collection of brain measurements by using devices such as an EEG. These signals are digitized and sent to the computer for feature extraction. Feature extraction refers to the processing of raw data to extract quantifiable features that can be further used for classification. An example of a feature may be the amplitude of an EEG signal at a specific time, such as the amplitude of the P300 component of the event-related potential. These features can then be classified and identified as target events or non-target events. This is how signals are distinguished between wanted or unwanted by the computer for the translation since it is usually the target signals that are translated into device output. Dynamic, complex algorithms and machine learning techniques are used to translate these features into commands that can be given to the output device [2]. The output devices then carry out the desired task, such as displaying the desired letter from a matrix. These brain-computer interfaces usually require training and are not intuitive to use. During training, the computer is calibrated to recognize the specific patterns of the user's brainwaves and adapts to perform its task. In doing so, the computer and the user become synchronized and can use each other to perform a task by bypassing the bodily function that would normally perform that task with their brain activity.

II. Background Research

II.A. Electroencephalogram

An EEG uses electrodes that are placed in contact with the scalp to measure the electrical potential changes due to electrophysiological activities of the brain. The electrophysiological activities measured would be voltage fluctuation on the scalp which is generated by postsynaptic potentials. These potentials represent the change in neuron membrane polarization that is caused by the opening and closing of voltage-gated ion channels in pyramidic cells [3, 4]. When an impulse arrives from a presynaptic neuron at a synapse, neurotransmitters are released, causing the opening of channels in the postsynaptic neuron, and allowing the flow of ions. This flow of molecules throughout the postsynaptic neuron changes the resting membrane polarization of the postsynaptic cell [4]. These postsynaptic potentials are exhibitory if they increase the likelihood of more postsynaptic potentials or inhibitory if they decrease the likelihood of more postsynaptic potentials [3]. Thousands of these postsynaptic neurons are measured around each electrode of the EEG.

Unlike other methods of measuring physiological changes in the brain, such as magnetic resonance imaging (MRI), which measures the brain's metabolic changes, EEG measures instantaneous electrophysiological changes, allowing for high temporal resolution. This makes EEG an ideal measurement device for this study as the high temporal resolution allows us to identify features in the measurements that appear at specific time intervals, such as an ERP and its different components. When tracking the P300 component of an ERP, the high temporal resolution is beneficial for being able to determine exactly where the P300 response is occurring in an EEG recording. The electrical signals recorded by EEG are constantly changing even during sleep and are measured at the microvolt level. This high temporal resolution allows the

BCI to immediately take action depending on how the EEG measurements are classified. One limitation of the EEG is that the signal must pass through the dura, skull, and scalp which can cause loss of information and worsened signal quality [2]. Nonetheless, the high temporal resolution allows the tracking of the electrical impulses in the brain with great accuracy. Even with low spatial resolution, the location of the electrical activity along the scalp can still be determined. For this study, a g-tec 16-channel EEG headset will be used. This headset follows the international 10-20 EEG scalp mapping scheme shown below in Figure 1 [5].

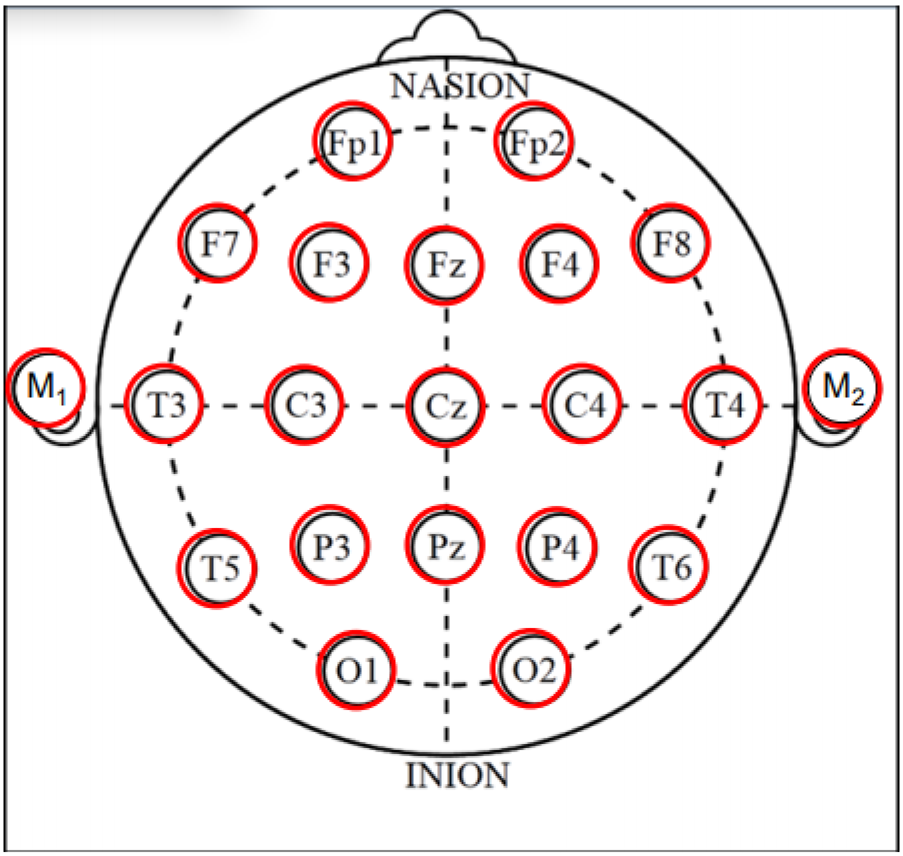


Figure 1. A map of the scalp showing the placement of the electrodes of an EEG. This EEG scalp map is held consistent with the international 10-20 standard [5].

The electrode placements shown in the international 10-20 EEG electrode map follow a naming scheme that allows users to identify where the electrode is placed. In this scheme, Fp, F, T, C, P, and O are used, defined as pre-frontal, frontal, temporal, parietal, central, and occipital, respectively. Along with each letter is a number or letter, z. The electrodes on the left side of the head are paired with an odd number; the electrodes on the right side are paired with an even number, and the electrodes down the midline sagittal plane are given the letter, z [5]. The further from the midline the electrode is, the larger the number associated with that electrode will be.

II.B. Event Related Potential Components

For the brain-computer interface to function, we need to create a measurable response elicited by the brain. In this study, the P300 component of an event-related potential (ERP) will be used. The P300 response is a component of an ERP, which is a brief fluctuation of EEG in response to an event [6]. The P300 is seen as a positive fluctuation in voltage in the EEG that occurs approximately 250ms to 500ms after the desired event [7]. In EEG signals, when a P300 response is evoked, the voltage fluctuates and the P300 component can be identified as the large positive fluctuation following the negative N170 response [7]. One way the P300 component of the ERP can be evoked is by using the “oddball” paradigm. The “oddball” paradigm assumes that all event outcomes are classified into two categories. One of the two events categories is considered “rare” with a lower probability of occurrence than the other. When the “rare” event occurs, an ERP is evoked, with the amplitude of the P300 response being proportional to how small of a probability the event has of occurring [6]. This proportionality is due to the P300 response being an anticipatory event. This means that as one anticipates the rare or “oddball” event, the intensity of the P300 response varies based on how often this event will occur [8].

When a stimulus is rare, the brain is “surprised” when the event does occur. This ERP is a measurable event with the typical P300 component of the ERP peaking between 5 and 10 μV [6]. This ERP and its components can be seen in Figure 2, shown below [9].

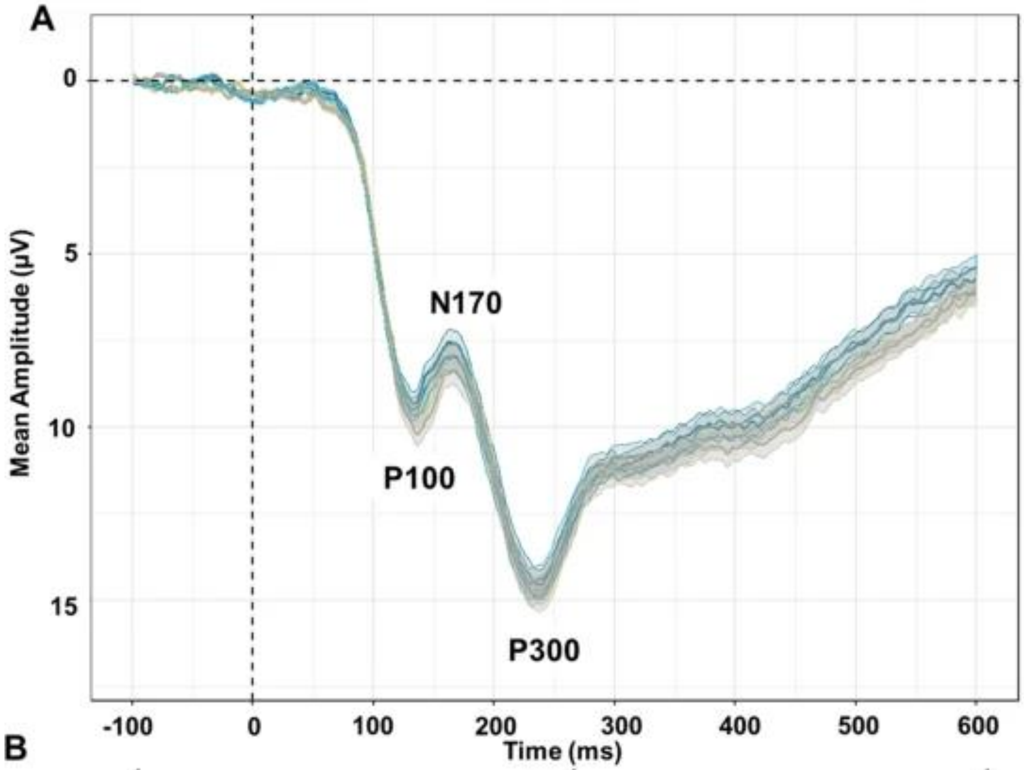


Figure 2. This image shows an ERP and the P100, N170, and P300 components. The P300 component has the largest amplitude in the ERP. The P300 component shown in this figure has an amplitude of around 15 microvolts [9].

Another component that could be considered for use is the N170. The N170 component of an ERP has been shown in several studies to be related to facial recognition and is easily reproducible [10, 11]. This component occurs between 140ms and 200ms after the presentation of a stimulus. Studies have shown that the amplitude and latency of this component are affected by the brain’s facial recognition processes. These are also affected by the emotions associated with the image [12]. The N170 component is the most studied component for studies regarding facial recognition indicating there is a consensus that the N170 component is a valid biomarker for facial recognition [11].

Although the N170 component is a viable marker for facial recognition, it does have pitfalls that can lead to inaccuracies in the results. For example, if the facial image that is shown is cropped or altered in any way, it can cause the amplitude of the N170 response to be altered. [10] This can cause the results to be misinterpreted and cause incorrect conclusions to be drawn. The N170 belongs to the N1 class of exogenous components. This means that the response is visually evoked. This is compared to the P300 response, which is an endogenous response, meaning that it is not evoked by the physical characteristics of a stimulus. Rather, the P300 response is linked to the individual's reaction to the stimulus [13]. The P300 response has been shown in hundreds of studies as consistently being produced with the introduction of the "oddball" event [6]. To reiterate, the P300 component does not require any verbal or physical response to be evoked [14]. This makes the P300 component of the ERP an extremely reliable, measurable source for communication via signals for this study.

II.C. Brain-Computer Interfaces

Brain-computer interfaces are computer-based systems that gather data from the brain, analyzes the data, and then translate them into commands that can be used to carry out desired tasks [2]. The most studied brain signals are the electrophysiological changes on the scalp which are measured with electrodes. Many different types of brain-computer interfaces can be used to perform desired tasks. Steady-state visually evoked potentials (SSVEP) is a BCI that uses signals that are natural responses to visual stimulation at specific frequencies [15]. When a visual stimulus is presented at a specific frequency, the brain generates an electrical response that is a harmonic of that frequency. This BCI is often used in vision and attention studies [16]. Another popular BCI is a movement imagery BCI. A movement imagery BCI involves detecting the intention to move a part of the body, which evokes an electrical response in the brain [17]. An example of this would be to have a subject think about raising their left or right hand. The thought of doing so would cause an electrical response in the brain which can be detected with measurement devices such as EEG. This BCI would not be ideal for this study as for locked-in individuals it may be difficult to create the intention of physical movement. This BCI also requires training sessions to be used. Although SSVEP, motor imagery, and P300 BCIs have been shown to evoke strong P300 responses, it is unlikely that a new method will be developed that can evoke a P300 response with a much larger amplitude [18]. This supports the need for studies to be conducted which can find methods that evoke P300 responses with larger amplitudes that are found to be significant.

II.D. BCI2000 Software

In P300 spelling systems, there has been a multitude of modifications made to optimize the speed and accuracy. All these modifications are performed with the BCI2000 software. This software allows the user to determine desired outputs by displaying them on a screen. Shown below in Figure 2 are the various displays that can be utilized for experiments using the BCI2000 software. Both the figure and the caption are taken directly from Schalke G. (2004) [19].

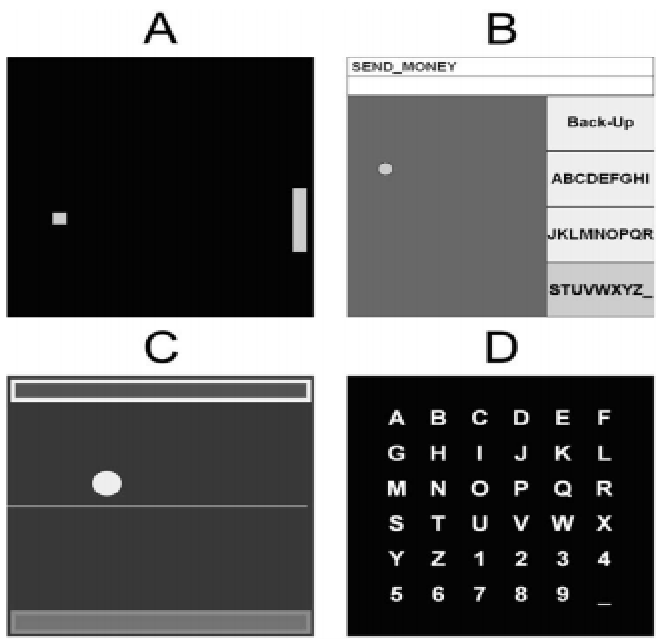


Figure 3. Output screens for several BCI2000 implementations tested to date. (A) Sensorimotor rhythm control of cursor movement to a variable number of selections (as in [35]). (B) Simple spelling application using sensorimotor rhythm control (as in [34]). (C) Slow cortical potential (SCP) control of cursor movement to two possible selections (as in [33]). (D) P300-based spelling application (as in [1] and [38]). In (A)–(C), the cursor moves from left to right at a constant rate with its vertical movement controlled by the user's brain signals. In (D), rows and columns of the matrix flash in a block-randomized fashion. [19]

From Figure 3, section D is the most utilized display of the BCI2000 software for the purpose of the P300 speller system. The matrix of letters and numbers can be altered to include varying amounts of characters, words, or commands for a subject to choose from. The rows and columns of this matrix is randomly intensified, causing a P300 response in the subject's brain when the desired character is intensified. The parameters of the experiment can be changed using the BCI2000 software as well. Some of these parameters include stimulus duration, sequence, and interstimulus interval, although, there are many more parameters that may be adjusted using the BCI2000 software [19].

II.E. Limitations Affecting Performance of the System

In a study by P. Brunner et al. (2010) [20], it was tested to see if the location of a subject's gaze affected the outcome of the system. In this experiment, the traditional six-by-six matrix of letters and numbers, shown in Figure 3 [19], was used but a focal point was added in the direct center of the matrix in the form of a cross. There were two experiments conducted for each subject. In the first experiment, the subjects could gaze freely at the characters. In the second experiment, the subjects were asked to focus their gaze on the cross. To ensure the gaze of the subject was fixed on the cross, an eye tracker was used. In both experiments, the subjects were assigned a target character and instructed to signal when the target letter flashed. The results of this experiment yielded that the first experimental data set (range: 80 to 100%) produced results with much higher accuracy compared to the second experimental data set (ranging from 2.8 to 90%). The significance of this study demonstrated that the focus of the gaze is important when attempting to produce a P300 response. Focusing directly on the target character creates a much larger ERP [20].

In a study by Fazel-Rezai R. (2007) [7], the P300 spelling system was tested to identify human errors that can cause inaccuracies. This study eliminated the use of advanced signal processing algorithms so that factors such as frustration, fatigue, level of attention, and other human factors could be accounted for. In addition, there was no training provided to the subject so that the data could be observed in its rawest form. The experiment was conducted using the Farwell and Donchin method of a six-by-six matrix with a randomized flashing of rows and columns. What was found was that the main source of error in the subject's results was the accidental targeting of adjacent letters in the matrix. By statistical analysis, it was found that in all error cases, 60% of the incorrectly detected characters were within one space vertically/horizontally from the target letter [7]. This error analysis is significant for the purposes of system optimization as eliminating this chance of error can significantly improve the overall accuracy.

II.F. Previous Works on System Optimization

In a study by Lu J. (2013) [21], an experiment was conducted to gain a better understanding of how stimulus presentation parameters affect the performance of the P300 spelling system. Using the traditional Farwell and Donchic six-by-six matrix, six subjects were tested by flashing the rows and columns of the matrix randomly. The parameters tested were stimulus off-time, interstimulus interval, flash duration, and the ratio between flash duration and the interstimulus ratio. The interstimulus interval (ISI) refers to the amount of time between flashes. The flash durations used were 32ms, 64ms, and 128ms. The results of the study showed that the lengthening of the stimulus off-time, increasing the ISI, and increasing the flash duration all individually improved the accuracy of the results produced by the system. It can also be noted that out of the three flash duration to ISI ratios (1:2, 1:3, 1:4), the 1:4 ratio caused significantly higher accuracy in overall system performance. With further research, the characters per minute produced by the system can be improved [21].

In a study by Rouja M. A. (2012) [22], it is proposed that using a hybrid of P300 signals and motor imagery signals may improve the performance of the P300 spelling system. In the system used for this experiment, motor event potentials (MEPs) are used together with P300 signals. MEPs are signals that are produced when one thinks about using a limb such as your right hand. These signals are near-instantaneous so it is hypothesized that they could improve the speed at which the P300 spelling system operates. Motor imagery signals, μ, can be detected with EEG using a bandpass filter of 10Hz to 22Hz. Two subjects were used for this experiment. The MEPs were used for a series of binary choices while the P300 signals were used for multiple choices. The subjects were tasked with spelling a given sentence. The first three consonants and vowels of each word were chosen by having the subject choose the desired column of characters using MEPs. After the column is chosen, the sequential intensification of letters method is used, and the subject chose the desired character using the P300 signals. After three letters are spelled, the columns then display suggested words using the T9 system. T9 is the "old school" method of texting via mobile phone that is conducted by choosing starting letters and then choosing a word from a set list that fits the chosen sequence of letters. The results of this study showed that the hybrid method reduced the overall time to spell the sentence by 70%. However, the accuracy of the spelling was reduced by 15%. This study does lay a path that can be improved upon. If the accuracy of this method is improved, it may be a viable alternative to the traditional Farwell and Donchic method [22].

Another alternative design to the Farwell and Donchic method was proposed by Guan C. (2005) [8]. This method follows the same six-by-six design; however, only one character was shown at a time. This method was coined the "Single Display Paradigm" or SD-speller. Instead of each row and column flashing randomly, each character in the matrix would randomly appear. For consistency, the flash rate of the SD-speller was kept the same as the conventional method described by Farwell and Donchic (1988) [23]. The results of this study conclude that the SD-speller displayed a better transfer rate of information than the traditional method. This was due to the differences in the P300 components of the ERPs between the methods. In the SD-speller, the P300 signals peaked from 0.6 to 2mV higher than the P300 signals produced by the traditional methods. This allows the BCI2000 data acquisition interface to better distinguish between desired and non-desired characters. The accuracies between the methods also showed a lower rate of error in the SD-spelling system. This may be due to the elimination of adjacent-character-error, outlined by Fazel-Rezai (2007) [7]. This study showed significant improvement to the P300 spelling system and outlines a new possible method for optimizing the system [8].

In a study by Lu et al. (2020), a paradigm was created using facial recognition. This study used the face of the subject and compared the results to a familiar face paradigm that used the face of a famous person. When using the self-face paradigm, it was found that ERP component amplitudes had increased from 340ms to 480ms in the parietal area (P300), from 480ms to 600ms in the parietal area (P600), and from 700ms to 800ms in the fronto-central area. The information transfer rate was also calculated and shown to have been improved by the self-face familiar face paradigm. In a study by Miyakoshi et al. (2008), it was shown that the P300 represents the recognition differences between the self-face and the famous face. This implied that the self-relevance of the image was proportional to the strength of the ERP response. This means that the self-face would produce the best results of any image because it will always have a high self-relevance [24]. This study also speculated that the reasoning behind the increase in ERP response between 700ms and 800ms in the fronto-central area may be due to subjects paying more attention to their own face than a famous face [25]. From this study, it can be surmised that when testing a familiar face paradigm using famous faces, the face used must be equally relevant to all subjects. This means that even though everyone may know a famous name, such as Michael Jordan, they may not be as familiar with the face as someone who is an avid sports enthusiast. The faces used in the paradigm should be carefully considered to not cause any bias in the results [25].

III. Research Questions

According to Lu J. (2015), a flash rate to interstimulus interval ratio of 1:4 showed improved accuracy and characters per minute. Based on this conclusion, this ratio will be kept static. A familiar face paradigm will be used as it is shown to produce larger ERP responses, resulting in better accuracy of the system (Lu et al., 2020). Along with the familiar face paradigm, the flash duration of the stimulus will be changed, having an interstimulus interval that is four times greater. If the flash durations of 32ms, 64ms, and 128ms are tested, will there be a difference in spelling time of the P300 spelling system and does the length of the flash duration cause significant changes in the accuracy of the results? Can spelling performance be improved if these flash durations are coupled with the presentation of familiar faces? If this experiment is conducted weekly for 24 weeks, will there be an improvement in the results over time.

IV. Hypothesis

If the flash rate is tested at 32ms, 64ms, and 128ms with a ratio of 1:4 between flash duration and interstimulus interval, the 64ms flash duration will have the fastest spelling speed without significant decrease in accuracy. The 32ms flash duration will have the fastest spelling time but will not be viable due to a significant decrease in accuracy. The 128ms flash duration will not have any significant decrease in accuracy but will not outperform the 64ms flash duration as it will not complete the spelling faster than the 64ms flash duration.

In comparison to the non-familiar face paradigm, it is hypothesized that the paradigm using the familiar face overlay will produce results with higher accuracy. It is predicted that this will hold true for all configurations of flash duration to ISI at a 1:4 ratio. The P300 response of the familiar face paradigm is hypothesized to have a higher amplitude than the non-familiar face paradigm, allowing for more accurate classification of the EEG data. This should allow for a better distinction between target and non-target events, resulting in higher accuracy. The amplitudes of the P300 responses in the familiar face paradigm will be larger than the amplitudes observed in the non-familiar face paradigm. It is hypothesized that the performance of the P300 spelling system will be best with a familiar face paradigm using a 64ms flash duration.

If the experiment is conducted over the course of twenty-four weeks, it is expected to see that the spelling performance will improve over time. It is expected that the spelling performance is to yield lower accuracy during the beginning weeks of testing and maintain high accuracy throughout the final weeks of testing. It is hypothesized that the speed of the P300 automatic spelling system will not be affected, but the accuracy will.

V. Significance and Rationale

The P300 spelling method is not as efficient as the conventional methods of communication such as oral speech, typing, or writing. The time it takes for the P300 speller to spell twenty characters has been shown to be in the order of minutes with non-perfect accuracy [21, 22]. For individuals who are "locked-in", this system optimization would put them one step closer to being able to communicate at a speed that is more conducive of normal conversation with spelling accuracy that would minimize frustration and confusion.

When we compare the speed of the different testing figurations, it is obvious that ideally, we would want the testing to take the least amount of time possible. However, if the flash duration is too fast or too slow, then the accuracy of the results may be reduced significantly. When comparing configurations that may be similar in speed, it is important to know if the difference in accuracy is significant. If one configuration shows slightly faster speeds but with reduced accuracy that is significant, then it may not be the ideal configuration for daily and regular use. Therefore, finding this balance between speed and accuracy is an important step toward the optimization of the basic P300 speller system. To measure the speed of the system, 32ms, 64ms, and 128ms flash durations will be used. This is so that they can be held consistent with flash durations used for testing found in recent literature. They also provide enough of a range to encompass what may be considered as too fast or too slow. This allows the results to show whether the increase or decrease of the flash duration will improve the performance of the spelling system.

The flash duration that results in the longest testing period will be 128ms. However, this should produce the largest ERPs, resulting in the most accurate results. This is because the flash duration will create a more "oddball' paradigm scenario due to the amount of time that is passing causing the target event to be rarer. It also gives the brain more time to recognize the target stimuli. This will cause a larger P300 response which can be more easily determined by stepwise linear discriminant analysis as a target ERP. However, this will not be the most efficient configuration of flash duration and ISI.

The 32ms duration flash will require more iterations than the 128ms flash duration. Although, due to the 1:4 ratio, the time required to complete the spelling of the characters will still be shorter than the 128ms flash duration. It is hypothesized that the 32ms duration does not give enough time to recognize target stimuli between consecutive flashes, producing an overlapping effect, resulting in reduced accuracy. This means that the number of iterations required by the software to determine the desired character will be significantly higher than the rest. Also, the 32ms flash duration will not cause an enhanced P300 response. Therefore, stepwise linear regression will not be able to classify target and non-target ERPs as easily, resulting in decreased accuracy.

Therefore, the 64ms flash duration will be the optimal flash duration for achieving the fastest spelling of the characters. This is because, with a 1:4 ratio of flash duration to ISI, there is at least 300ms total time given for the brain to elicit a P300 response without the overlapping of other desired stimuli ERPs. The P300 response observed should also be dramatic enough to be easily classified as a target ERP. It is estimated that same will hold true for the 128ms configuration. However, if the 128ms configuration results mimic those of the 64ms flash duration, then the time to complete the spelling will be longer.

The familiar face paradigm is predicted to also improve the speed and accuracy of the P300 speller. This paradigm has been shown to increase the amplitude of the P300 response, resulting in an easier classification of the target events as they will stand out more over non-target events. Li et al. showed that when testing the P300 spelling system with a familiar face paradigm, the P300 response was elevated in the frontal and parietal regions to levels higher than in the non-familiar face paradigm [26]. In a study by Spieler et al., the familiar face paradigm showed greater performance in all experimental trials compared to two non-familiar face paradigms. This was tested using both stepwise linear discriminant analysis and particle filtering [27]. In another study by Lu et al., it was shown that facial recognition could be improved by using faces that were relevant to the subject [25]. The faces used in the experiment will be very relevant to all subjects, such as the president of the United States [28]. This method of using familiar face paradigms has been shown to elicit stronger ERP responses as not only does it increase the amplitude of the P300 component, but it also evokes the N170 and N400 components [28]. This experiment seeks to use this familiar face paradigm in combination with the changing of flash duration and ISI times to find a paradigm with optimized parameters. The overall goal is to use this combination to produce a P300 spelling paradigm that can produce results faster without sacrificing accuracy. The combination of paradigm type with superior flash duration will be considered the more optimal parameter to be used for a P300 spelling system.

VI. Methodology

VI.A. Subject Recruitment and Consent

All the methods in this experiment, including the recruitment of the subject, have been approved by the East Carolina University's Institutional Review Board. Before testing can begin, each subject must provide his written informed consent before the start of testing. There will only be one subject recruited for this study. Due to the global pandemic, COVID-19, social distancing and precautions to maintain the safety and well-being of the subject's and researcher's health will be put in place. If the subject has a history of seizure disorders, visual impairment, or is on neuroleptic medications, they will be excluded from the study. The subject will only be allowed to participate in the study if they are able to give informed consent. A Montreal Cognitive Assessment test (MoCA test) will be given to the subject to determine if the subject shows any signs of cognitive impairment or illiteracy.

VI.B. EEG Recording

The subject will be fitted with a standard electrode cap containing 16 gold-plated, dry EEG electrodes (g.SAHARA). The electrodes' alignment will follow the 10-20 international system of locations [5]. The data from these electrodes will be obtained and amplified utilizing a g.tec g.USBamp amplifier. Amplified EEG signals are then transferred via USB to a secured laptop running BCI2000. All raw data will be stored within a xxxx.dat files in BCI2000 for future processing. The laptop used for storing data is password protected and has been approved by the ECU Institutional Review Board.

VI.C. P300 Spelling System

This P300 spelling system will work by collecting EEG data from the subject, filtering the data, classifying the data, and then deciding what character is to be displayed based on the classified data. First, the EEG data is collected from the subject using BCI2000 software. This data is then filtered using common average referencing (CAR) and artifact subspace reconstruction filtering (ASR). This filtering process removes unwanted artifacts from each EEG channel by keeping only valid EEG signals reflecting the brain's electrophysiological activities. This can help eliminate motion artifacts, electric noise, or other noise sources that pollute the data. The processed data is then classified using stepwise linear discriminant analysis (SWLDA). This method of classification allows the spelling system to determine if the stimulus presented to the subject is a target event or a non-target event. Once all the stimuli have been classified as target or non-target, the BCI2000 software uses the information to determine which character the subject wanted the computer to display.

The system uses a machine learning approach, which involves training and testing. Before the subject can be tested, a training session must be held so that the computer can be calibrated to the subject-specific data. Then, the testing session follows the same steps previously listed, but the data outputted will not be analyzed statistically. Now, when the computer processes the testing data, it will refer to this training data to confirm whether the data is target or non-target. This allows the computer to determine the desired results by comparing and validating the data to the training data that was originally fed into the system. This makes the testing process subject-specific as training data from the subject is not likely to work for anyone else.

VI.D. Paradigms

Two paradigms that will be used for testing. The first paradigm is the non-familiar face paradigm. A matrix consisting of five rows and columns, each containing five letters, flash individually in a randomized order. The letters “X” and “Z” have been removed from this matrix, although larger matrices can be created which incorporate all letters and other characters. The flash duration is based on the configuration that will be used. The time between the flashes is the interstimulus interval (ISI) and is maintained at a 1:4 ratio between flash duration and ISI. An example of this paradigm is shown below in Figure 4.

Figure 4. This image shows a screenshot of the non-familiar face paradigm. In this image, the phrase that is to be spelled is located at the top of the screen, and the third column of letters is flashing bright white.

The second paradigm used is the familiar face paradigm. This paradigm is like the non-familiar face paradigm as it consists of a matrix of letters. Between flashes (known as the ISI), the familiar face paradigm is identical to the non-familiar face paradigm. However, during the flash, instead of the row or column flashing bright white, a familiar face is flashed instead of the letter. For this experiment, the familiar face used was that of Tiger Woods. However, other familiar faces can be used as well such as the President of the United States' face. An example of this paradigm is shown below in Figure 5.

Figure 5. Figure 4. This image shows a screenshot of the familiar face paradigm. In this image, the phrase that is to be spelled is located at the top of the screen, and the first column of letters is flashed, showing Tiger Woods' face over each letter in the column.

VI.E. Training Session

A training session must first be held before the testing of the different configurations can begin. The BCI2000 software does not produce identical results for people as everyone has different brain waves. Therefore, the software must be trained to recognize the P300 responses that are specific to the user. Only after the training session is completed and information gathered is regressed and classified, can we begin to test the various flash durations.

Before training can begin, the software must be prepared. To do so, the parameters that are to be used for testing must be loaded in and then changed for the desired experimental design. Many of the parameters are prepared before the time of testing and saved into a parameter file that is loaded into the software. Also, the testing environment must be kept optimal to ensure there is minimal noise picked up by the EEG. Loud noises, distractions, changes in light, and other non-ideal environmental parameters could be picked up by the EEG or cause spikes in electrical activity in the brain which will then be measured by the EEG. Environmental factors, such as larger monitors for the paradigm, have also been shown to affect the size of the P300 component of the ERPs. Finally, the subject must be given clear instructions on how to perform the training. This is done to minimize any human error.

The training session must be performed for each paradigm before any testing sessions can be held. For each paradigm, the training session uses a 128ms flash duration with 10 iterations. It is believed that using the longest flash duration used in testing will elicit the strongest ERPs which allows the SWLDA classifier to better identify the P300 component. The parameter file is loaded into the BCI2000 software with the 128ms flash duration and the phrase "THEBIGDWARFONLYJUMPS" is spelled. This is the phrase that will be spelled for both training and testing sessions. Note that the BCI2000 software will output letters that do not match the desired phrase. This is because the training data has not yet been classified, therefore the BCI2000 software does not have the user's training data. Once this training session is completed, the training data is inputted into the SWLDA classifier. The classifier is checked to ensure that it is using the same parameters that are specified in the parameter file that was loaded into the BCI2000 software. Once the training data has been fed into the classifier, feature weights will be generated. These feature weights allow the BCI2000 system to determine if a stimulus is considered a target or non-target based on the ERP response generated by the user in response to the stimulus. These feature weights are loaded into the BCI2000 system and can now be used for testing.

After the training session has been concluded, the BCI2000 software generates a xxxx.dat file that can be loaded into the P300 classifier. This P300 classifier uses the common average referencing, artifact subspace reconstruction filtering, and stepwise linear discriminant analysis to determine which ERPs are significant, or target ERPs, and which ERPs are not significant, or non-target ERPs. This classifier uses the model for each flash iteration and displays the results of each iteration, along with its accuracy. For example, if the intended word is "PURPLE", the classifier may determine that the first iteration returned "PURJIT" which has a 50% accuracy. The next iteration, it may return "PURJIE" which is a 66% accuracy, and so on until the stepwise linear discriminant analysis has determined the intended characters with 100% accuracy. An example of this is shown below in Figure 6.

To generate these results, a xxxx.dat file that was generated from the previous training session is taken and loaded into the *Data* tab of the P300 classifier. Next, in the *Parameters* tab, ensure the following things: the spatial filter is set to CAR, the channel set includes all the relevant testing channels on the EEG headset, and the response window mimics the "Epoch Length" from the BCI2000 *Filtering* tab. In this case, it should be set to 800ms. After these parameters are set, click on "Generate Feature Weights" in the *Data* tab. This will run the stepwise linear discriminant analysis model and return the results after each iteration of flashes. This will also generate a xxxx.prm file that will be loaded back into BCI2000 software. This file contains the training and regression data that is specific to the user. When this file is loaded into the BCI2000 software, the software will be able to use this new training information to classify

ERPs as target and non-target. This will allow the P300 speller paradigm to be used in online-free mode, meaning that any character can be spelled at will.

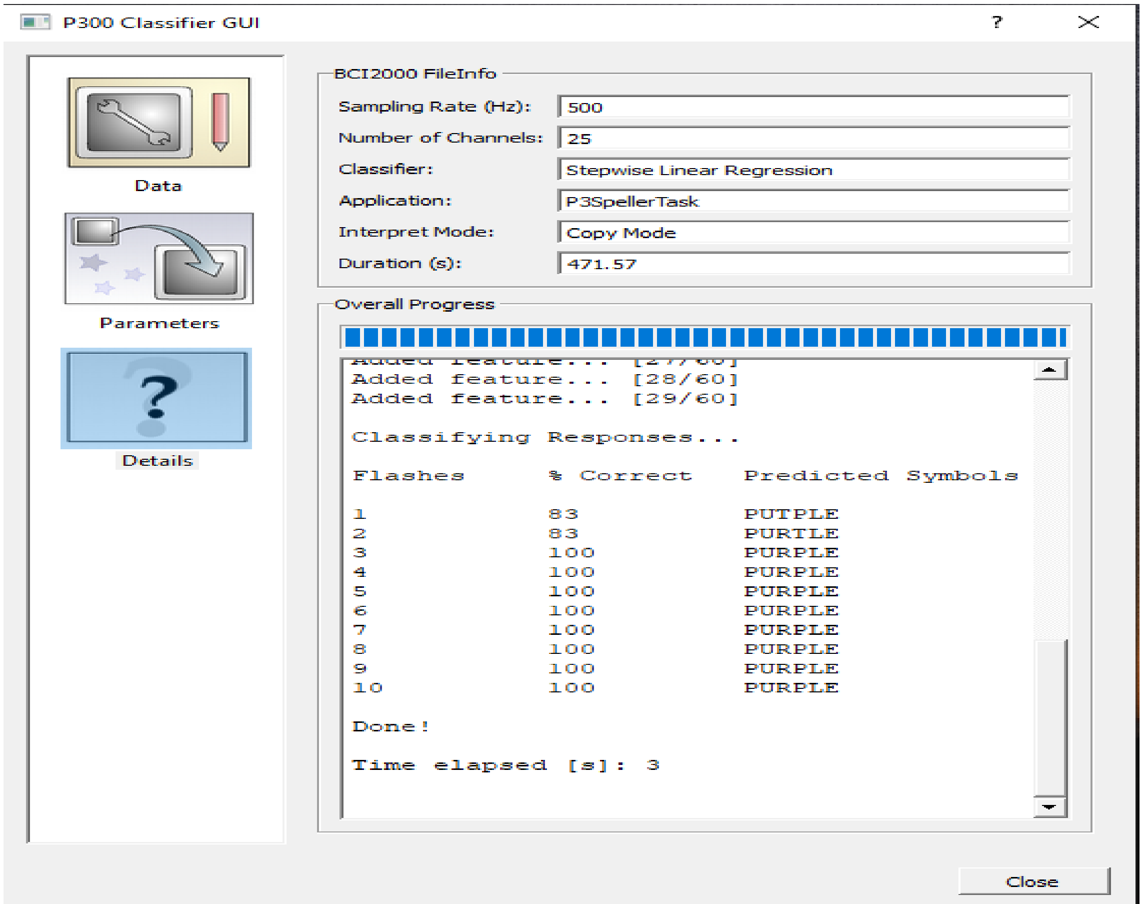


Figure 6. This image shows the details generated by the P300 classifier. Each iteration is tested using stepwise linear regression with the machine-learning algorithm allowing for increasingly accurate results. In this example, the minimum number of iterations required to achieve 100% accuracy is three. The time it took to complete the training is also shown at the top of the GUI. In this case, the training session took a total of 471.57 seconds.

VI.F. Common Average Referencing (CAR)

To explain what the P300 classifier is doing, the common average referencing spatial filter and stepwise linear discriminant analysis classification technique must be defined. Common average referencing can be used to minimize noise from a signal that can be generated from sources outside of the EEG's target zone. This noise would include motion artifacts and electrical noise [29]. To exclude these unwanted noises, we can use CAR. This is often very necessary as the signals generated are very weak (μV range) and can be overshadowed by these noise artifacts (mV range). The number of channels in an EEG recording can be represented by K [29].

$$d_{k,t} = s_{k,t} + w_k \times n_t \tag{1}$$

In Equation 1, $d_{k,t}$ represents the recorded signal channel, k, at time, t. $s_{k,t}$ represents the desired signal, w_k is the weighting coefficient between channels, and n_t is the noise artifacts. The weighting coefficient, w_k, can be assumed to be 1 as ideally all the channels would have the same level of noise. The CAR is generated by taking an average of each sample and using it as a global reference across all channels [29, 30].

$$s_{k,t} = d_{k,t} - \hat{n}_t \tag{2}$$

where

$$\hat{n}_t = \frac{1}{K}\sum_{k=1}^{K} d_{k,t} \tag{3}$$

This means that after the averaging, only the noise artifacts that are common amongst all channels will remain [29, 30]. After we have used CAR for spatial filtering, we can then move forward with the classification of the training data. This classification is done using stepwise linear discriminant analysis.

VI.G. Artifact Subspace Reconstruction Filtering (ASR Filtering)

To further remove artifacts such as noise from electrical, environmental, and motion noise ASR filters were used [31]. ASR is an automatic, adaptive, component-based method to correct or remove artifacts, including those that are transient or have a large-amplitude in data comprising multichannel EEG data. First, the ASR filter is calibrated to learn the data principal component space. In this step, it sets up a covariance matrix and uses it in future steps to determine whether a chunk of data is clear or contaminated by artifacts. To conserve the EEG data during calibration, the mean of the data is subtracted and pushed through an IIR filter before computing independent component analysis (ICA) [32]. A mixing matrix is then computed, and the root mean-square is taken of the principal components to set the thresholds of the filtering. Then, the data is analyzed in chunks to determine whether it contains artifact components. If it does, those components are removed and replaced with latent components taken from calibration [32].

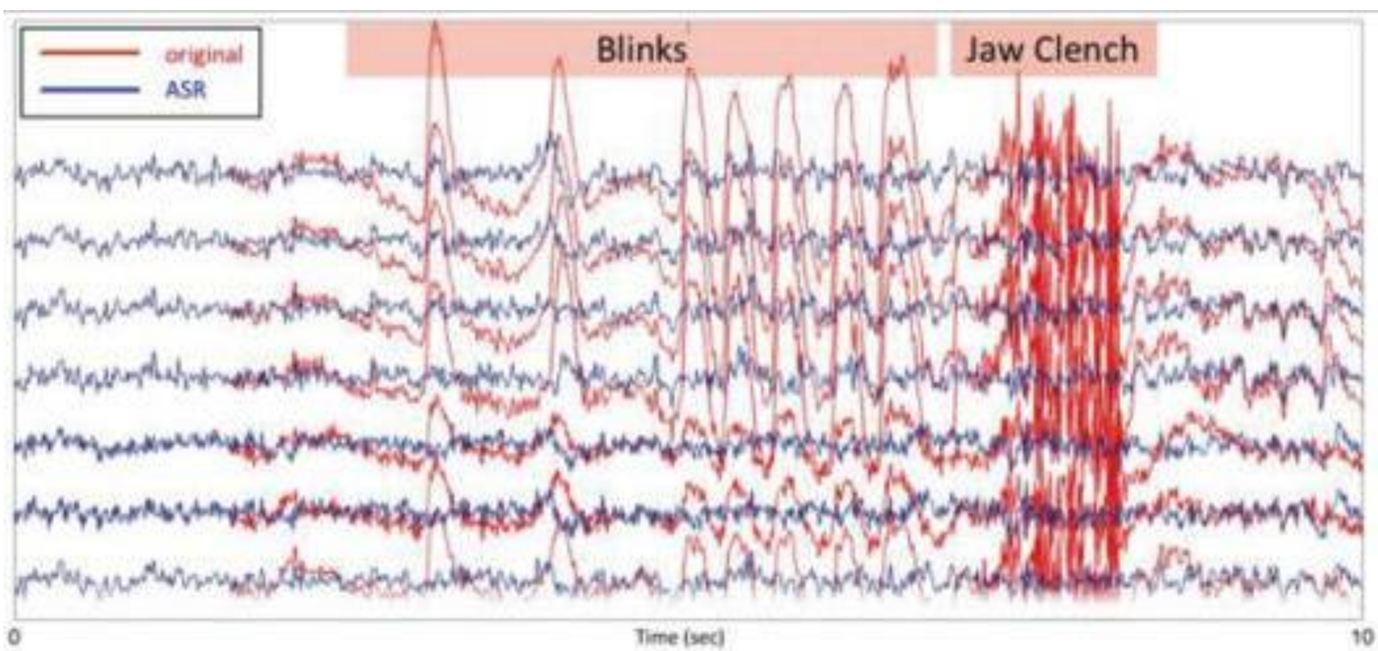


Figure 7. This image shows 10 seconds of EEG data following ASR data cleaning (blue trace) superimposed on original data (red trace) [32].

VI.H. Stepwise Linear Discriminant Analysis

To classify the testing data, a program is used that performs stepwise linear discriminant analysis (SWLDA). The linear discriminant analysis (LDA) is a two-variable classification technique used to determine the target and non-target stimuli. We can say that we have a set of observations, x, for each sample of an event, y [33]. This set of samples is the testing set. To find a predictor of y, we need to create a vector of coefficient estimates for the terms in the final model, b. These can also be referred to as the feature weights. When we use LDA, we assume that the probability density functions $P(x|y = 0)$ and $P(xy = 1)$ have a normal distribution with a mean and covariance of (μ_0, Σ_0) and (μ_1, Σ_1). LDA then moves to use Bayes optimal solution [34, 35, 36, 37].

The purpose of a linear discriminant analysis model is to create a predicted mean response variable, $\hat{y}$. This predicted response variable is what we use to classify each event as target or non-target. Here, x represents the predictor variables in a numeric, n-by-p matrix with n being the number of variables and p being the number of observations. The predictor response variable, $\hat{y}$, is returned as a n-by-1 logical array which is used to classify the results. An example of this logical array can be shown below:

$$\hat{y} = [1\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ ]$$

This logical array is representing 10 different events, with them having a predicted value of 1 or 0 which can be classified as target or non-target. Each entry in this logical array is the corresponding rows of x. To obtain the logical array of predicted responses, $\hat{y}$, we need to obtain the estimating coefficients, b. These coefficients are a numeric vector that corresponds to the terms in x. The method of least squares is used to determine these coefficients. If a term is included in the final model, then its estimated coefficient, b, is estimated from its resulting fitted model. If a term is excluded, then b is estimated from its resulting fitted model plus that term. This process of obtaining the feature weights, b, can be used to find the predicted responses, $\hat{y}$. The equation for the predicted responses can be seen below in Equation 4.

$$\hat{y} = bx \tag{4}$$

This linear equation outputs the predicted response, $\hat{y}$, based on the product of the input matrix, x, and the estimating coefficients, b. This produces a linear result that can be tuned using a stepwise fitting.

Stepwise LDA then goes to take LDA further by having the model change in steps. Using forwards and backwards regression, thresholds are set that determine which variables are to be held and which are to be thrown out. These thresholds are a p-value of an F-statistic [38]. This means that the correlation of a term in relation to a group of like terms is calculated and given a threshold. In the forwards regression portions of the stepwise process, an entrance threshold is set and the variables which meet this threshold are classified as a target variable. If a term is tested for correlation against a group of terms that have been deemed target variables and was found to have a p-value that is above the threshold, then that term would join the group of target variables. In the backwards regression portions of the stepwise process, an exit threshold is set which determines which variables are thrown out of the matrix of target variables due to not meeting the conditions of the exit threshold. The exit threshold is not allowed to be smaller than the entrance threshold because this would cause the stepwise process to enter an infinite loop of adding and removing the same variable repeatedly. Each iteration of this stepwise process changes the LDA threshold based on which variables are held in the target variable matrix that step. This means that as a variable is taken in or thrown out, the threshold will change. A variable that was not originally chosen as a target variable or a variable that was thrown out can then be re-added to the target variable matrix based on how the variables inside affect the threshold. The stepwise process terminates when no single step affects the outcome [38]. This stepwise process of LDA adds a layer of filtering to the classification, allowing for a narrower scope of what may or may not be considered a target variable [39].

This technique of classifying the data using SWLDA was shown to be superior to non-stepwise LDA [37]. In a study by Manyakov et al. (2011), it was shown that the stepwise version of LDA had outperformed the LDA, although the results were not significant. This was also compared to five other classification techniques with two other methods being shown as having better accuracy than the SWLDA, however, SWLDA is shown to be better suited for training data as it is adjustable [39].

VI.I. Experimental Design - Testing

After the training session is held and the training data has been classified to create the feature weights, the testing sessions can begin. First, the non-familiar face paradigm will be tested. The parameter file for the 128ms flash duration and non-familiar face paradigm is loaded into the BCI2000 software. Then, the file containing the feature weights that was created by the SWLDA classifier is loaded into the BCI2000 software. Now, the software will be able to differentiate between target and non-target stimuli based on the features of the ERPs elicited by the stimuli. The subject is instructed to spell the same phrase that was used in the training session, "THEBIGDWARFONLYJUMPS". This phrase is used because it contains 20 letters that are all used only once. Once the subject has been instructed how to conduct the testing procedure, they can begin. This testing procedure also uses 10 iterations for each letter. Once the phrase has been spelled, the testing session is ended. In the BCI2000 software, the flash duration is changed to 64ms and the ISI is changed to reflect a 1:4 ratio. The testing procedure is then repeated for this configuration. Once this is completed, the same testing procedure will be held for the 32ms flash duration configuration. After the testing session using the 32ms flash duration is completed, the testing sessions for the non-familiar face paradigm are concluded.

After the three testing sessions have been completed for the non-familiar face paradigm, the training session procedure must be repeated for the familiar face paradigm. Once the training session is completed and a feature weight file has been generated by the SWLDA classifier, testing sessions can be held. The parameter file for the 128ms flash duration and familiar face paradigm is loaded into the BCI2000 software. Then, the file containing the feature weights that were created by the SWLDA classifier is loaded into the BCI2000 software. The subject is instructed to spell the same phrase that was used in the training session, "THEBIGDWARFONLYJUMPS". Once the subject has been instructed how to conduct the testing procedure, they can begin. This testing procedure also uses 10 iterations for each letter. Once the phrase has been spelled, the testing session is ended. In the BCI2000 software, the flash duration is changed to 64ms and the ISI is changed to reflect a 1:4 ratio. The testing procedure is then repeated for this configuration. Once this is completed, the same testing procedure will be held for the 32ms flash duration configuration. After the testing session using the 32ms flash duration is completed, the testing sessions for the familiar face paradigm are concluded. This experimental procedure is conducted weekly over the course of six months.

VI.J. Data Analysis - MATLAB

Once we have the six xxxx.dat testing files, they are analyzed in MATLAB to determine the speed and the average height of the P300 responses generated by the different testing configurations. To do so, a code was written that loads in xxxx.dat files, then separates target ERPs from non-target ERPs, then epochs target ERPs and non-target ERPs, finds the mean of both, and finally saves the file as a .mat file for further analysis. This code is written to accommodate different length character combinations, different dimension matrices, and various numbers of iterations.

To begin using the code, the testing file generated by the BCI2000 software is inputted into the code as the desired file. Next, details about the file are entered to allow the code to adjust to different testing parameters. These details that are inputted are: the number of iterations, number of rows, number of columns, and number of characters. This is important to keep consistent with the testing data file. For example, if we have four iterations, five rows and columns, and six letters, then the code can calculate the number of events. Events refer to individual flashes. In this case, there would be 240 total events. This is because each row and column would flash four times (40 events), for each letter (six letters). The code then calculates the number of target and non-target ERPs there are per letter. To get the number of target ERPs, simply take the number of iterations times two. This is because if there are four iterations, then the correct row will flash four times and the correct column will flash four times. This is a total of eight target ERPs out of 40 total ERPs per letter. This will provide the basis for the code to separate and epoch the target and non-target ERPs. These specifics can all be adjusted to reflect the testing file.

Next, the code moves to separate the target and non-target ERPs. It does so by utilizing EEG-Lab toolbox to determine the position of each event. Each position corresponds to a specific row or column on the matrix. This allows for it to search for the events that occur at certain positions, separating the target and non-target ERPs. These are separated on a letter-by-letter basis so that we can average and observe many P300 responses per data file.

After the separation of target and non-target ERPs, the code has to epoch the data. Every event, both target and non-target, is associated with a specific timestamp that we can use to find where in the raw data the event is located. For example, if Event 1 has a time stamp of 1337, then the event happened at the 1337th data point recorded by each channel. If we have a sampling frequency of 500Hz, then we can determine that this event happened 2.674 seconds into testing. For each group, target and non-target, each event is associated with its timestamp and then epoched. This epoch is taken from 200ms before the event (200ms pre-stimuli) to 800ms after the event (800ms post-stimuli). This gives a one-second view around the event. We can expect to see the P300 response around 300ms after the event occurred, or the 500th data point in the epoch. Then, the average can be taken of the epoched target and non-target ERPs for each letter. This will find the mean between each of the events across the same channels. For example, if the target events for the first intended character are Event 1, Event 3, and Event 5, then the mean of these epoched ERPs will be the average ERP for that character. With a sampling frequency of 500Hz, this means that a 1-second epoch will be 500 data points per channel. With 16 channels, this means that the epoched data for each event will be a 16 by 500 matrix. When we take the mean of these events, we are taking the mean of each cell of the matrix across events. So, cell {1,1} of all three events will be averaged, and the same done for the rest of the cells. This will condense all the events in a group into one matrix containing the average target ERP or average non-target ERP per character. In the case of this project, ERPs will be grouped as target and non-target per letter.

The average target and non-target ERPs are then ready to be analyzed further. A second code to visualize the ERPs is created. As each row of the matrix represents a different EEG channel, we can plot the average target ERP against the average non-target ERP for each channel. A plot can be created for each letter with one subplot for each EEG channel. The peak can be easily obtained by using the "max" function in MATLAB. An example of this plot with the associated subplots is shown below in Figure 5, located in the *Results* section.

VI.K. Analysis of Results

To determine the performance of the different configurations, an accuracy of 95% will be used as the required accuracy to be determined as acceptable. This is because with modern spelling correction algorithms, we can accurately determine the desired word when a word with three letters or more is misspelled by 1 letter. For example, if the system determines the desired letters to be T, W, and E, we can determine the word to be either "two" or "the". However, if more than one letter is misspelled, it is exceedingly harder to correct the spelling error. Because the phrase used for testing has 20 letters, 95% accuracy is used because it represents 19 of 20 letters being spelled correctly. This cannot be more lenient because even though the 20-letter phrase is comprised of five words, it cannot be foreseen whether multiple misspelled letters will be consecutive or not.

VII. Results

VII.A Analysis of P300 and N170 Components

To visualize the ERP responses obtained, a figure was created with a subplot for each EEG channel. 16 subplots were arranged according to the 10-20 international mapping system which shows the location of the electrodes on the head. Each subplot contains a 1 second long epoch which shows the average ERP response for both target and non-target events. These plots were created for every testing session. An example of these subplots for the non-familiar face paradigm with a 128ms flash rate can be seen below in Figure 8.

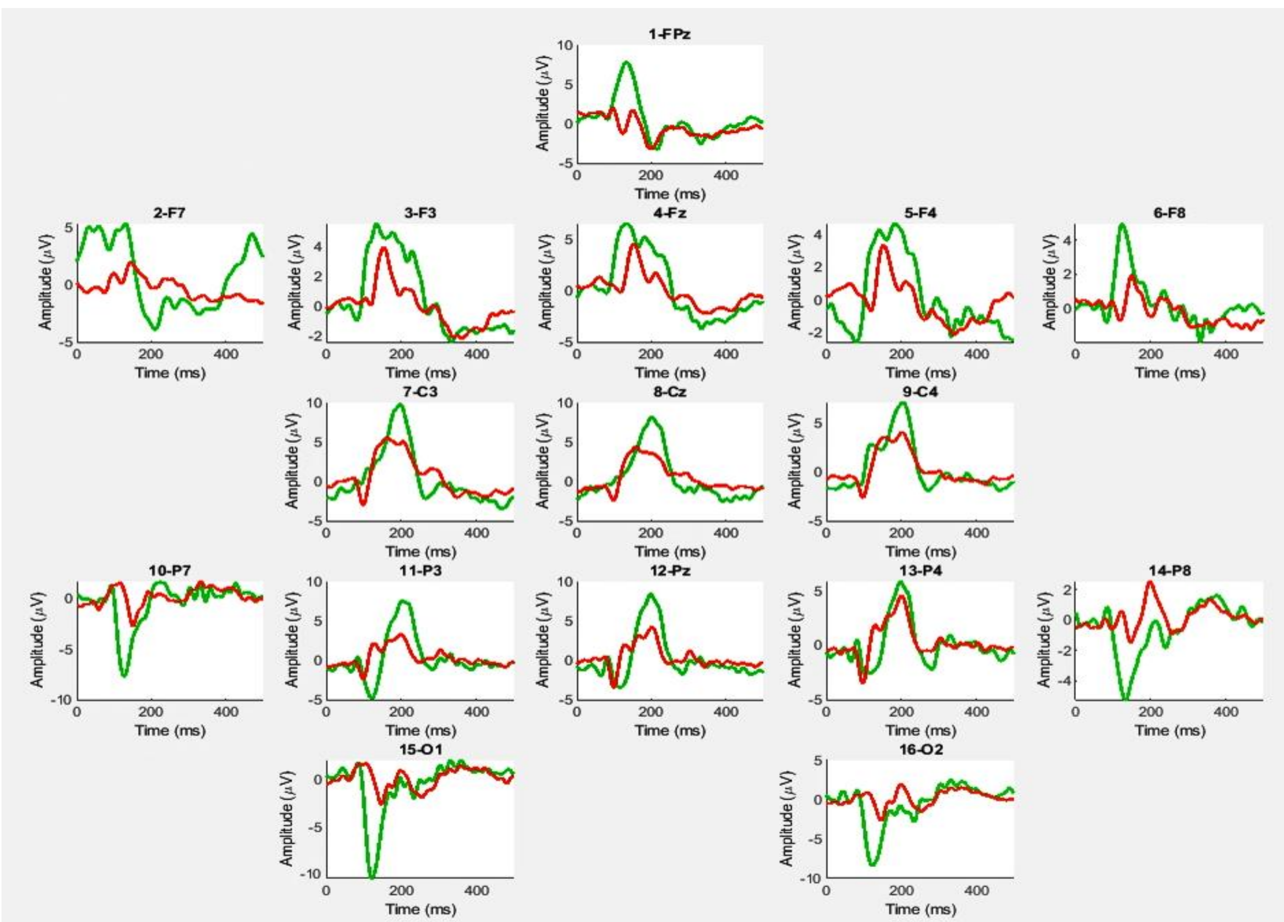


Figure 8. This figure shows the average P300 amplitude for both the target and non-target events using the non-familiar face paradigm with a 128ms flash rate. The green line represents the target events and red line represents the non-target events. This is calculated for all 16 channels and is displayed in the 10-20 international mapping system which shows the location of the electrodes on the head.

For target events, these subplots show the average peak P300 amplitude of the non-familiar face paradigm is 6.8059 μV with a range from 4.6212 μV to 9.8091 μV. The average N170 amplitude was -7.9175 μV. For non-target events, the average peak P300 amplitude of the non-familiar face paradigm is 3.8112 μV with a range from 2.179 μV to 5.6695 μV. The average N170 amplitude was -2.9161 μV.

This same analysis was done using for the familiar face paradigm. The results for the same testing session can be shown below in Figure 9. This figure shows the results for the familiar face paradigm with a flash rate of 128ms.

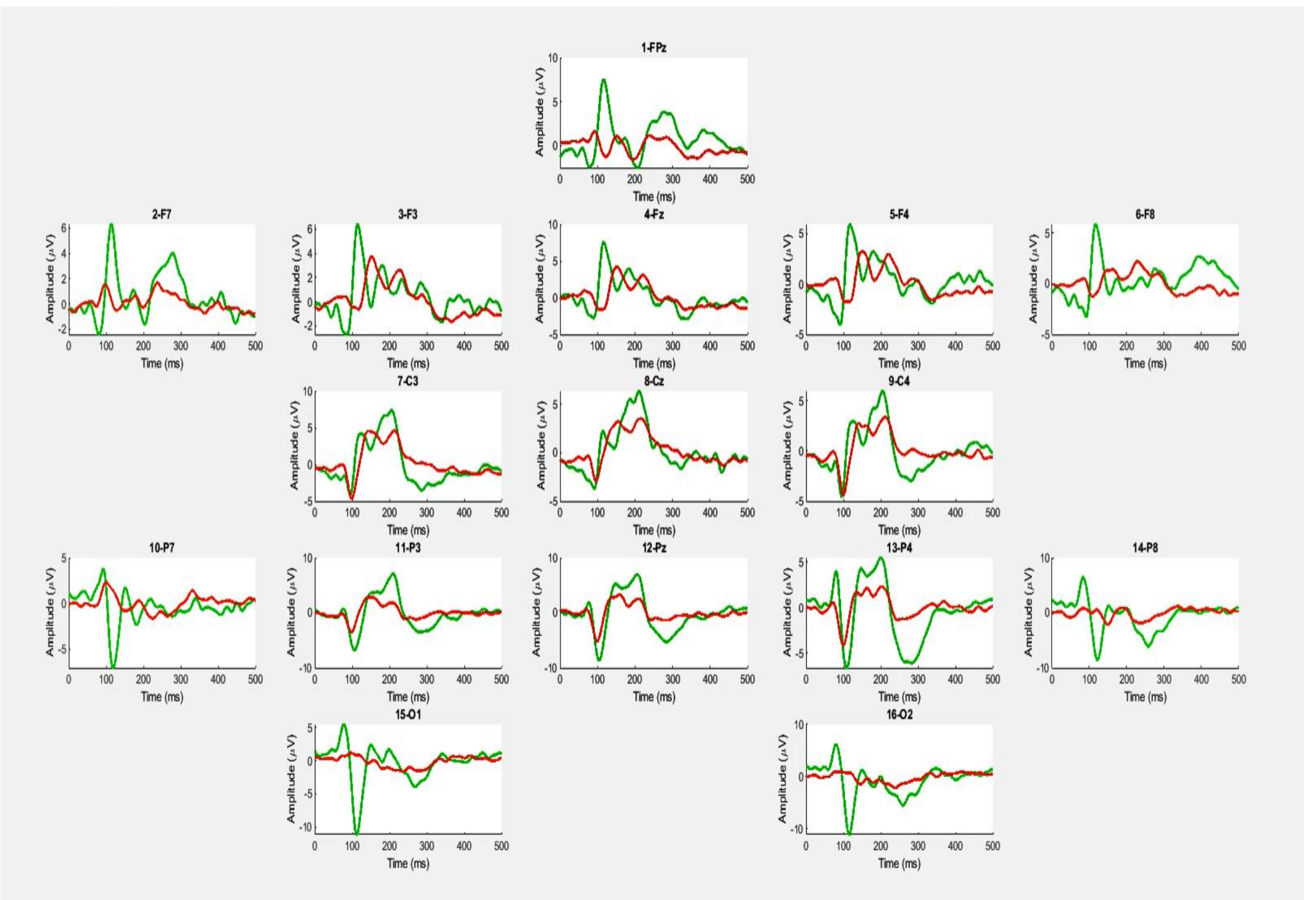


Figure 9. This figure shows the average P300 amplitude for both the target and non-target events using the familiar face paradigm with a 128ms flash rate. The green line represents the target events and red line represents the non-target events. This is calculated for all 16 channels and is displayed in the 10-20 international mapping system which shows the location of the electrodes on the head.

For target events, these subplots show the average peak P300 amplitude of the non-familiar face paradigm is 8.2898 μV with a range from 5.3189 μV to 11.5502 μV. The average N170 amplitude was -10.4522 μV. For non-target events, the average peak P300 amplitude of the non-familiar face paradigm is 5.0023 μV with a range from 1.6347 μV to 8.9445 μV. The average N170 amplitude was -2.5329 μV. These values are much larger than those produced using the non-familiar face paradigm.

The P300 amplitude averages were found for each testing configuration. These averages were taken only from the target events. Non-target events were not calculated for this portion of the analysis. A figure containing two subplots was created. Each subplot contains three lines which represent the three flash durations used. The first subplot represents the results of the non-familiar face paradigm, and the second subplot represents the results of the familiar face paradigm. These subplots can be seen below in Figure 10.

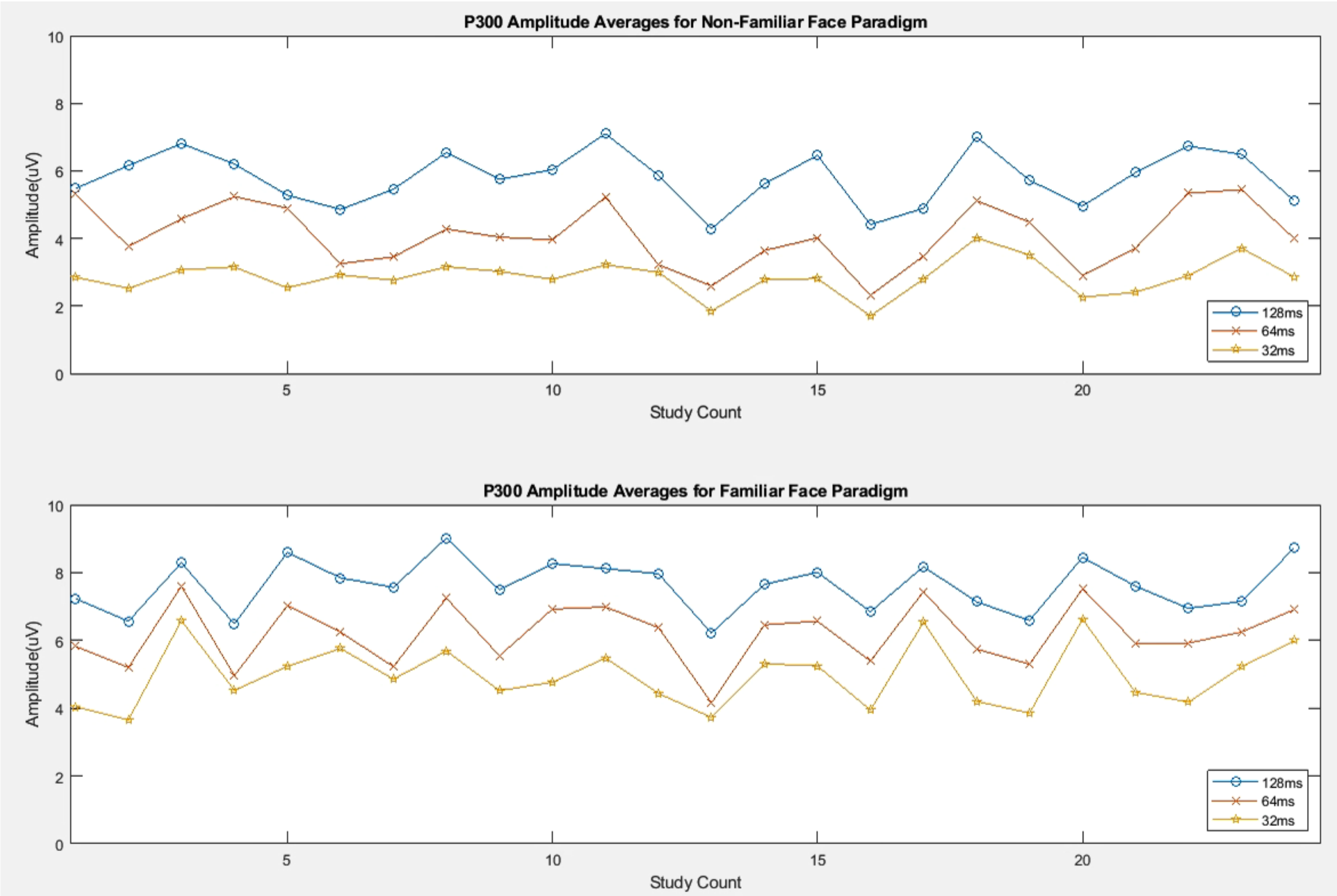


Figure 10. This figure contains two subplots that show the P300 amplitude averages for the non-familiar face paradigm(top) and the familiar face paradigm(bottom). Both paradigms exhibited P300 amplitude averages that were highest using a 128ms flash rate, followed by the 64ms flash rate, with the 32ms flash rate having the lowest average amplitudes.

For the non-familiar face paradigm, the average P300 amplitudes were highest with the 128ms flash duration, followed by the 64ms flash duration, and lowest with the 32ms flash duration. For the 128ms flash duration, the average P300 amplitude was 5.8311 µV with a range of 4.2862 µV to 7.1024 µV. The 64ms flash duration had an average P300 amplitude of 4.139 µV with a range from 2.3115 µV to 5.4414 µV. Last, the 32ms flash duration had an average P300 amplitude of 2.8341 µV with a range of 1.7035 µV to 3.7556 µV.

For the familiar face paradigm, the same trend was seen 128ms flash duration having the highest average P300 amplitudes, followed by 64ms flash duration, and the 32ms flash duration having the lowest. For the 128ms flash duration, the average P300 amplitude was 7.6656 µV with a range of 6.2141 µV to 9.0202 µV. The 64ms flash duration had an average P300 amplitude of 6.1007 µV with a range from 4.1568 µV to 7.5945 µV. Finally, the 32ms flash duration had an average P300 amplitude of 4.8667 µV with a range of 3.654 µV to 6.6218 µV.

For both paradigms, the 128ms flash duration consistently acquired a higher average P300 amplitude than the 64ms flash duration. The same is true for the 64ms flash duration and the 32ms flash duration. However, for all three flash durations, the average P300 amplitude is higher with the familiar face paradigm. A comparison of the average P300 amplitudes between the familiar face paradigm and the non-familiar face paradigm can be seen below in Figure 11.

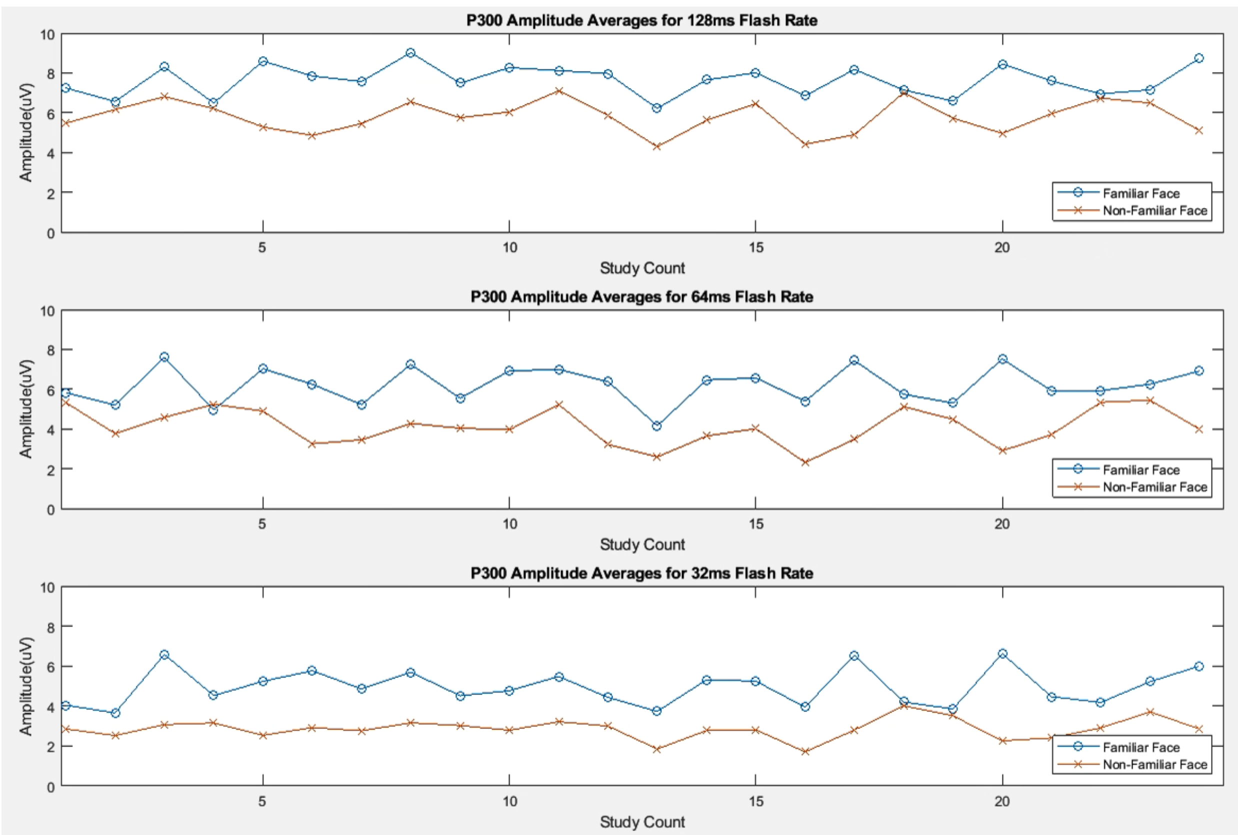


Figure 11. This figure contains three subplots that show the P300 amplitude averages for the 128ms flash rate, 64ms flash rate, and 32ms flash rate configurations. Each subplot contains the line plot for the non-familiar face paradigm and the familiar face paradigm.

For the 128ms flash duration, the familiar face paradigm consistently produced a higher average P300 amplitude than the non-familiar face paradigm. This amplitude was up to approximately 3 μV higher than the non-familiar face paradigm in some sessions.

For the 64ms flash duration, the familiar face paradigm consistently produced a higher average P300 amplitude than the non-familiar face paradigm, except for one testing session which produced a higher amplitude in the non-familiar face paradigm. Even for this one session, the difference between the non-familiar face paradigm and the familiar face paradigm is a mere 0.2801 μV.

The same trend that is seen in the 128ms flash duration comparison can be seen in the 32ms flash duration comparison. The familiar face paradigm consistently produced a significantly higher average P300 amplitude. This flash duration showed the largest amount of average difference between the familiar and non-familiar face paradigms.

VII.B Accuracy Over Each Iteration

Next, the performance of the system could be shown as a measure of accuracy over iterations. A figure was created with two subplots, one for each paradigm. Because there were 10 iterations used for each letter during testing, the x-axis shows 10 ticks which represent each of the 10 iterations. The accuracy is the average accuracy at that interval for all testing sessions and is displayed as a percentage. This figure and subplots can be seen below in Figure 12.

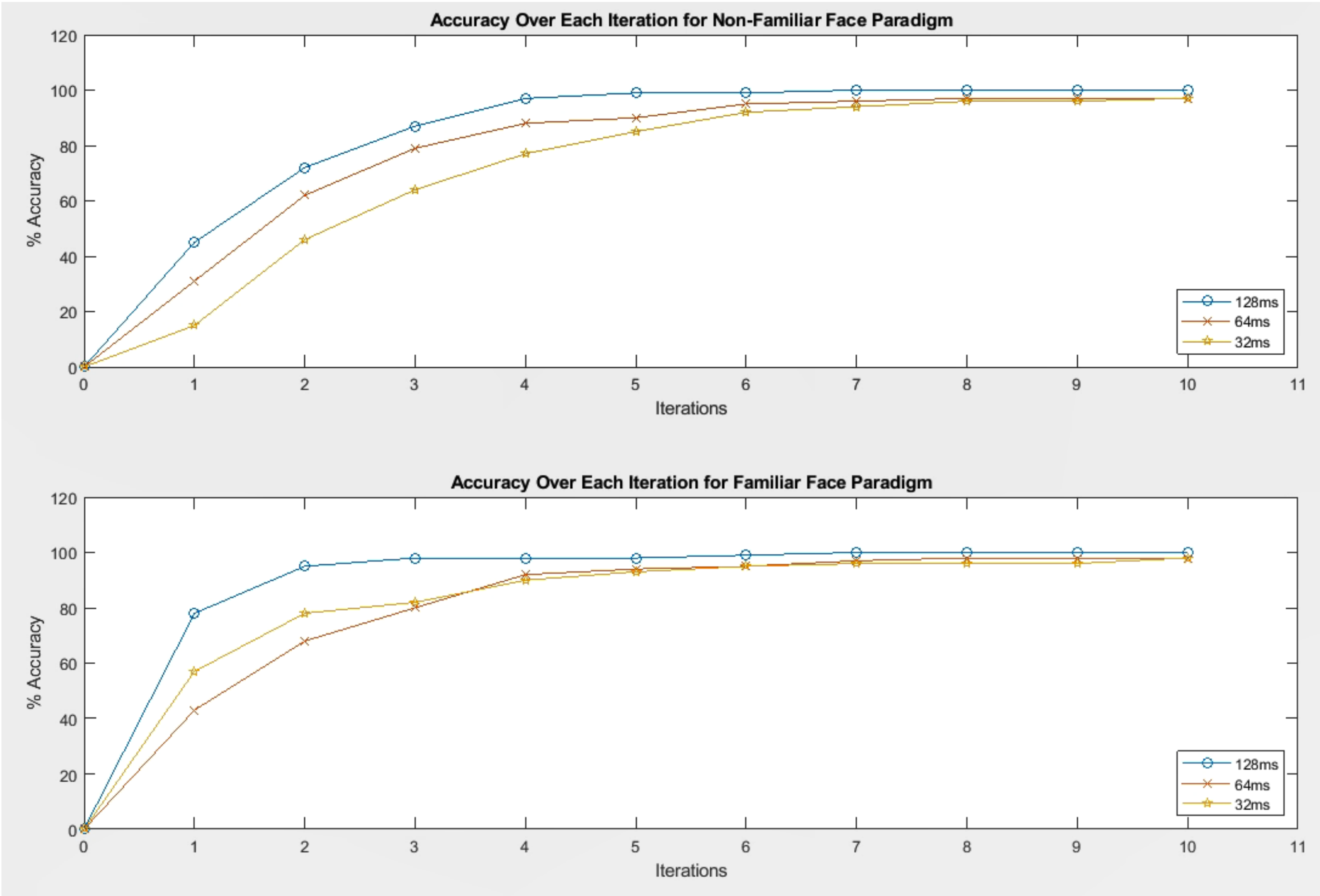


Figure 12. This figure contains two subplots that show the spelling accuracy over each iteration for the non-familiar face paradigm(top) and the familiar face paradigm(bottom). Both subplots contain three lines; one line for each flash rate configuration.

For the non-familiar face paradigm, the number of iterations required to reach 95% accuracy (the accuracy which represents only 1 letter misspelled) increases as the flash duration time decreases. For the 128ms flash duration, only 4 iterations were needed to reach the 95% accuracy needed. This is compared to the 64ms and 32ms flash durations which needed 6 and 8 iterations, respectively.

For the familiar face paradigm, the same trend is seen in the non-familiar face paradigm with the number of iterations required rising as the length of the flash duration decreases. For the 128ms flash duration, it only required 2 iterations to reach 95% accuracy. This is compared to the 64 and 32 ms flash durations which both required 6 iterations.

In comparison, the familiar face paradigm saw a steeper rise in accuracy with the results mimicking an exponential plateau as where the non-familiar face saw a more linear rise in accuracy. Also, the number of iterations required decreased for both the 128ms and 32ms flash durations when using the familiar face paradigm. However, the 64ms flash duration required the same number of iterations as the non-familiar face paradigm. This can be seen below in Figure 13.

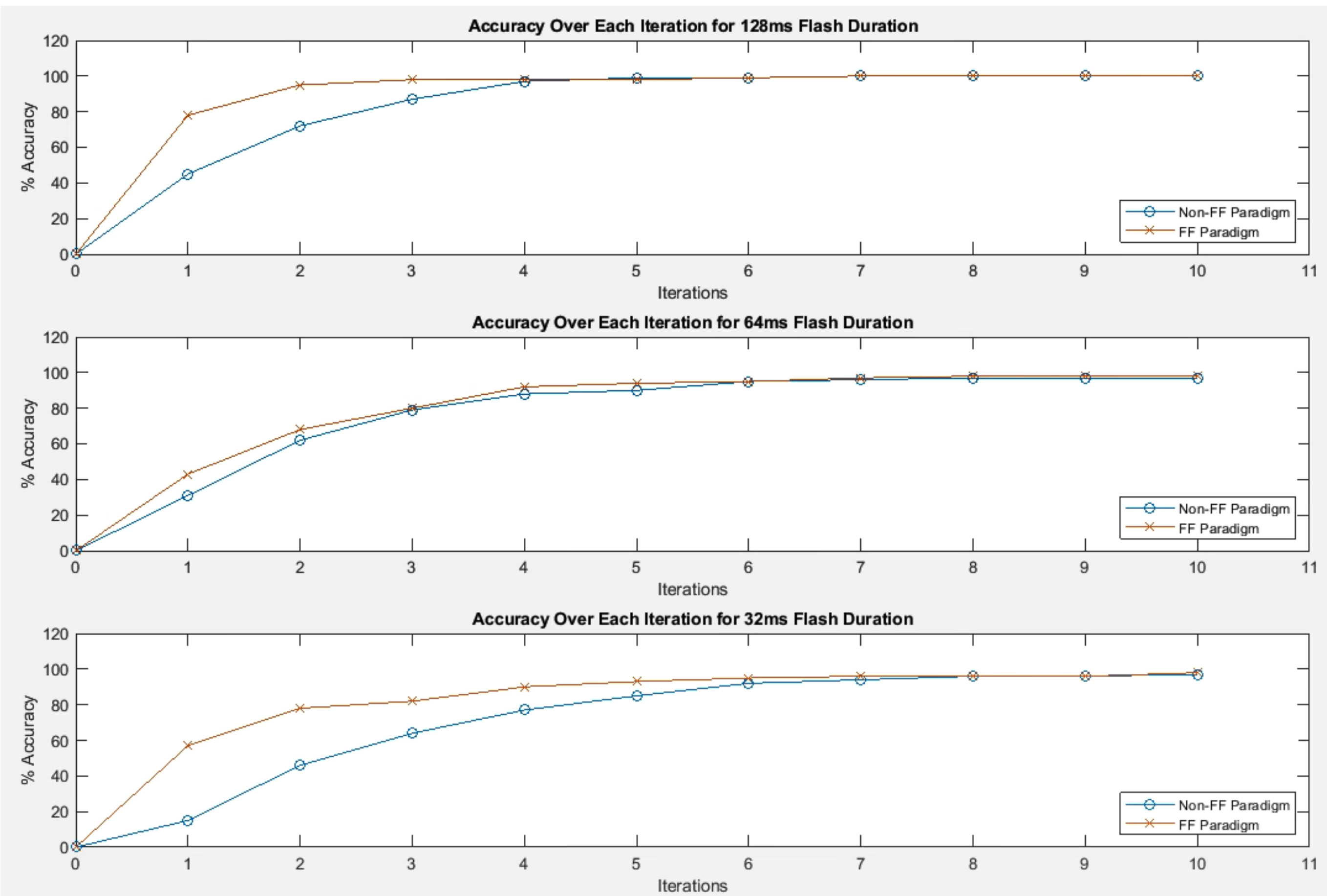


Figure 13. This figure contains three subplots that show the spelling accuracy over each iteration for the 128ms, 64ms, and 32ms flash durations. The subplots contain two lines; one line for each paradigm.

VII.C Accuracy Over Time

Finally, the performance of the system could be analyzed using the accuracy compared to time. A figure was created with two subplots, one for each paradigm. Each subplot shows the average accuracy for each flash duration as a function of time. Because there were 10 iterations used for each letter during testing, the x-axis shows the time needed to reach 10 iterations. For the 32ms flash duration, this time to complete 10 iterations is 16 seconds (1.6 seconds per iteration). For the 64ms flash duration, the time it takes to complete 10 iterations is 32 seconds (3.2 seconds per iteration). For the 128ms flash duration, the time it takes to complete 10 iterations is 64 seconds (6.4 seconds per iteration). Because the accuracy of the 64ms flash duration and the 128ms flash duration reaches 100% before the 10th iteration in both paradigms, the x-axis was able to be reduced to 30 seconds long to improve clarity. The accuracy is the average accuracy at that time for all testing sessions and is displayed as a percentage. This figure with both subplots can be seen below in Figure 14.

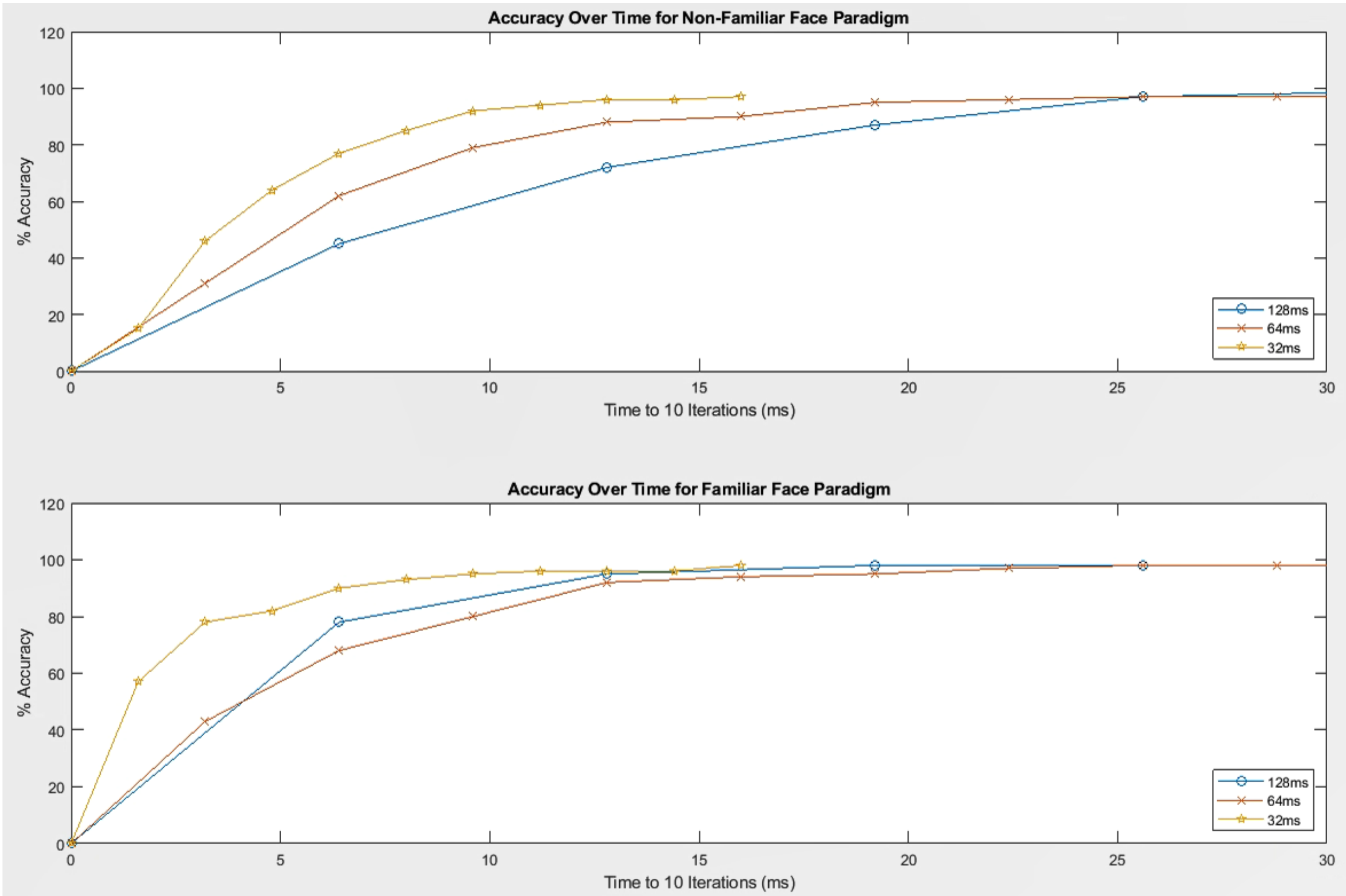


Figure 14. This figure contains two subplots that show the spelling accuracy over time for the non-familiar face paradigm(top) and the familiar face paradigm(bottom). Both subplots contain three lines; one line for each flash rate configuration. The x-axes end at 30 seconds due to the 64ms and 128ms flash rate configurations reaching 100% accuracy in less time.

For the non-familiar face paradigm, the first flash duration to reach 95% accuracy (accuracy at which only 1 letter is misspelled) is the 32ms flash duration. This is followed by the 64ms flash duration and then the 128ms flash duration. This same order is repeated when determining which flash duration achieves higher accuracy more quickly. The 32ms flash duration reaches 95% accuracy at roughly 12.8ms, or 8 iterations. This is much faster than the 64ms flash duration configuration which reaches this accuracy at 19.2 seconds, or 6 iterations. It is even faster than the 128ms flash duration configuration which produces a high accuracy of 95% at 25.6 seconds, or 4 iterations. Although the iterations required to reach this accuracy are reduced as the flash duration increases, the time to reach the high accuracy is reduced. This infers that the ideal flash duration to be used is the 32ms flash duration, then the 64ms flash duration, and lastly, the 128ms flash duration.

For the familiar face paradigm, the same trend can be seen. However, for this paradigm, the time required for the 64ms flash duration configuration to reach 95% accuracy is longer than that of the 128ms flash duration configuration. The 64ms flash duration reaches this milestone at 19.2 seconds, or 6 iterations. The 128ms flash duration reaches it at 12.8 seconds, or 2 iterations. Although the 64ms and 128ms flash durations had different results, it can be noted that their performances were still very similar. The 32ms flash duration performed the best, reaching 95% accuracy at 9.6 seconds, or 6 iterations. This again infers that the best flash duration to use for testing is the 32ms flash duration.

When we compare the two paradigms, for the 128ms flash duration, the time and iterations required to reach 95% accuracy for the familiar face paradigm are half that of the non-familiar face paradigm. This is most likely due to the stark differences in average P300 amplitude. The performance of the configurations using a 64ms flash duration was the same. Both paradigms took 19.2 seconds and 6 iterations to reach an accuracy of 95%. For the configurations using a 32ms flash duration, the familiar face paradigm again outperformed the non-familiar face paradigm. The time and iterations needed to spell 95% correctly was reduced by 3.2 seconds and 2 iterations when using the familiar face paradigm.

VIII. Discussion

When analyzing results, various trends were noticed that were expected of EEG recordings that include ERPs. When looking at the average P300 responses in all the channels in Figure 8 and Figure 9, the channels that measure the frontal lobe have a strong positive amplitude around 100-250ms. While the amplitudes of the peaks of these channels can be indicative of the effects that each paradigm and flash rate have on the elicited ERP, these channels do not represent a strong P300 response. Instead of having the typical form that the P300 response holds with a positive inflection, preceded by a negative inflection (N170), and finally the peak at around 300-500ms (P300), the frontal lobe has a simpler form. Most of the channels along the frontal lobe show large spikes in amplitude without the proper P300 form. However, these sudden jumps in amplitude in the frontal lobe channels appear much sooner than expected. In fact, most of them appear well before 300ms with some appearing nearly immediately after the stimulus is presented. They cannot be attributed to P300 because they appear too early.

The frontal lobe is not the only part of the brain to not elicit a P300 response to the presented stimulus. The temporal lobe (P7 and P8) and the occipital lobe (O1 and O2) both did not elicit any P300 response at all. These four channels did not have any positive spikes in amplitude. Instead, these four channels showed a very strong N170 response. This N170 response was consistent with what was expected from these channels. Because these channels produced only an N170 response, they were not used for calculations for the average peak P300 response. This is also true of the frontal lobe channels. The only channels that were used in the calculations for the results in Figures 10 and 11 were the channels in the central lobe (Cz, C3, C4) and the parietal lobe (Pz P3 P4). These six channels consistently produced the P300 response within the appropriate timeframe and held the proper P300 form.

The results that were generated did not support the first hypothesis. To reiterate, the first hypothesis that was given for the experiment was that the 128ms flash duration would produce the highest P300 response, but it would be the 64ms flash duration that would produce the fastest results without a loss in accuracy. It was believed that the 32ms flash duration would not be able to spell because a 32ms flash duration with an interstimulus interval ratio of 1:4 would mean that a letter would flash every 160ms, leaving no time for the P300 response and causing overlapping with non-target stimuli. It was also believed that the familiar face paradigm would be superior to the standard paradigm in terms of speed and accuracy. While the hypothesis was correct by assuming that the 128ms flash duration would produce the high amplitude P300 responses and by stating that the familiar face paradigm would provide improved performance, it was wrong about the efficiency of the flash durations. While the 128ms flash duration has an ISI of 512ms for a total of 640ms per flash, this still proved to be faster than the 64ms flash duration which had 320ms between flashes. This is because, in both paradigms, the 128ms flash duration produced a much larger P300 response that could be more easily classified by the SWLDA algorithm. As seen in Figure 12, the 128ms flash duration for the non-familiar face paradigm produced an accuracy of 95% or higher after 25.6 seconds. This is compared to the 64ms flash duration which achieved that accuracy in 19.2 seconds. However, when we look at the familiar face paradigm, the 128ms flash duration was able to achieve 95% accuracy after only 2 iterations. This equates to only 12.8ms. The 64ms flash duration saw no change in speed between the non-familiar face paradigm and the familiar face paradigm. For both, there was no change in speed or number of iterations needed to achieve high accuracy. Because the familiar face paradigm achieved consistently higher average P300 amplitudes than the non-familiar face paradigm, the results generated by this paradigm are accepted as more viable than the non-familiar face paradigm. This means that for the purposes of this study, it can be concluded that the 128ms flash duration yields better performance than the 64ms flash duration.

Although the 128ms flash duration has better performance than the 64ms flash duration, it is still not the best choice when using either paradigm. This may seem misleading because of the number of iterations needed to reach 95% accuracy is much lower with the 128ms flash duration. Looking at Figure 13, We can see that the number of iterations needed to reach 95% accuracy for 128ms flash duration is 4 using the non-familiar face paradigm and 2 using the familiar face paradigm. This is considerably lower than the 32ms flash duration which requires 8 for the non-familiar face paradigm and 6 for the familiar face paradigm. But the number of iterations needed to achieve this accuracy is irrelevant if the time it takes to spell accurately is increased. Even though the number of iterations needed to spell accurately is much higher with the 32ms flash duration, the time to spell accurately is lower. For the non-familiar face paradigm, the 32ms flash duration spelled with 95% accuracy after only 12.8ms compared to the 128ms flash duration which required 25.6ms. This is half the time needed. Using the familiar face paradigm, the 128ms flash duration performed much better but was still beaten by the 32ms flash duration. The 32ms flash duration boasted a speed of only 9.6ms. This is 25% faster than the 128ms flash duration. This is the fastest configuration of all six that were tested. Therefore, it can be concluded that the 32ms flash duration with a 1:4 ISI ratio using a familiar face paradigm is the best choice for the P300 automatic spelling system.

But what caused these results? First, when we look at Figures 11 and 12, the average P300 amplitudes were consistently higher using the familiar face paradigm. The N170 responses produced in the occipital and temporal lobes also had a more negative amplitude than with non-familiar face paradigm. These differences better allow the SWLDA algorithm to classify them. With the P300 response as the feature used for classification, the larger the amplitude is generated, the easier the algorithm can determine which portions of the EEG data represent that feature. Because the results generated by the familiar face paradigm showed a consistently higher P300 amplitude, the SWLDA was able to classify it using fewer iterations. This decrease in iterations needed improves the performance of the system by increasing the speed. With fewer iterations, the system can spell accurately with less time.

Although the familiar face paradigm had a stronger performance, it still had some issues. The biggest issue was the adjacency error. Adjacency error is when a letter is misspelled and the letter that is chosen by the system is adjacent to the desired letter. Most of all misspelled letters with the familiar face paradigm were adjacency errors. This is believed to be due to the familiar face paradigm producing much higher non-target P300 amplitudes than the non-familiar face paradigm. When a row or column flashes, there is still a response that is generated in the brain. This can be a minor fluctuation in voltage or its own P300 response. Either way, the average non-target ERP showed a much higher average amplitude. It is speculated that this can be caused by two things. First, when a row or column flashes with the non-familiar face paradigm, only the pixels that occupy the space of the letters turn from white to black. But with the familiar face paradigm, much more of the space on the screen changes. The facial overlay is larger than all the letters so significantly more of the screen flashes. This causes much more peripherally sighted events. This means that even though you may stare at one letter, because the flash caused by the familiar face overlay is so large, you can still see it and generate a P300 response. Second, it is difficult to focus solely on the desired letter and not have the eye wander at all. But, when you look at all the different letters, you can tell the difference between them. With the familiar face paradigm, when a row or column flashes, all the letters look the same. It is easier to produce a P300 response off a peripherally sighted event because the stimulus that was flashed is the image that you were waiting to see. The only difference is that that image is not in the correct location. This is believed to cause the adjacency error because even though you are not looking in that one spot on the screen, when the image you are waiting for flashes right beside the target letter, you still generate a P300 response which is classified as the desired letter. It can be noted too that the adjacency error was not common in the non-familiar face paradigm. The lack of these two reasonings may be the cause for the infrequency of the adjacency error in the non-familiar face paradigm.

Throughout the experiment, there were some issues that had to be worked around to generate an acceptable quality of EEG signals. The first, and most important, of these issues was the need for an electrostatically quiet area to conduct testing. The EEG is very sensitive to electrostatic noise that is present in the surrounding area. When there is electrostatic noise, the EEG recordings are polluted with a high-frequency noise that can muffle the true EEG measurements. Because of this noise, one of the measurement days produced outlier results that were not consistent with the other testing sessions. For that testing session, the system had difficulty spelling and yielded low accuracies. The amplitude of the EEG measurements when subjected to this noise was much higher than the measurements taken when there was little to no noise. This noise can cause many issues when filtering the data using ASR filtering and CAR filtering. For example, in a low-noise environment, the heartbeat can be clearly seen in the EEG recording and is filtered out due to it being consistent across all channels. However, since the heartbeat was not able to be spotted, it could further lower the accuracy because the filtering methods could not detect it. During the experiment, two testing locations were used. One had high electrostatic noise, and the other had little to no noise at all. The location with the high electrostatic noise was only used one time and produced results that contrasted the rest of the testing session results. The location with low noise was so electrostatically quiet that the P300 responses that were produced were exceedingly clear. This means that when the data was epoched into 1-second segments starting at the presentation of the stimulus, the P300 response could be clearly identified. This was not true for the one testing session that was taken in the noisy location. It should be noted that it would seem that the system could not spell using the 64ms or 32ms flash rate for either paradigm and that the 128ms flash rate spelled with undesired accuracy (~65%) for both paradigms in a noisy environment. This may be because the noise was strong enough to overshadow the P300 responses of the 64ms and 32ms flash durations but not strong enough to completely overshadow the P300 responses of the 128ms flash duration. Therefore, before using the P300 automatic spelling system, it is imperative that the location where testing is conducted is electrostatically quiet. The second issue is that the accuracy of the system is dependent on the alertness of the user. Testing sessions which the user could not focus well or was tired did not yield accuracies that were as high as other testing sessions. This is a limitation of the system as it is best to not use it on someone who is sleep deprived or has attention issues. It can even be said that it is best to not use the system immediately after waking up or before bedtime as the cognitive state of the individual may not be as strong at those times as it would be during the middle of the day. This lack of focus caused the results of the non-familiar face paradigm to be not as strong. There was an effect on the familiar face paradigm, but it was not as apparent as in the non-familiar face paradigm. This is because with the non-familiar face paradigm, the letters are bare and have no focal point. For example, the letter "O" does not have a focal point to stare at. Do you look at the top of the “O” or at the bottom? Often it was found that the eyes would wander the letters and sometimes would feel caught off-guard when the desired letter flashed. This is compared to the familiar face paradigm which uses faces. When you look at a face, it is natural to stare directly into the eyes. There is a focal point on the face for you to stare at that grabs and maintains the attention of the user. This made it much easier to be prepared for the letter to flash and made focusing much easier.

One thing that was noticed throughout the experiment was that there was no correlation between the number of testing sessions a user had conducted and the accuracy of the results. One question that was asked was if we conduct testing sessions weekly over the course of six months, will the accuracy of the testing session results improve over time? This does not seem to be the case. The results seem to be random depending on many factors such as focus, alertness, and other variables. But it does not appear that repetition makes the user better at spelling with the system. This was contrary to what was hypothesized.

This experiment raised other questions that could be tested in the future. First, this same experiment could be conducted to test the results using different machine learning techniques. Neural networks, support vector machines, or other machine learning methods could be used to see if the spelling speed could be improved without sacrificing accuracy. Another thing that could be tested is to add a focal point to each letter such as a dot at the top so that the eyes don’t wander. This could possibly improve focus and cause a stronger P300 response. A third question that was raised is the possibility of using both the P300 response and the N170 response as features of classification. As seen in the results, we can consistently produce an N170 response in the temporal and occipital lobes which have varying amplitudes that are dependent on the paradigm and flash duration. It is worth noting that the N170 response behaved similarly to the

P300 responses with the amplitudes of becoming more negative with higher flash durations and by using a familiar face paradigm. By using both simultaneously for classification, it may be possible to build a system that can better determine the difference between target and non-target stimuli. This may allow the system to determine the desired letters using a smaller number of iterations which would lower the time needed to spell. Lastly, to target the adjacency error that was seen in the familiar face paradigm results, it could be suggested to use one familiar face per letter. As explained previously, when the rows and letters adjacent to the target letter flash, a P300 response may still occur because the desired image still appeared even though it is not in the correct location. It may be worth testing this same experiment with a different face for each letter to combat this issue.

IX. Conclusion

In conclusion, a P300 spelling system is a brain-computer interface that can allow the P300 component to be converted to text digitally by using signal processing and machine learning techniques. This system has many parameters that can be modified to improve spelling or accuracy. In this experiment, two paradigms, the familiar face and non-familiar face paradigms were compared to determine which generated larger average P300 amplitudes and to see which could spell accurately more quickly. Each paradigm was tested with three different flash durations (128ms, 64ms, and 32ms). Of these flash durations, it was found that the 128ms flash duration consistently produced the highest amplitude P300 responses and that the 32ms flash duration produced the smallest P300 amplitudes. This means that the number of iterations needed to spell was lowest overall with the 128ms flash duration and highest with the 32ms flash duration. However, the 32ms flash duration was able to reach high accuracy faster than the 128ms and 64ms flash durations making it the ideal choice. The familiar face paradigm also decreased the time needed to spell accurately for both the 128ms and 32ms flash durations, meaning that the familiar face paradigm outperformed the non-familiar face paradigm. Therefore, the ideal configuration is to use a familiar face paradigm with a 32ms flash duration with an ISI ratio of 1:4.

This study has two main issues. First, the environment that testing is conducted in must be electrostatically quiet. This limitation is not ideal for those who would use the P300 automatic spelling system daily as you may not know if the location you are in is electrostatically quiet. The noise that is measured by the EEG can cause low accuracies which do not allow the system to spell the desired characters. Second, the alertness and focus of the user influences the accuracy of the spelling. The system is best used in the middle of the day after a good night's rest so that the user is awake and can focus.

This experiment also created other questions that should be further investigated. The first question is how does the performance of this system compare to using other machine learning techniques such as neural networks or support vector machines? Focal points could also be added to each letter of the non-familiar face paradigm to better hold the focus of the user. To decrease adjacency error, it could be suggested to use one face per letter instead of the same face for all letters. And finally, because the results of this study show consistent trends among the N170 responses in the temporal and parietal lobe, creating a system that uses both the P300 and N170 response for classification may improve the speed and accuracy.

X. References

[1] Pistoia, F., & Laureys, S. (2018, March 12). *Locked In Syndrome*. NORD (National Organization for Rare Disorders). https://rarediseases.org/rare-diseases/locked-in-syndrome/#:~:text=Additional%20conditions%20that%20can%20cause,amyotrophic%20lateral%20sclerosis%20(ALS).

[2] Shih, J. J., Krusienski, D. J., & Wolpaw, J. R. (2012). Brain-Computer Interfaces in Medicine. *Mayo Clinic Proceedings*, *87*(3), 268–279. https://doi.org/10.1016/j.mayocp.2011.12.008

[3] Purves D, Augustine GJ, Fitzpatrick D, et al., editors. Neuroscience. 2nd edition. Sunderland (MA): Sinauer Associates; 2001. Excitatory and Inhibitory Postsynaptic Potentials. Available from: https://www.ncbi.nlm.nih.gov/books/NBK11117/

[4] Britannica, T. Editors of Encyclopaedia (2008, May 22). Postsynaptic potential. Encyclopedia Britannica. https://www.britannica.com/science/postsynaptic-potential

[5] Morley, A., Hill, L., & Kaditis, A. G. (2016). 10-20 system EEG Placement. *European Respiratory Society, European Respiratory Society*.

[6] Donchin, E., Spencer, K., & Wijesinghe, R. (2000). The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 174-179. doi:10.1109/86.847808

[7] Fazel-Rezai, R. (2007). Human Error in P300 Speller Paradigm for Brain-Computer Interface. *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. doi:10.1109/iembs.2007.4352840

[8] Guan, C., Thulasidas, M., & Wu, J. (2005). High performance p300 speller for brain-computer interface. *IEEE International Workshop on Biomedical Circuits and Systems, 2004*. doi:10.1109/biocas.2004.1454155

[9] Decroix, J., Roger, C., & Kalénine, S. (2020). Neural dynamics of grip and GOAL integration during the processing of others' actions with objects: An ERP study. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-61963-7

[10] Eimer, M. (2011). The face-sensitivity of the n170 component. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00119

[11] Wang, Y., Huang, H., Yang, H., Xu, J., Mo, S., Lai, H., Wu, T., & Zhang, J. (2019). Influence of Eeg references on N170 component in human facial recognition. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00705

[12] Blau, V. C., Maurer, U., Tottenham, N., & McCandliss, B. D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behavioral and Brain Functions*, *3*(1), 7. https://doi.org/10.1186/1744-9081-3-7

[13] Başar, E., Başar-Eroglu, C., Rosen, B., & Schütt, A. (1984). A new approach to endogenous event-related potentials in man: Relation between eeg and p300-wave. *International Journal of Neuroscience*, *24*(1), 1–21. https://doi.org/10.3109/00207458409079530

[14] Papagiannopoulou, E. A., & Lagopoulos, J. (2016). P300 event-related potentials in children with dyslexia. *Annals of Dyslexia*, *67*(1), 99–108. https://doi.org/10.1007/s11881-016-0122-6

[15] Guger, C., Allison, B. Z., Großwindhager, B., Prückl, R., Hintermüller, C., Kapeller, C., Bruckner, M., Krausz, G., & Edlinger, G. (2012). How many people could use an SSVEP BCI? *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00169

[16] Beverina, F., Palmas, G., Silvoni, S., Piccione, F., & Giove, S. (2003). User adaptive BCIs: SSVEP and P300 based interfaces. PsychNology J., 1, 331-354.

[17] Li, J., & Zhang, L. (2012). Active training paradigm for motor imagery bci. *Experimental Brain Research*, *219*(2), 245–254. https://doi.org/10.1007/s00221-012-3084-x

[18] Allison, B. Z., & Neuper, C. (2010). Could anyone use a Bci? Brain-Computer Interfaces, 35–54. https://doi.org/10.1007/978-1-84996-272-8_3

[19] Schalk, G., Mcfarland, D., Hinterberger, T., Birbaumer, N., & Wolpaw, J. (2004). BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Transactions on Biomedical Engineering, 51*(6), 1034-1043. doi:10.1109/tbme.2004.827072

[20] Brunner, P., Joshi, S., Briskin, S., Wolpaw, J. R., Bischof, H., & Schalk, G. (2010). Does the 'P300' speller depend on eye gaze? *Journal of Neural Engineering, 7*(5), 056013. doi:10.1088/1741-2560/7/5/056013

[21] Lu, J., Speier, W., Hu, X., & Pouratian, N. (2013). The effects of stimulus timing features on P300 speller performance. *Clinical Neurophysiology*, *124*(2), 306-314. doi:10.1016/j.clinph.2012.08.002

[22] Roula, M. A., Kulon, J., & Mamatjan, Y. (2012). Brain-computer interface speller using hybrid P300 and motor imagery signals. *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. doi:10.1109/biorob.2012.6290944

[23] Farwell, L., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology, 70*(6), 510-523. doi:10.1016/0013-4694(88)90149-6

[24] Miyakoshi, M., Kanayama, N., Nomura, M., Iidaka, T., & Ohira, H. (2008). ERP study of viewpoint-independence in familiar-face recognition. *International Journal of Psychophysiology*, *69*(2), 119–126. https://doi.org/10.1016/j.ijpsycho.2008.03.009

[25] Lu, Z., Li, Q., Gao, N., & Yang, J. (2020). The Self-Face Paradigm Improves the Performance of the P300-Speller System. *Frontiers in Computational Neuroscience*, *13*. https://doi.org/10.3389/fncom.2019.00093

[26] Li, Q., Liu, S., & Li, J. (2015). Neural mechanism OF P300-Speller brain-computer interface using familiar Face Paradigm. 2015 International Conference on Network and Information Systems for Computers. https://doi.org/10.1109/icnisc.2015.99

[27] Speier, W., Deshpande, A., Cui, L., Chandravadia, N., Roberts, D., & Pouratian, N. (2017). A comparison of stimulus types in ONLINE classification of the P300 speller using language models. *PLOS ONE*, *12*(4). https://doi.org/10.1371/journal.pone.0175382

[28] Lu, Z., Gao, N., Zhou, W., Yang, J., Wu, J., & Li, Q. (2019). A comparison of facial p300-speller paradigm based on famous face and the familiar face. *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. https://doi.org/10.1109/cisp-bmei48845.2019.8965892

[29] Xinyu, L., Hong, W., Shan, L., Yan, C., & Li, S. (2017). Adaptive common average reference for in vivo multichannel local field potentials. *Biomedical Engineering Letters*, *7*(1), 7–15. https://doi.org/10.1007/s13534-016-0004-1

[30] Lepage, K. Q., Kramer, M. A., & Chu, C. J. (2014). A statistically robust EEG re-referencing procedure to mitigate reference effect. *Journal of Neuroscience Methods*, *235*, 101–116. https://doi.org/10.1016/j.jneumeth.2014.05.008

[31] Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., & Jung, T. P. (2020). Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. IEEE Transactions on Biomedical Engineering, 67(4), 1114-1121. https://doi.org/10.1109/TBME.2019.2930186

[32] Mullen, T. R., Kothe, C. A., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., … & Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. IEEE Transactions on Biomedical Engineering, 62(11), 2553–2567.

[33] Venables, W. N., & Ripley, B. D. (2010). *Modern applied statistics with S*. Springer.

[34] Hamsici, O. C., & Martinez, A. M. (n.d.). *Bayes Optimality in Linear Discriminant Analysis*. http://www2.ece.ohio-state.edu/~aleix/pami07.pdf.

[35] Xiaozhou, Y. A. N. G. (2020, October 7). *Linear Discriminant Analysis, Explained*. Medium. https://towardsdatascience.com/linear-discriminant-analysis-explained-f88be6c1e00b.

[36] Hastie, T., Friedman, J., & Tisbshirani, R. (2017). *The Elements of statistical learning: data mining, inference, and prediction*. Springer.

[37] Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, *30*(2), 169–190. https://doi.org/10.3233/aic-170729

[38] The MathWorks Inc. (2006). *stepwisefit*. Fit linear regression model using stepwise regression - MATLAB. https://www.mathworks.com/help/stats/stepwisefit.html#d123e787312.

[39] Manyakov, N. V., Chumerin, N., Combaz, A., & Van Hulle, M. M. (2011). Comparison of Classification Methods for P300 Brain-Computer Interface on Disabled Subjects. *Computational Intelligence and Neuroscience*, *2011*, 1–12. https://doi.org/10.1155/2011/519868

XI. Appendix

APPENDIX.A Training Procedure

1. Load up the BCI2000 software and choose “RDA client” for signal source, “P3signal processing” for signal processing, and “P3 speller” for application. Click “launch”, then click “configure”.
2. Begin by loading in the 5 by 5 matrix parameter file into the BCI2000 software. To do so, click “Load Parameters” on the right side of the *Parameter Configuration* menu and choosing the previously created parameter file.
3. Set the “number of sequences” in the *Application* tab to 10 as none of the configurations should require more than 10 iterations for the classifier to achieve 100% accuracy. In the *Filtering* tab, change “epochs to average” to 10, matching the “number of sequences”. Ensure that the “epoch length” is set to 800ms. Ensure that the “Spatial Filter Type” in the *Filtering* tab is set to common average referencing (CAR).
4. Ensure that the “Interpret Mode” in the *Application* tab is set to “copy mode” and that the box beside “Display Results” is checked. Change the “Stimulus Duration” to 32ms and the “Maximum Interstimulus Interval” and “Minimum Interstimulus Interval” to 128ms.
5. Change the subject identifiers in the storage section to match the session and run number.
6. Once the parameters have been set, click “set config” on the BCI2000 Operator. This will open the paradigm and give you the option to “Start” the training session.
7. Before beginning, ensure that the testing environment is electromagnetically dull. Turn off lights and ensure there are minimal vibrations. Remove electronics from the vicinity which could cause electromagnetic noise on any EEG recordings. Large monitors are preferred for testing, such as a television, or a large-screened computer monitor. White

noise will be played in the background to dampen any noise that may be unavoidable, such as a loud car on the road or a person/pet walking around.

8. Ensure that the subject is aware of the phrase that is to be spelled. For this experiment, the phrase will be "THEBIGDWARFONLYJUMPS". After each sequence of flashes, a pause, roughly six seconds, will signal that it is time to move on to the next character. Instruct the subject to focus directly on the intended character until they recognize this pause. Instruct the subject to remain as still and calm as possible for the duration of the training session. The training session can last for a considerable amount of time so offer the subject an opportunity to use the restroom or walk around before beginning the session.
9. Check the software associated with the *g-tec* EEG to ensure that all electrodes are in their correct location and are functioning properly with a reasonable amount of surface impedance.
10. On the BCI2000 Operator, click "Start" and click on the paradigm to move the BCI2000 windows behind the P300 speller paradigm (this allows the screen to show only the paradigm).
11. Wait until the subject has completed the testing. The P300 speller paradigm will show, in large letters, "TIME OUT", signaling the end of the training session.
12. Continue on to the classification portion of the experiment with the training file that has been generated.

APPENDIX.B Testing Procedure

1. Load up the BCI2000 software and choose “RDA client” for signal source, “P3signal processing” for signal processing, and “P3 speller” for application. Click “launch”, then click “configure”.
2. Begin by loading in the 5 by 5 matrix parameter file into the BCI2000 software. To do so, click “Load Parameters” on the right side of the *Parameter Configuration* menu and choosing the previously created parameter file. This time, select “Load Parameters” again and load in the xxxx.prm file that was generated by the P300 classifier.
3. Set the “number of sequences” in the *Application* tab to 10 as none of the configurations should require more than 10 iterations for the software to determine the target character. In the *Filtering* tab, change “epochs to average” to 10, matching the “number of sequences”. Ensure that the “epoch length” is set to 800ms. Ensure that the “Spatial Filter Type” in the *Filtering* tab is set to common average referencing (CAR).
4. Ensure that the “Interpret Mode” in the *Application* tab is set to “online free mode” and that the box beside “Display Results” is checked. This will allow us to compare the accuracy of each testing configuration. Change the “Stimulus Duration” to 32ms and the “Maximum Interstimulus Interval” and “Minimum Interstimulus Interval” to 128ms.
5. Change the subject identifiers in the storage section to match the session and run number.
6. Once the parameters have been set, click “set config” on the BCI2000 Operator. This will open the paradigm and give you the option to “Start” the training session.
7. Before beginning, ensure that the testing environment is electromagnetically dull. Turn off lights and ensure there are no vibrations. Remove electronics from the vicinity which could cause electromagnetic noise on any EEG recordings. Large monitors are preferred

for testing, such as a television, or a large-screened computer monitor. White noise will be played in the background to dampen any noise that may be unavoidable, such as a loud car on the road or a pet walking around.

8. Ensure that the subject is aware of the sequence of characters that are to be spelled. For this experiment, the phrase to be spelled is "THEBIGDWARFONLYJUMPS". After each sequence of flashes, a pause, roughly six seconds, will signal that it is time to move on to the next character. Instruct the subject to focus directly on the intended character until they recognize this pause. Instruct the subject to remain as still and calm as possible for the duration of each testing session. These testing sessions last for a considerable amount of time so offer the subject an opportunity to use the restroom or walk around before beginning.
9. Check the software associated with the *g-tec* EEG to ensure that all electrodes are in their correct location and are functioning properly with a reasonable amount of surface impedance.
10. On the BCI2000 Operator, click "Start" and click on the paradigm to move the BCI2000 windows behind the P300 speller paradigm (this allows the screen to show only the paradigm).
11. Wait until the subject has completed the testing. The P300 speller paradigm will show, in large letters, "TIME OUT", signaling the end of the training session.
12. Record the characters the software believes that the subject has spelled as this will be needed later for accuracy analysis.
13. Repeat the experimental procedure starting back at step 16. Change the "Stimulus Duration" and "Minimum Interstimulus Interval" and "Maximum Interstimulus Interval"

to the next testing configuring. Do this until all three of the testing configurations have been completed.

14. Once steps 1 through 25 have been completed, repeat them using the familiar face paradigm. This will require uploading the familiar face paradigm xxxx.prm file during steps 2 and 14.
15. Continue on to the data analysis portion of the experiment with the testing files that have been generated.

APPENDIX.C IRB Approval Letter

EAST CAROLINA UNIVERSITY
University & Medical Center Institutional Review Board
4N-64 Brody Medical Sciences Building· Mail Stop 682
600 Moye Boulevard · Greenville, NC 27834
Office **252-744-2914** · Fax **252-744-2284** ·
rede.ecu.edu/umcirb/

Notification of Initial Approval: Expedited

From:	Biomedical IRB
To:	Sunghan Kim
CC:	
Date:	11/15/2021
Re:	UMCIRB 21-002036 Using EEG to Improve Brain-Computer Interfaces

I am pleased to inform you that your Expedited Application was approved. Approval of the study and any consent form(s) occurred on 11/15/2021. The research study is eligible for review under expedited category # 4,6,7. The Chairperson (or designee) deemed this study no more than minimal risk.

As the Principal Investigator you are explicitly responsible for the conduct of all aspects of this study and must adhere to all reporting requirements for the study. Your responsibilities include but are not limited to:

1. Ensuring changes to the approved research (including the UMCIRB approved consent document) are initiated only after UMCIRB review and approval except when necessary to eliminate an apparent immediate hazard to the participant. All changes (e.g. a change in procedure, number of participants, personnel, study locations, new recruitment materials, study instruments, etc.) must be prospectively reviewed and approved by the UMCIRB before they are implemented;

2. Where informed consent has not been waived by the UMCIRB, ensuring that only valid versions of the UMCIRB approved, date-stamped informed consent document(s) are used for obtaining informed consent (consent documents with the IRB approval date stamp are found under the Documents tab in the ePIRATE study workspace);

3. Promptly reporting to the UMCIRB all unanticipated problems involving risks to participants and others;

4. Submission of a final report application to the UMICRB prior to the expected end date provided in the IRB application in order to document human research activity has ended and to provide a timepoint in which to base document retention; and

5. Submission of an amendment to extend the expected end date if the study is not expected to be completed by that date. The amendment should be submitted 30 days prior to the UMCIRB approved expected end date or as soon as the Investigator is aware that the study will not be completed by that date.

The approval includes the following items:

Name	Description
Informed Consent Form	Consent Forms
MoCA	Standardized/Non-Standardized Instruments/Measures
Questionnaire	Surveys and Questionnaires
Thesis Proposal	Study Protocol or Grant Application

For research studies where a waiver or alteration of HIPAA Authorization has been approved, the IRB states that each of the waiver criteria in 45 CFR 164.512(i)(1)(i)(A) and (2)(i) through (v) have been met. Additionally, the elements of PHI to be collected as described in items 1 and 2 of the Application for Waiver of Authorization have been determined to be the minimal necessary for the specified research.

The Chairperson (or designee) does not have a potential for conflict of interest on this study.

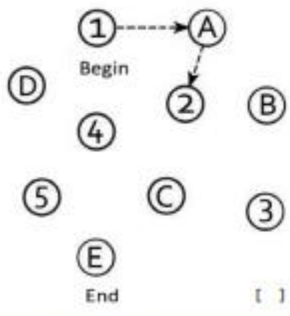
IRB00000705 East Carolina U IRB #1 (Biomedical) IORG0000418
IRB00003781 East Carolina U IRB #2 (Behavioral/SS) IORG0000418

APPENDIX.D MOCA Test Example

MoCA COGNITIVE ASSESSMENT Version 8.3 English

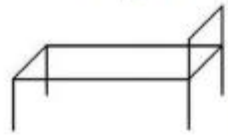
Name:
Education:
Sex:
Date of birth :
DATE :

VISUOSPATIAL / EXECUTIVE | **POINTS**



[ ]

Copy bed



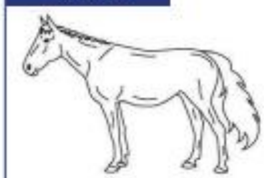
[ ]

Draw CLOCK (Five past ten)
(3 points)

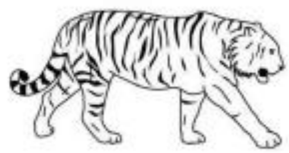
[ ] Contour [ ] Numbers [ ] Hands

__/5

NAMING



[ ]



[ ]

[ ]

__/3

MEMORY Read list of words, subject must repeat them. Do 2 trials, even if 1st trial is successful. Do a recall after 5 minutes.		LEG	COTTON	SCHOOL	TOMATO	WHITE	NO POINTS
	1st TRIAL						
	2nd TRIAL						

ATTENTION Read list of digits (1 digit/sec.). Subject has to repeat them in the forward order. [ ] 2 4 8 1 5
Subject has to repeat them in the backward order. [ ] 4 2 7 __/2

Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors.
[ ] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B __/1

Serial 7 subtraction starting at 60. [ ] 53 [ ] 46 [ ] 39 [ ] 32 [ ] 25
4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0 pt** __/3

LANGUAGE Repeat: The child walked his dog in the park after midnight. [ ]
The artist finished his painting at the right moment for the exhibition. [ ] __/2

Language Fluency. Name maximum number of words in one minute that begin with the letter B. [ ] ____ (N ≥ 11 words) __/1

ABSTRACTION Similarity between e.g. orange - banana = fruit [ ] hammer - screwdriver [ ] matches - lamp __/2

DELAYED RECALL	(MIS)	Has to recall words WITH NO CUE	LEG [ ]	COTTON [ ]	SCHOOL [ ]	TOMATO [ ]	WHITE [ ]	Points for UNCUED recall only	__/5
Memory Index Score (MIS)	X3								
	X2	Category cue						MIS = ___/15	
	X1	Multiple choice cue							

ORIENTATION [ ] Date [ ] Month [ ] Year [ ] Day [ ] Place [ ] City __/6

 www.mocatest.org
Training and Certification are required to ensure accuracy.

MIS: /15
(Normal ≥ 26/30)
Add 1 point if ≤ 12 yr education

TOTAL __/30

ADMINISTERED BY ALBANESE, THOMAS NOT MOCA CERTIFIED

APPENDIX.E Questionnaire

Questionnaire

Title of Research Study: Improving the Speed and Accuracy of the P300 Automatic Spelling System Through Facial Recognition

Principal Investigator: Sunghan Kim (Person in Charge of this Study)
Institution, Department or Division: Department of Engineering, East Carolina University
Address: East 5th Street, Greenville, NC 27858
Telephone #: (252) 737-1750
Study Coordinator: Thomas Albanese
Telephone #: (252) 737-1750

This is a questionnaire designed to gather information about you that can allow us to conclude whether you are fit for the study, whether the study is dangerous for you to participate in, or whether the data we gather from you will be of sufficient quality. For each question below, please circle the answer that fits your description (YES or NO). If you do not believe that a question can be answered with a simple yes or no, write in your response beside the field labeled "other:".

Medical History

Have you ever been diagnosed with attention deficit disorder (ADD)?

YES NO Other:

Have you ever been diagnosed with attention deficit hyperactive disorder (ADHD)?

YES NO Other: ____

Do you have any form of cognitive impairment?

YES NO Other:

Have you had five or more concussions or have had a traumatic brain injury (TBI)?

YES NO Other:

Do you suffer from epilepsy or any ailment that can cause harm when presented with rapidly flashing or bright stimulus?

YES NO Other:

Are you currently taking any neuroleptic medications?

YES NO Other:

Is there anything else that the principal investigator and study coordinator should be aware of (if so, please write it in the "other" section)?

YES NO Other:____

Recent 24 Hours

Did you sleep at least six hours the previous night?

YES NO Other:____

Do you feel awake and alert?

YES NO Other:

Have you eaten in the past 8 hours?

YES NO Other:

Have you consumed any caffeinated beverages?

YES NO Other:____

Are you under the effects of any stimulants or medications intended to enhance focus (example: amphetamines for those diagnosed with attention deficit hyperactive disorder)?

YES NO Other:____

Are you under the effects of any substances or medications that may dull the ability to focus or concentrate (example: certain pain medications or barbiturates)?

YES NO Other:

____

Participant's Name (PRINT) **Signature** **Date**