# Event-related potentials as a form of neurofeedback using low-cost hardware

by

James Derek Jacoby
BA, Rice University, 1995
MSc, University of Victoria, 2011

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy

in the Department of Computer Science

# Supervisory Committee

Event-related potentials as a form of neurofeedback using low-cost hardware

by

James Derek Jacoby
BA, Rice University, 1995
MSc, University of Victoria, 2011

## Supervisory Committee

Dr. Melanie Tory, Department of Computer Science
**Co-Supervisor**

Dr. James Tanaka, Department of Psychology
**Co-Supervisor**

Dr. Colin Swindells, Department of Computer Science
**Departmental Member**

# Abstract

**Supervisory Committee**
Dr. Melanie Tory, Department of Computer Science
**Co-Supervisor**
Dr. James Tanaka, Department of Psychology
**Co-Supervisor**
Dr. Colin Swindells, Department of Computer Science
**Departmental Member**

The studies reported in this dissertation demonstrate that low-cost hardware is capable of detecting neural responses to stimuli in the user's focus of attention, and that these responses increase in magnitude with training. Neurofeedback is a sub-category of biofeedback that is concerned with using brain signals as the source of training data in a feedback loop. The neurofeedback training procedures in this dissertation focused on the P300 component, a time-locked event-related potential (ERP) that reflects the cognitive processes of attention and context updating. The current work provides preliminary evidence that neurofeedback based on rewarding a P300-like ERP is effective in increasing the magnitude of this response.

Three main questions were examined: 1. Is the Emotiv Epoc, as an example of a low-cost consumer EEG, capable of reliably detecting the P300 component? 2. Is there a training effect whereby the P300 response gets stronger with practice? 3. To what extent is the P300 response affected by cognitive factors such as memory load and self-generation of prompts?

The studies employed an open source software framework—open source tools provide a transparent, crowd-supported means of conducting research, but are often difficult to initially use and the current dissertation provides a guide within this content domain. The Emotiv Epoc headset was capable of detecting P300-like ERP in a P3 speller task. The P3 speller is a well-studied paradigm in which users spell letters using only their thoughts as input, while the system determines the letter to be spelled by analyzing the strength and timing of the ERP. Although the analyzed ERP behaved functionally like a P300, and the timing was consistent, the spatial localization of the signal was more frontally dominant than a standard P300. In the training study, 12 participants completed five P3 spelling sessions. Although an ERP training effect was observed, participant motivation and fatigue modulated this effect. In an attempt to improve motivation and increase interest in the task, a novel card game task was introduced. In this task—a variant of the card game "Concentration," where players turn cards face-up one at a time to match pairs—the participants used an attentional mechanism to select cards. This allowed for attentional training while offering a task whereby cognitive difficulty could be manipulated. In these studies, the P300-like ERP proved itself to be robust in regards to changes in cognitive difficulty, as well as internal versus external generation of prompts. This led to confidence in the separation of underlying cognitive and attentional processes and validated the focus of the P300 ERP on the attentional process.

The results indicated that ERP-specific neurofeedback is effective in increasing ERP magnitude. This dissertation does not involve any clinical populations as study participants, but the long-term potential of this research is to directly train a brain response relevant to clinical conditions. The paradigm can be implemented using low-cost hardware as opposed to research-grade instruments, which increases the likelihood of further research by the clinical community and lowers the barrier of entry for future exploration of the techniques.

# Table of Contents

## List of Tables

## List of Figures

# Acknowledgments

Thanks to my committee members for their encouragement and support, to my lab-mates for their review and cheerleading, to my friends and family for putting up with the long hours in the lab, and to NSERC for providing the funding that made this work possible.

# 1 Introduction

This dissertation focuses on the use of electroencephalography (EEG) for neurofeedback. EEG is the measurement of oscillations of brain electrical potential as sensed by electrodes on the scalp, which represents the averaged post-synaptic activity of many individual neurons (Nunez and Srinivasan, 2006). The studies reported here demonstrate that low-cost hardware is capable of detecting neural responses to stimuli in the user's focus of attention, and that these responses increase in magnitude with training. Neurofeedback is a sub-category of biofeedback that is concerned with using brain signals as the source of training data in a feedback loop. In the clinical manual, "Getting Started with Neurofeedback," it is described as an effective treatment for such disorders as post-traumatic stress disorder (PTSD), anxiety, depression, attention deficit hyperactivity disorder (ADHD) and a number of other conditions which are described in more detail in section 2.1 (Demos, 2005.)

This dissertation proposes a novel approach that is particularly relevant to the use of neurofeedback in attentional conditions. Within EEG neurofeedback, there are two broad classes of electrical activity that are examined: spontaneous EEG, and event-related potentials (ERP), which are time-locked to the presentation of a stimulus (Nunez and Srinivasan, 2006). The spontaneous EEG signals of most interest are generally measured from the frontal cortex, and are also known as cortical slow waves. These spontaneous EEG signals are the usual target in neurofeedback. However, this dissertation will instead examine the use of ERP as a neurofeedback protocol. ERP components are specifically altered in individuals with attentional disorders, leading to the possibility that ERP neurofeedback can more specifically target these alterations than would spontaneous EEG neurofeedback. The specific ERP component that will be the primary focus of this work is the P300, a positive deflection at the top and center of the head at approximately 300 milliseconds after a stimulus presentation. This is the first discovered, and largest, of the ERP components, and correlates to attention and novelty (Sutton et al, 1965; Picton, 1992; Polich, 2007). The current work provides preliminary evidence that neurofeedback based on rewarding correct P300 ERP (that is, P300 ERP which corresponds to the stimulus the user was instructed to focus upon) is effective in increasing the magnitude of this response, which may address a specific reduction in the P300 response seen in individuals with ADHD (Prox, 2007).

The paradigm used to study the P300 in these studies, the P3 speller (Farwell and Donchin, 1988), makes use of the P300 ERP to allow the user to spell words using only the EEG signal in response to stimuli on the computer screen. In using this paradigm as a neurofeedback technique, attention must be paid to cognitive factors to maintain the user’s focus and attention. The studies reported here provide evidence that the P300 is most closely associated with attentional processes rather than other cognitive processes. Through manipulations of cognitive difficulty and internal versus external generation of prompts, the P300 ERP responses in these studies remain robust.

The EEG system used in these studies, the Emotiv Epoc, is a low-cost wireless EEG headset intended for consumer use. It is generally used to capture EEG which is not tied to a specific stimulus, but the research in this dissertation shows that it is also capable of detecting ERP. The potential of using a low-cost device in a neurofeedback protocol focused on the strengthening of an ERP component raises the possibility of future clinical applications of the approach. The Epoc is an example of a consumer EEG device which is relatively slow as compared to many research grade systems (128 samples per second as compared to 256 or 512) and has easily fitted, but not as sensitive, saline-coupled electrodes.

## 1.1 Research questions and contributions

This research examines the application of low-cost hardware, and predominantly open source software, for training the P300 ERP neurofeedback response in cognitively normal young adults. The studies vary cognitive factors of the tasks and explore the internal versus external generation of prompts to answer the question of whether these elements can be modified in the creation of tasks more suitable for longer-term neurofeedback training.

**The first research question (RQ1) is whether the low-cost, consumer-grade Emotiv Epoc EEG headset is capable of reliably detecting P300 ERP.** This headset is representative of a class of recent low-cost hardware as detailed in section 2.6, and due to the design and number of electrodes, appears to be the candidate most likely to be able to provide access to the appropriate electrode locations and detect the changes in electrical potential characteristic of P300 ERP. This dissertation provides proof by demonstration of the efficacy of the Emotiv Epoc headset in obtaining reliable P300 ERP data.

The type of neurofeedback that these consumer EEG headsets were generally designed to use focuses on training via the cortical slow waves of the frontal lobe. These cortical slow waves, which include alpha waves and beta waves, are susceptible to neurofeedback training, but are periodic waves across the entirety of the frontal cortex. In contrast to cortical slow waves, which are repeating waveforms that are not prompted by a stimulus, ERPs happen in direct response to a specific stimulus. As a result, the precision of the recording device becomes more critical, since there are fewer readings over which to calculate the signal.

**The second research question (RQ2) is whether the P300 ERP component is trainable through neurofeedback to increase the magnitude of the P300 response**. This work shows experimental evidence of a P300 training effect in response magnitude resulting from neurofeedback using a P3 speller training protocol. Targeting specific brainwave patterns tied directly to a stimulus presentation promises a specificity of attentional training in these ERP neurofeedback paradigms that is not possible in the training of cortical slow waves. There are very few mentions of ERP training effects in the literature, and this intentional use of ERP as a neurofeedback mechanism is a novel contribution to this work.

**The third research question (RQ3) is whether the P300 ERP response can be maintained through task variations in cognitive complexity and internal versus external generation of prompts**. This dissertation provides preliminary data on the efficacy of a modified P3 speller task paradigm in the form of a card game that has different cognitive characteristics but which still produces a robust P300 response. Task variations that explore prompts generated by the user (“internal generation of prompts”) as compared to prompts provided by the experimenter (“external generation of prompts”) are also examined. The impact of these manipulations in creating paradigms capable of being used in longer-term neurofeedback studies are discussed. Demanding attentional tasks pose challenges in terms of fatigue and motivation, and having established that a robust P300 response can be maintained through task changes, it is important to design tasks that can remain interesting in longer-term studies.

The answer to these research questions are a contribution to the understanding of the potential for ERP to be used as a form of neurofeedback. The exploration of these questions on low-cost consumer devices and open source software provide a roadmap for subsequent researchers to pursue extensions of these questions with lower barriers to entry than have previously been present.

## 1.2 Scope

This dissertation is inherently cross-disciplinary in nature. There are theory contributions from the psychology discipline in terms of ERP neurofeedback and factors affecting the production of P300 ERP responses. The contributions from a computer science standpoint are in systems design and integration. Assembling the tools from Matlab and open source tools, and evaluating the hardware and analytical approaches to working with this noisy biological data source are a proof by demonstration of the approach, but there are no pure theory contributions to the domain of computer science presented by this work. It is not intended to convey the impression that there are novel algorithmic techniques being applied to this problem. There are, however, significant domain-specific contributions in terms of how to collect and analyze this type of data. In particular, the characteristics of the headset and the software used to collect this data are critical to ensuring good experimental results, and these decisions are more complex when dealing with a varied set of open source tools rather than the product of a single company. The statistical analysis of the data also exposes some complexities, which are detailed in this work. This combination of computational expertise, statistical rigor, and domain expertise are characteristic of many data science problems, or problems where large amounts of noisy data must be carefully processed to attain a specific goal.

Several goals of this line of research are explicitly excluded from this work. There are no clinical populations sampled as part of these studies. All investigations are based on the performance of cognitively normal young adult subject populations in an effort to gain preliminary data on the viability of the technique. The effects of fatigue and motivation are complex and highly task-dependent. While a discussion of these effects as they apply to this particular experimental protocol is appropriate, it is left to other venues to generalize this into a more comprehensive theory of fatigue and motivation in the P300 ERP and some of the references presented in section 2.3 begin to do this. The only experimental training results presented are the results of short-term studies of up to five sessions.

## 1.3 Structure

This dissertation presents an introductory chapter reviewing the literature on which this work depends (chapter 2), and reviews some of the hardware decisions that were made at the outset of the project (chapter 3). These hardware decisions have some ramifications in terms of software tools and the methodology that can be employed in the studies. An initial study is presented in which users were brought back over two sessions to see if there was a training effect (chapter 4 and 5). The main training effect study is then presented (chapter 6) and the results of this 5-session, 12-subject study are discussed in relation to how task performance changed over time (chapter 7). The memory attention study is then presented (chapters 8 and 9), followed by a discussion section integrating the three studies (chapter 10). This work is highly preliminary, and while it offers some promising possibilities, it is far from showing any clinical utility, so a ‘future work’ section speculates on possible directions that the research could take (chapter 11). A conclusion reflects on the research questions and their result (chapter 12). In the appendices, the specific code changes to the analysis pipeline are presented in Appendix A, while the code modifications to the presentation framework appear in Appendix B. Some experimental suggestions are provided in Appendix C regarding factors effecting EEG studies. The detailed questionnaires and other study materials are presented in Appendix D.

# 2 Literature Review

In this section, the dissertation will review the principles of biofeedback, neurofeedback and EEG recording, and will review the assumptions of event-related potentials and introduce the P300 brainwave component. The chapter will then describe how the P300 ERP component is applied in the P3 speller paradigm. It has a history of assistive uses, and an existing large body of work focuses on how to make the P3 speller a faster and more accurate assistive device, though with only sporadic mention of user training effects. A discussion of data preparation for EEG is necessary in order to understand how to make use of noisy data. This is followed by an analysis of previous attempts to use consumer-grade devices for research studies. Some specific techniques will be reviewed that make neurofeedback protocols sufficiently engaging for long-term use. Finally, the scant literature on ERP training will be examined.

## 2.1 Biofeedback and Neurofeedback

Biofeedback is the process of establishing a feedback loop with biological signals, with the intention of exerting some level of control over the processes being monitored. Since many human processes are largely unconscious, the effect of presenting an individual with feedback regarding the nature of those processes allows them to learn to exert some measure of conscious control over them. It became popular in the 1960s and 1970s, and was found effective for muscle retraining, resolving cardiac arrhythmias, lowering blood pressure, and reducing the frequency of epileptic seizures (Blanchard, 1974). It was popularized and turned into an alternative therapy in many situations in which it did not have proven effectiveness, such as pain relief (Jessup et al, 1979). As a result, it became associated with sham practitioners and has had a difficult time being adopted by mainstream medicine. In the popular account, “A Symphony in the Brain,” Jim Robbins describes being introduced to brain biofeedback, or neurofeedback, and how he had to separate the science from the charlatans (Robbins, 2000). Today, there are serious and effective biofeedback and neurofeedback techniques for a variety of conditions, such as psychiatric conditions including anxiety and depression (Shoenberg and David, 2014), and physical conditions such as muscle control after stroke (Barcala et al, 2013). With the advent of fMRI-based neurofeedback, pain relief has also found its way back into reputable studies (Hawkinson et al, 2011) and has surprising efficacy with few adverse effects (Sulzer et al, 2013). There are now several comprehensive guides to practical protocols for EEG-based neurofeedback, focusing on cortical slow wave neurofeedback (Demos, 2005; Budzynski, 2009; Larsen, 2012). However, while some of these guides define the term ERP, none use it as a basis for neurofeedback.

The use of neurofeedback for attentional conditions is of particular interest for the current dissertation. Direct attentional training, where participants are given tasks that require sustained attention, has been shown to be an effective intervention in children with ADHD (Kerns et al, 1999), suggesting that neurofeedback-mediated mechanisms of attentional training have a strong basis for success. Vanessa Prox has shown that individuals with ADHD have markedly different patterns of ERP response, including an attenuated P3 response as compared to individuals without ADHD (the P300 ERP is also known as the P3 or P3b, as is described more fully in section 2.3). In this study, participants were given a go/no-go task, in which target and non-target stimuli were presented, and the participants were to press a button for target stimuli. Results showed an increased amplitude in ADHD subjects in N1 and N2 waves (N1 is the earliest attentional ERP component and is thought to reflect initial attentional focus, while N2 waves are thought to be an inhibitory response to those initial brain activations in N1), followed by decreased P3 responses (interpreted as impaired later attentional processing) (Prox, 2007). EEG neurofeedback for ADHD has proven to be an effective intervention using sensorimotor rhythm (SMR) neurofeedback and alpha wave neurofeedback (Loo and Barkley, 2010; Loo and Makeig, 2012; Arns, 2014). Since the focus of the current dissertation is ERP neurofeedback on a specific brainwave pattern known to be altered in individuals with ADHD (Prox, 2007), it is hoped that ERP neurofeedback will be a more effective attentional intervention than existing spontaneous EEG neurofeedback techniques, which do not target specific ERP components but rather train overall brain patterns not tied to a specific response.

## 2.2 EEG and the event-related potential

Biofeedback is a general term for any type of biological feedback loop, while neurofeedback is biofeedback based on brain signals. Neurofeedback can take many forms, but this dissertation will focus on neurofeedback using EEG. The potential of a single EEG electrode measured at the scalp is the summed potential of between 100 million and 1 billion neurons (Nunez and Srinivasan, 2006), and the precise cortical localization of these signals can be challenging. More precise localization of signals could be provided through functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), and precisely localized fMRI brain imaging has also been used for successful neurofeedback (Weiskopf, 2012). However, the temporal precision of these approaches is not as effective as with EEG (Menon et al, 1999; Debener et al, 2005.) In addition, a significant benefit of EEG is the low cost and ease of use of the technology. While this dissertation will discuss difficulties, such as electrode contact, sample rate, receiver precision, and signal loss, these technical challenges are encountered while using a several-hundred-dollar piece of hardware that can be used in any quiet office, whereas studies using fMRI or PET require several orders of magnitude more capital investment and are much less portable. Magnetoencephalography (MEG) is another technique which can access data comparable to EEG in it’s temporal resolution, and has been used for a different view of P300 ERP (Mecklinger et al, 1997), but is cumbersome from an equipment perspective and not substantially more precise than EEG in regards to spatial resolution. Functional near infrared spectroscopy (fNIR) is an interesting complementary modality to EEG, and one could imagine useful neurofeedback protocols designed around this modality. fNIR has been used to look at P300 signals and even to produce a hybrid BCI (Liu et al, 2013) but will not be examined further in this dissertation.

In his book, “Introduction to the Event-Related Potential Technique,” Stephen Luck describes the history of Event-Related Potentials (ERP) and provides a comprehensive manual for the use of ERP in psychology studies. Since it is a brain response to a specific stimulus, ERP is a technically challenging brain electrophysiology technique as compared to spontaneous EEG. Any EEG is the detection of the averaged electrical potential of many thousands or millions of individual neurons, which set up an electrical dipole moment. If these individual dipole moments are spatially aligned, then an electrical potential can be detected at the scalp. In some cases, this is complicated by the physical folding of the cortex. ERP, specifically, is the detection of EEG, but with the additional complication of having to precisely locate a marker of the timing of a specific stimulus presentation within the EEG record (Luck, 2005, pp29-31). Given the precision needed to analyze EEG in general, and ERP in particular, its history is surprisingly long, far pre-dating the computer age. In 1929, Hans Berger first published an account of EEG (Berger, 1929), and a decade later the field became highly active, along with implanted electrode studies in animals. In 1939, the first description of something similar to an ERP was published (Davis, 1939), using an audio stimulus protocol whereby the frequency of the stimulus also provided the capability to record the time between stimulus presentation and neural response. It wasn’t until the mid-1960s that computerized stimulus presentations made commonplace a full exploration of stimulus-locked responses, soon called evoked response potentials (Sutton et al, 1965). Even during this period, the majority of the field was engaged in the study of spontaneous EEG rather than ERP, and it is spontaneous EEG that has remained the focus of neurofeedback techniques.

### 2.3 P300 and P3 Speller

The P300 is a specific ERP component that starts approximately 300 milliseconds after a stimulus. It is measured using EEG, and since it is strongly associated with attention and is attenuated in individuals with ADHD it will be the focus of the remainder of this dissertation (other conditions also attenuate and delay the P300 and are important in the design of P300 studies) (Gray et al, 2004). A comprehensive review in the early 1990s summarized the factors leading to a robust P300 response, including novelty, attention, and the absence of certain medical and attentional conditions (Picton, 1992). The following figure illustrates a typical P300.

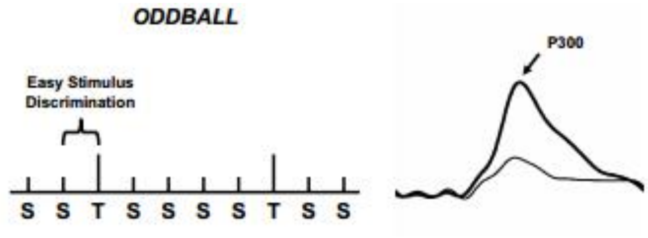


**Figure 2.1: The P300 response, showing a greater positive deflection in the case of an infrequent target stimulus (Polich, 2007)**

Although the underlying cognitive basis of the P300 is not completely known, the most accepted theory is that of “context updating,” which is the theory that the wave is involved in updating the individual’s representation of the current environment (Luck, 2005, pp42-43). In the context updating theory, when a new stimulus arrives that modifies an individual’s current mental state, a wave of electrical activations occur that reflect this updating of the brain state. A rare stimulus that the individual has been waiting for has a particularly large context updating effect. In this dissertation, and commonly elsewhere, the P300 response is recognized as being made up of two subcomponents, the P3a and the P3b. The P3a begins slightly earlier than the P3b and is more frontally located. The P3a is more associated with perceptual novelty and the P3b with context updating (Dundon, 2015). In the studies reported in this dissertation, the distinction between these two subcomponents is not definitively determined and spatial localization results will be further discussed in sections 5.1.1 and 7.2.1.

The P300 is primarily associated with measures of attention, and therefore, manipulations that can impact attention appear to have the greatest effect on its magnitude. This has also allowed it to be used as an indicator of impending errors when attention wanes; when the P300 response to target stimuli reduces in magnitude, then it is a sign that attention is wandering and the likelihood of a task error goes up (Datta et al, 2007). Factors that improve attention, such as mindfulness meditation, also increase the magnitude of the P300 response (Lakey et al, 2011). Attentional and motivational factors strongly affect P300 strength and reliability (Gray et al, 2004), and this is likely a reason that a marked decrease in the P300 response is often seen in individuals with attentional deficits (Brandeis et al, 2002; Loo and Makeig, 2012). This reduced P300 response has been suggested as a diagnostic criterion for ADHD (and also Alzheimer's dementia, Parra et al, 2012). Intriguingly, in adults with ADHD, adaptations and training strategies are postulated as being able to at least partially overcome these deficits (Prox et al, 2007). The attenuated P300 response in individuals with ADHD is fairly robust; it has been suggested that this is due to deficits in the ventral attentional network, which is a later stage of sensory event processing following processes in the pre-frontal cortex and prior to activation of the dorsal attentional network involved in behaviourally relevant stimuli processing (Szuromi, 2011).

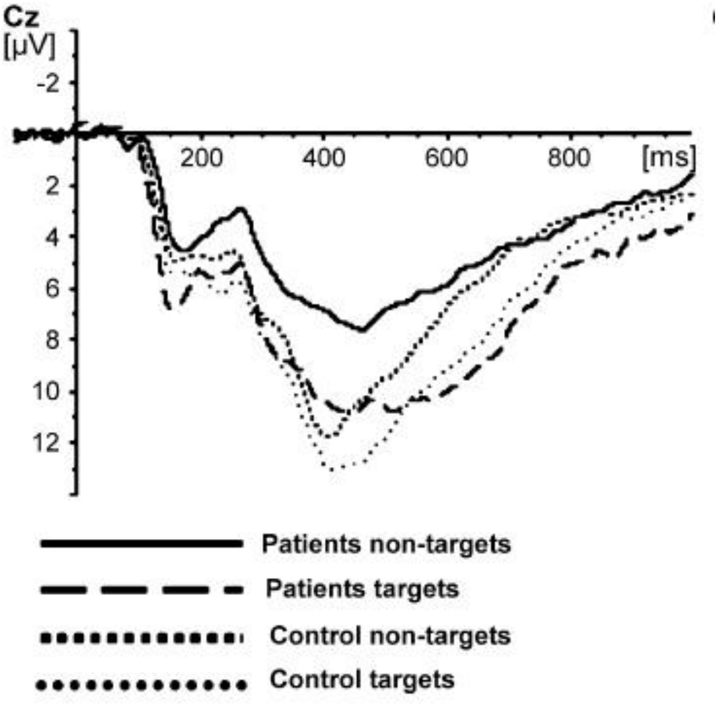
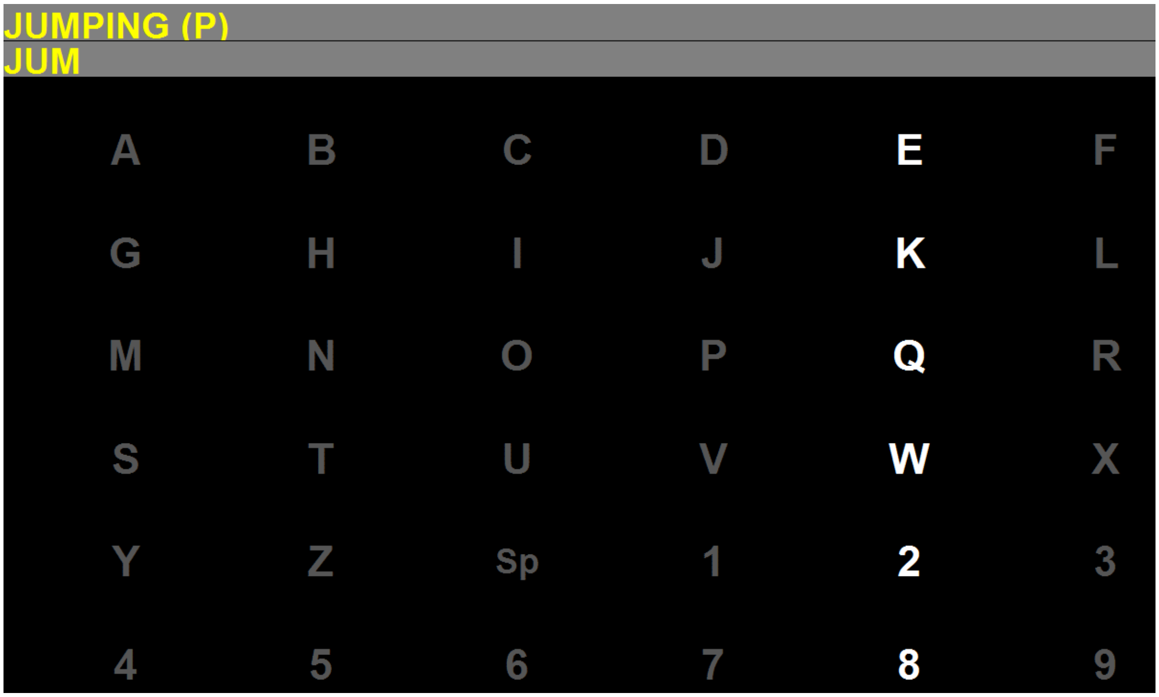


**Figure 2.2: P300 response in ADHD (“patients”) and non-ADHD (“controls”), for target items and non-target items (Prox, 2007)**

The P300 is one of the most popularized ERP components, having been used for highly-publicized studies on mind reading (Wang et al, 2012; Meixner and Rosenfeld, 2014), and even finding its way into courtroom trials as an indicator of the significance of information that a suspect did not want to admit to (Danaher, 2015). The strength and robust nature of the component make it an attractive target for studies, particularly on consumer-grade hardware, which may have more signal quality issues than more precise research equipment would have.

Although a variety of P300 paradigms exist, the simplest is the “oddball” paradigm, where the user is presented with two types of stimuli: a target and a non-target.

The target stimulus appears at a low frequency (generally 10% of the time or less). By aligning the EEG recordings following stimulus presentations (also known as epochs), the electrical potential of target epochs can be subtracted from the electrical potential of non-target epochs to, on average, show a differential effect of the target stimulus presentation.



JUMPING (P)
JUM

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| G | H | I | J | K | L |
| M | N | O | P | Q | R |
| S | T | U | V | W | X |
| Y | Z | Sp | 1 | 2 | 3 |
| 4 | 5 | 6 | 7 | 8 | 9 |

**Figure 2.3: Grid of letters flashing at a 120 msec rate, currently spelling 'P'**

The P3 speller paradigm, shown in figure 2.1, is a variant of the oddball paradigm, in that the letter the user is intending to spell is illuminated only a small fraction of the time, such that when it is illuminated the user produces a P300 response. The P3 speller paradigm was originally introduced as an assistive device, and was the first ERP-controlled brain computer interface (BCI) where users focus attention on flashing stimuli and where the P300 response is used for the system to identify the focus (Farwell and Donchin, 1988). It has become a prominent experimental paradigm due to the reliability of the P300 ERP and because the P300 is one of the most easily detected ERP components. However, the P3 speller has largely been supplanted by eye tracking solutions as an assistive device, due to the time it takes to spell each individual letter and owing to the necessity of repeating presentations so as to reliably detect the P300 response (Marchetti and Priftis, 2014). Some other EEG-based spellers, specifically ones using motor imagery, have shown themselves to be more efficient than the P3 Speller (Al-Negheimish, 2013). Being such an important research paradigm, however, a great deal of attention has been focused on how to get it to work as well as possible by optimizing variables such as presentation rate, target frequency, and variable delays (Citi et al, 2010; Mak et al, 2011). These task optimizations have allowed an established experimental paradigm to be adapted for a new use as a neurofeedback protocol without an extensive period of ensuring that the basic paradigm itself works. The P3 speller is an example of a task-relevant oddball paradigm, making it more prone to generate a P3b response than a task-irrelevant oddball paradigm (Bowman et al, 2013).

The P300 is only one of several EEG-derived measures correlated with attention. Recent work has focused on a technique known as Steady-State Visual Evoked Potentials (SSVEPs) where the repeating flashing of a stimulus is linked with ERP responses and the known frequency of the flashing can help extract perturbations of the signal from the background noise. Centro-Parietal Positivity (CPP) is another measure correlated with attention (O’Connell et al, 2012). Work beyond that presented in this dissertation will have to establish which technique, and perhaps which ERP sub-components, are most beneficial as targets of EEG neurofeedback.

## 2.4 EEG Data Acquisition and Analysis

The process of obtaining and analysing data from ERP involves four main steps. First, obtaining the data requires the selection of the appropriate electrodes and electrode locations, including identifying the proper reference electrode locations and careful fitting of the electrodes to the participant. Next, collecting the data requires that the stimulus presentation be tied to the EEG recording so as to maintain a reliable record of the presentation epochs (an epoch is one cycle of presentation and neural response). Third, the data is averaged, the artifacts are rejected, and then they are possibly re-referenced and filtered. Finally, the cleaned data is plotted and statistical analysis is performed (Luck, 2005).

Although it is possible to create an EEG application from scratch, there are a great number of decisions involved; with the existence of several open source frameworks, the effort of doing so is generally not warranted. The most recent framework, which the current studies were originally intended to be based upon, is a set of libraries built on top of Matlab and EEGlab, called BCILab (Kothe, 2013). Designed at the University of California in San Diego, BCILab offers the most up-to-date integration with analysis routines, and the simplest programming model. However, the driver layer is somewhat incomplete, and the low-cost headset used for these studies had to pass through several intermediate layers before being surfaced into Matlab. One of these layers introduced unacceptable levels of signal jitter, which led to the rejection of this framework.

An alternative framework that was considered was a package called OpenVIBE (Renard et al, 2011). However, it also suffered from driver problems, as its data acquisition component was part of the driver chain with BCILab. In addition, the programming model was not as straightforward as BCILab. Thus, the presentation framework that was eventually decided upon was BCI2000 (Schalk, 2004), which is a set of C++ libraries built on top of a large set of driver interfaces, including one for the Emotiv Epoc.

The process of offline analysis of data frequently begins with re-referencing the data. This is a technique used to correct signal drift and noisy reference electrodes, where the baseline that signals are compared to is comprised of the average of all channels. This common average reference (CAR) is a standard approach (Dahlwi and Hadi, 2012), and the runtime training of the classifier uses this approach. Average reference approaches have some issues, however; in particular, when a given wave is broadly distributed and sampled by a limited number of electrode locations, the resulting signal can be greatly attenuated (Luck, 2005, pp111-112). In the current studies, with only 14 electrodes covering a fairly small range of locations, more usable data was provided through the selection of specific reference electrodes rather than depending on an average reference.

An essential step in EEG data analysis is the rejection of artifacts. These artifacts primarily come from motion (which disrupts the electrode contact), from muscle movement (muscle EMG has far greater electrical potential than does EEG), from distraction of the participant (for instance, by a noise in the experiment room), and from signal drift (where the EEG amplifier changes baseline values independent of any participant electrical potential changes, or where skin contact and other non-neural changes cause that baseline to move). The rejection of trials is a complicated topic, and the particular approach in this work was most informed by chapters in the "Event-Related Potential Technique" (Luck, 2005) and "Analyzing Neural Time Series Data" (Cohen, 2014). There are several fully automated approaches to trial rejection (Ahmadi et al, 2013, Arico 2014), and some fully manual approaches that involve inspecting the data by eye, but the current studies used the built-in rejection functions in EEGLab in a semi-automated fashion. This involved examining the output datasets to set specific rejection criteria. These criteria are detailed in the methods sections of this dissertation, but largely involve the rejection of epochs where drift is too great by setting parameters for linear change over time, as well as the rejection of epochs with potential changes outside the magnitude that would be expected by neural responses.

The statistical analysis of EEG data can be complicated; since the collection of the data is assisted by some standard formats, tools for analyzing data have emerged to standardize the treatment. Software packages exist in several languages, including a python-based framework called pyEEG (Bao, 2010). One of the most popular is a package built on top of Matlab called EEGLab, which emanates from the University of

California in San Diego (Delorme and Makeig, 2004). In addition to the base EEGLab package, extensions were contributed by the BCI2000 developers that aided the loading of data into EEGLab (Schalk and Mellinger, 2010), including a package for the analysis of ERP data called ERPLab (Lopez-Calderon and Luck, 2014). These tools allow for the treatment of ERP data in a relatively standardized manner, including artifact rejection, plotting of the data, and statistical analysis. Many experiment-specific decisions remain, but this toolset of analysis options provides a strong starting point.

## 2.5 Runtime classification of ERP results

Since the P3 speller must give a result to the user in near-real time, some of the data analysis steps mentioned above are approximated in the runtime system. This will not necessarily produce data that is clean enough to report for publication, but does allow the system to learn from the user responses so as to make a good prediction of which letter they were trying to spell. The determination at runtime as to whether a particular response is a P300 ERP is, in these studies, dependent on training a linear classifier. This is not training the user, but instead training the computer to know on which electrodes to place the most weight (and which electrodes are most predictive of noise, see Haufe et al, 2014), and which time intervals are most predictive of a P300 ERP. A linear classifier sums the inputs in a high dimensional space and returns a classification in a lower dimensional space. This is accomplished by having the user spell known words so that the system can align the predicted and actual responses. In this case, the EEG values aligned by time of stimulus onset, in an array of 14 electrode channels, is trained through a comparison of the known responses to produce an array of weights relating inputs to an output classification with just two states – P300 ERP or no P300 ERP.

The approach used in these studies requires known stimuli to allow for supervised training of the linear classifier. The most common approach is a step-wise linear discriminant analysis technique for training a linear classifier, which is what is natively implemented in BCI2000 and used in these studies. This approach may not always be the fastest or the absolute best performer, but is an approach that provides good performance, as shown in a comparison of P300 classification methods (Krusienski, 2006). A more detailed recent explanation of the approach is found in a paper applying this technique to the classification of N170 ERP responses (Lee et al, 2012). Non-trained approaches, also known as unsupervised learning, have been more recently developed (Lage-Castellanos et al, 2010). However, since these techniques are more experimental and the supervised methods function well in most circumstances, they were not further explored in these studies.

## 2.6 EEG Hardware and Consumer EEG Devices

There are three primary factors that differentiate consumer-focused EEG systems from the more expensive research models. The first is electrode construction. Most research EEG systems still use gel-coupled electrodes in a swim-cap style headset, whereas consumer devices often use dry or saline coupled electrodes, although the quality of signal from dry electrodes is such that they are beginning to make inroads into research devices. The second factor is that most consumer-focused EEG systems are wireless, as compared to needing a tether of wires to the amplifier. Lastly, consumer-focused systems often have fewer electrodes, and those electrodes tend to be in a frontally-oriented montage as they are often designed for cortical slow wave data.

In the last five years, there have been a growing number of attempts to use these low-cost headsets for collecting research data. The first attempt was presented in 2011. It showed that it was possible to use the Neurosky Mindwave single sensor headset to detect P300 responses (Grierson and Keifer, 2011). This headset was determined by other researchers to be too limited in its capabilities. The Emotiv Epoc has received more study, with the general conclusion that while it generates reliable data, it suffers from long-term comfort issues (Mayaud et al, 2013), reduced signal quality (as compared to research grade instruments), and a lack of precision in headset fitting (Duvinage et al, 2012). Nonetheless, it does produce usable ERP data, even when compared directly to data from research systems, despite some data issues that require correcting during the analysis phase, such as motion artifacts (Badcock et al, 2013) and jitter issues (Ries, 2014). There has also been a study using the Emotiv Epoc to successfully detect the N170 ERP, which is more difficult to detect and is associated with facial recognition (DeLissa, 2015). Although the saline-coupled electrodes of the Emotiv Epoc are problematic in regards to motion artifacts, the wireless headset has been modified such that it is coupled to a gel-cap electrode set for a study of walking users (Debener et al, 2013). The Emotiv headset has also been modified to accept marker data to more precisely time stimulus onset (Thie, 2013). Given the range of studies on the Emotiv Epoc at this point in time, it is clear that while there are trade-offs to be made, it is successfully acquiring reliable ERP data for a number of researchers.

## 2.7 Gamification of Neurofeedback

In a short psychology study, the P3 speller task is acceptably interesting to participants; however, the novelty soon wears off. This limits its utility as a long-term training mechanism. If future work shows utility for ERP-based neurofeedback, protocols must be devised that can sustain interest over a longer period of time. Some individuals can wear computers full-time if they show sufficient value (Mann, 2013), and the integration of attentional training into everyday routines allows for long-term progression. However, the task has to hold the user’s interest. Recent work has looked at using P300 signals as an input modality in computer gaming (Kaplan et al, 2013; Ganin et al, 2013). This approach appears to be an excellent start in developing stimulus materials and test methodologies that are usable for a sufficiently long period in order to see a training effect develop. With the advent of consumer virtual reality headsets, some early work has looked at the integration of P300 brain-computer interfaces (BCIs) in a virtual environment (Käthner et al, 2015). The card game proposed in this dissertation is a step above the base P3 speller in maintaining participant interest; however, when virtual environments and games integrate truly engaging protocols with underlying utility, we can expect significant changes in the applicability of neurofeedback.

## 2.8 ERP Neurofeedback and Training Effects

Although general training for attentional conditions has been studied, the specific question of ERP training has received little coverage. The operant conditioning literature shows examples of P300 ERP training that suggest that training responses are possible, though the line of research has not been carried through to its logical conclusion of investigating the effect as a type of neurofeedback. Operant conditioning is a form of reinforcement learning in which desired behaviours are rewarded. As early as 1963, the operant conditioning of ERP was shown in animals (Olds, 1963), but operant conditioning of the visual-stimulus P300 response in humans was not demonstrated until 1981 (Roger and Galand, 1981). This study focused on positive and negative training using financial rewards in a task very similar to what is now known as the oddball paradigm. These results were later confirmed in a more extensive auditory P300 training study (Sommer and Schweinberger, 1992), which investigated the relative ease of training a reduced P300 response as compared to increasing P300 responses. These studies demonstrate that ERP responses can be conditioned, but do not explicitly investigate multi-session training or the visual attention paradigm of the P300 speller. Similar to much of the operant conditioning literature, these studies focused on the physiological aspects of response conditioning. No other mention of ERP training appeared in the literature until a 2011 study titled, “In Search of New Protocols of Neurofeedback: Independent Components of Event-Related Potentials.” However, this study uses only a single 20-minute training session with a small pool of subjects, sees no significant effect, and is not further mentioned in the literature (Kropotov et al, 2011).

In the BCI literature, one of the few mentions of ERP training is a short workshop paper that attempts to train users to improve P300 stability, which is a method of identifying ERP responses using the timing of the signal with respect to cortical slow waves rather than the overall magnitude of the P300 peak (Aloise et al, 2011). A small training effect was found, but the study involved only seven participants who used a P3 speller paradigm over the course of two short training sessions. Although some of the co-authors later investigated time of day as a factor in P300 performance (Arico, 2012), and the research is later mentioned in a Ph.D. dissertation (Arico, 2014), there appears to have been no further follow-up of the training effect. This result should be seen broadly as being in support of the existence of a magnitude training effect, but P300 magnitude and P300 stability are not identical measures and the small sample size in this workshop paper make it most comparable to the small two-session study reported in chapters 4 and 5 of the current dissertation.

## 2.9 Systems Design and Integration

A contribution of the current work is the integration of several low cost tools to provide researchers with the capability to perform EEG research, specifically the investigation of ERP neurofeedback, in a way that is considerably cheaper and easier than previously enabled. While it would have been possible to perform these studies on a research-grade instrument, perhaps the Brain Vision Systems or a Cognionics headset, the overall costs would have been far greater than the current study which used a $500 Emotiv device and primarily open source analysis tools. The benefit of tying together these tools is to allow researchers with a primary focus in the psychology domain rather than computer science to follow on with related work. In this sense, the current dissertation is an example of a systems paper – tying together disparate resources in a way that enables an application area, in this case ERP neurofeedback. A blog post from Tessa Lau described some of the elements of a good systems paper, and these are applied to the current work (Lau, 2010). First, this dissertation describes a potential treatment modality of ERP neurofeedback which is not captured by current research tools, and describes its use in enough detail for replication. Second, the studies shown in this dissertation demonstrate the potential applicability of this technique to external conditions and explore alternate approaches, both hardware and tasks, that could be used to similar ends. Finally, the studies show that the technique is effective, at least in a preliminary analysis, and analyzes some barriers to use and limitations of the technique. In other domains, some good examples of systems papers include Dr. Lau's own recent paper on a reprogramming environment for service robots, and a paper by James Landay on Gestalt, an integrated environment for implementing and assessing machine learning algorithms (Patel et al, 2010; Huang, Lau, and Cakmak, 2016). In both of these examples of systems papers, the process of assembling tools for a particular application, assessing their use for that application, and discussing limitations and future enhancements was followed.

# 3 Hardware Review

When this line of research begun, there were very few consumer EEG headsets on the market—primarily only the Neurosky Mindwave and the Emotiv Epoc. The OpenEEG project had been around since 2006, but significant effort and electronics expertise was required to turn that system into a viable headset. Now, however, there are at least a dozen consumer headsets on the market, with more planned for the future. Most do not provide sufficient access to the locations of the head required for ERP, though several can overcome this issue with some work. This section is intended to give a brief survey of the available devices, describe those evaluated as part of this work, and go into detail on the Emotiv Epoc, which is the headset on which the studies presented in this dissertation were conducted.

A consumer EEG headset has several characteristics which distinguish it from research-grade devices. The first is price, whereas a research-grade headset can cost many tens of thousands of dollars, consumer EEG devices are generally under $1000. The second element is ease of use; most research-grade EEGs use gel-coupled electrodes in a swim-style cap. These offer many more electrodes and better contact, but require laborious skin preparation and cleanup after use. The use of wired electrodes is also standard, so that a large bundle of wires must connect the user to the computer. The downside of this convenience is a paucity of electrode locations, and more data quality issues than on research-grade devices. The hardware capabilities of research-grade devices are changing, of course, and as wireless headset and dry electrodes make their way into these devices the research scenarios described in this dissertation should become even more accessible, both to well-funded researchers and those on a budget. The remainder of this chapter will focus on several of these consumer-grade systems and explore some of the trade-offs in functionality.

Table 3.1 presents a list of the EEG models that were evaluated as part of the work for this dissertation. The OpenEEG was only briefly looked at—the author had previously constructed the device, but as it is a wired amplifier with poor driver support, the construction of a headset for it was deemed too great a barrier to entry for other researchers. OpenBCI is an updated version of the OpenEEG, with a completely new amplifier design, better driver support, and a wireless headset connection; however, a fundamental problem is that it requires a headset designed by the user. The Neurosky Mindwave was the first consumer headset to be evaluated. However, with only a single sensor placed on the forehead, and an electrode that did not read through hair effectively so couldn’t be moved to another location, it was quickly discarded as a candidate for the ERP studies reported in this paper. The Mindwave also produced noisier data than the other consumer headsets. The Interaxon Muse is a nicely performing device but it is primarily focused on frontal locations and its built-in electrodes are not amenable to reading through hair. An additional electrode may be added, and placed in any location, and the signal quality is excellent. However, since only one non-frontal electrode would have been available, it would have been limiting for this study and wasn’t available when the line of research was started. The newest headset on the list, the Emotiv Insight, is compatible with the Emotiv Epoc on a driver level, but has fewer electrodes and is less expensive. Its electrodes are not as flexibly positioned as the Epoc, however, and it shipped after most of the studies in this dissertation were completed, so a decision was made not to include it in the studies. The Lab Streaming Layer (LSL) column in table 3.1 is relevant in that this is the mechanism used to surface EEG data in real-time to matlab and EEGlab, which offers a presentation layer that is considerably easier to use than BCI2000, although the OpenVibe acquisition layer introduced too much jitter in the early investigations, so a native LSL driver is preferable.

| Name | Number of channels | Date released | Price (USD) | LSL driver | BCI2000 driver | Suitability for ERP |
|---|---|---|---|---|---|---|
| OpenEEG | 4 | 2006 | $100 + electrodes and headset | No | Yes | Depends on headset. Wired. |
| OpenBCI | 8 or 16 | 2014 | $500-$900 + electrodes and headset | Through OpenVibe, native forthcoming | No | Good |
| Neurosky Mindwave | 1 | 2011 | $100 | No | No | No |
| Emotic Epoc | 14 | 2009 | $500 | through OpenVibe | Yes | Good |
| Emotiv Insight | 5 | 2015 | $300 | through OpenVibe | No | Unknown |
| Interaxon Muse | 4 | 2014 | $300 | Yes | No | Good, except the montage of available electrode locations is constrained |

**Table 3.1: EEG units evaluated for this dissertation**

Table 3.2 lists the electrode locations that are accessible using the various headsets evaluated. The only headset that could be easily reversed in orientation was the Emotiv Epoc, and in these studies it was used exclusively in that reversed orientation. These locations are based on the standard 10-10 system (Luck, 2005, p.118), as shown in figure 3.1. The P300 is detected most strongly at parietocentral locations (Picton, 1992), which are at locations C3, Cz, C4, P3, Pz, and P4 in figure 3.1.

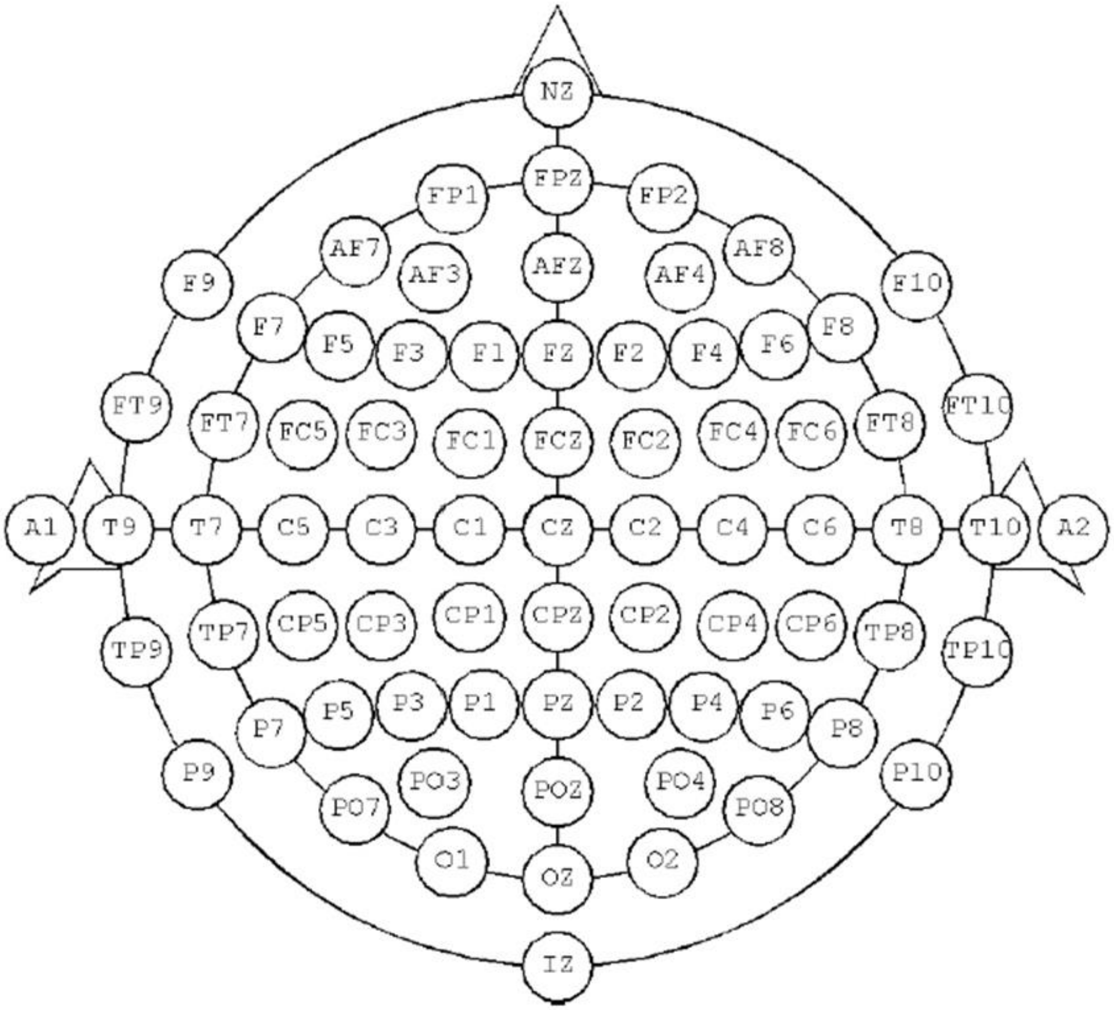


**Figure 3.1: Common electrode locations under the 10-10 system**

| Headset | Locations (normal) | Locations (reversed) |
|---|---|---|
| Neurosky Mindwave | Fp1 | N/A |
| Interaxon Muse | TP9, Fp1, Fp2, TP10 | N/A (extra electrode possible) |
| Emotiv Insight | AF3, AF4, T7, T8, Pz | N/A |
| Emotiv Epoc | AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 (with DRL and CMS at P3 and P4) | FC5, FC1, FC2, FC6, T7, T8, CP5, CP1, CP2, CP6, P3, P4, P7, P8 (with DRL and CMS at C3 and C4) |
| OpenBCI | Depends on electrodes | N/A |
| OpenEEG | Depends on electrodes | N/A |

**Table 3.2: Electrode locations accessible in normal and reversed orientations for each headset**

The Emotiv Epoc, in its reversed orientation, is the hardware on which the remainder of this dissertation will focus. The Epoc uses saline-coupled electrodes—this is a simpler system for casual use than a standard gel-coupled headset, but the conductivity is not as effective, particularly in participants with thick hair. It is particularly sensitive to motion artifacts, although in its reversed position in the studies reported here, those artifacts are more pronounced from full head movement than they are from simple eye blinks or facial movements. This particular headset is also sensitive to the size of the head. Unlike a standard EEG cap, which comes in multiple sizes, the Emotiv is available in only a single size. Some participants with smaller-than-average heads had a difficult time achieving an adequate headset fit; in particular, two of the women in this study required, on average, an extra 10 minutes of fitting time to ensure good scalp conductivity. The Epoc is also a relatively slow headset, recording only 128 samples per second as compared to 256 or even 512 from some of the faster research grade EEG headsets.

This 14-channel headset is normally worn with the majority of the sensors on the frontal area, but it was reversed in order to access the desired head locations. The electrodes were positioned as closely as possible to the locations described in figure 3.2, although head shape variability and a relatively non-adjustable headset did not allow these locations to be as precisely targeted as would have been the case with a cap-style electrode set. After positioning, the onscreen Emotiv control panel was used to ensure that all electrodes were shown as green in the user interface (see figure 3.3), indicating an impedance of 20K Ohm or less (“Emotiv.com forum”, 2015). In the case of inadequate connectivity, the electrode was pressed down into the hair and moistened with additional sterile saline solution. The rejection criterion for headset fitting was that no more than 2 of the electrodes could be yellow (i.e. no more than two electrodes at an impedance greater than 20K Ohms). This attention to headset fitting reduced the rejection rate substantially from early piloting, but required up to 10 minutes of fitting time for some subjects. This careful attention ensured optimal connectivity and was re-checked in the middle of the session and whenever a participant took a break. If connectivity was inadequate at any time, the electrode was pressed down into the hair and moistened with additional sterile saline solution, ensuring the location of the electrode was not changed.

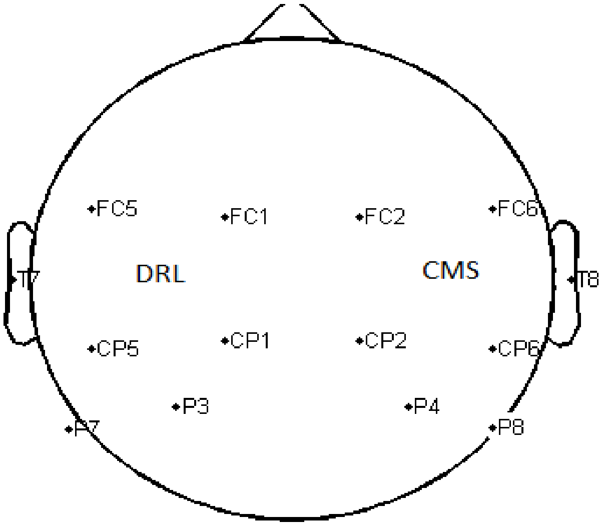
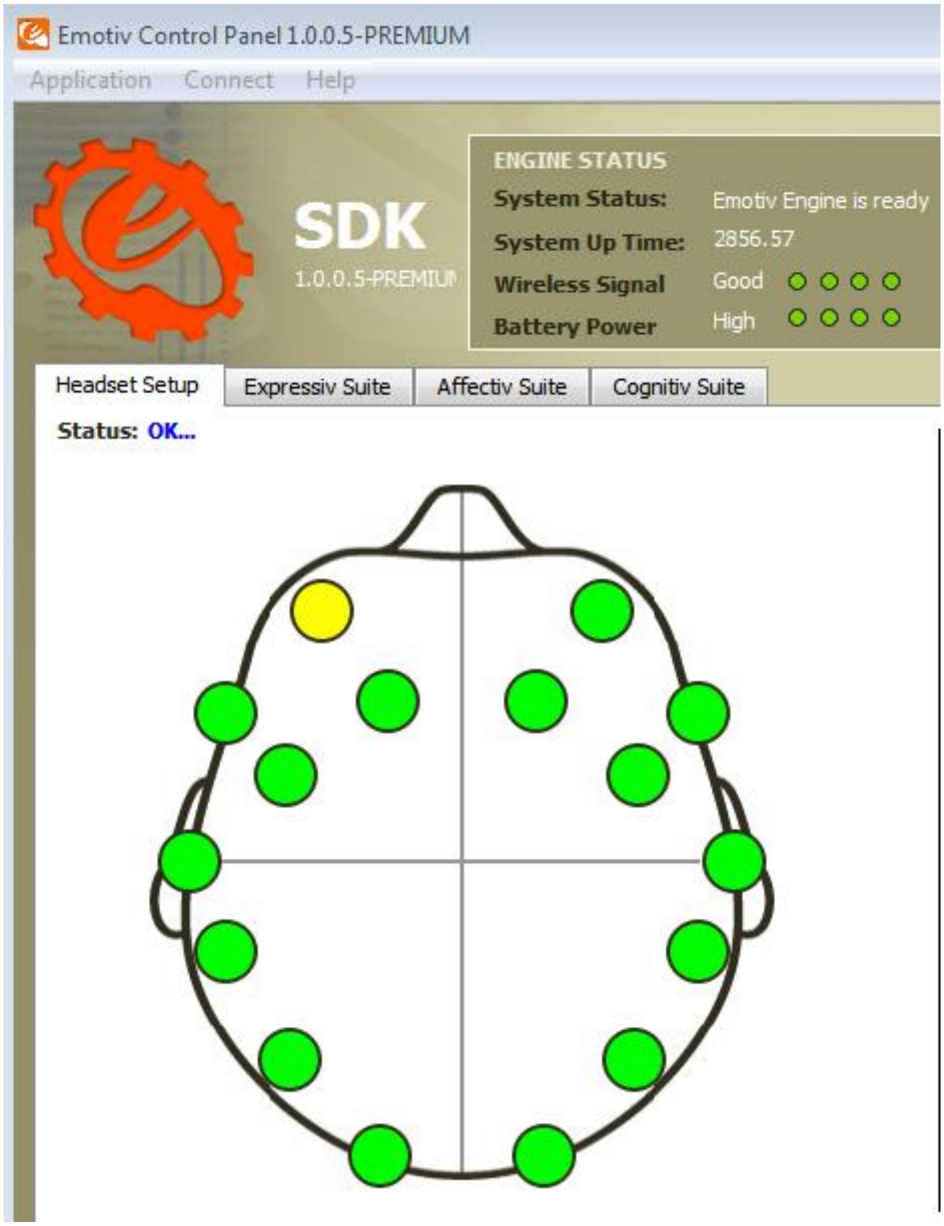


**Figure 3.3: Approximate electrode positions; DRL is a noise cancellation electrode and CMS is the common reference**



ENGINE STATUS

System Status: Emotiv Engine is ready

System Up Time: 2856.57

Wireless Signal Good

Battery Power High

Status: OK...

**Figure 3.4: Headset fitting display showing one electrode with less than ideal connection and the remainder showing good connection. This would meet the fitting criteria, but it was usually possible to ensure that all electrodes were green.**

Although there is a storage case for the Emotiv Epoc's electrodes that allows them to be moistened, the electrodes use a plastic connector, which is fragile. It was found that removing them each day led to a high breakage rate. The storage case also led to corrosion in the electrodes, which eventually compromised their effectiveness. Consequently, the electrodes were left attached to the headset and re-moistened with contact lens saline solution before each session. Approximately once a week, all felts were removed from the electrodes and soaked in water to remove excess salt buildup. After wetting the electrodes, the headset was placed on the participant. If sufficient connectivity was not achieved, the first step was to move as much hair as possible out of the way. The next step was to re-wet the electrode while it was in place. Some electrodes are difficult to fit on small heads, and long hair makes this more difficult. It was discovered that using a chopstick on each side underneath the long electrodes at the back of the head, and above the shorter electrodes, provided enough downward pressure to allow good scalp connection on the shorter electrodes without unduly disrupting the performance of the longer electrodes. These short electrodes are the ones originally designed for locations F7 and FC5, and F8 and FC6.

The Emotiv Epoc, then, is an example of a class of consumer EEG devices that are becoming more performant, cheaper, and more common. The studies here will examine the type of ERP neurofeedback that is possible in a device with saline-coupled electrodes, a moderately slow sampling rate, and a limited electrode montage. As the consumer-grade headsets improve over time, the measures used in this dissertation will only become easier to achieve.

# 4 Two-session study

## 4.1 Introduction

This initial study was the subject of a presented paper at the 2015 IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing entitled, “Evoked Response Potential Training on a Consumer EEG Headset.” Although the material has been greatly reorganized and edited, portions of this chapter have been previously published there. The purpose of this study was to establish that the basic P3 speller protocol was effective and properly parameterized, and to give an initial idea as to whether training of the P300 ERP was occurring. The primary research question in this study is whether P300 ERP results can be demonstrated on the P3 speller task using the Emotiv Epoc headset. Secondly, if the P300 ERP responses can be obtained on this hardware, to determine whether there is a training effect such that P300 magnitudes are greater on the second session.

## 4.2 Methods

### 4.2.1 Participants

Eight University of Victoria undergraduates participated in exchange for experimental credit in their psychology classes. Two of the eight subjects had to be removed due to rejection criteria, leaving 6 subjects in the study. The rejection criteria were determined in advance to exclude participants unable to achieve consistent performance, in order to enable the linear classifier to be trained. Participants returned for two sessions. Five of the six subjects were female. Selection criteria included an absence of attentional disorders as well as not having taken part in related EEG studies; both selection criteria were assessed by self-report.

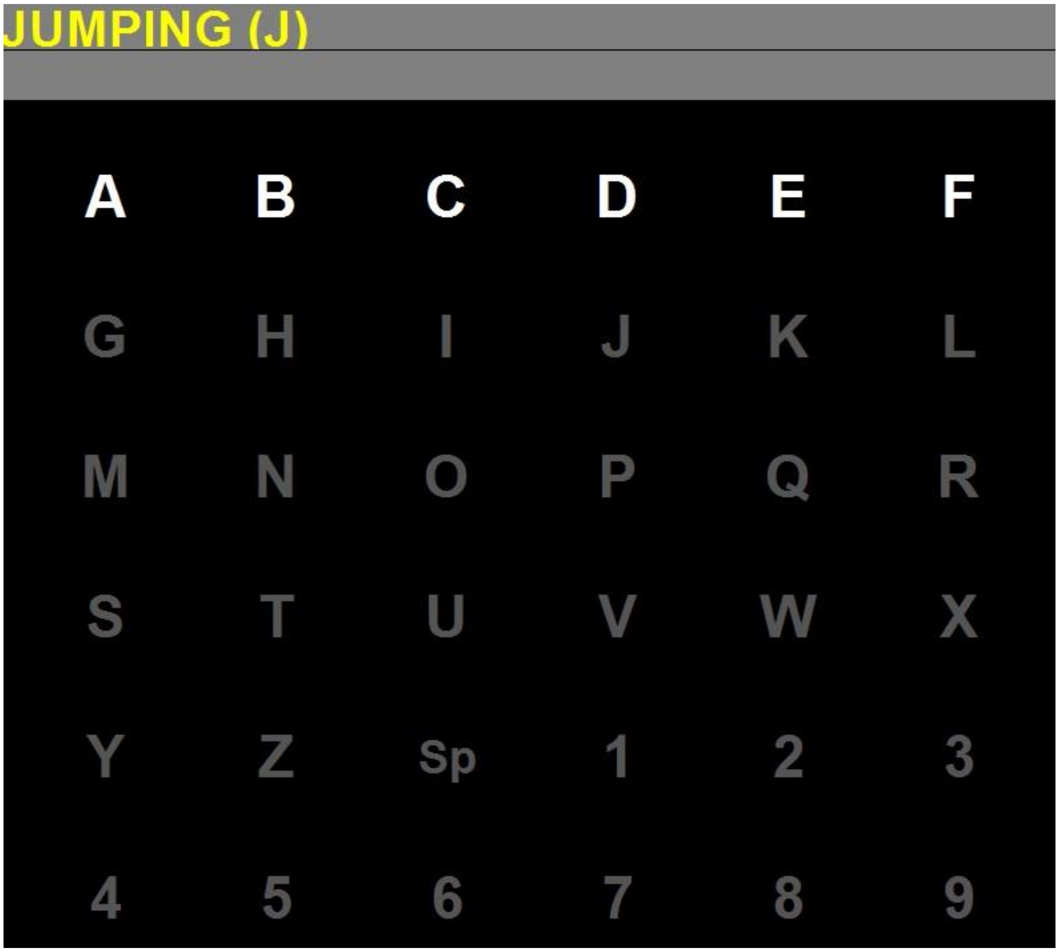
### 4.2.2 Apparatus

All current studies were performed on the Emotiv Epoc with a reversed headset position as described in the hardware chapter.

The P3 speller task uses P300 ERP responses to determine which letter the user is attempting to spell. In the version of the task described here, rows or columns of letters were intensified at a rate of 125 msecs per intensification (62.5 msec intensified, then 62.5 msec pause before the next intensification) as shown in figure 4.1. For each letter to be spelled, the letter was intensified 30 times (15 times in row intensifications and 15 times in column intensifications), alternating rows and columns in a random presentation. The participant was instructed to count flashes of the letter or number that they were trying to spell, although participants were encouraged to develop their own strategies to maintain concentration if they thought that would be more effective. In the example in figure 4.1, the participant was spelling the 1st letter of the word “JUMPING.” In looking for the P300 responses, the current study used an epoch size of 796 msecs, and since the rate of intensification is every 125 msecs, this meant that the epochs were overlapping. Individual variation in terms of the maximum presentation rate is large, and the rate of

125 msecs per intensification was chosen during pilot trials to allow most participants to achieve success at the task.

The monitor was at a viewing distance of 80cm, causing it to subtend a visual angle of approximately 34 degrees. In this study, the distance was precisely fixed due to the use of a chin rest and an Eyelink eye tracker. The eye tracking data was not used except to provide a visual cue to the experimenter that the participant was looking at the correct letter; this marginal benefit did not lead to it being used in later studies in this dissertation. The participants did not use a mouse or keyboard, and all interaction with the system was through the EEG.



JUMPING (J)

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| G | H | I | J | K | L |
| M | N | O | P | Q | R |
| S | T | U | V | W | X |
| Y | Z | Sp | 1 | 2 | 3 |
| 4 | 5 | 6 | 7 | 8 | 9 |

**Figure 4.1: Interface for P3 speller task**

It is worth noting that the presentation timing was influenced by the sample rate of the headset being used. In the case of the Emotiv Epoc, this sample rate was 128 samples per second, and samples were transmitted in a block size of 4. Consequently, each sample block was 31.25 msecs in duration, which defines the most precise temporal resolution of stimulus presentation possible. This sample rate is a trade-off with the Epoc, and while most research EEGs offer a 256 Hz sample rate, in working with the P300, the current sample rate did not present a significant limitation except in limiting the number of samples available to be analyzed. The slower sample rate could have limited frequency domain analyses if those had been the focus, however.

### 4.2.3 Procedure

The study was broken into two sessions. On the first day, subjects arrived and had the P3 speller task explained to them, then had the headset fitted. Participants were told that each letter would be intensified for a moment, in rows and columns, and that they were to mentally count the number of times the letter they wished to spell turned bright. When participants understood, the first session began. This was with a very short phrase, “BIRD,” in order to get the participant used to the task. Following this, they spelled three phrases ranging from 10 to 15 characters per phrase which took 5-7 minutes per phrase (e.g., “CHERRY BLOSSOMS,” full prompts given in Appendix D). On the second session, the protocol was the same, except instructions were abbreviated and the short orienting phrase was not presented.

On each day, linear classifier training was attempted after the first and third phrase in order to maximize subject accuracy on the task. This training consisted of using the prompted phrase paired with the results generated from the user. This alignment of known correct responses and user-generated responses allowed the system to train a linear classifier to be able to associate the user EEG inputs to more accurately predict user intention in future trials. This training adjusts the inputs by electrode location and P300 timing to correctly weigh the inputs on a per-user (and per-session) basis.

A fourth phrase was given to test the fully-trained model. The subject was then allowed to spell a word of their choosing (free spelling), by writing it on paper before the experimenter began the session. A short break was then given (minimum two minutes, up to 5 minutes), after which another free spelling trial was performed. Finally, another prompted trial was given. All participants were given the same phrases, and the phrases were the same on both days (with the exception of the free spelling trials).

Participants returned for a second session 24 hours later (plus or minus 3 hours). The procedure on day two was identical, with the exception that task instructions were abbreviated and the orienting phrase was not given. A questionnaire was filled out on both days asking for the participant's subjective effort ratings, comfort, and general alertness level.

### 4.2.4 Data Analysis

The data analysis was done within Matlab, using the EEGlab analysis package (Delorme and Makeig, 2004). A few significant changes needed to be made to the data before it could be imported into EEGlab. The first was to map the event format of the results from the BCI2000 standard into a format amenable to analysis in EEGlab, as described in Appendix A. A more significant theoretical point is that the epochs in the P3 speller in BCI2000 are overlapped, meaning that with an epoch length of 796 msecs and am inter-stimulus interval (ISI) of 125 msec, each epoch contains multiple stimuli. This is not a problem, but does mean that the independent component analysis (ICA) features of EEGlab are not appropriate to use on this epoched data, as each epoch contains data from several stimuli and therefore have strong cross-dependencies (Cohen, 2014, p.77). For this reason, all analyses below are in the time domain rather than more complex decompositions, although a reviewer noted that ICA decomposition could have been performed prior to epoching. All non-target epochs are those in which there is no target event anywhere in the epoch. The reference signal on the Emotiv also requires a brief reminder; since the reference channel in the reversed electrode position is itself neurally active at location C4, the channels are all relative references to this position. In the case of a broadly distributed signal like the P300, this is not as convenient as an absolute signal, but still allows reasonable data interpretation.

Two analyses were of primary interest. The first was whether we could find a significant difference between target and non-target items, broadly distributed across electrode locations, and centered in the time range that we would expect a P300 response. The second was whether a training effect was shown from day 1 to day 2. To identify the baseline, non-target trials were subtracted from target trials, so subsequent figures show the response at each time point (averaged across trials) of target trials compared to the same response for non-targets. The comparison at each test electrode to the reference electrode determines what potential difference is reported at each location. Unlike in the subsequent study reported in chapters 6 and 7, trial rejection in this study was only on a per-user basis as described in section 4.2.1, so users who were not able to achieve sufficiently reliable results to train a model were rejected (2 subjects out of 8) but unlike in the later study individual trials were not rejected.

The Emotiv headset, in its reversed positon as used in this study, has its reference electrodes at locations C3 and C4, so the analyses were performed on electrodes P3, P4, P7, and P8 and a negative deflection is expected rather than the standard positive deflection at more central locations. This is discussed further in sections 5.1.1 (for this study) and at more length in section 7.2.1. Additionally, a few words must be said about baselining. The EEGLab package has facilities for removing a baseline which were used in the production of these figures. The manual suggests removing the average value of the signal throughout the epoch (in the present case, that is from 0 to 800 msecs following each stimulus presentation). An alternative way of baselining, which was not used here, would be to remove the value at time 0 and then allow the rest of the epoch to appear naturally. This second approach was not used because it tends to accentuate the differences between the target and non-target responses when signal drift is responsible, whereas baselining with the average value seems more neutral in high-drift situations.

# 5 Two-session study results and discussion

## 5.1 Results

Figure 5.1 is a scalp map displaying the activations of overall electrode positions and comparing the non-target trials on the left (where the letter being intensified was not the one being spelled), versus target trials on the right (where the letter being intensified was the one the participant was spelling). Due to the central location (c4) of the Emotiv reference channel, the P300-like ERP was not accessible from the typical front-central channels. The channels chosen for these studies (P3, P4, P7, and P8) reflect the negative dipole of the positive frontal P300 component.

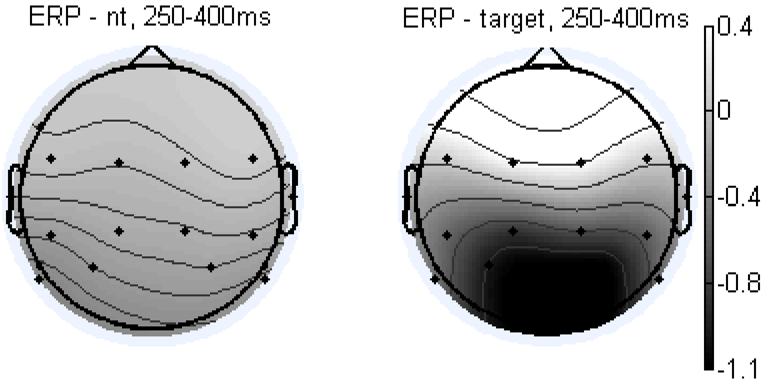


**Figure 5.1: Topographical view of all electrode locations showing minimal activity on non-target trials (left) and a band of activity across the top of the head on target trials (right), between 250 and 400 msec in each epoch. Differences between target and non-target trials are significant at the p<.05 level.**

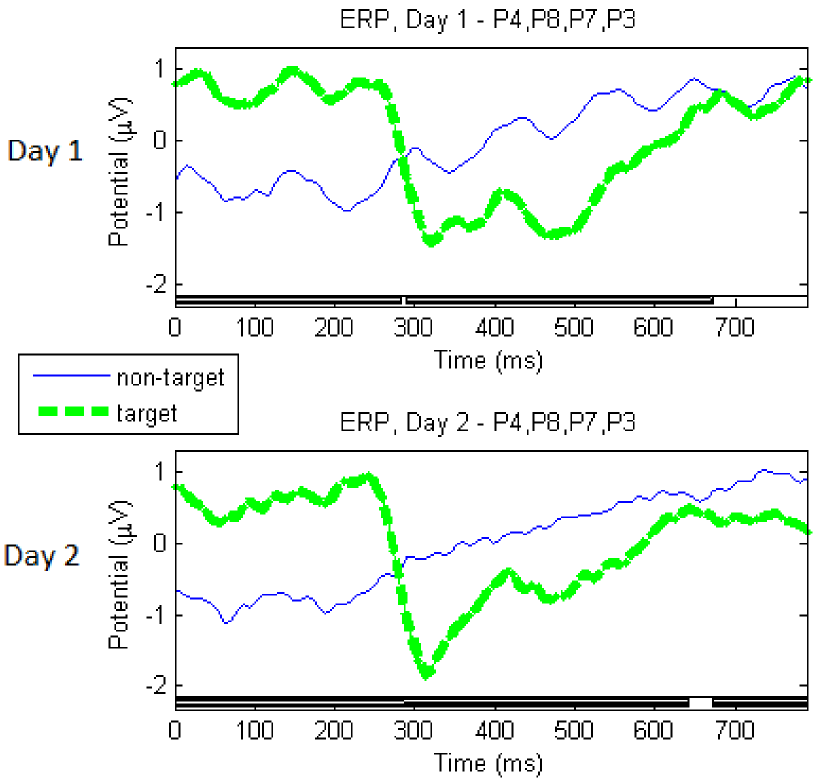


**Figure 5.2: Non-target (blue) vs. target (green) ERP responses on day 1 (top) and day 2 (bottom) Dark bars at base of charts indicate regions where target and non-target are significantly different at the .05 level, showing a robust P300 signal on both days.**

For the analysis of the training effect on the ERP, the maximum peaks were picked at an interval of 250 msecs to 400 msecs post-stimulus-presentation, on an average of the electrode locations at p3, p4, p7, and p8. These results are shown in figure 5.2. The peaks differed significantly on day 1 vs. day 2, with a day 1 mean of -1.66 microvolts and a day 2 peak of -2.31 microvolts (t-test on peak values shows p<.05). These results were broken down by subject, with figure 5.3 showing a suggestive trend of improvement over time for four out of six subjects. Subjects 6 and 7 began day 1 at a high level of performance, and this reversed on day 2, presumably due to motivational factors as they became bored with the task, although there is not enough data in this study to confirm this hypothesis and further discussion of fatigue and motivation is provided in section 7.5.

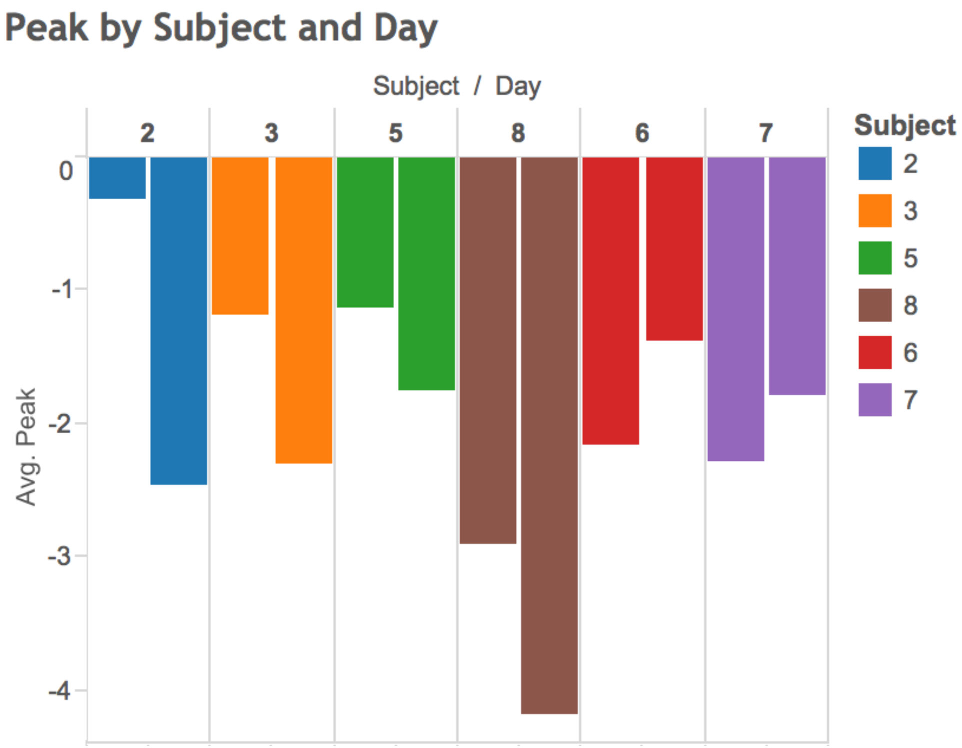


**Figure 5.3: Average peak height on untrained model trials for day 1 (left bar in each pair) and day 2 (right bar in each pair) across subjects (coloured), showing the strength of the P300 response increasing through the training. Y-axis is in microvolts at the peak of the P300 response (range 250-400 msecs).**

Overall, accuracy rates ranged from 4% on the first untrained phrase to 90% on a phrase of their own choosing. On day 2, the first untrained phrase had an average of 10% correctness. Since the number of subjects in the study was chosen to show statistically significant results on individual ERP responses, the overall word error rate did not have enough samples to be statistically significant.

On the post-test questionnaire, subjects reported headset comfort (1=very uncomfortable, 7=very comfortable) and saline solution irritation (1=very irritating, 7=no irritation), with average values of 6.2 and 7, respectively, indicating that headset comfort was not seen as an issue in this study.

### 5.1.1 Validating the P300-like ERP with additional analyses

Although this study was not intended to delve deeply into the spatial localization of the P300 signal, it is necessary to ensure that the signal was occurring where our predictions would have suggested and to attempt to explain some differences in spatial localization from a classic P300 response. As mentioned in section 2.3, the P300 is made up of P3a and P3b subcomponents, with the P3a subcomponent occurring earlier and more frontally than the 3b subcomponent (Polich, 2007). The placement of the reference electrodes at locations C3 and C4 made localization of the signal somewhat challenging without further analysis. Figure 5.4 shows the same results as figure 5.2, but with a difference reference. The common average technique is used, where the average of all the electrodes are taken and used to compare each individual electrode. As mentioned in section 2.4, in a broadly distributed ERP like the P300 this has the effect of attenuating the signal, but it does allow a clearer picture of the signal localization than does a single reference electrode, particularly one in an active location like C3/C4. As can be seen in figure 5.4 there is a positive deflection at the frontal electrodes, and in figure 5.5 one can observe a negative deflection at the parietal electrodes. This is a more frontally dominant signal than would be expected from the task. The P3 speller is generally considered to be a task-dependent oddball paradigm variant, which is expected to produce a P3b response

(Bowman et al, 2013). The fact that the peak was not seen at the parietal electrodes is unexpected and may be due to hardware, headset fitting factors, or unexpected task characteristics. Further analyses will be carried out on the main training study in section 7.2.1.

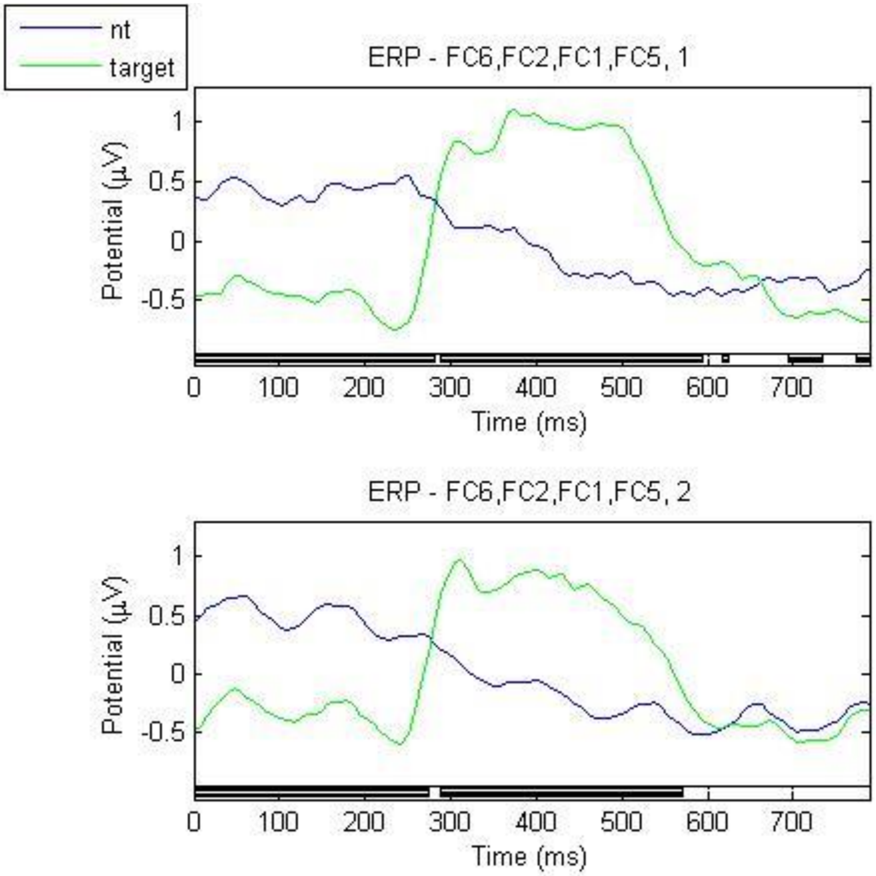


**Figure 5.4: Frontal electrodes using a Common Average Reference (CAR)**

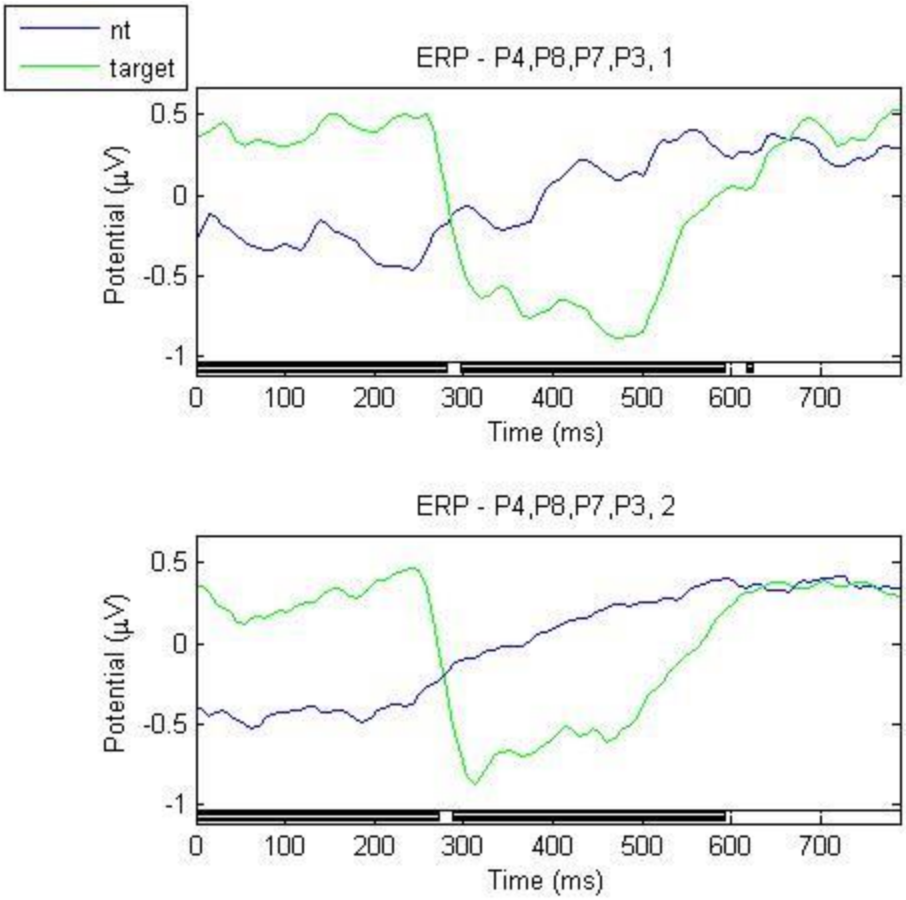


**Figure 5.5: Parietal electrodes using a Common Average Reference (CAR)**

## 5.2 Discussion

This first study was important in validating several aspects of the research plan. First, it confirmed that participants could wear the Emotiv Epoc headset in the reversed position to access the desired electrode locations. The study showed that this could be done comfortably and effectively, unlike in prior studies from other labs (Mayaud et al, 2013). The second element validated that the headset and analysis was capable of reliably detecting a P300 ERP. Finally, some preliminary evidence of a P300 ERP training effect was shown.

A two session study, however, was not sufficient to examine the trend in training over several days. The reversals in this study opened questions in regards to motivation and longer-term practice effects. Given the number of participants in this initial study, the training effect also required replication. This was the purpose of the next study.

# 6 Training Effect Study

The purpose of this study was to identify whether there was a training effect inherent in the use of the P3 speller across all days of the study, and to confirm that the low-cost EEG hardware was robustly detecting P300 ERP signals. Previous studies described in the literature review (Olds, 1963; Roger and Galand, 1981; Aloise et al, 2011) looked at performance within a single session, or at most 2 sessions, so this study was designed to look at progression over 5 sessions within a week to see whether performance improvements continued throughout the training period. The question of internally versus externally provided stimuli was also considered.

## 6.1 Research questions and hypotheses for the training effect study

Two research questions were of primary interest. The first was whether the P3 speller data on its own showed a robust P300 ERP. Was there a difference between target and non-target items that was significant, broadly distributed across electrode locations, and centered in the time range that we would expect a P300 response? Secondly, was there an effect that causes stronger P300 ERP responses on later sessions for a particular subject? Additionally, we were interested in whether there was a difference between prompted trials and those in which the participant provided their own phrase. Anecdotally, the participants enjoyed providing their own phrase, with some even hiding the phrase from the experimenter to prove that the system was actually spelling from their thoughts; therefore, one might expect a higher level of motivation on these self-provided prompts.

The hypotheses were that a robust P300 would be shown, that performance would increase with training, and that self-prompted phrases would result in better performance. Specifically, we hypothesize **(H1) that a clear P300 signal will be present in those trials where a target stimulus was presented, but not in the non-target trials**. Data from the two session study leads to the clear expectations that the hardware can detect a P300 response. Secondly, **(H2), we expect that there will be an improvement in P300 magnitude from earlier sessions to later sessions**. The two-session study showed that a training effect was present from day one to day two, but to extend that, an improvement is expected across all days. Thirdly, **(H3), it is anticipated that there will be stronger performance on free spelling trials than trials where the prompt to be spelled is given to the participant**. This hypothesis is based on the observation that participants enjoy the free spelling trials more than those in which the prompt is given to them. The supposition is that motivation will increase task performance. Lastly, **(H4), it is expected that trials earlier in a particular session will have better performance than those later in the same session**. The P3 speller is a tiring task; on longer phrases, maintaining attention is difficult, and over a long session it is expected that attention will waver.

## 6.2 Methods

### 6.2.1 Participants

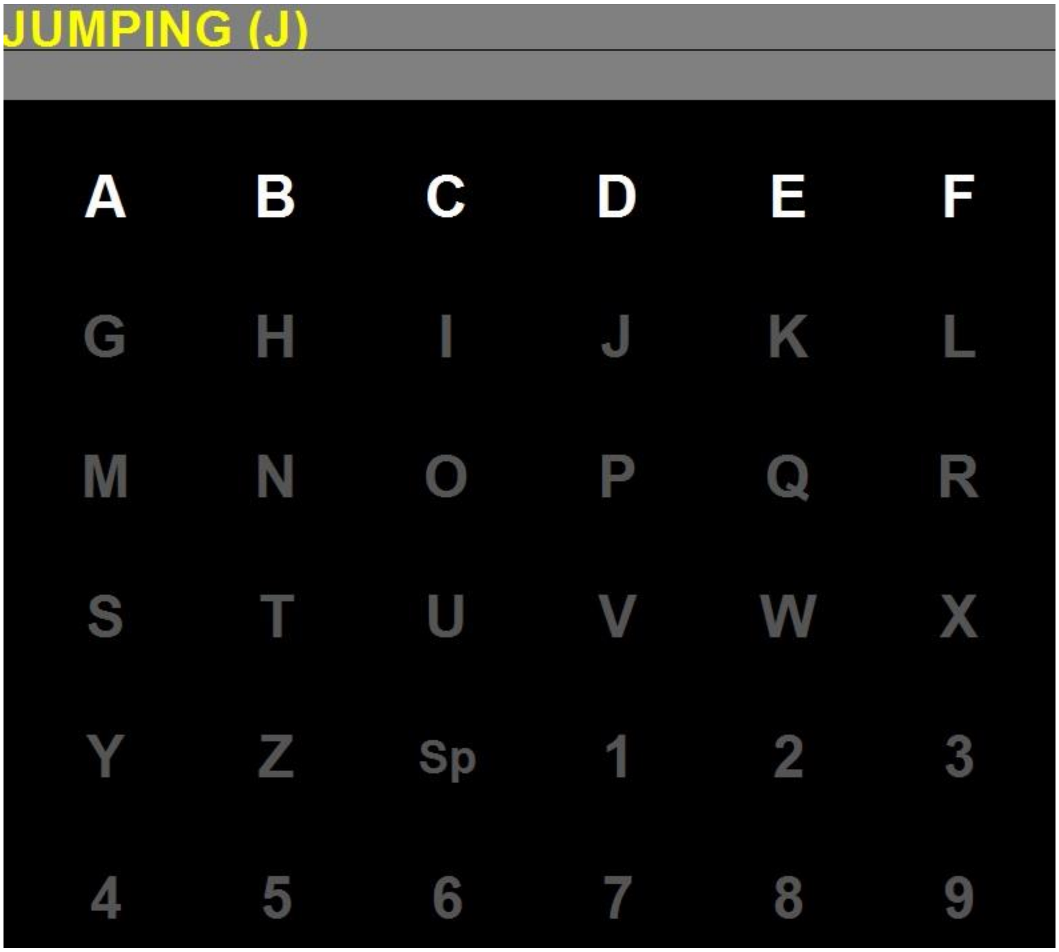
Thirteen University of Victoria undergraduates participated in exchange for experimental credit in their psychology classes, or for $50. One participant became ill and was unable to complete the sessions, and was therefore removed from the analysis, leaving twelve participants. Rejection criteria were established in advance to exclude participants who were unable to achieve consistent performance, in order to enable the classifier to be trained, but in this round of study, no participant had to be excluded. Participants returned for five sessions over the course of seven days. Six of the twelve participants were female. The selection criteria were: an absence of attentional disorders (by self-report), and lack of experience with related EEG studies. All participants were between 18 and 29 years of age and had normal or corrected vision. Vision problems and attentional disorders were assessed by participant self-report.

### 6.2.2 Apparatus

The studies were performed on the Emotiv Epoc with a reversed headset position as described in the hardware chapter.

The P3 speller task uses P300 ERP responses to determine which letter the user is attempting to spell. In the version of the task described here, rows or columns of letters were intensified at a rate of 120 msecs per intensification (62.5 msec intensified, then 62.5 msec pause before the next intensification) as shown in figure 6.1. For each letter to be spelled, the letter was intensified 30 times (15 times in row intensifications and 15 times in column intensifications), alternating rows and columns in a random presentation. The participant was instructed to count the flashes of the letter or number that they were trying to spell, although participants were encouraged to develop their own strategies to maintain concentration if they thought that would be more effective. In the example in figure 6.1, the participant was spelling the 1st letter of the word “JUMPING.” In looking for the P300 responses, the current study used an epoch size of 796 msecs, and since the rate of intensification is every 120 msecs, this meant that the epochs were overlapping. Individual variation in terms of the maximum presentation rate is large, and the current rate of 120 msecs per intensification was chosen during pilot trials to allow most participants to achieve success at the task.

The monitor was at a viewing distance of approximately 80cm, causing it to subtend a visual angle of 34 degrees. Unlike in the two-session study, in this study the distance was not precisely fixed, and the chin rest and eye tracker were not used. The participants did not use a mouse or keyboard, and all interaction with the system was through the EEG.



JUMPING (J)

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| G | H | I | J | K | L |
| M | N | O | P | Q | R |
| S | T | U | V | W | X |
| Y | Z | Sp | 1 | 2 | 3 |
| 4 | 5 | 6 | 7 | 8 | 9 |

**Figure 6.1: Interface for P3 speller task**

As in the two-session study, the sample rate was 128 samples per second, and samples were transmitted in a block size of 4 such that each sample block was 31.25 msecs in duration. This defines the most precise temporal resolution of stimulus presentation possible and stimuli were presented for 2 sample blocks, then a 2 sample block blank interval before the next illumination to reduce visual persistence of the previous illumination.

### 6.2.3 Procedure

The study was broken into five sessions. On the first day, subjects arrived and had the P3 speller task explained to them, then had the headset fitted. Participants were told that each letter would be intensified for a moment, in rows and columns, and that they were to mentally count the number of times the letter they wished to spell turned bright. When participants understood, the first session began. This was with a very short phrase, “BIRD,” in order to get the participant used to the task.

On each day of the study, following the introduction on the first day, the participants spelled 3 phrases ranging from 10 to 15 characters per phrase. A linear classifier training was attempted after the first and third phrase in order to maximize subject accuracy on the task. This training consisted of using the prompted phrase paired with the results generated from the user, and is described in detail in Appendix B when discussing the configuration of the BCI2000 system. This alignment of known correct responses and user-generated responses allows the system to train a linear classifier so as to be able to associate the user EEG inputs and more accurately predict user intention in future trials. This training adjusts the inputs by electrode location and P300 timing to correctly weigh the inputs on a per-user (and per-session) basis.

On each day, a fourth phrase was then given to test the fully-trained model. The subject was then allowed to spell a word of their choosing (free spelling), and write it on paper before the experimenter began the session. A different prompt for that word was given on each of the five days (“favorite month,” “favorite animal,” “favorite city,” etc.).

On day 3 through 5 of the study, as participants were becoming more practiced and requiring fewer breaks between trials, a sixth prompted phrase was given to keep the overall length of the sessions consistent across all 5 sessions. Each session averaged 45 minutes in length.

In addition to the exhortation to try as hard as they could, an additional incentive was provided in the form of a draw for an iPod, with three entries into the draw being provided for each high-confidence correctly identified letter, two entries for each medium-confidence letter, and one entry for each low-confidence letter. This was intended as an inducement to continued concentration, even when accuracy rates had increased to levels where the effort required to obtain correct responses had become reduced.

### 6.2.4 Data Analysis

#### 6.2.4.1 Preparing data for analysis

Although BCI2000 is a very adequate presentation environment, and an effective place to create real-time brain-computer interfaces, it is less effective for analysis of the results. The EEGLab package is a set of Matlab libraries designed for the analysis of EEG data. There are several alternative analysis packages in existence, but EEGLab is open source and very well maintained.

The first issue to discuss is the data format. Although there is an import plugin for BCI2000 data included with EEGLab, version 13.4.4b, it proved difficult to differentiate target and non-target trials, so this was instead done on the BCI2000 side. This code modification is presented in more detail within Appendix B. Additionally, the BCI2000 data format uses markers, which the Matlab code that inputs data into EEGLab must transform into events—in this case, events that are adorned with a property indicating whether they are target or non-target trials. These must be converted into a form that is more usable for EEGLab analysis. The free-spelled phrases also required some special processing since the system did not know in advance which were correct responses. Appendix A presents the Matlab code required to perform these transformations.

The more interesting analytical problem is posed by the nature of the data frames in the BCI2000 P3 speller task. It would be possible, but very inefficient, to design a P3 speller that did not have overlapping frames. It would ensure a presentation rate slow enough that after each intensification the system would wait long enough for the P300 to occur and return to baseline before presenting the next event. Instead of the presentation rate of 125 msecs in the current studies, this might take 600msecs, or even the full 800msecs, which is the epoch length chosen for the data analysis here (see Cohen, 2014, pp75-77 for a full discussion of epoch lengths). Spelling would take much longer in this case, however, it would allow for analyses that are simply not statistically correct in the overlapped epoch case.

#### 6.2.4.2 Analyzing ERP data

A maximum of ten percent of the overall presentations should be target trials (Picton, 1992), although later researchers sometimes use 20% as the cutoff. It is also the case that for many statistical analyses, it is beneficial to equalize the number of trials in target versus non-target events (Cohen, 2014). Therefore, many non-target trials must be thrown away. In the current studies, this was done by first deleting all non-target events that had a target anywhere within its epoch. This reduced the number of non-target events enough to do statistical analyses that don't depend on a strict independence assumption, which is what was done for the current studies. This retained approximately 360 target presentations, and 360 non-target presentations, per trial (for a twelve-character phrase, presented at 30 repetitions per character, as described in the apparatus section).

One of the more powerful analysis techniques in EEGlab is the ability to do independent component analysis (ICA) on the ERP data. This combines the signals from different electrodes and, based on the concordance of the waveforms, calculates a set of computed channels that are independent of each other. However, this independence assumption is not entirely true in the best of times, and is even less true in the case of overlapping epochs, since a portion of the data from one response is being used in another response (Cohen 2014, p. 77). There is a way of correcting this after the fact, but it greatly reduces the amount of data available. Instead of throwing away those non-target trials necessary to equalize target and non-target trial numbers, one can keep only the non-target events that fall outside each other's sliding window of epoch times, and also keeping only those target events that do not have another target event within its epoch window. We would then have retained only those trials, both target and non-target, which were non-overlapping and could perform analyses such as ICA. In the end, it was decided that the vast reduction in data needed to perform ICA properly was not warranted, in that with the reduced sample size, the loss of power in the analyses was a greater detriment than the benefits of the analysis technique. It was later noted that the ICA decomposition could have been performed on the data prior to epoching, but this was not done.

The trials that were retained were used in a magnitude analysis of the P300 signal. First, artifact rejection took place to reject trials that were disrupted by noise, often caused by head movements. These steps are best presented with the data itself, so are described in detail in the study results section. After rejection, the remaining trials were grouped into non-target and target trials, grouped by session, and then an average was taken of the signal amplitude across trial epochs in each group. At each time-point, non-target trials were subtracted from target trials to show only those average signal changes which differed between conditions. These averages were plotted, and statistics performed using an ANOVA to show overall differences, with a *t*-test performed between individual pairs of sessions in a post-hoc comparison and applying a correction to account for multiple comparisons. The multiple corrections test chosen is the Holm’s correction, which is recommended and included as part of the EEGLab package (Delorme and Makeig, 2004).

# 7 Training Effect Study Results and Discussion

## 7.1 Electrode locations and artifact rejection

It is necessary to revisit several issues pertaining to the Emotiv Epoc headset. As described in the hardware chapter, the headset is limited in terms of the electrode locations that it can access. In particular, the location where we would expect to see the strongest P300 response is actually where the reference electrodes sit. Consequently, the effects reported are actually looking for a negative deflection from the electrodes furthest away from where where the strongest positivity is seen. More discussion of localization is found in section 7.2.1.

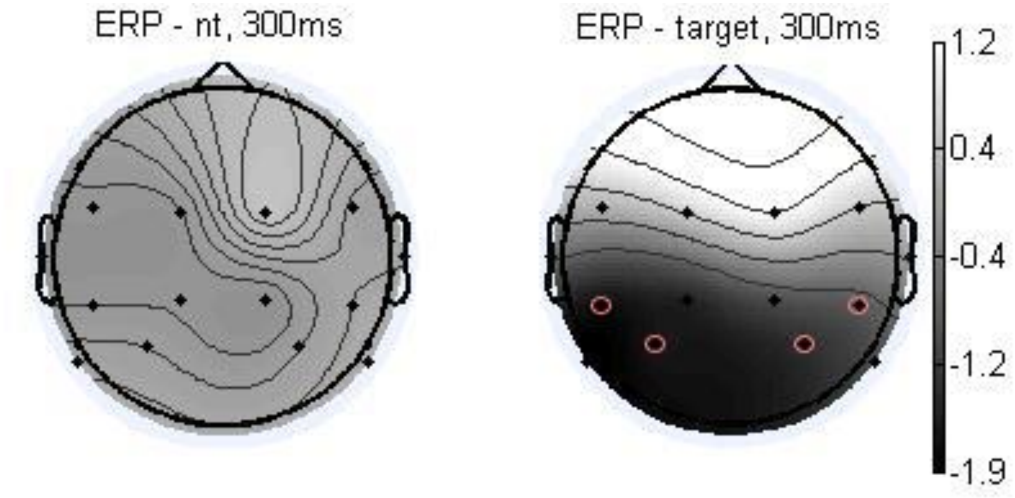


**Figure 7.1: non-target (on the left) and target trials (right) at 300 msecs into each epoch showing the pattern of electrode activation. The parietal electrodes (at the bottom of the figures, circled) show a negative deflection as compared to the reference electrodes. This data is from the 4th phrase of each session and the scale is in microvolts.**

The next Emotiv issue to be looked at was the quality of data coming from the electrodes. There were two primary problems. The first was drift. This is the tendency of the system to return values that are not directly comparable. Since we are subtracting the non-target trials from the target trials, this drift is largely controlled for, though if there is too much time between those trials, or if the rate of drift is too high, then the trials are no longer directly comparable and drifting trials must be rejected. The next issue was that of point source irregularities. These can be most commonly caused by movement, although as the sensors dry out, the artifacts become more frequent, so the electrode couple also plays a role. The data quality issues are individualized—one person may have excellent data quality, while another (or the same person on a different day), may have data that is almost unusable. The careful attention to electrode fitting at the start of each session helped in this matter, but was not a complete solution.

As an example of the artifact rejection steps, the following figures show data from the third phrase of each day and are displayed in a visualization where each response draws one line on the chart. This is not an effective visualization to show final results, but gives a good indication of the variability of the data. The other element of these charts to note is the scale of the data, in microvolts; though the charts show -1000 to +1000 microvolts, our effect is in the 1 to 2 microvolt range. This is due to the amount of noise in the signal, which is obscuring the signal of interest. In this section we will show how this noise is removed.

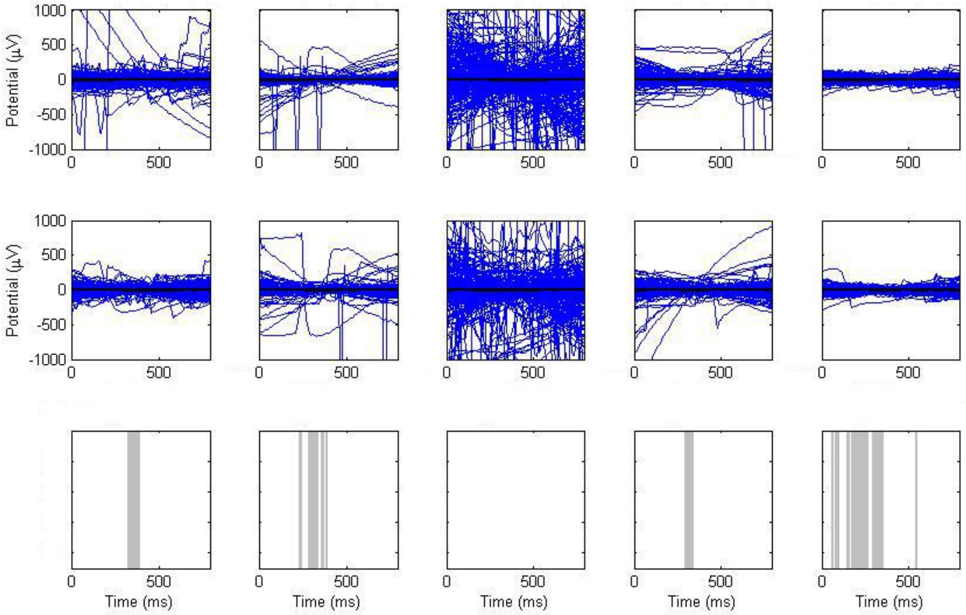


**Figure 7.2: uncleaned data for non-target trials (on top) and target trials (in the middle), showing days 1 through 5 (left to right) for the parietal electrode locations: P3, P4, P7, and P8. The bottom charts show regions where targets are different from non-targets at a significance level of *p*<.05 using a Holms approach to multiple measures correction.**

As shown in figure 7.2, the large variability of results on day 3 has entirely prevented the observation of a significant difference between target and non-target items. In addition, figure 7.3 shows the same chart organization with the 3 participants with the greatest variability removed from the analysis.

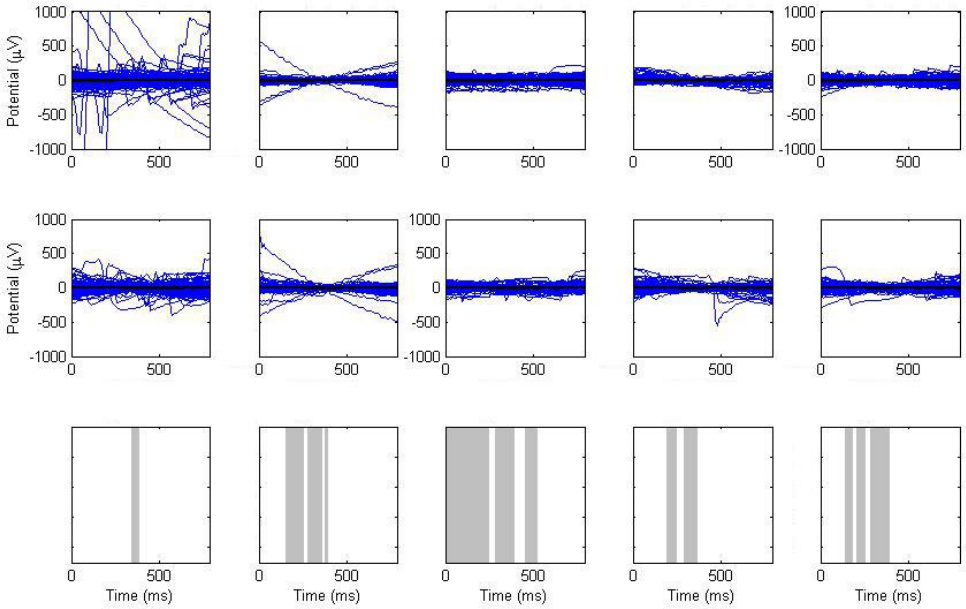


**Figure 7.3: With participant numbers 2, 6, and 8 removed. Non-target trials (on top) and target trials (in the middle) showing days 1 through 5 (left to right) for the parietal electrode locations. The bottom charts show regions where targets are different from non-targets at a significance level of *p*<.05 using a Holms approach to multiple measures correction.**

Removing a quarter of the participants due to an observation after the fact that their data was of poor quality is not an ideal solution. As described in figures 7.4 and 7.5 an alternate solution was therefore used which rejected data on a trial by trial basis since even in those participants, there are many trials where good data was present. This alternate rejection strategy ended up removing approximately 12% of trials rather than the one quarter which the naïve rejection approach would have caused. Electrode drift appears to be the largest contributor to the data quality problems, so some systematic rejection of individual trials was performed to minimize drift. Figure 7.4 shows the same data as in figure 7.2, but with the trials removed in which a drift of more than 50 microvolts over 50 data points was observed. There are 160 data points in an 800 msec epoch, so this corresponds roughly to a drift of 150 microvolts over an epoch, where our effect size is between one and two microvolts. The rejection technique is a standard one (Cohen, 2014; Luck, 2006), but the rejection values are dependent on the specific study, and these rejection parameter values were derived empirically based on the size of the effect versus the noise in the data. This rejection criterion varies by subject and trial as to how many epochs are rejected, but results in the removal of approximately 10% of the epochs.

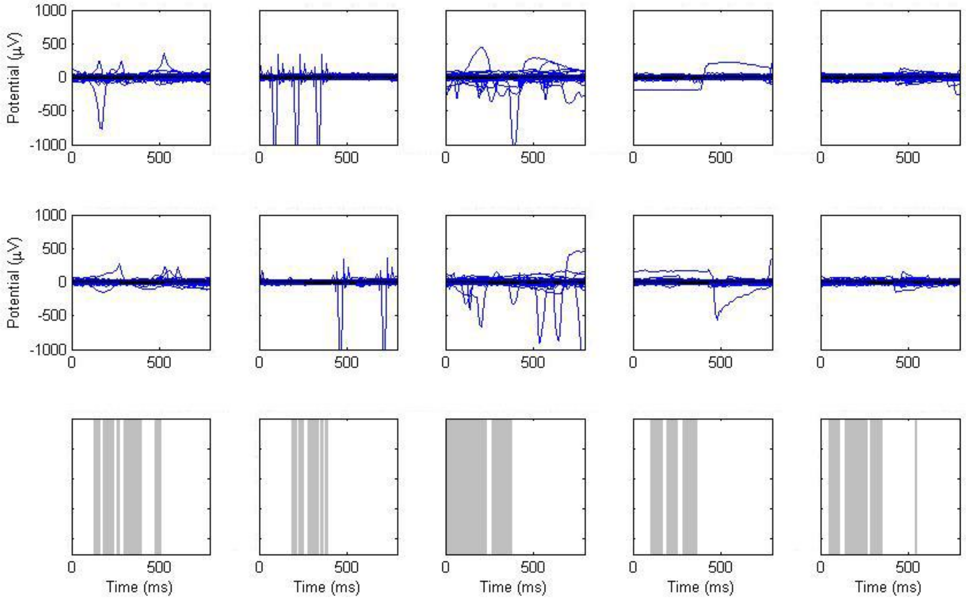


**Figure 7.4: Data with drift rejection for non-target trials (on top) and target trials (in the middle) showing days 1 through 5 (left to right) for the parietal electrode locations: P3, P4, P7, and P8. This data has been processed using the drift-rejection technique mentioned in the text. The bottom charts show regions where targets are different from non-targets at a significance level of *p*<.05 using a Holms approach to multiple measures correction.**

This data is far cleaner than the initial set, but note that in figure 7.4 there are still a large number of spikes in the data. The second step in individual trial rejection is to remove data points that are statistically improbable. In this case, through empirical testing, it was decided to remove all those trials whose results differed by more than 3 standard deviations from the norm (see Luck, 2006, pp.151-173 for a complete discussion on choosing rejection thresholds). Figure 7.5 shows the results of this, added sequentially to the drift rejections previously discussed. This rejection criterion resulted in the removal of only a few trials in some cases, none in others, and up to 2% of epochs in others.

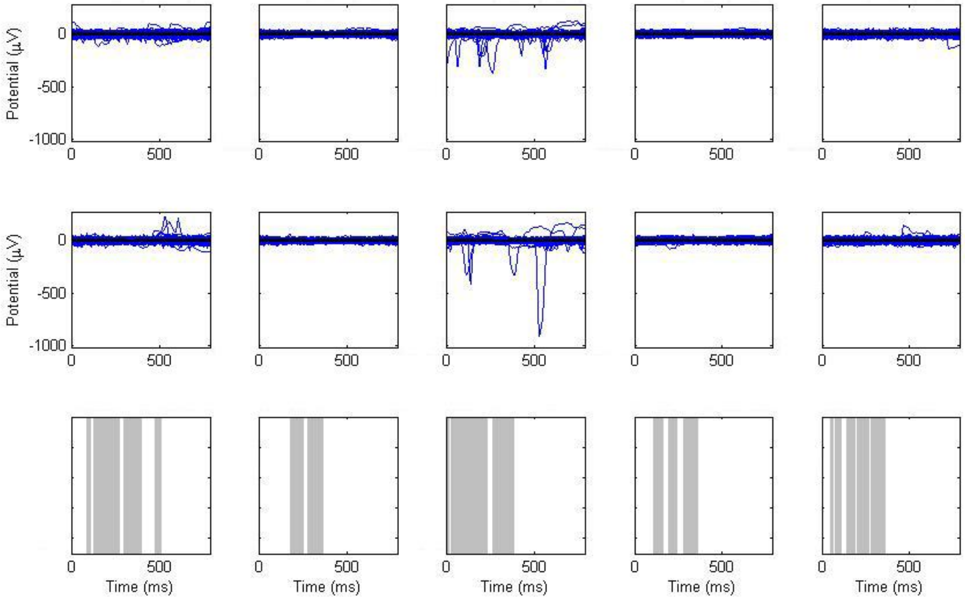


**Figure 7.5: Cleaned data for non-target trials (on top) and target trials (in the middle) showing days 1 through 5 (left to right) for the parietal electrode locations: P3, P4, P7, and P8. The bottom charts show regions where targets are different from non-targets at a significance level of *p*<.05 using a Holms approach to multiple measures correction.**

There were still some problematic trials, particularly from the 3 participants previously noted on day three, though the data is significantly more analyzable than before rejection measures were taken. The details of this trial rejection process, as well as the rest of the Matlab code for the data processing, are presented in Appendix A.

## 7.2 P300 ERP results

The final summarized data for phrase 3 appear in figure 7.6 (top). Note that this presentation format, where individual responses are averaged into a single line but each session is presented separately, and where target and non-target trials are presented on the same chart, is how the remainder of the magnitude data will be presented. The ERP data is shown in each trial in the study, and is apparent in many of the results in following sections, so phrase 3 is selected here as an example of results that appear throughout the study.

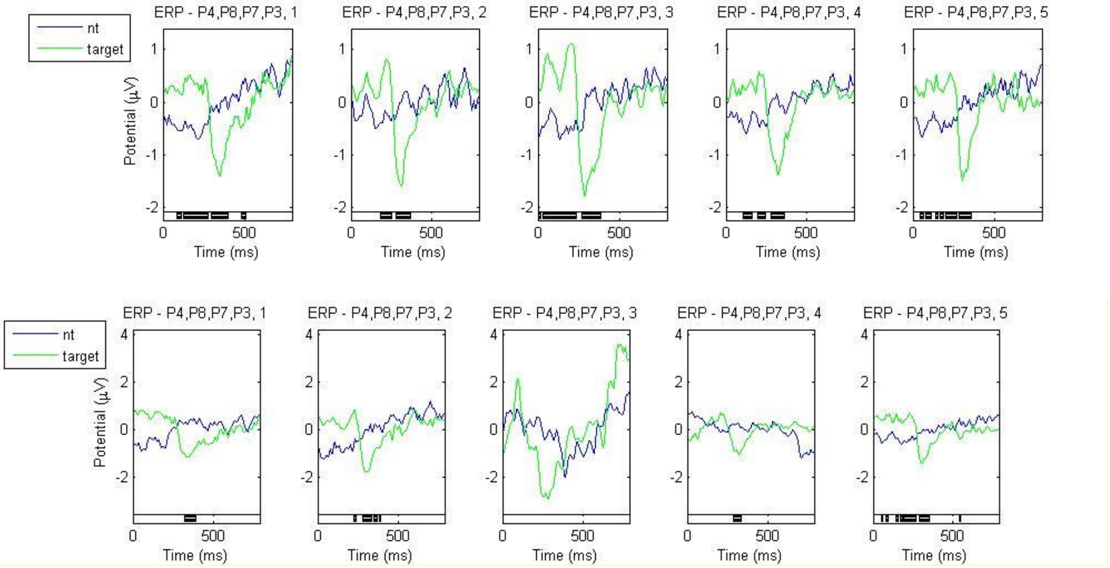


**Figure 7.6: cleaned data from phrase 3 is averaged across participants and presented by session (left to right) and with target and non-target trials shown on the same chart (top is cleaned data). Black bars at the bottom are regions where the target trials are different from non-target trials at a significance of *p*<.05 with a Holms correction applied. The original uncleaned data is shown in the bottom series of charts.**

Figure 7.6 compares the cleaned data (top) to the same data prior to the artifact rejection steps (bottom). Notice that the same trend is still present but that the existence of a high amount of random noise greatly reduced the size of the effect that we were able to observe. Note also the effect of this random variation on the areas that reach a *p*<.05 significance.

On the cleaned data in particular, it is observed that there was a peak at approximately 300 ms post-stimulus presentation, which is present in the target trials but not in the non-target trials.

### 7.2.1 P300-like ERP localization and validation

Similar to the two-session study presented in section 5.1, this study was also re-analyzed using a common average reference. As previously mentioned, this does have the effect of attenuating the P300 signal shown, but helps to localize the response as compared to a single reference electrode at C3/C4. The frontal electrode locations shown in figure 7.7 show a P300-like response, and unlike the analyses presented in the majority of this chapter it is a positive deflection as compared to the negative dipole used in most of the analyses.

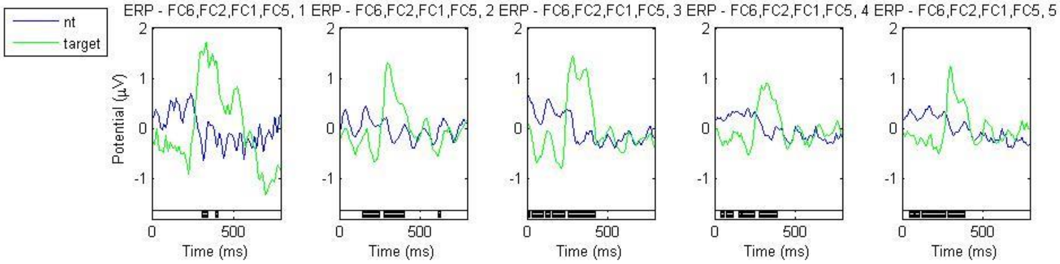


**Figure 7.7: Frontal electrode locations for phrase 3 with common average reference (CAR)**

The next analyses to be performed in examining the localization of signals was to infer a Cz electrode. This was done in EEGLab using the surrounding electrodes to calculate what the Cz response would have looked like if there had been an electrode placed there. This does have the effect of introducing some noise over what would have been seen with an actual electrode at that location (Cohen, 2014) but gives an approximation of the expected signal. In this case, there was not a strong signal seen at Cz, as is shown in figure 7.8, but some positivity was seen at the predicted time.

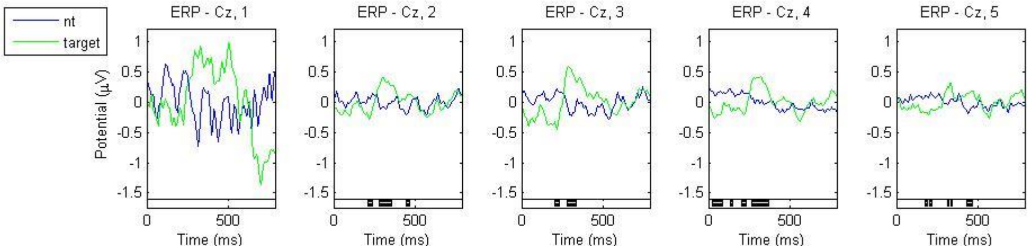


**Figure 7.8: Interpolated Cz electrode location for phrase 3 with common average reference (CAR)**

The same interpolation technique was used to infer a signal at Pz. In this case, the signal seen was a negative deflection, just as seen in most of the analyses using the parietal locations presented in this chapter.

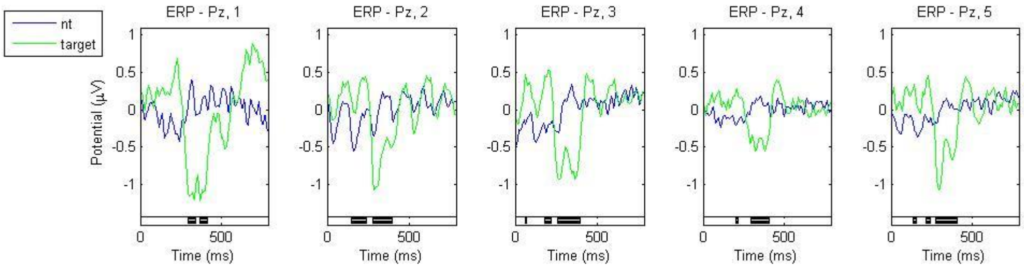


**Figure 7.9: Interpolated Pz electrode location for phrase 3 with common average reference (CAR)**

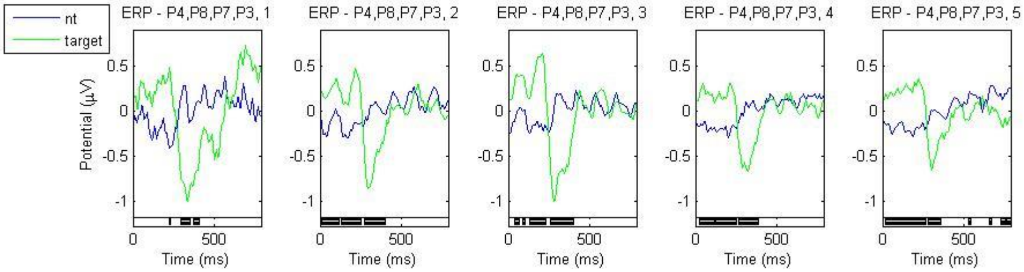


**Figure 7.10: Parietal electrode locations for phrase 3 with common average reference (CAR)**

The last analysis presented in this section is the average of the four parietal electrodes, but this time shown using the common average reference rather than the original reference electrodes at C3/C4. This is presented in figure 7.10. In comparing with the results previously shown in figure 7.6 it can be observed that very little difference in the resultant signal is shown when using a common average reference rather than the original C3/C4 reference location. There is certainly some attenuation of the signal but the pattern and timing remains largely the same.

## 7.3 ERP magnitude across sessions

Phrase 3 was the final phrase prior to the linear classifier being fully adapted in each session. The first three phrases were also used for linear classifier adaptation, but it was not until phrase 4 that participants achieved their best performance of each session (with ‘best performance’ referring to letter error rate, not necessarily in terms of overall P300 magnitude, though the two are related). In particular, the feedback of seeing words properly spelled was inherently rewarding, which is important to note as motivational factors appear to be a significant aspect of this task. In figure 7.10, for phrase 3, the basic training effect is visible; we will now examine that effect in more detail for phrase 4.

Figure 7.11 presents the cleaned data for phrase 4. We can see that there is a distinct difference between day 1 and day 2 performance, though not as significant on subsequent days. This pattern, with some variations, is true with each of the phrases.

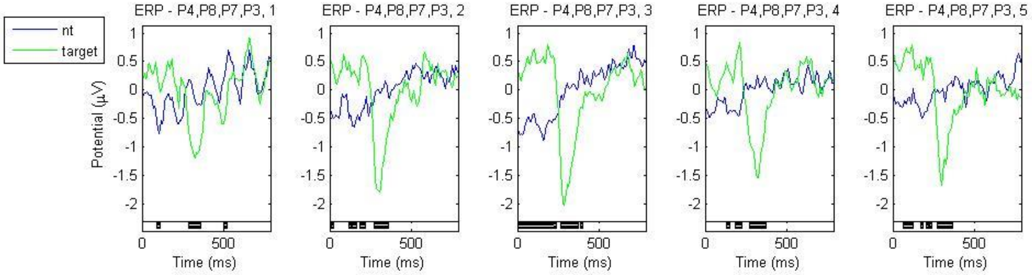


**Figure 7.11: Target versus non-target trials for the fourth phrase on sessions one through five. The black bars represent a difference at p<.05 with a Holms correction applied.**

Figure 7.12 shows all 5 sessions of target trials, along with the results of an ANOVA testing magnitude of response at each time point by session, showing where statistically significant differences occurred. The black bars at the base of figure 7.12 show the results of a post-hoc comparison using t-tests to determine which individual pairs of days showed significant differences. Figures 7.13 and 7.14 show individual pairs of these comparisons, to emphasize where differences exist between days. In the session 2 versus session 3 chart (figure 7.13), we can see that there is no statistically significant difference. The day 3 versus day 5 graphic actually shows a reversal of the trend predicted (figure 7.14), with day 5 showing a lesser magnitude P300 than day 3. Although not shown in these figures, in no cases was there a significant difference between non-target trials, as we would expect.

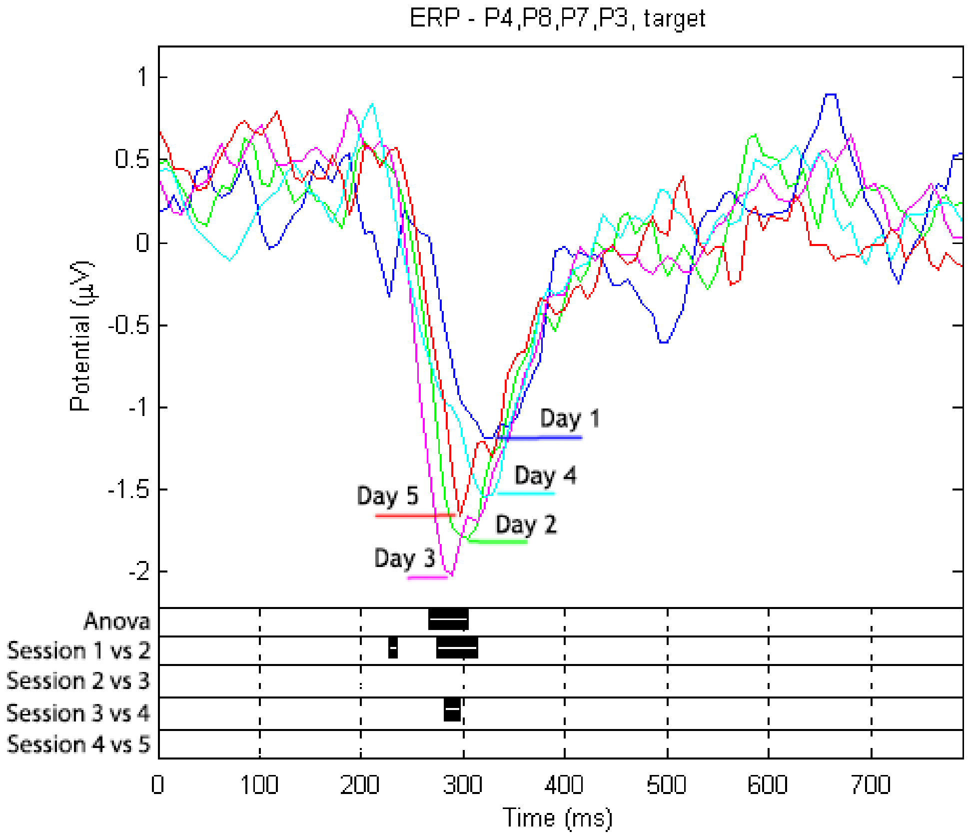


**Figure 7.12: Target trials for the fourth phrase on the five sessions. The black bars represent a difference at *p*<.05 with a Holms correction applied as calculated by an ANOVA followed by post-hoc t-tests to show which days were different from each other.**

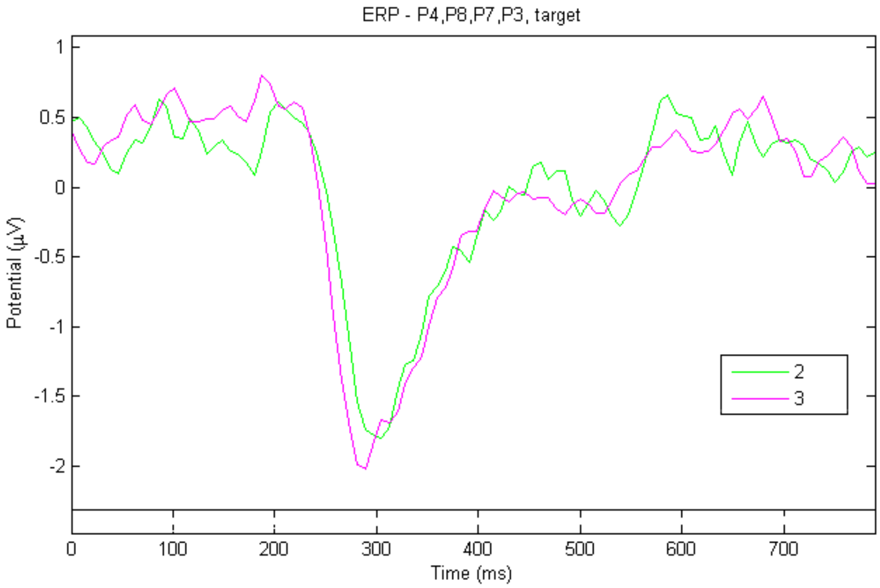


**Figure 7.13: Target trials for the fourth phrase on session two versus session three. The black bars represent a difference at p<.05 with a Holms correction applied; in this case there were no significant differences.**

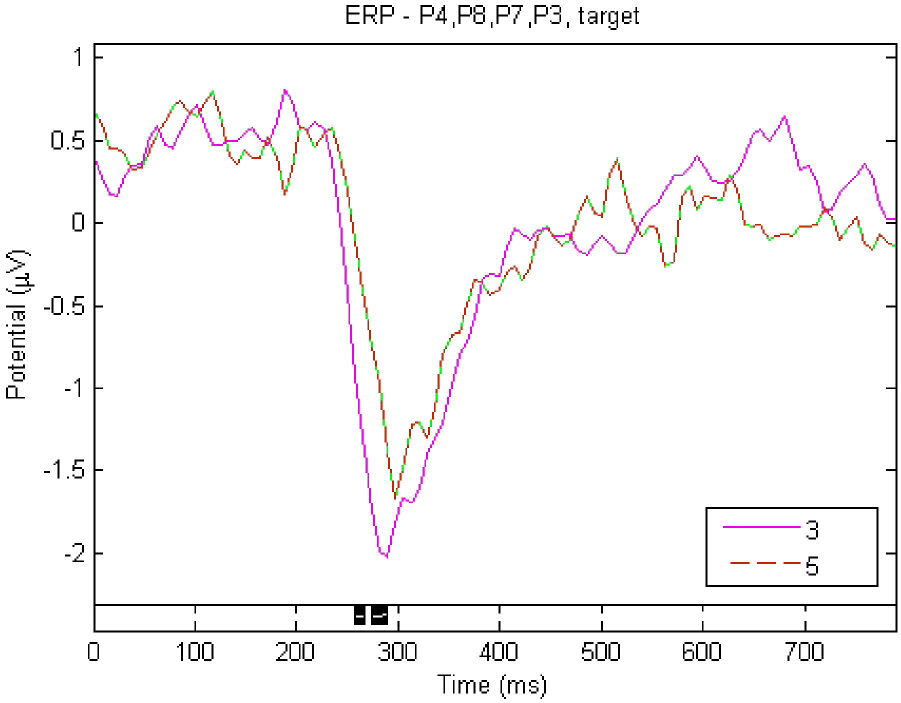


**Figure 7.14: Target trials for the fourth phrase on session three versus session five. The black bars represent a difference at p<.05 with a Holms correction applied.**

Similar results, with day 3 showing a significant improvement over day 1, were found for phrases 3 and 5. These results are shown in figures 7.15 and 7.16.

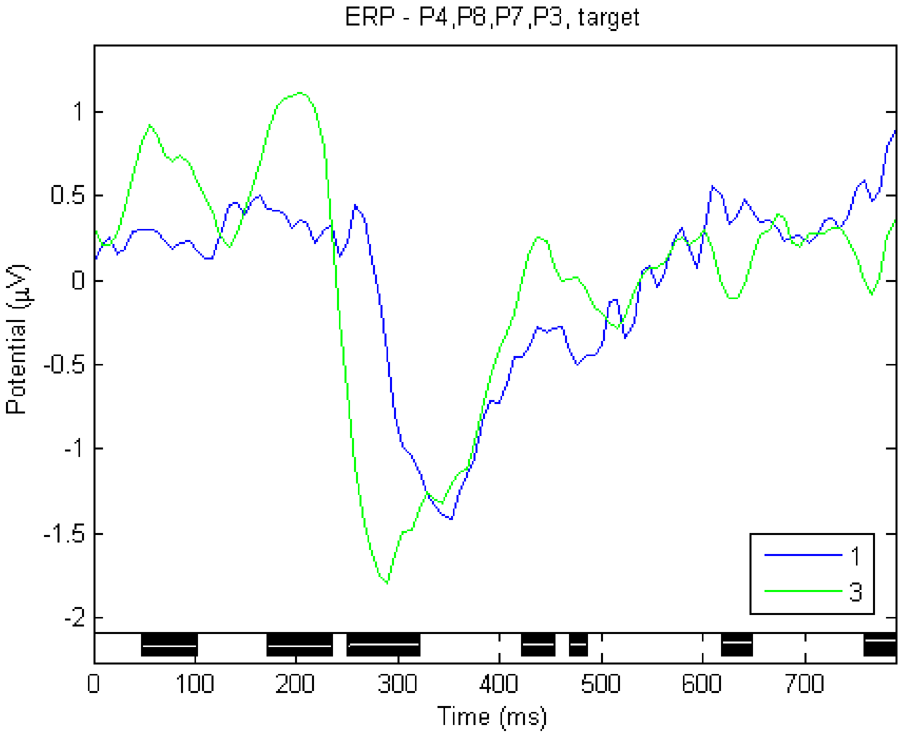


**Figure 7.15: Target trials for the third phrase showing a significant improvement from day one to day three. Black bars are regions that are significantly different at a p<.05 with Holms correction.**

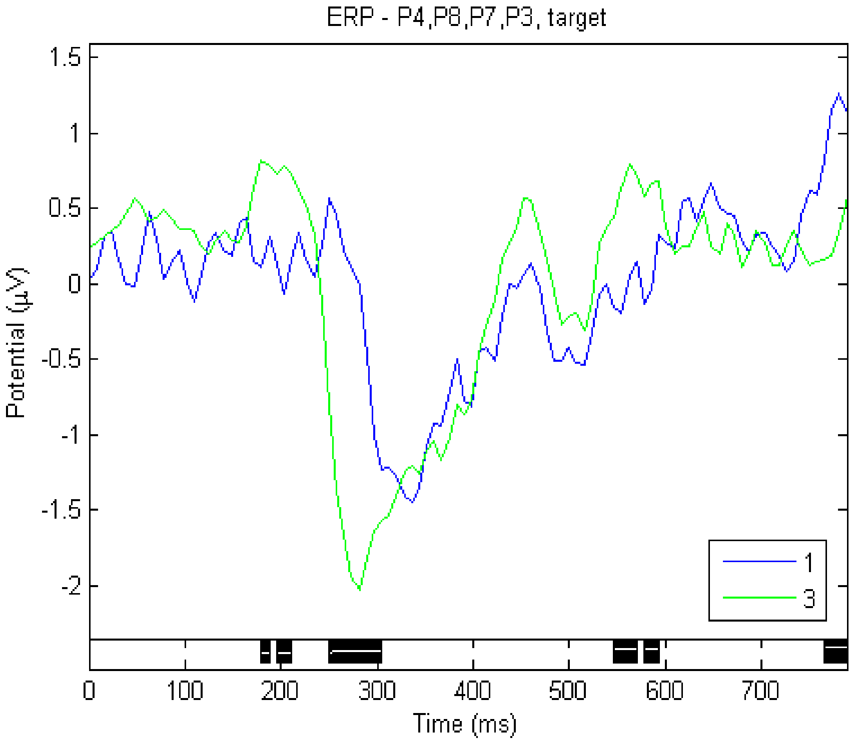


**Figure 7.16: Targets for the fifth phrase, provided by the participant, showing a significant improvement from day one to day three. Black bars are regions that are significantly different at a p<.05 with Holms correction.**

These effects are complicated by the fact that the peak of the P300 appears to be slightly earlier on day 3, so is not the precise peak being compared in the statistical analysis. In addition, in phrase one, where no such P300 magnitude difference exists but where the peak is earlier, the regions are judged to be statistically distinguishable, as shown in figure 7.13.

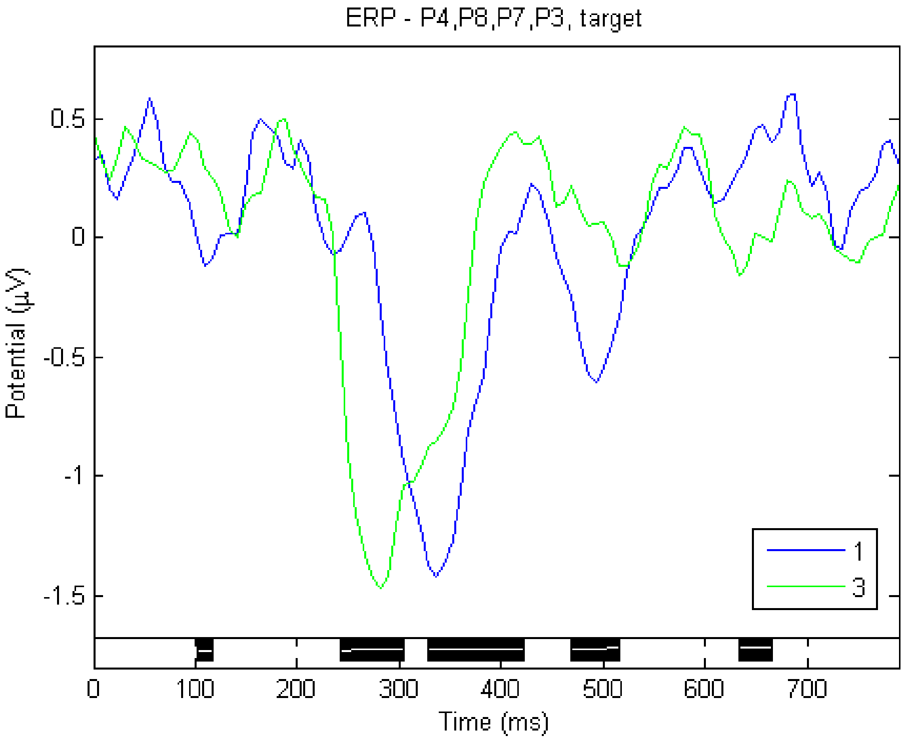


**Figure 7.17: Targets for the first phrase showing a significant difference from day one to day three but with no magnitude difference of the peak. Black bars are regions that are significantly different at a p<.05 with Holms correction.**

Since the previous figures did not look at peak alignment, it is useful to look at the maximum P300 peaks. Figure 7.18 presents data from each individual participant, in the form of the peak P300 value for each session. This peak value is shown with the corresponding time point in the non-target trials subtracted so as to roughly correct for baseline. This is also the cause of some negative readings in these peaks, where the P300 peak is minimal enough that random variation in the non-target data overpowers it. Despite this noise, it is possible to see the general trend wherein the second and subsequent session peaks are stronger than in the first session. This trend is even easier to see on the grand average results in figure 7.19

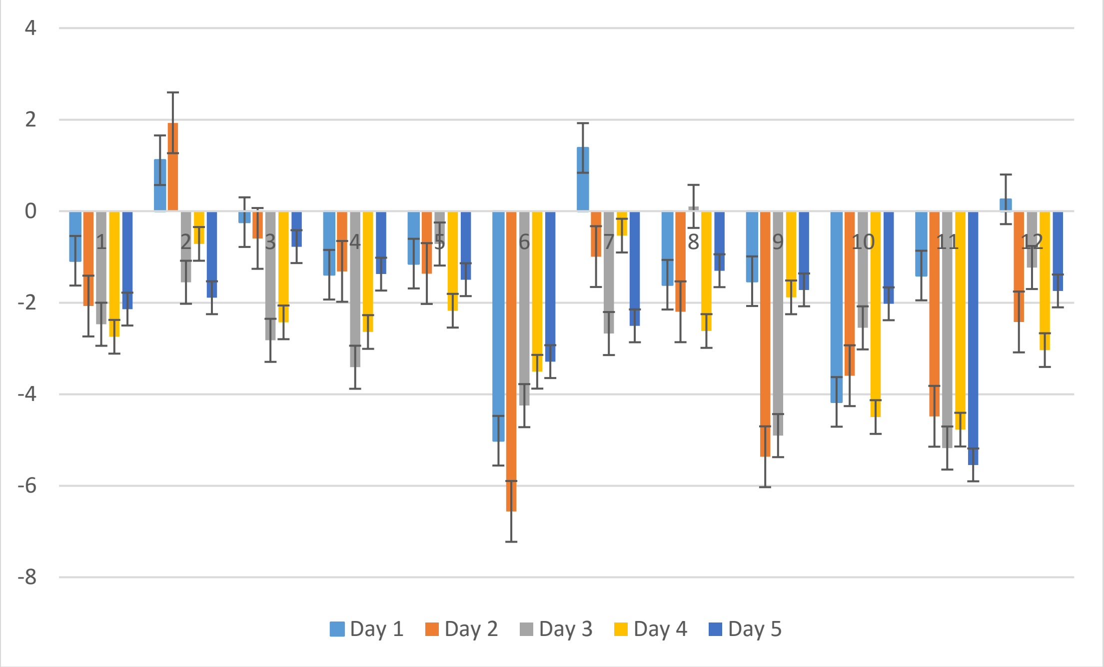


**Figure 7.18: Mean peaks picked from target trials on phrase 4, showing each of the 12 participants with non-target trial means subtracted, showing peak voltage in µV. The few positive results are present because random variations in non-target response overpowered a weak P300 response. Error bars show standard error.**

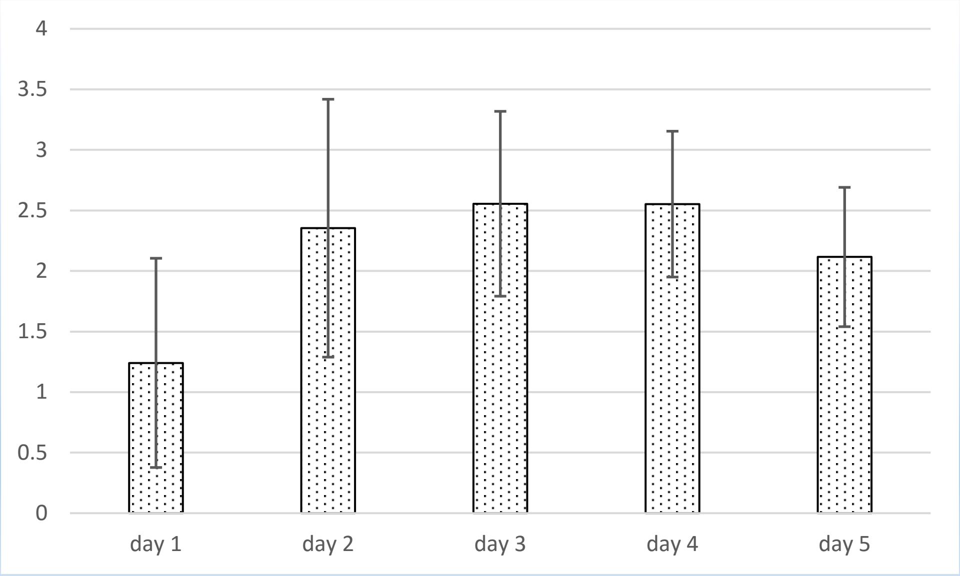


**Figure 7.19: Grand average P300 magnitude in microvolts by day on phrase 4 error bars represent one standard deviation**

## 7.4 User generated prompts

In phrase 5, the participants chose their own text to spell. From the pilot data, this was known to be intrinsically interesting to participants. In the first session, it was noted on several occasions by participants that they did not believe that the task was working, until they participated in this user-generated prompt task. Participants were told during the break when the experimenter was doing the last linear classifier training to write down a phrase on the questionnaire sheet. Two of the participants made a game of actively hiding the questionnaire sheet from the experimenter so that they were the only ones who knew what word they were attempting to spell, ensuring that the system was not rigged. From this intrinsic interest, it was expected that motivation, and therefore performance, would be increased. As shown in figure 7.20, it was observed that the general pattern in terms of P300 remained, and an improvement from session one through session three can be seen.

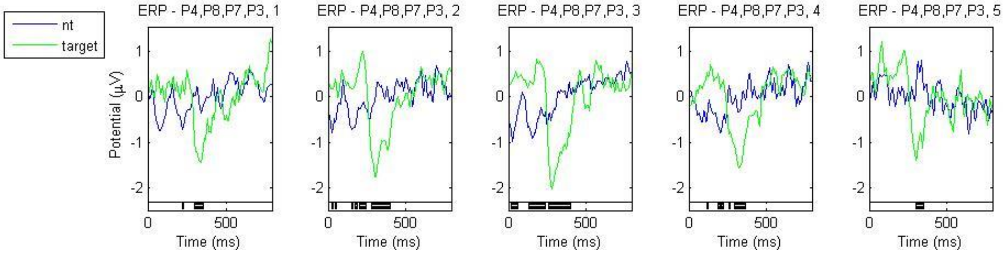


**Figure 7.20: Target versus non-target trials for the fifth phrase, provided by the participant, on sessions one through five. The black bars represent a difference at p<.05 with a Holms correction applied.**

In figure 7.21, the fourth phrase was compared with the fifth phrase. Rather than being prompted for a specific phrase, the fifth phrase was supplied by the participant. Although the patterns of results were very similar, no statistical difference was found between the performances on these two phrases.

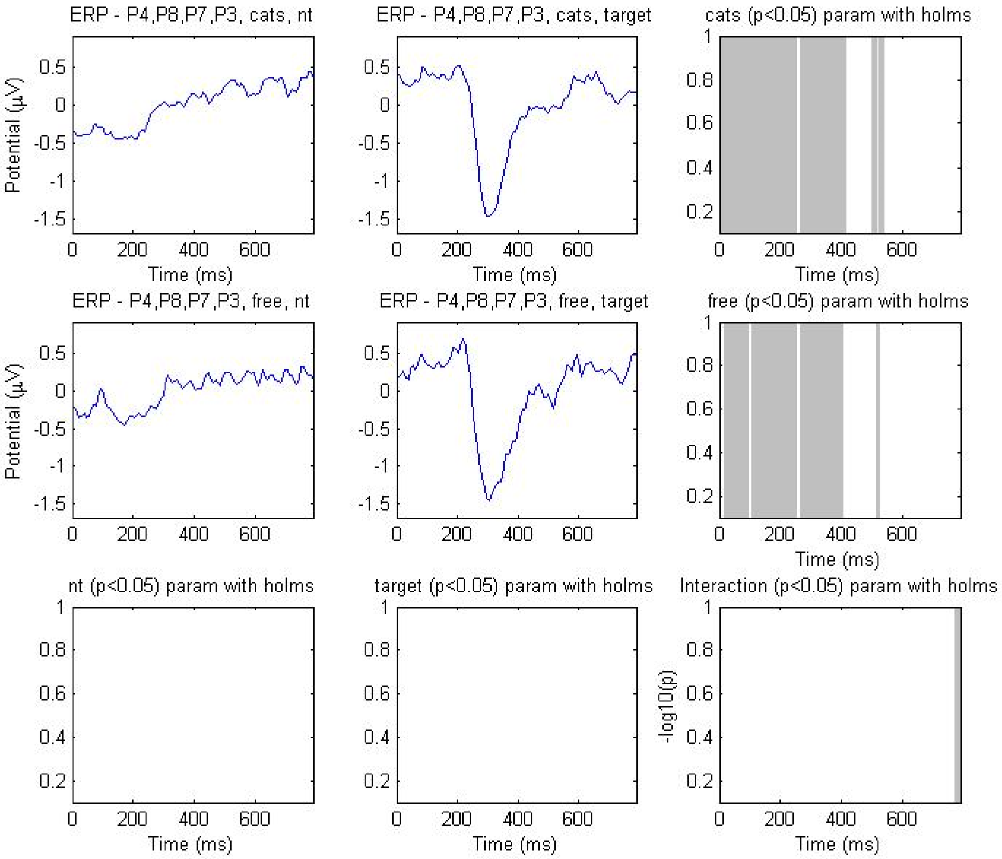


**Figure 7.21: The fourth phrase (top) is compared to the fifth phrase, supplied by the participant. Shown with non-targets on the left and targets on the right. In each case, there was a significant difference between target and non-target, but the two phrases were not statistically distinguishable from each other.**

## 7.5 Fatigue

Sessions were kept short at under an hour, with the expectation that in such a repetitive task, a large fatigue effect would be present. In fact, the reverse was shown to be true, with performance on later items marginally greater than performance on earlier items, as shown in figure 7.22. A confounding factor is that on the first items, the linear classifier was not yet fully trained, causing the letter error rate to be much higher, which participants appeared to find demotivating. Through self-reporting, participants indicated that it was easier to focus when they were getting letters correct than when the system appeared to be choosing letters randomly. Figure 7.22 compares performance on the first phrase and the fourth phrase. It was observed that the magnitude of the P300 response was larger on the fourth phrase, and that this difference was small but statistically significant.

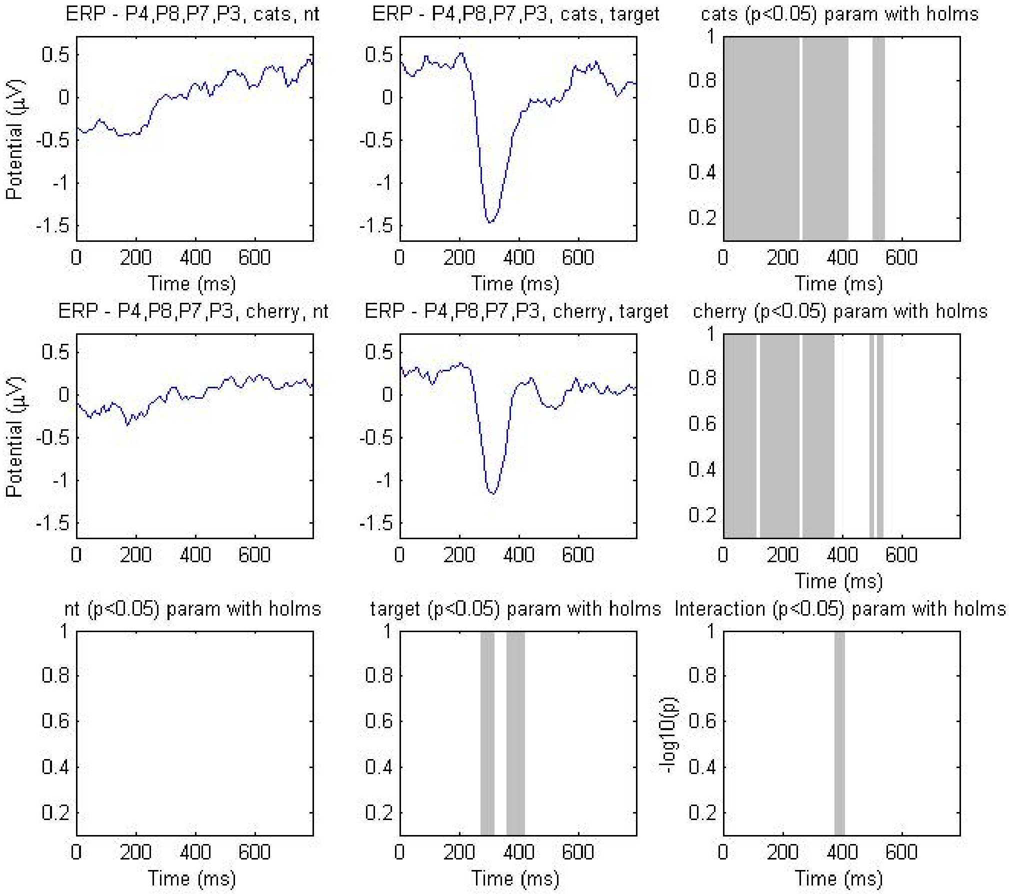


**Figure 7.22: The fourth phrase (top) is compared to the first phrase. Shown with non-targets on the left and targets on the right. In each case there was a significant**

**difference between target and non-target, and the P300 magnitude was greater in phrase 4. T-test significance results are shown in the panels on the bottom and left at a value of p<.05 with Holms correction applied.**

## 7.6 Questionnaire, characters-correct and self-report data

Participants were asked to fill out a questionnaire on each day to assess mood, caffeine consumption, and sleep quality from the night prior to the session. Complete questionnaires and responses are shown in Appendix D. The questionnaires were given each day, in part to assess changes over the course of the study, but also as a cue to trigger discussion with the participant regarding comfort level and what strategies they were employing.

The analysis of correct responses is of secondary importance in this study, since it is primarily designed to show statistical significance in the EEG responses where we have many more repetitions of each ERP event. There was, however, an increase in accuracy from the first to the fifth session in terms of accuracy of recognition. On the free-spelled phrases in the first session, only 64% of the characters were correct, on average, whereas on the 5th session, 78% of the characters were correct, which is a significant difference at the $p<.05$ level.

## 7.7 Training study discussion

This five-day training study served several roles in this line of research. First, it was a replication of the training findings of the two-day study, and the successful capture of the P300 signals on which that depends. It also showed that self-generated and prompted trials gave P300 ERP results and provided data on fatigue. These issues are discussed in greater detail in the sections below.

### 7.7.1 P300 ERP and training effects

With a dozen participants, this second study provided a replication to the training effects seen in the two-session study. The reliable capture of P300-like ERP was robust in this study, as well as all other studies in this dissertation. This was consistent with the first hypothesis (H1) from this study, and the first research question presented in the introduction (RQ1), which stated that a P300 ERP would be observable in those trials containing a target stimulus.

In regards to the specific pattern of training, it both answered some questions and raised others. The early training from session one to session three was strong and consistent, which supported the second hypothesis (H2) stating that the P300 magnitude would improve in later sessions. However, the session four and five P300 results actually showed reduced magnitude from the third day, on average. This is puzzling from a pure training perspective, but can be explained by motivational factors. As the training sessions went on, the novelty wore off the task and subject performance dropped. In a long-term training program, this would counter the goals, and is part of the motivation for the more cognitively-challenging task discussed in the next study.

The individual variation in this training pattern was also consistent with a motivational explanation. In particular, among those participants who had a high level of performance, there was a greater likelihood that they would not see a training effect. This can be seen in figure 7.14, where those participants who showed a large P300 magnitude in their initial sessions had a higher propensity to show a reduced P300 magnitude on later sessions. This is also consistent with observational data showing that the participants who were achieving high levels of performance quickly became bored (although this was collected only informally through user comments rather than systematically, so deserves future study).

### 7.7.2 User generated phrases

The hypothesis (H3) stating that free-spelling trials would provide a larger magnitude P300 response than prompted trials was inconclusive. Subjective reports on the self-prompted phrases clearly indicated that subjects enjoyed these trials more, along with the fact that subjects hid their written answers from the experimenter to ensure that the test was not biased. This indicates that these questions had a higher degree of self-motivation than prompted phrases. However, the performance data did not show these answers to be statistically distinguishable from the prompted phrases. With only one self-prompted phrase per session, it is possible that the sample size was not large enough to show a significant difference.

Although there was no performance benefit for user-generated phrases, it is important to note that there was also no performance detriment. In fact, the P300 ERP results in this study show no evidence of being effected at all by the cognitive processes surrounding phrase generation. This is useful in considering how to create tasks with longer-term interest.

### 7.7.3 Fatigue, medications and user strategies

The hypothesis (H4) that fatigue would make earlier phrases within a session show better performance than later in the session was also inconclusive. In fact, the reverse was shown, although with a very small effect. The confound is present that training works in opposition to fatigue. There is also a confound in motivational factors, since the linear classifier was not yet fully trained for the earlier phrases, and participants reported trying harder when the system appeared to be identifying the intended selection.

Informally collected data from the interviews hints at the possibility of anxiety medication or Benadryl having an adverse impact on P300 response. Both of these comments were from two participants who performed much worse than usual on the particular session where they had mentioned having taken medication. None of this information is anything but anecdotal in this study.

The information regarding what strategies participants were using was also relevant to motivation, but was not at a level whereby it could be statistically analyzed. All participants were initially told to count the number of times the letter they were trying to spell was flashed. This was intended to cause them to retain focus and notice if their attention was wavering. Some participants continued with this strategy throughout the entire study, while other participants changed their strategy as they progressed through the study. Three participants felt too rushed while counting and instead began silently repeating the letter to themselves each time it flashed. One participant was learning Chinese and found it more interesting to repeat the letter in Chinese. One participant was interested in the factors that increase P300 responses and decided that he could increase the emotional valence of this response by thinking of swear words that started with the letter he was trying to spell, then repeating those words silently to himself. One participant decided that singing was more interesting and had a complicated scheme to change notes with each flash. All of these other strategies were carried out silently, without head motion, as participants were instructed that speaking or motion could cause EEG artifacts. These strategies were of inconclusive effectiveness, but did appear to allow the participants to find approaches which they felt would hold their interest better than the counting strategy that was initially suggested to them.

# 8 Memory attention study

One of the problems with the P3 speller task is that it is not sufficiently cognitively demanding to maintain interest, causing participants to get bored easily. In the task's original incarnation as an assistive device, this was not an issue, as the users of the paradigm were attempting to input text and had no other means to do so. When necessity no longer drives the use of the interface, however, boredom is a significant detractor to performance. In an attempt to use the P3 speller for neurofeedback training, this presents a concern with adherence to the training protocol, particularly if it were to be used outside of a supervised setting. The solution is to make the task more interesting. The approach taken in this study was to increase the cognitive demands of the task while still requiring the same type of focused visual attention as in the P3 speller. The purpose of this brief study was to establish that the alternate card game task would generate P300 responses similar to the previously described P3 speller. This is research question RQ3 from the introduction.

There are several possible approaches that can be used to make the task more interesting and cognitively challenging. Gamification of the current speller interface is one such approach, and a primitive form of this involves telling the user how many correct responses they are getting at any point in time. In the training study, participants competed with themselves from day to day to optimize their performance. Some known techniques take a cue directly from video games, such as scoreboards and audio visual feedback for correct answers, making this self-competition more prevalent. The approach taken in this study, however, was to make the task more cognitively difficult: to add a memory component to the task that required the user to concentrate not just on the item that they wanted to select, but to think about which item that was. The vehicle for this was the card game *Concentration*, also called *Memory*, in which a set of cards were initially shown face-up for the user to study. After a period of time, the cards were turned face-down, and the participant was instructed to select matching pairs to turn them back over. The game ended when all the pairs were matched. In this paradigm, the cognitive difficulty can be manipulated by means of increasing the number of cards in a set and by reducing the amount of time the cards are shown to the participant. In this initial study, only one set of timing and set size parameters were chosen, since the goal of the study was to show that the P300 effect remains unchanged when cognitive factors are added.

## 8.1 Methods

### 8.1.1 Participants

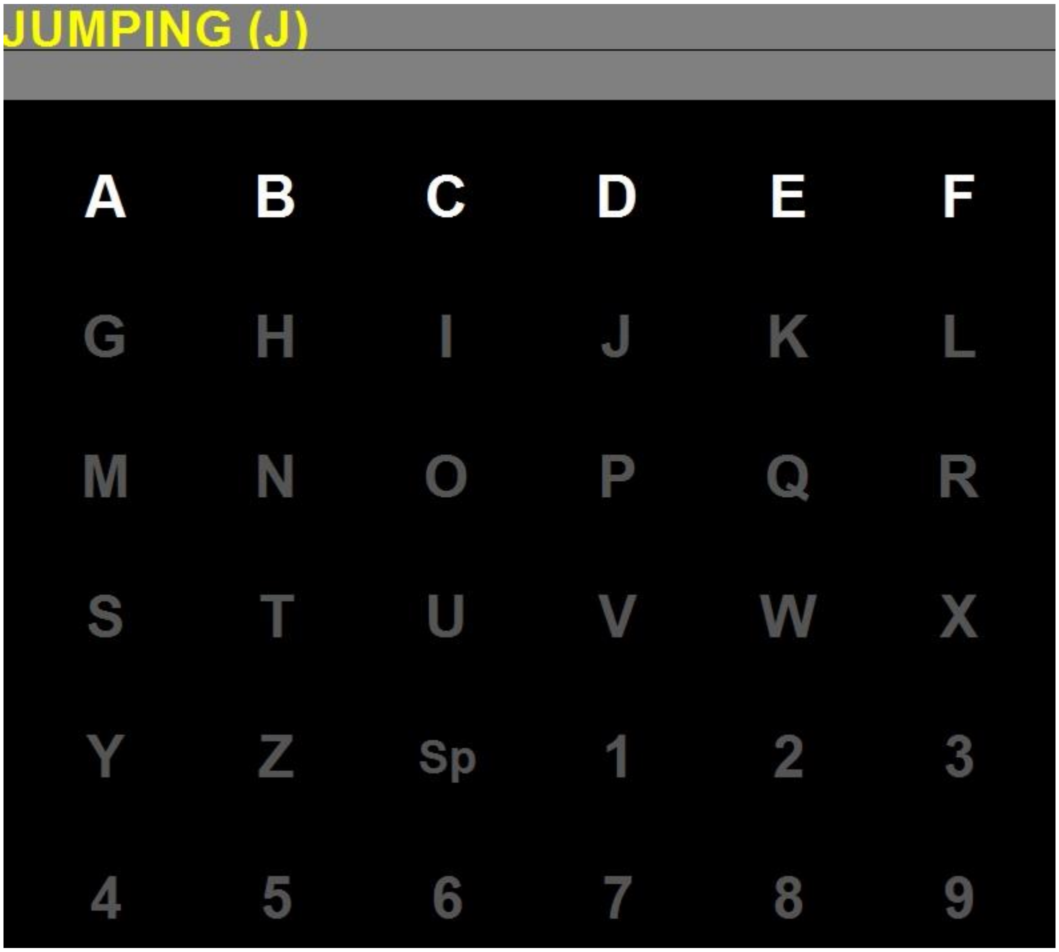
Four University of Victoria undergraduates participated in exchange for experimental credit in their psychology classes. Rejection criteria were determined in advance to exclude participants who were unable to achieve consistent enough performance to enable the linear classifier to be trained. No participants from this study had to be rejected. Three of the four subjects were female. Selection criteria included an absence of attentional disorders, as well as lack of experience with related EEG studies.

### 8.1.2 Apparatus

All studies were performed on the Emotiv Epoc with a reversed headset position, as described in the hardware chapter. In addition to 3 trials of the P3 speller task previously described in chapters 4 and 6, 2 additional trials per user were run on an alternate card game task. This was done after the P3 speller tasks, which were used for the training of the linear classifier.

The P3 speller task uses P300 ERP responses to determine which letter the user is attempting to spell. In the version of the task described here, rows or columns of letters were intensified at a rate of 120 msecs per intensification (62.5 msec intensified, then 62.5 msec pause before the next intensification) as shown in figure 8.1. For each letter to be spelled, the letter was intensified 30 times (15 times in row intensifications and 15 times in column intensifications), alternating rows and columns in a random presentation. The participant was instructed to count the flashes of the letter or number that they were trying to spell, although participants were encouraged to develop their own strategies to maintain concentration if they thought that would be more effective. In the example in figure 8.1, the participant was spelling the 1st letter of the word “JUMPING.” In looking for the P300 responses, the current study used an epoch size of 796 msecs, and since the rate of intensification is every 120 msecs, this means that the epochs were overlapping. Individual variation in terms of the maximum presentation rate is large, and the current rate of 120 msecs per intensification was chosen during pilot trials to allow most participants to achieve success at the task.

The monitor was at a viewing distance of approximately 80cm, causing it to subtend a visual angle of 34 degrees. Similar to the training study, but unlike in the two-session study, in this study the distance was not precisely fixed and the chin rest and eye tracker were not used. The participants did not use a mouse or keyboard, and all interaction with the system was through the EEG.



JUMPING (J)

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| G | H | I | J | K | L |
| M | N | O | P | Q | R |
| S | T | U | V | W | X |
| Y | Z | Sp | 1 | 2 | 3 |
| 4 | 5 | 6 | 7 | 8 | 9 |

**Figure 8.1: Interface for P3 speller task**

As in the two-session study described in chapter 4, the sample rate was 128 samples per second, and samples were transmitted in a block size of 4 such that each sample block was 31.25 msecs in duration. This defines the most precise temporal resolution of stimulus presentation possible and stimuli were presented for 2 sample blocks.

In the alternate card game task, shown in figure 8.2, the same attentional mechanism was used to select matching pairs of cards. The participants were first shown all of the cards, and then they were flipped over and revealed one by one as the participant concentrated on trying to form pairs. If a pair was not matched, both cards were flipped back over and the process started again. Although the presentation rate was the same, in this variant of the tasks, each card was illuminated separately to keep the overall target presentation rate low (11%, which is slightly above the guideline of 10% or lower (Picton, 1992), but not as frequent as the 33% rate that flashing in rows and columns would have caused). Since the participant was entirely in control of which card was being chosen, the back of each card was marked with a letter and the participant was instructed to say the letter of the card they were focusing on during the pause before each sequence, allowing for this to be manually recorded in the experimenter’s notebook and used for assessing system recognition accuracy.

Figure 8.2: Alternate card game task

### 8.1.3 Procedure

Subjects arrived and had the P3 speller task explained to them, then had the headset fitted. Participants were told that each letter would be intensified for a moment, in rows and columns, and that they were to mentally count the number of times the letter they wished to spell turned bright. When participants understood, the session began. This was with a very short phrase, “BIRD,” in order to get the participant used to the task. Following this, they spelled three phrases ranging from 10 to 15 characters per phrase (e.g., “JUMPING JACKS;” full prompts are given in Appendix C).

As in the previous studies, linear classifier training was attempted after the first and third phrase in order to maximize subject accuracy on the task. This training consisted of using the prompted phrase paired with the results generated from the user. This alignment of known correct responses and user-generated responses allowed the system to train a linear classifier to be able to associate the user EEG inputs to more accurately predict user intention in future trials. This training adjusts the inputs by electrode location and P300 timing to correctly weigh the inputs on a per-user (and per-session) basis.

Following the linear classifier training, participants were shown the card game. All of the cards were shown face-up for 30 seconds to allow the participant to memorize their positions. Then all the cards were turned face-down. Participants were instructed to focus on a single card for 30 flashes, as they had done in the speller task. Before each card, they were instructed to say the letter on the back of the card to the experimenter, who recorded which card they were attempting to turn over. The game continued until all the cards but one had been turned over (as it was a 3 by 3 grid of cards, with 9 cards total, there was one card that was a joker that remained unmatched). Participants played either one or two rounds of the card game while the system collected EEG data, depending on how long the session had taken up to that point (sessions were limited to 1 hour total, so the second card game task was not commenced unless there was at least ten minutes left).

After the card game, one additional P3 speller trial was attempted, time allowing; due to the one-hour limit on the study session, only two participants had time to do so.

### 8.1.4 Data Analysis

The data analysis for the P3 speller was very similar to the training effects study described in chapter 6, and was done within Matlab, using the EEGlab analysis package (Delorme and Makeig, 2004) and using the same import scripts as described in the training study (described in detail in Appendix A).

The card game data required an analysis similar to the free-spelling tasks from the training effects study in chapter 6, since the system did not know the intended target in advance. In this case, the participants were instructed to tell the experimenter which card it was that they were trying to turn over during the pause between trials. They referred to the cards via the letter shown on the back. The experimenter recorded this information and then used it during analysis to differentiate between target and non-target tasks. The card games also had a variable number of trials, since the game continued until all possible matches were made. The same artifact rejection steps and analysis were then performed on the card game data.

The primary question in this study, RQ3, was whether the card game task generates a P300 response in the same way that the P3 speller task does.

# 9 Memory attention Study Results and Discussion

## 9.1 Results

### 9.1.1 P3 speller data

The first results examined were the P3 speller data. The task was identical to the previous studies, and the P300 ERP results are presented in the same way as in previous studies, with the target trials compared to non-target trials across the 800 msec trial epoch. The results for the third training phrase on the P3 speller showed that there was a P3 response consistent with previous studies, and can be seen in figure 9.1.

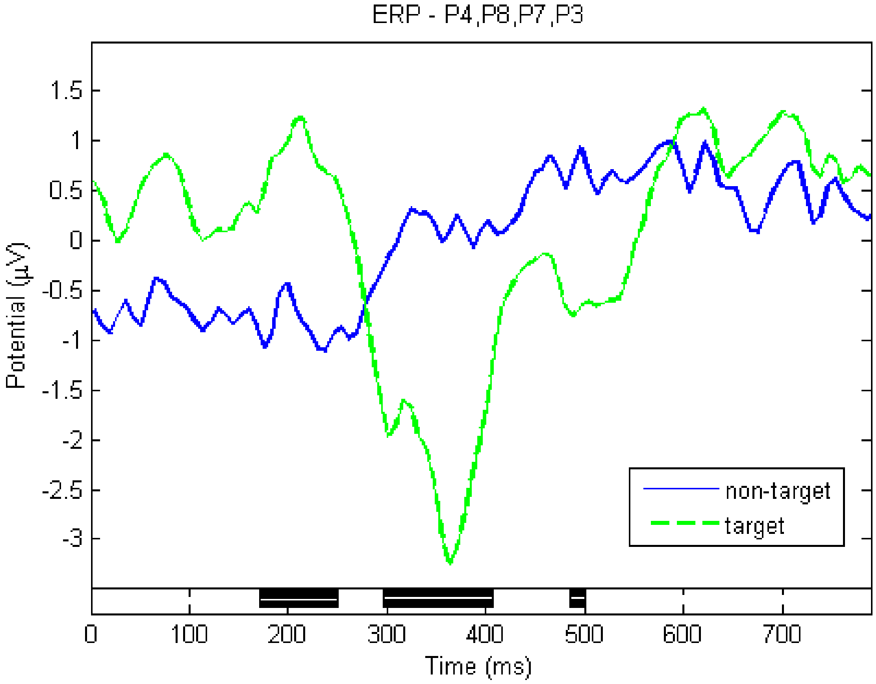


**Figure 9.1: Target vs. non-target ERP responses in the P3 speller at electrode locations P3, P4, P7, and P8 for the third training phrase showing a P300 response which is significant at *p*<.05.**

### 9.1.2 Card game data

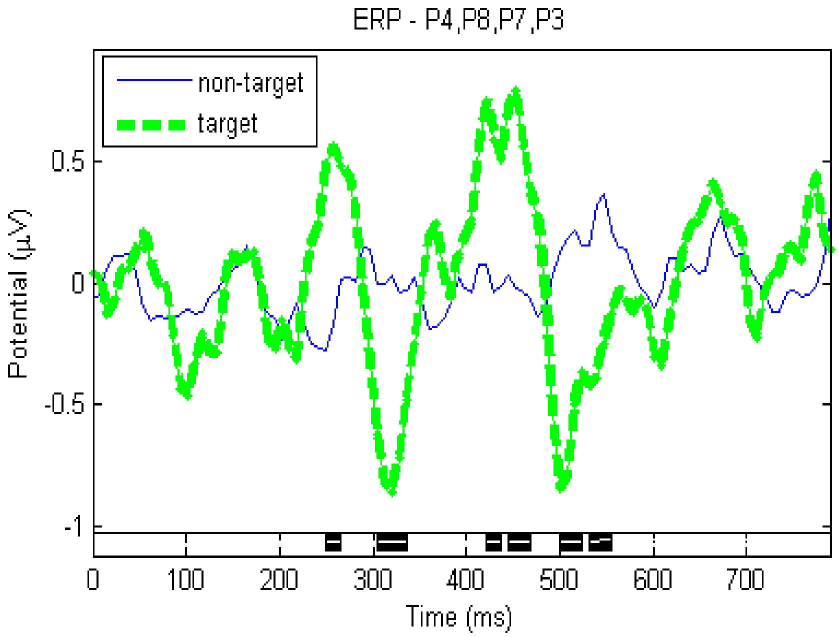


**Figure 9.2: Target vs. non-target ERP responses at locations P3, P4, P7, P8 (averaged) on the card game task. The P300 for target presentations is significantly different than the non-target presentations at *p*<.05.**

The card game task was treated in the same manner as the P3 speller data and shows a strong P300 peak in figure 9.2, although the pattern of the data following the P300 is somewhat different between the two tasks, as discussed in section 9.2.1.

The data on correct responses is shown in table 9.1. In regards to the correct cards, if a card was not a match but was the card the user intended to select, this was registered as a correct card. In general, games in which the percentage correct was approaching one hundred percent would be much shorter, since an error on either of the two cards would extend the number of cards required to complete a game.

| Participant | # games | Correct cards | Total cards |
|---|---|---|---|
| 1 | 2 | 17 | 18 |
| 2 | 1 | 9 | 12 |
| 3 | 2 | 18 | 18 |
| 4 | 1 | 12 | 16 |

**Table 9.1: Correct cards in the card game by participant**

### 9.1.3 Validation

As in the previous studies, the results were reanalysed using a common average reference. This again resulted in a positive deflection at the frontal electrodes. Figure 9.3 shows this for the card game task.

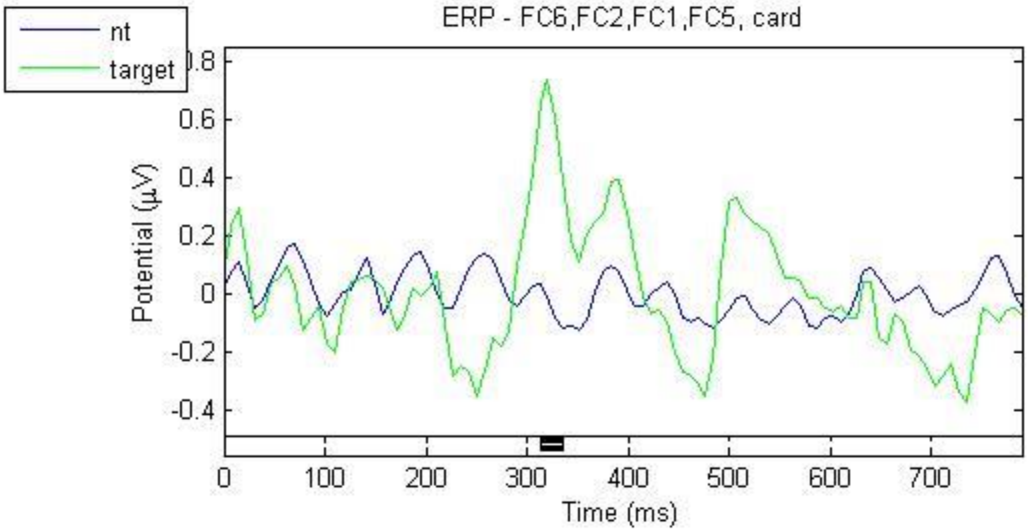


**Figure 9.3: Common average reference P300 results for the card game task**

The next two figures, figure 9.4 and 9.5, show the results for the P3 speller task using a common average reference. This shows the now familiar pattern of a positive deflection at the frontal electrodes and the negative dipole at the parietal electrodes.

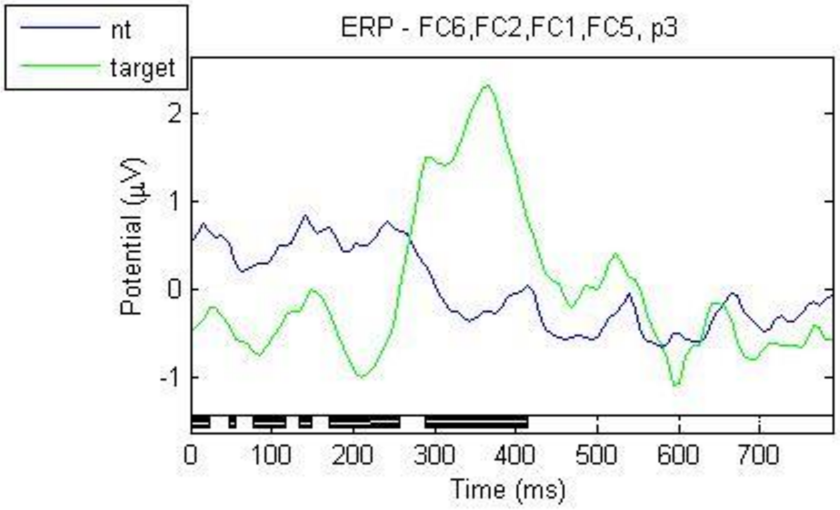


**Figure 9.4: frontal electrodes showing a P300 using a common average reference**

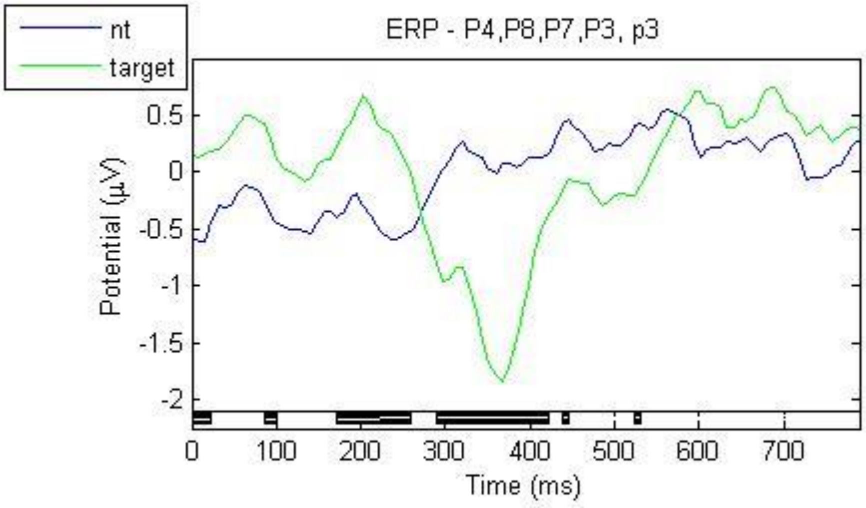


**Figure 9.5: Parietal electrodes using a common average reference**

## 9.2 Memory attention study discussion

It was the ability of the cognitive parameters to be manipulated while retaining the attentional characteristics of the task that made this an interesting paradigm for using P300 training as a neurofeedback mechanism. In this preliminary test of the paradigm, however, we did not vary any of the parameters that would allow for independent manipulation of cognitive difficulty—the number of cards in the game remained at 9, and the time to view the cards remained at 30 seconds. The intention of this first exploratory study was only to compare the card game to the standard P3 speller in order to address research question RQ3, which asks whether cognitive parameters of the task can be varied while maintaining a P300 response.

### 9.2.1 P300 pattern in the speller and the card game

A clear P300 ERP peak was seen in both the P3 speller data and the card game data. This suggests that a similar visual attention mechanism is being used in the case of this game, although the cognitive parameters of the task differ. There were a number of differences in the remainder of the trial epoch. These other components are likely related to the task differences, however, due to the small sample size in this study, no analysis of these other differences was possible. In addition, due to the small size of this study, and the fact that only one session was performed, no analysis of training data was completed. This was an exploratory study to determine if the new paradigm would allow for cognitive parameters to be manipulated while still generating a robust P300 ERP.

Following this proof of concept, subsequent studies (described in the future work section, chapter 11) could elaborate on the paradigm by maintaining interest through varying cognitive difficulty and using other gamification techniques. These techniques do not appear to disrupt the attentional mechanisms that lead to the production of a robust P300 response (although the remainder of the waveform has a number of differences), so the capability of using the P300 as a neurofeedback technique should still be effective even if the task was one with more intrinsic interest and the ability to remain engaging.

# 10 Discussion

The three studies reported in this dissertation serve to show that the Emotiv Epoc, as an example of a class of low-cost wireless devices, is capable of detecting P300 event-related potentials (RQ1). Furthermore, the studies show that the strength of these P300 responses increase with training (RQ2), although with some qualifications discussed in section 10.2. Finally, the P300 response is robust in the studies reported here, despite task changes that vary the cognitive difficulty of the task, and which vary depending on whether the user is generating their own prompts or having them provided by the experimenter. This final point is relevant both in terms of the ability to develop an interesting neurofeedback protocol for training P300 responses, and as an example of the attentional factors underlying the P300 and their separation from the cognitive processes involved.

## 10.1 P300 ERP detection on the Emotiv Epoc

The initial research question in this dissertation was whether the Emotiv Epoc is sufficiently sensitive to detect ERPs. Our results show that the Emotiv, despite some limitations in the reference electrode location and signal quality, does provide usable ERP data. In each of the three studies reported here, there were usable P300-like results, and the literature review presents other work supporting this conclusion (Badcock et al, 2013; De Lissa et al, 2015). Although the artifact rejection steps were more complicated than they would have been on a more precise amplifier (see the detailed discussion of this process in the training study results section), it did not prevent usable signals from being extracted.

The Emotiv Epoc was an effective research tool in these studies. The first research question (RQ1) merely asked whether it was capable of detecting ERP; however, other researchers (Duvinage, 2012; Mayaud, 2013) had expressed concern over the accuracy of the instrument, the precision of headset fitting, and the long-term comfort of the headset. None of these issues presented themselves to such a degree that they fundamentally interfered with the research studies. Although the issue of the reference electrode being in a very active location and causing the P300 to appear in a negative fashion on uninvolved locations is cumbersome from a data analysis perspective, it did still allow for reasonable signals to be captured. Although there was significant noise in the captured EEG signals, judicious processing of those signals was able to extract clean data suitable for ERP analysis. The current studies also alleviate some concerns over the comfort and usability aspects of the Emotiv Epoc headset. Given the decreasing price of consumer EEG headsets and the strengthening evidence of a training effect, further study of these low-cost devices seems warranted. The authors are hopeful for wider adoption opportunities for EEG-mediated biofeedback approaches.

The predominantly frontal aspect of the signal shown in the additional localization analyses in sections 5.1 and 7.2.1 are a decidedly unexpected result. Despite the localization issues with the P300 the timing remains consistent with expectations. The fact that there is a strong difference between target and non-target tasks that the classifier is using to allow for proper spelling performance is evidence that we are looking at an ERP that is behaving substantially like a P300, even if our localization results are non-standard.

It is likely the case that the current tasks and hardware are unable to distinguish between P3a and P3b subcomponents and that this will have to be the subject of future extensions to this research, described in chapter 11. The P3a subcomponent is generally seen frontally, but it is unexpected that this task would have generated a predominantly P3a response. The task characteristics were certainly chosen with reference to a cognitive task which would be expected to invoke a P3b response in that the P3 speller has been shown to be a task-dependant P300 task rather than a task-independent P300 task, and task-dependence is a major factor in elucidating the P3b subcomponent - the more the target stimulus was task-dependant, the more P3b subcomponent was found in the resulting ERP (Bowman et al, 2013). In the studies presented in this dissertation, no significant differences were seen in self-generated prompts versus words that the participants themselves generated. These types of cognitive elements, however, could have an impact on which P300 subcomponents predominate. In the version of the P3 speller where participants are being presented with the word to spell, there is certainly a cognitive component to the task, but there is less decision-making going on than in task variants where the participant is providing their own word, or making their own decisions in the card game task.

The Emotiv hardware is another significant limitation in regards to making this P3a versus P3b distinction. Because the headset offers only fixed electrode locations, they do not fit the same on all head sizes. Since the headset was in a reversed orientation, the fit is even harder to assess. In particular, on participants with small heads two of the parietal electrodes reliably made poor contact and had to be propped with chopsticks to achieve any connectivity at all. On these small-headed participants the electrode position was also generally further back on the head. The connectivity was assessed at the start of each session and during classifier training breaks, as described in chapter 3, so there is reason for confidence in the results in general, but the parietal electrodes in particular were harder to keep in effective scalp contact. This means that the spike and electrode drift artifact rejection steps rejected proportionally more measurements from the parietal electrodes.

A final factor to be considered are participant screening characteristics. It is unlikely that this would have had an overall biasing effect on the data, but studies have shown that anxiety and depression cause more frontal P300 response and an overall attenuation of response (Li et at, 2015). There were two participants in the 5 session study who reported anxiety conditions, and one who had a particularly troublesome day after taking anxiety medication earlier in the day. Although this is not likely to have had a systematic effect on the results, in future studies more careful participant screening would be warranted.

## 10.2 P300 Training effects

The existence of a P300 ERP training effect (research question RQ2) requires a nuanced analysis. There was an effect of training, which was shown clearly in both the two-session study and the main training study where performance improved over time; however, it was not uniform across the sessions. There was a rapid adaptation from the first day, or in some subjects the second day, and then performance peaked on the third day. This was true in both overall magnitude data as well as in the limited data on error rate shown in the training study (where the correct-characters rate on the first day was 64% and rose to 78% for both day 3 and day 5). That there was a clear training effect, however, raises the possibility of using this type of ERP protocol as a form of neurofeedback.

The P300 magnitude peaked on day 3, while the correct-characters rate did not improve after day 3, which requires some further discussion. One possibility, explored in more detail in section 10.3, is that it was a motivational problem caused due to participants losing interest in the task and therefore choosing to no longer strive for success. It is also possible that the day one through day three training effect was a rapid adaptation that then quickly plateaued into a level of performance that was better than the first day, but which no longer continued to improve. This would limit the utility of this technique as a long-term neurofeedback approach. Anecdotal data suggests that further improvement does occur, though this was not systematically studied in this dissertation. It is also possible that the plateau was simply a short-term phenomenon where, after a period of rapid adaptation, there was a longer slow period of adaptation which these studies were neither long enough, nor sensitive enough, to detect.

The possibility of very precise neurofeedback based on ERP components offers different potential benefits from the standard neurofeedback techniques involving cortical slow waves (spontaneous EEG). When using a consumer EEG for its intended purpose—often the training of alpha state relaxation—the training changes are slow and broad. Similar to learning how to meditate without the help of neurofeedback, the ability to progressively learn to relax and practice mindfulness meditation confers benefits in many areas, including anxiety reduction, relief of depression, and attenuation of ADHD symptoms (Demos, 2005).

ERP neurofeedback, by contrast, has yet to be shown to have general benefits, but can be expected to be more task specific than spontaneous EEG neurofeedback. Rather than training the entirety of the frontal cortex over an averaging period of seconds or minutes, ERP training provides feedback on a neural response to a specific stimulus. In these studies, this has the effect of keeping participants engaged in a sustained visual attention task and results in an increased P300 magnitude, which is an ERP component that is generally attenuated in attention deficit disorders (Brandeis et al, 2002; Loo and Makeig, 2012). It is, as yet, unclear whether training of the P300 has clinical utility, though the effect of the current studies shows that there is a magnitude increase in an ERP component of clinical significance, which is a strong initial step. In addition, there are ERP components that correlate with other conditions (such as the N170 for face recognition), so the conceptual approach of neurofeedback training on ERP components has potentially broad applicability.

### 10.3 Motivation and impact of cognitive factors on P300 ERP performance

Past research has shown an effect of motivation on P300 responses (Gray et al, 2004). This was supported by our participants' self-reported strategies. Most subjects continued with the "count the number of flashes" strategy described in the instructions, though other participants found ways of making the task more interesting. As described in the training study results in chapter 7, several subjects played word association games, with one inventing swear words that started with the letter they were looking for in an attempt to make a stronger association.

However, insufficient data was collected on alternate strategies to assess a performance impact. The inherent repetitiveness of the P3 speller paradigm is one possible reason that performance did not continue to improve after the third session in the training study; participants were potentially getting bored and realizing that they could achieve moderately acceptable recognition results without expending as much effort on the task. Non-motivational reasons are explored more fully in section 10.2, but if we posit that motivation is the reason for the performance drop, then the development of a task that has more inherent interest is necessary to incentivize longer-term use of the training paradigm. Gamification of computer tasks is one way of achieving this, so that the participant competes with their prior results or with others (Ganin et al, 2013). Another approach is to increase the cognitive complexity of the task to maintain participant interest. This last approach is the focus of the third research question in this dissertation: whether cognitive complexity of the task can be manipulated while still seeing a robust P300 ERP.

In the card game in the memory attention study, task difficulty can be manipulated by means of the card grid size, the number of pairs, and the time given to memorize them. This study has shown that the cognitive challenge of the task can be manipulated while preserving the visual attention focus of the task. In this preliminary study, all that has been shown is that a more complex cognitive task is still capable of producing P300 results similar to a less cognitively demanding visual attention task. In longer-term training studies, this independent manipulation of cognitive load versus attentional load may become beneficial for keeping users interested. The discussion of P300 for gaming interfaces is directly relevant to this approach (Kaplan et al, 2013). One problem with the current card game paradigm is that it does not function as a mechanism to do supervised model training; in the P300 speller, since the subject is being given the letters to spell, the system can align participant responses with the known correct responses in order to train the linear classifier to recognize responses better. In the case of the card game, the system does not know a priori which card is the correct response, so this training cannot occur. In this experiment, the subject was asked to verbally state which card they were trying to turn over so that the experimenter could keep track of “correct” responses for offline analysis, but other paradigms are being considered to alleviate this difficulty. Nonetheless, the fact that the P3 speller trials could be used to train a model for the card game task (allowing it to work at an effective degree of accuracy) is confirmatory evidence that the card game and P3 speller share a common mechanism. The presence of the characteristic P300 response pattern in the data is also suggestive that the card game and standard P3 speller share a similar underlying attentional mechanism.

Internal versus external prompt generation could not be shown to have any effect on P300 ERP generation in these studies. This is a positive sign for the ability to use this technique in neurofeedback, since the existence of cognitive factors and self-generation of prompts is one potent way to ensure that users remain motivated and invested in the task. In the original incarnation of the P3 speller task, this was central to its use as an assistive device. To have the P300 ERP show itself as a pure attentional component, unaffected by cognitive or memory factors, will require further research. However, the evidence presented in these studies can lead to a provisional conclusion that these factors may be manipulated without interference with the P300 ERP or the attentional nature of the task.

It is worth noting that these are preliminary results. The memory attention paradigm has been shown on only a handful of participants to have a P300, which appears similar to that in the P3 speller. This must be confirmed. The training effect is seen most strongly very early in the process; if it does not continue over a longer time period, then its utility as a neurofeedback target is questionable. These studies have been performed on cognitively normal young adults. If the P300 is so attenuated in clinical populations that the P3 speller, or alternate variants, cannot be completed at all, then its utility as a neurofeedback technique will fail to help the population it is most designed to assist. Finally, although the P300 is attenuated in individuals with ADHD, and these results suggest that ERP neurofeedback may strengthen that response, an important question to be posed is whether that will make any practical difference. In the literature review, it was noted that individuals with ADHD may have already developed other coping mechanisms (Prox, 2007), and if this is the case, then restoring a stronger P300 may have no discernable effect at a behavioural level.

Despite these caveats, the evidence that ERP components are trainable, and that a more precise form of neurofeedback may be possible, is an initial suggestion of an exciting new treatment modality.

# 11 Future Work

Despite its suitability for these training studies, the Emotiv electrode montage—the positioning of electrodes—is not optimal for P300 studies. The reference channel is inconvenient and the saline-coupled electrodes are cumbersome. There is a new option from Emotiv: the Emotiv Insight, a new capacitively coupled headset that deserves further evaluation, although with some electrode montage problems of its own. There are also some open source alternatives, such as the OpenBCI project, which appears to result in good data, although the drivers to use it in BCI2000 are not yet available. Some other consumer headsets have been recently introduced to the market, such as the Interaxon Muse, but these do not seem to be easily modifiable in order to give access to the electrode locations that would be most useful for ERP studies. The predominant use of these consumer devices remains for frontal lobe slow wave activity, which, since it is averaged over a much longer time, is not as technically challenging to detect as ERPs.

The question of which P300 subcomponent is being seen is a critical one moving forward. ERP is being detected, and the training effects are clear, but the theoretical difference between P3a and P3b components is quite significant (Dundon et al, 2015). Determining which subcomponent is being trained could make a significant difference in any behavioural effects that might result from the training. This may require both task changes and electrode montage and reference electrode location changes to definitively determine the P300 subcomponent and develop a training protocol specific to P3b.

A direct comparison study between the Emotiv headset and a more standard gel-cap headset is the first experimental element to addressing the confusion among subcomponents. As mentioned in section 10.1, there is some concern over the ability of the Emotiv headset to access precisely enough the locations of interest. There is also the question of electrode noise profiles at those locations. Being able to more accurately identify the P3a and P3b subcomponents on a more precise research-grade instrument would allow separation of hardware concerns from task concerns. The second element in separating these concerns is to look at the P3 speller task in more detail. As described in section 10.1, the more the task involves decision making rather than simply identification, the more P3b subcomponent appears in the resultant ERP (Bowman et al, 2013). Identifying task differences and their impact on P300 subcomponents would be a useful line of research in this regard. The question of whether the Emotiv is capable of detecting ERP has been answered in this dissertation, but the precision of the signal and the capability of this instrument to identify P300 subcomponents is in the realm of future research.

With capable and inexpensive devices becoming available, it makes sense to work on strengthening the case for neurofeedback of ERP components as being effective and useful. Effectiveness of the training seems straightforward to undertake through longitudinal studies and longer-term use, but will require tasks that are intrinsically more interesting to keep participants motivated. Assessing utility will require tying those longitudinal ERP training results to measures that correlate with known conditions. ADHD is the obvious first candidate, and the reason for the focus within this paper on visual attention; however, as inexpensive and accessible headsets improve in capability, other ERP targets are also very interesting to consider. Using validated external measures—questionnaires, performance data, brain imaging data, etc.—and correlating positive changes in those measures to ERP magnitude increases resulting from training would be the gold standard of assessing the utility of this type of neurofeedback approach.

It is likely that some of the factors leading to the successful use of the P3 speller are not strictly due to the training of the brainwaves being produced, but rather in training how to fit the headset properly, how (and when) it is necessary to maintain attention, and how to avoid movement artifacts. Since the experimenter took care of headset fitting and gave instructions and feedback to minimize artifacts, these current studies were designed to show training effects on only the actual P300 brainwave production. In an unsupervised use situation, however, all of these other factors must be trained as well. A more comprehensive longitudinal study would offer the opportunity to disentangle these factors from the underlying training of the ERP component over a longer time scale.

The ability of participants to reliably use P300 neurofeedback depends on their engagement with the task. If the task is too uninteresting, such that they cannot force themselves to attend to it properly, then no training will occur. The line of research that was begun with the memory attention task can be taken much further. Gamification can promote competition both within oneself and with others. The card game is merely an example to show that cognitive factors do not preclude the generation of a robust P300.

There is a long history in the videogame industry of making games challenging and engaging, and these techniques merely need be added to an attentional paradigm that rewards the type of P300 response that is desired for training purposes. The combination of video game levels of engagement, and therapeutic approaches to brain training, offer exciting treatment modalities, although with a significant level of research required to realize their potential.

In addition, there are other far-reaching studies that take P300 training into naturalistic environments. One of the benefits of the generation of consumer EEG headsets described here is that they are generally wireless, easily installed, and compared to a standard EEG gel cap, fairly inconspicuous. This raises the possibility of their use in a more naturalistic situation. Determining the stimulus onset is a major issue for naturalistic studies, but one can imagine wearing some of these devices throughout the day, along with a camera or audio recorder, and performing an analysis of which events that should have triggered an attentional focus actually did so. Measuring P300 ERP in a classroom situation, for instance, could generate useful data on learning styles, the best time of day for learning, ADHD treatment strategies, and various other factors relating to attention in the classroom. At that point, ERP neurofeedback effectiveness itself could be, in part, assessed by these naturalistic EEG measures. The assessment of EEG in these naturalistic situations is made vastly simpler through user-friendly headsets and amplifiers, though it is still by no means easy. The difficulty of accurately assessing stimulus onset time is a difficult problem with no obvious solution. The degree of timing precision necessary to compare ERP events is difficult even when one is able to completely control the stimulus.

When both the stimulus and the timing differs, it would be difficult to perform analyses. There are some efforts underway to correct this. There is some work on plotting and analysis techniques that can be applied for single-trial ERP epochs (Delorme et al, 2015). A recent study looked at the cocktail party effect, where people can separate out the conversation they want to focus on out of the background noise of other conversations, and used EEG to understand the processing involved in focusing attention on a single stimulus (O'Sullivan et al, 2015). However, the area of naturalistic ERP analysis is on the edge of what is feasible with current EEG analysis techniques.

The general paradigm of ERP neurofeedback could potentially go beyond the P300 as well. Is it useful to train other ERP components? The possibility is very speculative, but other ERP components have certainly been associated with specific cognitive processes that could be subject to upregulation. The N170, for instance, is associated with the processing of faces and some research shows differences in N170 responses in individuals with developmental prosopagnosia (face blindness) (Towler et. al, 2012). There is little evidence that someone with prosopagnosia would benefit from ERP neurofeedback training on N170 responses, but in any condition where a specific ERP component shows a reduction in a clinical population there is at least the possibility that neurofeedback to upregulate that ERP response could be an interesting approach to investigate.

# 12 Conclusion

The Emotiv Epoc is a suitable device for ERP neurofeedback in settings where a full research EEG is cost-prohibitive or inconvenient. Within this study, we see some initial results that indicate that brain training is effective in increasing the strength of P300-like responses over multiple sessions, but not consistently through all the sessions. We also demonstrate that an alternate paradigm, the card game task, shares an underlying P300 mechanism, or at least produces P300-like responses that appear very similar. This, along with similar tasks, may be more interesting than the P3 speller task for long-term ERP neurofeedback training.

First, we answer research question RQ1 and show that the low-cost wireless Emotiv Epoc EEG is capable of reliably detecting P300 ERP responses. In each of the studies reported here, it was possible to separate ERP responses from noise and other confounding factors. The Emotiv Epoc is not as robust as a true research-grade EEG, has more noise, and has less convenient electrode placement. However, the device makes up for these characteristics with a low price, ease of fitting, and ease of cleanup. For a device that would be used in non-traditional research settings, or cost-constrained settings, the Emotiv Epoc is a highly viable option. The research on the Epoc is indicative of the capabilities of a class of consumer-grade EEGs that are becoming more common in the marketplace and the constraints of relatively slow EEG devices with convenient, but somewhat noisy, electrodes is an indication that this work may become more accessible to researchers in the future.

Secondly, relating to research question RQ2, these studies show that there is a P300 ERP training effect, but that this effect is not consistent throughout all sessions. There is definitely training occurring, though earlier sessions are not producing robust P300 responses as reliably as later sessions. However, the training progression is not linear. Engagement and effort may be required to continue progressing, and boredom with the task is an issue, but it is also possible that training rapidly plateaus and will not progress further regardless of the interest in the task. Finally, there is a high degree of variability, both between subjects and within a subject between sessions, which complicates the training. More research is certainly needed to disentangle these possibilities.

Finally, the studies here present a new P300 training paradigm based on a card-game task that provides further cognitive challenge through the use of a task requiring memory and decision-making while still relying on a comparable underlying visual attention mechanism. These studies also show that internal versus external generation of prompts cannot be shown to affect the generation of a robust P300 ERP. Future studies may show differences between internal and external prompt generation, but for now, participants' greater interest in internal prompt generation is a potentially viable way to increase the inherent interest of a P300 task. The card-game task is a cognitively challenging task as compared to the P3 speller; it engages memory, strategy, and choice in a way that the P3 speller does not. It also results in a P300 response very similar to the less cognitively challenging P3 speller task. It is clear that visual attention can be separated or combined with cognitive complexity, memory, and internal prompt generation, yet still retain the essential character of a P300 visual attention task.

Many questions remain regarding the long-term trainability of ERP components, and the potential clinical benefits of doing so. The studies in this dissertation provide an initial basis for the concept of ERP neurofeedback as a biofeedback modality separate from spontaneous EEG neurofeedback, and that is feasible to undertake on low-cost hardware.

# 13 Bibliography

Ahmadi, M., & Quian Quiroga, R. (2013). Automatic denoising of single-trial evoked potentials. *NeuroImage*, *66*, 672–680. http://doi.org/10.1016/j.neuroimage.2012.10.062

Alhaddad, M. J., Kamel, M., Malibary, H., Thabit, K., Dahlwi, F., & Hadi, A. (2012). P300 Speller Efficiency with Common Average Reference. In M. Kamel, F. Karray, & H. Hagras (Eds.), *Autonomous and Intelligent Systems* (pp. 234–241). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-31368-4_28

Al-Negheimish, H., Al-Andas, L., Al-Mofeez, L., Al-Abdullatif, A., Al-Khalifa, N., & Al-Wabil, A. (2013). Brainwave Typing: Comparative Study of P300 and Motor Imagery for Typing Using Dry-Electrode EEG Devices. In C. Stephanidis (Ed.), *HCI International 2013 - Posters' Extended Abstracts* (pp. 569–573). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-39473-7_113

Aloise, F., Arico, P., Schettini, F., Lucano, E., Salinari, S., Babiloni, F., & Cincotti, F. (2011). Can the P300-based BCI training affect the ERPs? *International Journal of Bioelectromagnetism*, *13*(3), 148–149.

Arico, P. (2014). *Mental states monitoring through passive Brain-Computer Interface systems*. University of Bologna. Retrieved from http://amsdottorato.unibo.it/6557/

Aricò, P., Aloise, F., Schettini, F., Soragnese, V., Salinari, S., Mattia, D., & Cincotti, F. (2012). Variability of ERPs–based Brain Computer Interface performance across repeated sessions in a day. In *3rd Workshop of the TOBI Project: Bringing BCIs to End-Users: Facing the Challenge, Evaluation, User Perspectives, User Needs, and Ethical*

*Questions. Würzburg, Germany Mar 20th-22th*. Retrieved from http://www.tobi-project.org/sites/default/files/public/Publications/TOBI_WS3_059_Final.pdf

Arns, M., Heinrich, H., & Strehl, U. (2014). Evaluation of neurofeedback in ADHD: The long and winding road. *Biological Psychology*, *95*, 108–115. Retrieved from http://doi.org/10.1016/j.biopsycho.2013.11.013

Arun, M. P. (2015). EEG Neurofeedback – A Novel Treatment Modality for Children with Attention Deficit Hyperactivity Disorder. *Kerala Journal of Psychiatry*, *28*(1), 59–65.

Badcock, N. A., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., & McArthur, G. (2013). Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. *PeerJ*, *1*, e38. Retrieved from http://doi.org/10.7717/peerj.38

Bao, F. S., Liu, X., & Zhang, C. (2011). PyEEG: An Open Source Python Module for EEG/MEG Feature Extraction. *Computational Intelligence and Neuroscience*, *2011*, 1–7. Retrieved from http://doi.org/10.1155/2011/406391

Barcala, L., Grecco, L. A. C., Colella, F., Lucareli, P. R. G., Salgado, A. S. I., & Oliveira, C. S. (2013). Visual Biofeedback Balance Training Using Wii Fit after Stroke: A Randomized Controlled Trial. *Journal of Physical Therapy Science*, *25*(8), 1027–1032. Retrieved from http://doi.org/10.1589/jpts.25.1027

BCI2000 wiki, P3SpellerTask. (n.d.). Retrieved from http://www.bci2000.org/wiki/index.php/User_Reference:P3SpellerTask

Berger, P. D. H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*(1), 527–570. Retrieved from http://doi.org/10.1007/BF01797193

Bishop, D. (2011). Bishop_techbits: Simplified introduction to time-frequency analysis in EEGLAB. Retrieved from http://bishoptechbits.blogspot.ca/2010/07/simplified-introduction-to-time.html

Blanchard, E.B., & Young, L.D. (1974). Clinical applications of biofeedback training: A review of evidence. *Archives of General Psychiatry, 30*(5), 573–589. Retrieved from http://doi.org/10.1001/archpsyc.1974.01760110003001

Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A., & Wyble, B. (2013). Subliminal Salience Search Illustrated: EEG Identity and Deception Detection on the Fringe of Awareness. *PLOS ONE*, *8*(1), e54258. http://doi.org/10.1371/journal.pone.0054258

Brandeis, D., van Leeuwen, T. H., Steger, J., Imhof, K., & Steinhausen, H.C. (2002). Mapping brain functions of ADHD children. In *International Congress Series,1232*, 649–654. Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/S0531513101006732

Budzynski, T. H., Budzynski, H. K., Evans, J. R., & Abarbanel, A. (2009). *Introduction to Quantitative EEG and Neurofeedback: Advanced Theory and Applications*. Academic Press.

Citi, L., Poli, R., & Cinel, C. (2010). Documenting, modelling and exploiting P300 amplitude changes due to variable target delays in Donchin's speller. *Journal of Neural Engineering*, *7*(5), 056006. Retrieved from http://doi.org/10.1088/1741-2560/7/5/056006

Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.

Danaher, J. (2015). Responsible Innovation in Social Epistemic Systems: The P300 Memory Detection Test and the Legal Trial. In B.J. Koops, I. Oosterlaken, H. Romijn, T. Swierstra, & J. van den Hoven (Eds.), *Responsible Innovation, 2,* 245–263. Springer

International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-17308-5_13

Datta, A., Cusack, R., Hawkins, K., Heutink, J., Rorden, C., Robertson, I. H., & Manly, T. (2008). The P300 as a Marker of Waning Attention and Error Propensity. *Computational Intelligence and Neuroscience*, *2007*, e93968. Retrieved from http://doi.org/10.1155/2007/93968

Davis, P. A. (1939). Effects of acoustic stimuli on the waking human brain. *Journal of Neurophysiology*, *2*(6), 494–499.

Debener, S., Minow, F., Emkes, R., Gandras, K., & de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, *49*(11), 1617–1621. Retrieved from http://doi.org/10.1111/j.1469-8986.2012.01471.x

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Cramon, D. Y. von, & Engel, A. K. (2005). Trial-by-Trial Coupling of Concurrent Electroencephalogram and Functional Magnetic Resonance Imaging Identifies the Dynamics of Performance Monitoring. *The Journal of Neuroscience*, *25*(50), 11730–11737. Retrieved from http://doi.org/10.1523/JNEUROSCI.3286-05.2005

de Lissa, P., Sörensen, S., Badcock, N., Thie, J., & McArthur, G. (2015). Measuring the face-sensitive N170 with a gaming EEG system: A validation study. *Journal of Neuroscience Methods*, *253*, 47–54. Retrieved from http://doi.org/10.1016/j.jneumeth.2015.05.025

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. Retrieved from http://doi.org/10.1016/j.jneumeth.2003.10.009

Delorme, A., Miyakoshi, M., Jung, T.P., & Makeig, S. (2015). Grand average ERP-image plotting and statistics: A method for comparing variability in event-related single-trial EEG activities across subjects and conditions. *Journal of Neuroscience Methods*, *250*, 3–6. Retrieved from http://doi.org/10.1016/j.jneumeth.2014.10.003

Demos, J. N. (2005). *Getting Started with Neurofeedback*. W. W. Norton & Company.

Dundon, N. M., Dockree, S. P., Buckley, V., Merriman, N., Carton, M., Clarke, S., … Dockree, P. M. (2015). Impaired auditory selective attention ameliorated by cognitive training with graded exposure to noise in patients with traumatic brain injury. *Neuropsychologia*, *75*, 74–87. http://doi.org/10.1016/j.neuropsychologia.2015.05.012

Duvinage, M., Castermans, T., Dutoit, T., Petieau, M., Hoellinger, T., Saedeleer, C. D., & Cheron, G. (2012). A P300-based Quantitative Comparison between the Emotiv Epoc Headset and a Medical EEG Device. ACTAPRESS. Retrieved from http://doi.org/10.2316/P.2012.764-071

Emotiv.com forum. (n.d.). Retrieved from https://emotiv.com/forum/forum4/topic1369/messages/

Farwell, L. A., & 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.

Freyer, F., Becker, R., Dinse, H. R., & Ritter, P. (2013). State-Dependent Perceptual Learning. *The Journal of Neuroscience*, *33*(7), 2900–2907. Retrieved from http://doi.org/10.1523/JNEUROSCI.4039-12.2013

Ganin, I. P., Shishkin, S. L., & Kaplan, A. Y. (2013). A P300-based Brain-Computer Interface with Stimuli on Moving Objects: Four-Session Single-Trial and Triple-Trial Tests with a

Game-Like Task Design. *PLoS ONE*, *8*(10), e77755. Retrieved from http://doi.org/10.1371/journal.pone.0077755

Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology*, *40*(2), 216–224.

Grierson, M., & Kiefer, C. (2011). Better brain interfacing for the masses: progress in event-related potential detection using commercial brain computer interfaces. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, 1681–1686. Retrieved from http://dl.acm.org/citation.cfm?id=1979828

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. http://doi.org/10.1016/j.neuroimage.2013.10.067

Hawkinson, J. E., Ross, A. J., Parthasarathy, S., Scott, D. J., Laramee, E. A., Posecion, L. J., & deCharms, R. C. (2011). Quantification of Adverse Events Associated with Functional MRI Scanning and with Real-Time fMRI-Based Training. *International Journal of Behavioral Medicine*, *19*(3), 372–381. Retrieved from

Jessup, B. A., Neufeld, R. W. J., & Merskey, H. (1979). Biofeedback therapy for headache and other pain: An evaluative review. *Pain*, *7*(3), 225–270. Retrieved from http://doi.org/10.1016/0304-3959(79)90083-6

Kaplan, A., Shishkin, S., Ganin, I., Basyul, I., & Zhigalov, A. (2013). Adapting the P300-based brain–computer interface for gaming: a review. *Computational Intelligence and AI in Games, IEEE Transactions on*, *5*(2), 141–149.

Käthner, I., Kübler, A., & Halder, S. (2015). Rapid P300 brain-computer interface communication with a head-mounted display. *Frontiers in Neuroscience*, *9*. Retrieved from http://doi.org/10.3389/fnins.2015.00207

Kerns, K. A., Eso, K., & Thomson, J. (1999). Investigation of a direct intervention for improving attention in young children with ADHD. *Developmental Neuropsychology*, *16*(2), 273–295.

Kothe, C. A., & Makeig, S. (2013). BCILAB: a platform for brain–computer interface development. *Journal of Neural Engineering*, *10*(5), 056014. Retrieved from http://doi.org/10.1088/1741-2560/10/5/056014

Kropotov, J. D., Pronina, M. V., Ponomarev, V. A., & Murashev, P. V. (2011). In Search of New Protocols of Neurofeedback: Independent Components of Event-Related Potentials. *Journal of Neurotherapy*, *15*(2), 151–159. Retrieved from http://doi.org/10.1080/10874208.2011.570696

Krusienski, D. J., Sellers, E. W., Cabestaing, F., Bayoudh, S., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2006). A comparison of classification techniques for the P300 Speller. *Journal of Neural Engineering*, *3*(4), 299. Retrieved from http://doi.org/10.1088/1741-2560/3/4/007

Lage-Castellanos, A., Nieto, J. I., Quiñones, I., & Martinez-Montes, E. (2010). A zero-training algorithm for EEG single-trial classification applied to a face recognition ERP experiment. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4209–4212). Retrieved from http://doi.org/10.1109/IEMBS.2010.5627395

Lakey, C. E., Berry, D. R., & Sellers, E. W. (2011). Manipulating attention via mindfulness induction improves P300-based brain–computer interface performance. *Journal of Neural Engineering*, *8*(2), 025019. Retrieved from http://doi.org/10.1088/1741-2560/8/2/025019

Larsen, S. L. (2012). *The Neurofeedback Solution: How to Treat Autism, ADHD, Anxiety, Brain Injury, Stroke, PTSD, and More*. Inner Traditions / Bear & Co.

Lau, T. (2010). What Makes a Good HCI Systems Paper? | blog@CACM | Communications of the ACM. Retrieved May 30, 2016, from http://cacm.acm.org/blogs/blog-cacm/86066-what-makes-a-good-hci-systems-paper/fulltext#comments

Lee, W. L., Tan, T., Falkmer, T., & Leung, Y. H. (2012). Single-Trial Multi-channel N170 Estimation Using Linear Discriminant Analysis (LDA). In T. Huang, Z. Zeng, C. Li, & C. S. Leung (Eds.), *Neural Information Processing*, 347–355. Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-34478-7_43

Li, Y., Wang, W., Liu, T., Ren, L., Zhou, Y., Yu, C., … Hu, Y. (2015). Source analysis of P3a and P3b components to investigate interaction of depression and anxiety in attentional systems. *Scientific Reports*, *5*, 17138. http://doi.org/10.1038/srep17138

Liu, Y., Ayaz, H., Curtin, A., Onaral, B., & Shewokis, P. A. (2013). Towards a Hybrid P300-Based BCI Using Simultaneous fNIR and EEG. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of Augmented Cognition* (pp. 335–344). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-39454-6_35

Loo, S. K., & Barkley, R. A. (2005). Clinical utility of EEG in attention deficit hyperactivity disorder. *Applied Neuropsychology*, *12*(2), 64–76.

Loo, S. K., & Makeig, S. (2012). Clinical Utility of EEG in Attention-Deficit/Hyperactivity Disorder: A Research Update. *Neurotherapeutics*, *9*(3), 569–587. Retrieved from http://doi.org/10.1007/s13311-012-0131-z

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. Retrieved from http://doi.org/10.3389/fnhum.2014.00213

Luck, S. J. (2006). *By Steven J. Luck - An Introduction to the Event-Related Potential Technique: 1st (first) Edition*. MIT Press.

Mak, J. N., Arbel, Y., Minett, J. W., McCane, L. M., Yuksel, B., Ryan, D., & Erdogmus, D. (2011). Optimizing the P300-based brain–computer interface: current status, limitations and future directions. *Journal of Neural Engineering*, *8*(2), 025003. Retrieved from http://doi.org/10.1088/1741-2560/8/2/025003

Mann, Steve. (n.d.). What I’ve learned from 35 years of wearing computerized eyewear. Retrieved from http://spectrum.ieee.org/geek-life/profiles/steve-mann-my-augmediated-life?

Marchetti, M., & Priftis, K. (2014). Effectiveness of the P3-speller in brain–computer interfaces for amyotrophic lateral sclerosis patients: a systematic review and meta-analysis. *Frontiers in Neuroengineering*, *7*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4013458/

Mayaud, L., Congedo, M., Van Laghenhove, A., Orlikowski, D., Figère, M., Azabou, E., & Cheliout-Heraut, F. (2013). A comparison of recording modalities of P300 event-related potentials (ERP) for brain-computer interface (BCI) paradigm. *Neurophysiologie*

*Clinique/Clinical Neurophysiology*, *43*(4), 217–227. Retrieved from http://doi.org/10.1016/j.neucli.2013.06.002

Mecklinger, A., Maess, B., Opitz, B., Pfeifer, E., Cheyne, D., & Weinberg, H. (1998). A MEG analysis of the P300 in visual discrimination tasks. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(1), 45–56. http://doi.org/10.1016/S0168-5597(97)00092-0

Meixner, J. B., & Rosenfeld, J. P. (2014). Detecting Knowledge of Incidentally Acquired, Real-World Memories Using a P300-Based Concealed-Information Test. *Psychological Science*, *25*(11), 1994–2005. Retrieved from http://doi.org/10.1177/0956797614547278

Menon, R. S., Kim, S.-G., Menon, R. S., Kim, S.-G., Menon, R. S., Kim, S.-G., & Kim, S.-G. (1999). Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends in Cognitive Sciences*, *3*(6), 207–216. Retrieved from http://doi.org/10.1016/S1364-6613(99)01329-7

Nunez, P. L., & Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735. http://doi.org/10.1038/nn.3248

Olds, J. (1962). Mechanisms of instrumental conditioning. *Electroencephalography and Clinical Neurophysiology*, Suppl 24:219+–219+.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., & Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706.

Parra, M., Ascencio, L., Urquina, H., Manes, F., & Ibanez, A. (2012). P300 and neuropsychological assessment in mild cognitive impairment and Alzheimer dementia. *Dementia*, *3*, 172. http://doi.org/10.3389/fneur.2012.00172

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, *9*(4), 456.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. http://doi.org/10.1016/j.clinph.2007.04.019

Pop-Jordanova, N. (2011). Brain Rate as an Indicator of the Level of Consciousness. In D. Cvetkovic & I. Cosic (Eds.), *States of Consciousness*, 187–201. Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-18047-7_8

Prox, V., Dietrich, D. E., Zhang, Y., Emrich, H. M., & Ohlmeier, M. D. (2007). Attentional processing in adults with ADHD as reflected by event-related potentials. *Neuroscience Letters*, *419*(3), 236–241. Retrieved from http://doi.org/10.1016/j.neulet.2007.04.011

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., & Lécuyer, A. (2010). OpenViBE: an open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. *Presence: Teleoperators and Virtual Environments*, *19*(1), 35–53.

Ries, A. J., Touryan, J., Vettel, J., McDowell, K., & Hairston, W. D. (2014). A Comparison of Electroencephalography Signals Acquired from Conventional and Mobile Systems. *Journal of Neuroscience and Neuroengineering*, *3*(1), 10–20. Retrieved from http://doi.org/10.1166/jnsne.2014.1092

Robbins, J. (2008). *A Symphony in the Brain: The Evolution of the New Brain Wave Biofeedback*. Grove Press.

Roger, M., & Galand, G. (1981). Operant Conditioning of Visual Evoked Potentials in Man. *Psychophysiology*, *18*(4), 477–482. Retrieved from http://doi.org/10.1111/j.1469-8986.1981.tb02485.x

Ruxton, G. D., & Beauchamp, G. (2008). Time for some a priori thinking about post hoc testing. *Behavioral Ecology*, *19*(3), 690–693. Retrieved from http://doi.org/10.1093/beheco/arn020

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., & Wolpaw, J. R. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, *51*(6), 1034–1043. Retrieved from http://doi.org/10.1109/TBME.2004.827072

Schalk, G., & Mellinger, J. (2010). Contributed Modules. In *A Practical Guide to Brain–Computer Interfacing with BCI2000*, 219–247. Springer London. Retrieved from http://link.springer.com/chapter/10.1007/978-1-84996-092-2_11

Schoenberg, P. L. A., & David, A. S. (2014). Biofeedback for Psychiatric Disorders: A Systematic Review. *Applied Psychophysiology and Biofeedback*, *39*(2), 109–135. Retrieved from http://doi.org/10.1007/s10484-014-9246-9

Sommer, W., & Schweinberger, S. (1992). Operant conditioning of P300. *Biological Psychology*, *33*(1), 37–49. Retrieved from http://doi.org/10.1016/0301-0511(92)90004-E

Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Blefari, M. L., & Sitaram, R. (2013). Real-time fMRI neurofeedback: Progress and challenges. *NeuroImage*, *76*, 386–399. Retrieved from http://doi.org/10.1016/j.neuroimage.2013.03.033

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188.

Szuromi, B., Czobor, P., Komlósi, S., & Bitter, I. (2011). P300 deficits in adults with attention deficit hyperactivity disorder: a meta-analysis. *Psychological Medicine*, *41*(7), 1529–1538. Retrieved from http://doi.org/10.1017/S0033291710001996

Thie, J. (2013). A wireless marker system to enable evoked potential recordings using a wireless EEG system (EPOC) and a portable computer. *PeerJ PrePrints*, *1*, e32v1.

Towler, J., Gosling, A., Duchaine, B., & Eimer, M. (2012). The face-sensitive N170 component in developmental prosopagnosia. *Neuropsychologia*, *50*(14), 3588–3599. Retrieved from http://doi.org/10.1016/j.neuropsychologia.2012.10.017

Wang, C., Xiong, S., Hu, X., Yao, L., & Zhang, J. (2012). Combining features from ERP components in single-trial EEG for discriminating four-category visual objects. *Journal of Neural Engineering*, *9*(5), 056013. Retrieved from http://doi.org/10.1088/1741-2560/9/5/056013

Weiskopf, N. (2012). Real-time fMRI and its application to neurofeedback. *NeuroImage*, *62*(2), 682–692. Retrieved from http://doi.org/10.1016/j.neuroimage.2011.10.009

# Appendix A – EEGLab code

There are 3 scripts which will be reviewed in detail in this Appendix. All 3 are analysis-time matlab scripts. The display-time C++ code changes are shown in Appendix B.

## A.1 – changes to the BCI2000 data import plugin, pop_loadBCI2000

The first change is to edit the function definition to allow for a third option parameter to be used to pass in an array of correct values.

```
function EEG = pop_loadBCI2000(fileName, events, correctin)
```

Secondly, the fields must be edited to have a parameter named StimulusType attached to each data point. This takes the input from the display-side C++ code, which identifies whether a particular stimulus is a target or a non-target. In the case of free-spell data, of course, this is superseded by the correct response array that is input to the function.

```
    these_events = struct(...
        'latency', num2cell(states.(stateNames{k}).latency),...
        'position', num2cell(states.(stateNames{k}).value),...
        'duration', num2cell(states.(stateNames{k}).duration),...
        'StimulusType', num2cell(states.(stateNames{k}).StimulusType));
```

Thirdly, for the free-spell data, the vector of correct input results must be transformed into a row and a column, since each letter appears in one of each. This is done in two steps: first, the input characters are transformed into a value of 1 through 36 representing where on the presentation array they fall. Then, that value is transformed into a row and a column, as shown in the following code:

```
%No GUI, always ignore events provided.
    events = stateNames; %Select all events.
    if ~isempty(correctin)
        for j = 1:length(correctin)
            if (abs(correctin(j))>64)
                correctcode(j) = abs(correctin(j))-64;
            elseif (correctin(j)=='-')
                correctcode(j) = 27;
            else
                correctcode(j) = abs(correctin(j))-21;
            end
            correctcodeR(j) = ceil(correctcode(j)/6);
            correctcodeC(j) = mod(correctcode(j),6)+6;
            if (correctcodeC(j)==6)
                correctcodeC(j) = 12;
            end
        end
```

Finally, for each data value, go through and set a StimulusType corresponding to whether or not it is a target presentation. This particular algorithm could be vastly improved in terms of runtime efficiency, but since it had to run only once per data set, and the overall time was not prohibitive, the runtime optimization has not been done.

```
if (freespell == 1)
    TargetData = states.('SelectedTarget');
    TargetTimes = find([false;diff(TargetData) ~= 0]); %times of new letter presentations
     StimCodeData = states.('StimulusCode');

    for j = 1:length(TargetData) % starting from the first letter

        for k = 1:length(TargetTimes)
            if (j < TargetTimes(k))
                if (k<=length(correctcode))
                    TargetNowR = correctcodeR(k); %pulling actual correct row from command
line
                    TargetNowC = correctcodeC(k); %pulling actual correct column from
command line
                else
                    TargetNowR = 0; %if there are characters after the last spelled char it just
means that the experimenter waited too long to stop it, disregard
                    TargetNowC = 0;
```

```
            end
            break;
         end
      end

      if (StimCodeData(j) > 0) %if the stimulus code is equal to either the row or column
that is to be recognized, set stimulus type to true
         if ((StimCodeData(j) == TargetNowR) || (StimCodeData(j) == TargetNowC))
            StimTypeData(j) = 1;
         end
      end

   end
```

In addition, the file hard codes the scalp positions to be the reversed position of the Emotiv:

```
EEG.chanlocs = struct('labels',
rot90({'P4','P8','CP2','CP6','T8','FC6','FC2','FC1','FC5','T7','CP5','CP1','P7','P3'},3));
```

## A.2 – Script to process the data in EEGLab for each participant

A different version of this script was used for each phrase, but it differed mainly in the input file and output name. The runtime system named all the data files, and then the data files were annotated with the correct responses as indicated on the participant questionnaires. The non-free-spell entries were similarly processed, but without needing to input the correct responses. All other processing steps were the same.

First, the function is declared. Initially, some of the addpath statements happened in the script, but it proved easier to do those once per session, so they are commented out and left here to show that they are necessary:

```
function argout = processstudy3free()
```

```
%make sure paths below are added, eeglab started, and the pop_loadBCI2000 version set
properly to free
%addpath('F:\\MATLAB\\signal\\signal');
%addpath('F:\\MATLAB\\stats\\stats');
%addpath('F:\\MATLAB\\eeglab13_4_4b');
cd F:\\MATLAB\\eeglab13_4_4b
%eeglab
pause(10);
cd plugins\bCI2000import0.36
```

Next, the output directory is declared and input files and correct responses are drawn from a file:

```
fp_out = 'F:\\eegstudy\\study3\\processed\\';
% loop over files and folders
datafile = 'D:\\studydata\\study3\\out_free.txt';
fid = fopen(datafile);
while ~feof(fid)
    file = fgetl(fid);
    if ~isempty(file)

        % extract just the filename (fn) and filepath (fp)
        t = findstr('\',file);
        t1 = findstr(',',file);
        fn = file((t(length(t))+1):(length(file)-4));
        fp = file(t1(1)+1:t(length(t)));
        correctin = file(1:t1(1)-1);
```

The following operations are now performed for each file in the data set, resulting in a .set file for non-targets, targets, and combined:

**Table A.1: EEGLab data analysis and rejection steps**

| | |
|---|---|
| Load BCI2000 data file | z1 = pop_loadBCI2000([fp fn '.dat'],'',correctin); |
| Load the standard scalp positions definition file | z1 = pop_chanedit(z1, 'lookup','F:\\MATLAB\\eeglab_13_2\\eeglab13_2_2b\\plugins\\dipfit2.3\\standard_BESA\\standard-10-5-cap385.elp'); |
| Select only those events that are needed | z1 = pop_selectevent( z1, 'type',{'SelectedColumn' 'SelectedRow' 'SelectedTarget' 'SpellerMenu' 'StimulusBegin' 'StimulusCode' 'StimulusCodeRes' 'StimulusType' 'StimulusTypeRes'},'deleteevents','on'); |

| | |
|---|---|
| Save out a combined data set. This isn't used for the analysis, but could be useful is other operations are desired | z2 = pop_saveset( z1, 'filename',[fn '.set'],'filepath',fp_out); |
| For each StimulusBegin event, start a new data epoch. Note that these are overlapping | z2 = pop_epoch( z2, {  'StimulusBegin'  }, [0 0.8], 'newname', 'Imported BCI2000 data set epochs', 'epochinfo', 'yes'); |
| Clean out non-target epochs in which there is too great of a slope in the data (too much drift) or in which there are highly improbable data points. | % reject slopes of greater than 50 uV over 50 points in the 4 channels of interest<br>z2 = pop_rejtrend(z2, 1, [1,2,13,14], 50, 50, .3, 0, 1 );<br>% reject improbable data - outside 3 SD<br>z2 = pop_jointprob(z2, 1, [1,2,13,14], 3, 3, 0, 1, 0); |
| Delete those epochs that contain target events | z2 = pop_selectevent( z2, 'StimulusType',1,'deleteevents','off','deleteepochs','on','invertepochs','on'); |
| Rebase the epochs | z2 = pop_rmbase( z2, [0      789.0625]); |
| Save these epochs out as non-target events | z3 = pop_saveset( z2, 'filename',[fn '_nontargets.set'],'filepath',fp_out); |
| Note that we are starting from the combined set, delete those events that are not target events | z3 = pop_selectevent( z1, 'type',{'StimulusBegin'},'StimulusType',1,'deleteevents','on'); |
| Re-epoch based on the events as they now exist – note that some overlap will still occur but it will be minimal since target events are relatively infrequent | z3 = pop_epoch( z3, {  'StimulusBegin'  }, [0 0.8], 'newname', 'Imported BCI2000 data set_ones epochs', 'epochinfo', 'yes'); |
| Clean out target epochs in which there is too great of a slope in the data (too much drift) or in which there are highly improbable data points. | % reject slopes of greater than 50 uV over 50 points in the 4 channels of interest<br>z3 = pop_rejtrend(z3, 1, [1,2,13,14], 50, 50, .3, 0, 1 );<br>% reject improbable data - outside 3 SD<br>z3 = pop_jointprob(z3, 1, [1,2,13,14], 3, 3, 0, 1, 0); |
| Rebase the epochs | z3 = pop_rmbase( z3, [0      789.0625]); |
| Save this out as the set of target events | z3 = pop_saveset( z3, 'filename',[fn '_targets.set'],'filepath',fp_out); |

### A.3 – Create an EEGLab study from the results

This script uses the same input file as above to identify the individual data files, along with their meta-data (subject number, session number, etc.). It then uses those to build an

EEGLab study structure that can be used for analysis. There seems to be a requirement that the study file be initialized with a valid file and all of the study-level information input first. We will then discard the data file as we start to loop through the input files from (in this case) out_free.txt:

```
[study1 z1] = std_editset([],[], 'commands', {{'index' 1 'load'
'F:\eegstudy\study3\processed\subj1S001R05_targets.set' 'subject' 'subj1' 'condition' 'nt'
'session' 1 }},'name', 'study 3 crazy cats', 'filename', 'study3cats', 'filepath',
'F:\\eegstudy\\study3', 'updatedat', 'on', 'savedat', 'on');
index = 0;

% loop over files and folders

datafile = 'D:\\studydata\\study3\\out_free.txt';

fid = fopen(datafile);
while ~feof(fid)
    file = fgetl(fid);
    if ~isempty(file)
```

For each file that is not empty, find the appropriate details in the filename, extract those into separate variables, and then create the entry in the study structure for target and non-targets, respectively:

```
       % extract just the filename (fn) and filepath (fp)
        t = findstr('\',file);
        fn = file((t(length(t))+1):(length(file)-4));
         fp = 'F:\eegstudy\study3\processed\';
        file_nt = strcat(fp,fn,'_nontargets.set');
        file_t = strcat(fp,fn,'_targets.set');

        t = findstr('S00',fn);
        subject = fn(1:t-1);
        s = fn(t+3);
        cond = 'target';
        condnt = 'nt';

            index = index + 1;
                                    [study1 z1] = std_editset(study1, z1, 'commands', {{'index' index
'load' file_nt 'subject' subject 'condition' condnt 'session' s }});
            index = index + 1;
```

```
                              [study1 z1] = std_editset(study1, z1, 'commands', {{'index' index
'load' file_t 'subject' subject 'condition' cond 'session' s }});
disp(file_nt);

   end
end
```

Finally, the study is saved and the function returned:

```
[study1] = pop_savestudy(study1, z1);
fclose(fid);
argout = 1;
return;
```

## A.4 – operations in the EEGLab GUI

From this point, the analysis took place manually rather than through scripts. The following operations were performed:

- Fill out the study details
- Create a study design with condition as the first independent variable and session as the second
- Pre-compute channel measured using single trial data, and saving ERPs
- Plot the channel data, first selecting the electrodes of interest and all users, then selecting the desired statistics parameters (Holms correction, p value cut-off, etc.), then selecting the chart parameters to put the results on the same table and average across channels
- Format and save the images

# Appendix B – BCI2000 modifications and configuration parameters

There are three sets of changes to the P3 speller code in BCI2000 that are important for these studies. The first are the configuration settings for presentation. Many of these have been reported in the main text of the dissertation, but are gathered here for completeness. The second section of this Appendix are the changes needed to pass whether a particular stimulus presentation was a target or non-target stimulus. The third is a discussion of the changes necessary to present the card game. The code for the card game is too long to be worth including in its entirety, but can be retrieved from: https://github.com/derekja/P3Speller.

## B.1 BCI2000 configuration settings

The BCI2000 presentation UI requires that you set a large number of parameters in order to successfully complete a study. The parameters for the P3 speller are reasonably well documented, but since the card game is new, it is not documented. The .prm file defines these parameters, and is loaded from the BCI2000 user interface by the experimenter. The complete parameter files are in the github repository, but a few items are worth noting explicitly here:

The first set of configuration settings are defined by the headset being used. The Emotiv Epoc has a 128Hz sample rate and 14 channels, and transmits 4 blocks of data at a time.

```
Source:Signal%20Properties:DataIOFilter int SourceCh= 14 16 1 % // number of digitized and stored channels
Source:Signal%20Properties:DataIOFilter int SampleBlockSize= 4 32 1 % // number of samples transmitted at a time
Source:Signal%20Properties:DataIOFilter int SamplingRate= 128 256Hz 1 % // sample rate
Source:Signal%20Properties:DataIOFilter list ChannelNames= 14 AF3 F7 F3 FC5 T7 P7 O1 O2 P8 T8 FC6 F4 F8 AF4 // list of channel names
```

The next settings are to establish a naming scheme and a data directory for the resulting files. In this study, the study name and subject number were used for the SubjectName, the day was used for the session number, and the run was the trial.

```
Storage:Data%20Location:DataIOFilter string DataDirectory= C:\Users\derek\Desktop\data ..\data % % // path to top level data directory (directory)
Storage:Session:DataIOFilter string SubjectName= nfa_pilots Name % % // subject alias
Storage:Session:DataIOFilter string SubjectSession= 001 001 % % // three-digit session number
Storage:Session:DataIOFilter string SubjectRun= 01 00 % % // two-digit run number
```

Some filtering is beneficial. The powerline notch filter helps reduce interference from the 60Hz powerline frequency. A high pass filter was set to remove any very low frequency noise. No low pass filter was used in these studies. All channels were aligned automatically.

```
Source:Source%20Filter:SourceFilter int NotchFilter= 2 0 0 2 // Power line notch filter: 0: disabled, 1: at 50Hz, 2: at 60Hz (enumeration)
Source:Source%20Filter:SourceFilter int HighPassFilter= 1 0 0 2 // Source high pass filter: 0: disabled, 1: at 0.1Hz, 2: at 1Hz (enumeration)
Source:Source%20Filter:SourceFilter int LowPassFilter= 0 0 0 4 // Source low pass filter: 0: disabled, 1: at 9Hz, 2: at 30Hz, 3: at 40Hz, 4: at 70Hz (enumeration)
Source:Alignment:AlignmentFilter int AlignChannels= 1 0 0 1 // align channels in time (0=no, 1=yes)
```

The spatial filtering was performed using a common average reference, with all channels being averaged. This was only for run-time classification. If any channels were omitted (for instance if a sensor lost connectivity), then the system was to ignore that rather than flagging it as an error.

```
Filtering:SpatialFilter int SpatialFilterType= 3 2 0 3 // spatial filter type 0: none, 1: full matrix, 2: sparse matrix, 3: common average reference (CAR) (enumeration)
Filtering:SpatialFilter:SpatialFilter intlist SpatialFilterCAROutput= 0 // when using CAR filter type: list of output channels, or empty for all channels
Filtering:SpatialFilter:SpatialFilter int SpatialFilterMissingChannels= 0 0 0 1 // how to handle missing channels 0: ignore, 1: report error (enumeration)
```

Set the epoch length to 796 ms and the number of epochs to average. This must correspond to the number of epochs actually presented in the display timings section.

```
Filtering:P3TemporalFilter int EpochLength= 796ms 500ms 0 % // Length of data epoch from stimulus onset
Filtering:P3TemporalFilter int EpochsToAverage= 15 1 0 % // Number of epochs to average
```

The window sizes were set based on the monitor being used, in order to allow the display to take up the whole screen against a black background.

```
Application:Application%20Window:ApplicationWindow int WindowWidth= 1280 640 0 % // width of Application window
Application:Application%20Window:ApplicationWindow int WindowHeight= 1024 480 0 % // height of Application window
Application:Application%20Window:ApplicationWindow int WindowLeft= 0 0 % % // screen coordinate of Application window's left edge
Application:Application%20Window:ApplicationWindow int WindowTop= 0 0 % % // screen coordinate of Application window's top edge
Application:Application%20Window:ApplicationWindow string WindowBackgroundColor= 0x000000 0x505050 % % // Application window background color (color)
```

Set a four second delay the start and end of each session – half a second before, and 3.5 seconds after each card to be presented. Stimulus is presented for 40ms with 80ms in between presentations, for an overall rate of 120 ms for each stimulus presentation. Since nothing is intensified in the period between 40 ms intensifications, visual persistence increases the perceived duration of each presentation. If the intensification is a larger proportion of each time period, participants have a harder time defocusing their attention for the next presentation than if there is a blank gap. The number of sequences in a set of intensifications must match the epochs to the average parameter described earlier.

```
Application:Sequencing:StimulusTask float PreRunDuration= 4s 1 0 % // pause
preceding first sequence
Application:Sequencing:StimulusTask float PostRunDuration= 4s 0 0 % // pause
following last sequence
Application:Sequencing:StimulusTask float PreSequenceDuration= 0.5s 2s 0 % // pause
preceding sequences/sets of intensifications
Application:Sequencing:StimulusTask float PostSequenceDuration= 3.5s 2s 0 % // pause
following sequences/sets of intensifications
Application:Sequencing:StimulusTask float StimulusDuration= 2 40ms 0 % // stimulus
duration
Application:Sequencing:StimulusTask float ISIMinDuration= 2 80ms 0 % // minimum
duration of inter-stimulus interval
Application:Sequencing:StimulusTask float ISIMaxDuration= 2 80ms 0 % // maximum
duration of inter-stimulus interval
Application:Result%20Processing:StimulusTask int InterpretMode= 1 0 0 2 //
interpretation of results: 0 none, 1 online free mode, 2 copy mode (enumeration)
Application:Result%20Processing:StimulusTask int DisplayResults= 0 1 0 1 // display
results of copy/free spelling (boolean)
Application:Sequencing:P3SpellerTask int NumberOfSequences= 15 15 1 % // number of
sequences in a set of intensifications
```

## B.2 Target vs. non-target stimuli

This is a distinction which the system can only know in some of the cases. In free spelling, where the participant has provided their own phrase to spell, and in the card game, where the participant is deciding which card to turn over, there is no way for the system to know a priori what the intended target is. For these cases, the experimenter must record the intended target and then enter it during the analysis phase as described in Appendix A. In the case of copy mode spelling, though, which is the main presentation mode used for training the linear classifier, the system knows the correct answer in advance and can supply that to the analysis scripts. Due to the differences in event structure between the BCI2000 format and that used in EEGlab, it is useful to explicitly record target events within the BCI2000 results file and pass that on to the conversion program described in Appendix A. The OnNextStimulusCode function is where this occurs. The code sections are contiguous, but broken up in this Appendix section with some explanatory comments beyond those in the file itself.

```
int
P3SpellerTask::OnNextStimulusCode()
{
  // Return values of this function determine sequencing in the following way:
  //  A zero stimulus code ends the current sequence of stimuli.
  //  A null sequence (no nonzero stimulus codes between two zero codes) ends
  //  the run.
  int result = 0;
  bool outputMarker = false;
```

Since the first sequence of intensifications is actually the cards turned upside down, we do not want to record stimulus targets during that trial. This is a workaround to the problem of the BCI2000 programming model not easily supporting throwaway trial blocks. This next section is simply determining which stimulus is going to be intensified in the next round, or if the sequence is finished.

```
  outputMarker = !mFirstSequence;
  if( !mSequence.empty() )
  {
    if( mSequencePos == mSequence.end() )
    { // During a run, we always use the same sequence object and re-shuffle it.
      //
      // Sequences should fulfil the constraint that no stimulus
      // presented on the previous sequence's last stimulus presentation
      // may be presented on the next sequence's first stimulus presentation
      // (unless this is impossible by the way stimuli are grouped).
```

```
      int prevStimulusCode = *mSequence.rbegin();
      do
      {
        random_shuffle( mSequence.begin(), mSequence.end(), RandomNumberGenerator
);
      } while( mAvoidStimulusRepetition
               && Associations().StimuliIntersect( *mSequence.begin(), prevStimulusCode )
);

      mSequencePos = mSequence.begin();
      if( ++mSequenceCount == mNumberOfSequences )
      {
        result = 0;
        mSequenceCount = 0;
                        mFirstSequence = true;
                        outputMarker = false;
      }
      else
      {
        result = *mSequencePos++;
                        mFirstSequence = false;
      }
    }
    else
    {
      result = *mSequencePos++;
              mFirstSequence = false;
    }
  }
```

Here begins the section where the determination of target versus non-target is made. A presentation is a target if it is in the row and the column of the letter to be spelled. This “target” versus “non-target” marker is sent to both the data file and the log, which is presented to the experimenter as the study is underway.

```
  // here want access to copy text to be spelled (in terms of which two sequence positions
it occupies, send target/non-target on that basis

  if (outputMarker) {

              if( mInterpretMode_ == InterpretModes::Copy ) {
```

```
                        const char *Markertypes[] = {"start", "target", "non-target"};

  int targetID = NULL;
//  int i = 0;
  string mrk = "xx";

  int pRow    = targetID ? ( targetID - 1 ) / mNumMatrixCols + 1 : 0;
  int pCol = targetID ? ( targetID - 1 ) % mNumMatrixCols + 1 : 0;

  if ((result==pRow) || (result==pCol)) {
                        mrk = Markertypes[1];
  }
  else {
            mrk = Markertypes[2];
  }

                        mStreamOutlet.push_sample(&mrk);

                         AppLog << "mSequencePos: "
                                   << result
                                   << "        mrk: "
                                   << mrk
                                  << endl;
            }
  }

  return result;
}
```

## B.3 Card game presentation changes

There are two primary sections that are different between the speller and the card game. The first is the stimulus display, while the second is the game logic. A smaller change is the intensification pattern of individually intensifying cards rather than in a row or column. Please see the github repository for the details of the intensification and display changes, as they are well documented in the source code. Note that some discussion of the game logic will be made below; again, the code is contiguous but is broken here by comments.

The card game takes place in the free spell mode only – as mentioned above, since there is no *a priori* knowledge by the system of what response is correct, the copy mode is not workable.

```
// code for game logic of flipping and matching is in free mode only
  if( mInterpretMode_ == InterpretModes::Free ) {
                          // variable to hold whether we need to flip everything back
                                    bool fnd = false;
                          // if firstresult then flip card up appropriately
```

The first set of intensifications is handled differently so as to present the cards to the user face-up in order to be remembered. This next code block applies to all EXCEPT the first run. First, one needs to see if this is the first card to be matched, or if there is already a current selection. If there is no current selection, then the next time through this code there will be, so set mFirstMach to 1. The next block of code is checking to make sure that the card selected is not already turned over (i.e. it was not from a previous match). If it is, then complain to the user and return to an initial match attempt by setting mFirstMatch back to 0. Otherwise, turn the card over.

```
                          if (mFirstMatch == 0) {
                                      mFirstMatch = 1;
                                    for( SetOfStimuli::const_iterator i = mStimuli.begin(); i !=
mStimuli.end(); ++i )
                                    {
                                       ImageStimulus* ims = dynamic_cast<ImageStimulus*>( *i );
                                       TextStimulus* txt = dynamic_cast<TextStimulus*>( *i );
                                       if (ims != NULL) {
                                                // if the tags match, flip the card up
```

```
                if (pTarget->Tag()==ims->Tag()) {
                    if (mAsocFile[ims->Tag()].Tag == 2) {
                        DisplayMessage("already matched!");
                        mFirstMatch = 0;
                    }
                    else {
                        ims->SetFile( mAsocFile[ims-
>Tag()].Name );
                        mAsocFile[ims->Tag()].Tag = 1;
                    }
                }
            }
        }
          }
```

Assuming that there is already a selection, and that the user has just turned over another card, see if it matches the first. The indenting was slightly too deep for easy readability here, so interior loops and ‘if statements’ have been unindented for legibility. If the second card chosen is already flipped up, then it has been matched previously; tell the user, flip the cards back down, and go back to a first selection. If the same card was chosen for the second selection as was selected first, tell the user and go back to a first selection. If this is a new card to be matched and it is not the same as the first selection, then see if they match. If so, then keep them flipped up and tag the cards as matched.

```
else if (mFirstMatch == 1) {
     mFirstMatch = 0;
    for( SetOfStimuli::const_iterator i = mStimuli.begin(); i != mStimuli.end(); ++i )
        {
           ImageStimulus* ims = dynamic_cast<ImageStimulus*>( *i );
           TextStimulus* txt = dynamic_cast<TextStimulus*>( *i );
            if (ims != NULL) {
                // if the tags match, flip the card up
                if (pTarget->Tag()==ims->Tag()) {
                    if (mAsocFile[ims->Tag()].Tag == 2) {
                        DisplayMessage("already matched!");
                    }
                    else if (mAsocFile[ims->Tag()].Tag == 1) {
                        DisplayMessage("you just chose that!");
```

```
                    // will depend on unflip loop at end to do this
                    //ims->SetFile( "images\\redback.gif" );
                    //mAsocFile[ims->Tag()].Tag = 0;
                }
                    else {
                        for (int j=1;j<90;j++) {
                        if (mAsocFile[j].Name == mAsocFile[ims-
>Tag()].Name) {
                        if (mAsocFile[j].Tag == 1) {
                            DisplayMessage("Good match!");
                            ims->SetFile( mAsocFile[ims-
>Tag()].Name );

                            mAsocFile[ims->Tag()].Tag = 2;
                            mAsocFile[j].Tag = 2;
                            fnd = true;
                        }
                        }
                        }
                    }
                    // found our match, break out of iterator
                    break;
                    }
                    }
            }
```

If any problems occurred in the match, i.e. the card was already selected or didn’t match the first selection, then the two cards need to be turned back to a face-down position to allow the user to select them in another round. This is also where the check is made which stops the game—if there are less than two cards that are still face-down, then there is nothing left to be matched.

```
        // flip all cards that aren't already matched back over, set any 1's in the
mAsocFile back to zero
            int cardup = 0;
            for( SetOfStimuli::const_iterator i = mStimuli.begin(); i !=
mStimuli.end(); ++i )
         {
            ImageStimulus* ims = dynamic_cast<ImageStimulus*>( *i );
            if (ims != NULL) {
                if (mAsocFile[ims->Tag()].Tag == 1) {
```

```
                    ims->SetFile( "images\\redback.gif" );
                    mAsocFile[ims->Tag()].Tag = 0;
                }
                if (mAsocFile[ims->Tag()].Tag == 0) cardup++;
            }
        }
        if (cardup<2) StimulusTask::StopRun();
    }
}
```

# Appendix C: Experimental Factors

## C.1 Running the experiment

Electrode fitting is critical, and it is difficult to get good performance with a poorly-fitted headset. Some factors that make fit more difficult are long hair, overly dry hair, and a small head size. For this reason, the effort to balance gender in the current studies was also important in balancing headset fitting issues; male subjects tended to have shorter hair and larger heads.

On the first day, after the headset was fitted, a short phrase was spelled as an orienting task. On all other days, a long phrase was spelled, and then the linear classifier was trained in order to emphasize the most diagnostic times and electrodes in predicting a P300 signal. After this initial linear classifier training, most participants were able to achieve recognition levels where the results no longer appeared random, and it was easier to get them to exert sufficient effort to continue the training. Three phrases (between 9 and 15 letters per phrase) seemed to be sufficient data to build well-performing models for most participants; however, if the training phrases were not fully attended to, the models were not effective, so the positive feedback of an intermediate training session was important. If a model could not be trained, for example, if there was too much inconsistency in the data to come up with a single set of weights for the model, then that participant was removed from the study. In earlier rounds of data collection, this was a common occurrence; however, once sufficient attention was paid to headset fitting, this no longer occurred during the final 12 participant study.

Even with a properly fitted headset and sufficient attention to model training, there are factors which greatly influence accuracy and P300 strength. Material design factors, as outlined in the literature review, apply; such target trials should be no more than 10% of the overall trials (Picton, 1992). There are also cognitive factors that have been less rigorously studied that may influence performance (Picton, 1992; Datta et al, 2008). Time of day can also affect results: too early can be difficult, but more often, too late in the day is problematic. In the pilot studies, attentional capacity appeared to be lower later in the day, particularly after other classes or exams. This was not rigorously studied in this research, but in the training study, an attempt was made to have participants test at the same time each day.

On long phrases, performance often tends to be better towards the beginning of the phrase—attention is progressively more difficult to maintain the longer the phrase is. This was noted verbally by participants and corresponds with previous studies (Datta, 2008). Long sessions appear to cause the same effect; the last phrase of the session is more difficult than phrases in the middle of the session, both from a perceived effort standpoint and from a letter recognition accuracy perspective. Due to these inter-trial factors, most comparisons in this study are made on the same phrase on different days, and those that by necessity must be across trials are made on phrases of approximately the same length.

Caffeine and other drugs may also play a role. This was not looked at rigorously in the current studies, but deserves further examination. Amount of coffee, in particular, was a variable that appears potentially relevant in the mornings. Benadryl and anti-anxiety medications were mentioned in 2 cases where a participant who had previously shown strong results had a day in which they were unable to achieve effective letter recognition rates. Participants were asked about these potentially relevant factors on the questionnaire, shown in Appendix D.

Two factors that are further under experimenter control are also worth noting. It appears to be very hard for a participant to get back into providing correct recognitions once a mistake is made within a phrase. Avoiding external distractions is important in not having these recognition mistakes begin. Loud noises or speech in the background that captures a participant's attention are particularly disruptive. It is also useful to wait when starting each phrase until the display of the participant's brainwaves is quite settled; starting too early can lead to a mistake on the first letter of a phrase, which is then hard to recover from. These causes of poor performance are summarized in table 3.

These effects are not a certainty, and are not claimed as study results in this dissertation. In an effort to maintain consistent study parameters among individuals, however, they were factors on which some consistency and control was exerted. The more consistency exerted on these factors, and others noted in the literature (Picton, 1992; Datta, 2008; Duvinage, 2012), the more comparable performance may be between study participants, and with the same participant in different sessions.

| Problem | Type | Solution |
|---|---|---|
| Poor electrode connectivity | Headset fitting | Wiggle electrodes under hair, apply more saline solution |
| Dry hair | Headset fitting | More saline solution was applied |
| Long phrases | Attention | Phrases of 15 characters or less were used |
| Long sessions | Attention | Sessions were kept under 1 hour, 45 minutes |
| Recovering after a mistake | Attention | Study materials focused on short phrases and care was taken to eliminate distractions |
| Evenings or when tired or nervous | Mental state | Tests were scheduled at the participants preferred time, and where possible at the same time on each day of the study |
| Caffeine and other drugs | Mental state | Caffeine use was reported on each day’s questionnaire. |
| Starting too soon before brainwaves have settled | Distractions | The experimenter would wait until brainwaves settle before starting a trial |
| Loud noises or speech | Distractions | Testing took place in a quiet, distraction-free environment |

**Table C.1: Potential causes of poor EEG performance that were either controlled for or noted on questionnaire data**

# Appendix D – Study materials

This Appendix contains the stimulus materials, questionnaires, and questionnaire responses for the main training study. The two-session study and memory attention study used the same questionnaires, but differed in the number of days in the study and the stimulus presentation order, these differences are detailed here. This Appendix also contains the ethics application for these studies.

## D.1 Training effect study

### D.1.1 Stimulus presentation order

day 1

Favourite colour

phrase 1 "cherry blossoms"

phrase 2 "jumping fish"

phrase 3 "raging rivers"

train model (do pre-test questionnaire)

phrase 4 "crazy cats"

prompt 1 (write answer then spell)

day 2

favourite food

phrase 1 "cherry blossoms"

phrase 2 "jumping fish"

phrase 3 "raging rivers"

train model (do pre-test questionnaire)

phrase 4 "crazy cats"

prompt 1 (write answer then spell)

day 3
favourite city

phrase 1 "cherry blossoms"

phrase 2 "jumping fish"

phrase 3 "raging rivers"

train model (do pre-test questionnaire)

phrase 4 "crazy cats"

prompt 1 (write answer then spell)

phrase 5 “double rainbow”

day 4
favourite month

phrase 1 "cherry blossoms"

phrase 2 "jumping fish"

phrase 3 "raging rivers"

train model (do pre-test questionnaire)

phrase 4 "crazy cats"

prompt 1 (write answer then spell)

phrase 5 “double rainbow”

day 5

favourite animal

phrase 1 "cherry blossoms"

phrase 2 "jumping fish"

phrase 3 "raging rivers"

train model (do pre-test questionnaire)

phrase 4 "crazy cats"

prompt 1 (write answer then spell)

phrase 5 “double rainbow”

### D.1.2 Questionnaire

(Only the day 1 questionnaire is shown. Subsequent days were identical except for the prompt for the participant’s own phrase, which went what is your favourite colour, food, city, month. Animal for days 1 through 5 respectively).

Participant number:

Day 1

Pre-test questions (actually answered as a break between tasks)

1) Do you generally concentrate well? 1- no, not at all, 7- yes, always

2) How do you feel today? 1- horrible, 7-fantastic

3) Do you have any conditions that could affect today’s test? Neurological conditions, ADHD, etc. yes, no, prefer not to say (discuss with experimenter if yes)

4) What is your gender? Male, female, other or prefer not to say

5) How much did you sleep last night? 1- less than 5 hours, 7- 8 or more hours

6) How good a night's sleep was that for you? 1- horrible, 7- fantastic

7) Have you had caffeine today? 1- none, 7-7 or more cups of coffee (or equivalent)

8) Do you generally drink this amount of caffeine? 1- usually much less, 7-usually much more

9) Do you feel tense today? 1- totally mellow, 7- way stressed

10) Are you dehydrated? 1- not at all, 7- very dehydrated

Prompt #1 (complete when asked)

What is your favourite colour? ________________________________________

Post-test questions

1) How hard did you have to concentrate to make the system respond well? 1- not at all, 7 – concentrated very hard

2) Did it respond to your brainwaves well? 1- not at all, 7 – amazingly well

3) How hard was it to maintain attention? 1- easy, 7- very hard

4) Did you feel tired? Did your ability to pay attention fall off at some point in the study? 1-never, 7- frequently

5) Was the headset comfortable? 1- very uncomfortable, 7- very comfortable

6) Was the saline solution irritating? 1- very irritating, 7- not at all

7) Would you wear the headset in public? 1- no way, 7- sure, no problem

### D.1.3 Questionnaire responses

| | | pre-test | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Participant number | day | Ques 1 | Ques 2 | Ques 3 | Ques 4 | Ques 5 | Ques 6 | Ques 7 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 5.5 | 4 | no | f | 8 | 5.5 | 1 |
| 1 | 2 | 4.5 | 5 | no | f | 7 | 5 | 1 |
| 1 | 3 | 4.5 | 6 | no | f | 8 | 6.5 | 3 |
| 1 | 4 | 4.5 | 6 | no | f | 7 | 7 | 5 |
| 1 | 5 | 4.5 | 2.5 | no | f | 8 | 3 | 2 |
| 2 | 1 | 5 | 6 | no | f | 3 | 6 | 1 |
| 2 | 2 | 3 | 5 | no | f | 2 | 3 | 1 |
| 2 | 3 | 5 | 3 | no | f | 2 | 3 | 1 |
| 2 | 4 | 4 | 7 | no | f | 5 | 5 | 2 |
| 2 | 5 | 5 | 7 | no | f | 3 | 6 | 1 |
| 4 | 1 | 4 | 4 | anxiety | f | 8 | 6 | 2 |
| 4 | 2 | 4 | 5.5 | anxiety | f | 7 | 6.5 | 2 |
| 4 | 3 | 4 | 5.5 | anxiety | f | 6.5 | 5 | 2 |
| 4 | 4 | 4 | 3 | anxiety | f | 7 | 5 | 3 |
| 4 | 5 | 4 | 5.5 | anxiety | f | 7 | 5 | 3 |
| 5 | 1 | 4 | 4 | no | m | 2 | 5 | 1 |
| 5 | 2 | 3 | 4 | no | m | 7 | 7 | 2 |
| 5 | 3 | 4 | 3 | no | m | 4 | 6 | 2 |
| 5 | 4 | 4 | 4 | no | m | 5 | 5 | 1 |
| 5 | 5 | 4 | 3 | no | m | 6 | 5 | 1 |
| 6 | 1 | 5 | 6 | no | m | 8 | 4 | 2 |
| 6 | 2 | 5 | 6 | no | m | 8 | 5 | 1 |
| 6 | 3 | 5 | 5 | no | m | 8 | 5 | 1 |
| 6 | 4 | 5 | 5 | no | m | 8 | 6 | 2 |
| 6 | 5 | 5 | 5 | no | m | 8 | 6 | 2 |
| 7 | 1 | 5 | 6 | no | m | 4 | 3 | 2 |
| 7 | 2 | 5 | 6 | no | m | 4 | 3 | 1 |
| 7 | 3 | 5 | 6 | no | m | 4 | 3 | 2 |
| 7 | 4 | 5 | 6 | no | m | 4 | 3 | 2 |
| 7 | 5 | 5 | 3 | no | m | 3 | 3 | 2 |
| 8 | 1 | 5 | 4 | fatigue | f | 5 | 6 | 1 |
| 8 | 2 | 5 | 6 | hungry | f | 7.5 | 6 | 1 |
| 8 | 3 | 5 | 5 | fatigue | f | 7 | 6 | 1 |
| 8 | 4 | 5 | 6 | no | f | 7 | 6 | 1 |
| 8 | 5 | 5 | 3 | tired | f | 5 | 4 | 2 |
| 9 | 1 | 5 | 5 | no | f | 8 | 6 | 2 |
| 9 | 2 | 5 | 5 | no | f | 6 | 6 | 1 |
| 9 | 3 | 5 | 4 | no | f | 7 | 6 | 1 |
| 9 | 4 | 5 | 4 | no | f | 7 | 6 | 1 |
| 9 | 5 | 5 | 6 | no | f | 1 | 2 | 2 |
| 10 | 1 | 6 | 7 | no | m | 1 | 2 | 1 |
| 10 | 2 | 6 | 5 | no | m | 7 | 7 | 1 |
| 10 | 3 | 6 | 6 | no | m | 7 | 7 | 3.5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 10 | 4 | 6 | 6 | no | m | 7 | 6 | 3.5 |
| 10 | 5 | 6 | 7 | no | m | 7 | 7 | 1 |
| 11 | 1 | 5 | 6 | no | m | 10 | 6 | 2 |
| 11 | 2 | 6 | 6 | no | m | 7 | 6 | 2 |
| 11 | 3 | 6 | 5 | no | m | 6 | 6 | 1 |
| 11 | 4 | 5 | 6 | no | m | 7 | 6 | 2 |
| 11 | 5 | 5 | 6 | no | m | 7 | 6 | 1 |
| 12 | 1 | 4 | 4 | no | m | 8 | 4 | 0 |
| 12 | 2 | 4 | 4 | no | m | 8 | 5 | 1 |
| 12 | 3 | 4 | 2 | no | m | 5 | 2 | 0 |
| 12 | 4 | 4 | 4 | no | m | 8 | 5 | 1 |
| 12 | 5 | 4 | 4.5 | no | m | 7 | 3 | 1 |
| 13 | 1 | 3 | 5 | no | f | 2 | 4 | 2 |
| 13 | 2 | 6 | 6 | no | f | 6 | 6 | 2 |
| 13 | 3 | 5 | 6 | no | f | 5 | 5 | 2 |
| 13 | 4 | 5 | 6 | no | f | 6 | 5 | 2 |
| 13 | 5 | 5 | 5 | no | f | 7 | 6 | 3 |

**Table D.1: Pre-test questionnaire results**

| Participant number | day | Ques 8 | Ques 9 | Ques 10 | Phrase 4 length | Phrase 4 correct | % correct |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 5 | 5 | 6 | 9 | 7 | 0.777778 |
| 1 | 2 | 4 | 3 | 3 | 9 | 7 | 0.777778 |
| 1 | 3 | 4 | 4.5 | 6 | 9 | 4 | 0.444444 |
| 1 | 4 | 3 | 5 | 6 | 9 | 7 | 0.777778 |
| 1 | 5 | 4 | 5 | 1.5 | 9 | 5 | 0.555556 |
| 2 | 1 | 3 | 3 | 3 | 9 | 3 | 0.333333 |
| 2 | 2 | 2 | 3 | 2 | 9 | 6 | 0.666667 |
| 2 | 3 | 2 | 4 | 1 | 9 | 6 | 0.666667 |
| 2 | 4 | 2 | 2 | 3 | 9 | 7 | 0.777778 |
| 2 | 5 | 2 | 3 | 2 | 9 | 7 | 0.777778 |
| 4 | 1 | 3 | 6 | 2 | 9 | 6 | 0.666667 |
| 4 | 2 | 4 | 6.5 | 6.5 | 9 | 1 | 0.111111 |
| 4 | 3 | 4 | 6.5 | 1 | 9 | 5 | 0.555556 |
| 4 | 4 | 4 | 6 | 2 | 9 | 6 | 0.666667 |
| 4 | 5 | 4 | 6 | 2 | 9 | 4 | 0.444444 |
| 5 | 1 | 4 | 2 | 3 | 9 | 4 | 0.444444 |
| 5 | 2 | 3 | 3 | 4 | 9 | 4 | 0.444444 |
| 5 | 3 | 4 | 3 | 5 | 9 | 1 | 0.111111 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5 | 4 | 3 | 3 | 5 | 9 | 8 | 0.888889 |
| 5 | 5 | 2 | 2 | 3 | 9 | 6 | 0.666667 |
| 6 | 1 | 4 | 5 | 4 | 9 | 6 | 0.666667 |
| 6 | 2 | 5 | 5 | 4 | 9 | 8 | 0.888889 |
| 6 | 3 | 3 | 5 | 4 | 9 | 6 | 0.666667 |
| 6 | 4 | 4 | 4 | 3 | 9 | 9 | 1 |
| 6 | 5 | 4 | 4 | 3 | 9 | 8 | 0.888889 |
| 7 | 1 | 3 | 3 | 2 | 9 | 9 | 1 |
| 7 | 2 | 2 | 4 | 3 | 9 | 8 | 0.888889 |
| 7 | 3 | 4 | 4 | 3 | 9 | 9 | 1 |
| 7 | 4 | 3 | 4 | 3 | 9 | 9 | 1 |
| 7 | 5 | 3 | 5 | 3 | 9 | 9 | 1 |
| 8 | 1 | 2 | 6 | 4 | 9 | 4 | 0.444444 |
| 8 | 2 | 2 | 6 | 4 | 9 | 9 | 1 |
| 8 | 3 | 2 | 6 | 3 | 9 | 9 | 1 |
| 8 | 4 | 2 | 3 | 3 | 9 | 7 | 0.777778 |
| 8 | 5 | 2 | 6 | 4 | 9 | 8 | 0.888889 |
| 9 | 1 | 1 | 4 | 1 | 9 | 0 | 0 |
| 9 | 2 | 2 | 3 | 1 | 9 | 6 | 0.666667 |
| 9 | 3 | 2 | 4 | 1 | 9 | 7 | 0.777778 |
| 9 | 4 | 2 | 5 | 1 | 9 | 9 | 1 |
| 9 | 5 | 2 | 5 | 1 | 9 | 3 | 0.333333 |
| 10 | 1 | 3 | 4 | 5 | 9 | 9 | 1 |
| 10 | 2 | 3 | 3 | 7 | 9 | 7 | 0.777778 |
| 10 | 3 | 3.5 | 2 | 3 | 9 | 8 | 0.888889 |
| 10 | 4 | 3.5 | 6.5 | 6 | 9 | 9 | 1 |
| 10 | 5 | 5 | 4 | 3 | 9 | 8 | 0.888889 |
| 11 | 1 | 3 | 1 | 3 | 9 | 9 | 1 |
| 11 | 2 | 3 | 2 | 3 | 9 | 8 | 0.888889 |
| 11 | 3 | 5 | 2 | 4 | 9 | 9 | 1 |
| 11 | 4 | 3 | 2 | 4 | 9 | 7 | 0.777778 |
| 11 | 5 | 4 | 2 | 3 | 9 | 7 | 0.777778 |
| 12 | 1 | 0 | 5 | 3 | 9 | 9 | 1 |
| 12 | 2 | 5 | 2 | 5 | 9 | 9 | 1 |
| 12 | 3 | 4 | 3 | 5 | 9 | 9 | 1 |
| 12 | 4 | 5 | 2 | 5 | 9 | 9 | 1 |
| 12 | 5 | 5 | 4 | 3 | 9 | 9 | 1 |
| 13 | 1 | 4 | 5 | 1 | 9 | 7 | 0.777778 |
| 13 | 2 | 4 | 3 | 5 | 9 | 8 | 0.888889 |
| 13 | 3 | 5 | 3 | 3 | 9 | 7 | 0.777778 |
| 13 | 4 | 3 | 5 | 5 | 9 | 8 | 0.888889 |
| 13 | 5 | 3 | 5 | 5 | 9 | 7 | 0.777778 |

**Table D.2: Pre-test questionnaire results(continued)**

| Participant number | day | prompt | prompt # | prompt_correct | prompt% |
|---|---|---|---|---|---|
| 1 | 1 | pink | 4 | 3 | 0.75 |
| 1 | 2 | pasta | 5 | 4 | 0.8 |
| 1 | 3 | rome | 4 | 4 | 1 |
| 1 | 4 | june | 4 | 4 | 1 |
| 1 | 5 | horse | 5 | 5 | 1 |
| 2 | 1 | gray | 4 | 1 | 0.25 |
| 2 | 2 | sashimi | 7 | 6 | 0.857142857 |
| 2 | 3 | taipei | 6 | 0 | 0 |
| 2 | 4 | august | 6 | 3 | 0.5 |
| 2 | 5 | bulldogs | 8 | 6 | 0.75 |
| 4 | 1 | green | 5 | 2 | 0.4 |
| 4 | 2 | cheesecake | 10 | 3 | 0.3 |
| 4 | 3 | victoria | 8 | 5 | 0.625 |
| 4 | 4 | october | 7 | 6 | 0.857142857 |
| 4 | 5 | cats | 4 | 3 | 0.75 |
| 5 | 1 | yellow | 6 | 0 | 0 |
| 5 | 2 | chilli | 6 | 4 | 0.666666667 |
| 5 | 3 | paris | 5 | 1 | 0.2 |
| 5 | 4 | may | 3 | 2 | 0.666666667 |
| 5 | 5 | cat | 3 | 1 | 0.333333333 |
| 6 | 1 | blue | 4 | 4 | 1 |
| 6 | 2 | banmahi | 7 | 5 | 0.714285714 |
| 6 | 3 | lima | 4 | 4 | 1 |
| 6 | 4 | may | 3 | 3 | 1 |
| 6 | 5 | dog | 3 | 3 | 1 |
| 7 | 1 | blue | 4 | 4 | 1 |
| 7 | 2 | spaghetti | 9 | 9 | 1 |
| 7 | 3 | calgary | 7 | 6 | 0.857142857 |
| 7 | 4 | july | 4 | 4 | 1 |
| 7 | 5 | elephant | 8 | 8 | 1 |
| 8 | 1 | red | 3 | 2 | 0.666666667 |
| 8 | 2 | pizza | 5 | 5 | 1 |
| 8 | 3 | tokyo | 5 | 5 | 1 |
| 8 | 4 | june | 4 | 4 | 1 |
| 8 | 5 | dog | 3 | 2 | 0.666666667 |
| 9 | 1 | yellow | 6 | 1 | 0.166666667 |
| 9 | 2 | stirfry | 7 | 5 | 0.714285714 |
| 9 | 3 | chitwan | 7 | 5 | 0.714285714 |
| 9 | 4 | september | 9 | 8 | 0.888888889 |

| | | | | | |
|---|---|---|---|---|---|
| 9 | 5 | elephant | 8 | 0 | 0 |
| 10 | 1 | purple | 6 | 6 | 1 |
| 10 | 2 | all-food | 8 | 6 | 0.75 |
| 10 | 3 | baldurs-gate | 12 | 12 | 1 |
| 10 | 4 | februrary | 9 | 7 | 0.777777778 |
| 10 | 5 | mouse | 5 | 5 | 1 |
| 11 | 1 | blue | 4 | 4 | 1 |
| 11 | 2 | ricebowl | 8 | 7 | 0.875 |
| 11 | 3 | victoria | 8 | 8 | 1 |
| 11 | 4 | august | 6 | 5 | 0.833333333 |
| 11 | 5 | sloths | 6 | 5 | 0.833333333 |
| 12 | 1 | red | 3 | 3 | 1 |
| 12 | 2 | steak | 5 | 5 | 1 |
| 12 | 3 | kelowna | 7 | 7 | 1 |
| 12 | 4 | october | 7 | 7 | 1 |
| 12 | 5 | lion | 4 | 4 | 1 |
| 13 | 1 | pink | 4 | 2 | 0.5 |
| 13 | 2 | sushi | 5 | 2 | 0.4 |
| 13 | 3 | paris | 5 | 5 | 1 |
| 13 | 4 | august | 6 | 6 | 1 |
| 13 | 5 | penguin | 7 | 7 | 1 |

**Table D.3: Percent correct results**

| | | post-test | | | | | |
|---|---|---|---|---|---|---|---|
| Participant number | day | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 1 | 6 | 4 | 6 | 6 | 7 | 6 |
| 1 | 2 | 5 | 6 | 4.5 | 4.5 | 6 | 6 |
| 1 | 3 | 7 | 2.5 | 6.5 | 5.5 | 7 | 7 |
| 1 | 4 | 5.5 | 6 | 4.5 | 5 | 7 | 7 |
| 1 | 5 | 5.5 | 4 | 5 | 4 | 7 | 7 |
| 2 | 1 | 5 | 2 | 4 | 7 | 5 | 7 |
| 2 | 2 | 6 | 1 | 5 | 5 | 7 | 7 |
| 2 | 3 | 3 | 2 | 3 | 4 | 7 | 7 |
| 2 | 4 | 5 | 3 | 3 | 2 | 7 | 7 |
| 2 | 5 | 2 | 3 | 2 | 2 | 7 | 7 |
| 4 | 1 | | | | | | |
| 4 | 2 | 0 | 1.5 | 0 | 4 | 7 | 7 |
| 4 | 3 | 6 | 4 | 6 | 2 | 7 | 5 |
| 4 | 4 | 6 | 6.5 | 5 | 4 | 7 | 6 |
| 4 | 5 | 6 | 6.5 | 5 | 4 | 7 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5 | 1 | | | | | | |
| 5 | 2 | 7 | 7 | 5 | 5 | 4 | 4 |
| 5 | 3 | 7 | 2 | 4 | 6 | 6 | 6 |
| 5 | 4 | 6 | 5 | 4 | 4 | 6 | 6 |
| 5 | 5 | 6 | 5 | 5 | 4 | 5 | 5 |
| 6 | 1 | 6 | 6 | 6 | 6 | 6 | 7 |
| 6 | 2 | 6 | 5 | 6 | 5 | 7 | 7 |
| 6 | 3 | 6 | 6 | 6 | 7 | 7 | 7 |
| 6 | 4 | 6 | 6 | 5 | 5 | 7 | 7 |
| 6 | 5 | 6 | 6 | 6 | 7 | 6 | 6 |
| 7 | 1 | 5 | 7 | 3 | 3 | 7 | 7 |
| 7 | 2 | 7 | 6 | 5 | 5 | 7 | 7 |
| 7 | 3 | 7 | 6 | 6 | 4 | 7 | 7 |
| 7 | 4 | 5 | 6 | 4 | 3 | 7 | 7 |
| 7 | 5 | 6 | 7 | 4 | 2 | 7 | 7 |
| 8 | 1 | | | | | | |
| 8 | 2 | 6 | 5 | 3 | 3 | 2 | 6 |
| 8 | 3 | 5 | 6 | 4 | 3 | 2 | 6 |
| 8 | 4 | 5 | 3 | 3 | 4 | 2 | 2 |
| 8 | 5 | 6 | 3 | 5 | 6 | 3 | 5 |
| 9 | 1 | 7 | 1 | 5 | 3 | 7 | 7 |
| 9 | 2 | 6 | 5 | 5 | 3 | 6 | 7 |
| 9 | 3 | 6 | 4 | 6 | 6 | 7 | 7 |
| 9 | 4 | 6 | 5 | 4 | 5 | 7 | 7 |
| 9 | 5 | 6 | 2 | 3 | 2 | 7 | 7 |
| 10 | 1 | 2 | 7 | 1 | 1 | 6 | 6 |
| 10 | 2 | 6 | 5 | 5 | 3 | 2 | 7 |
| 10 | 3 | 6 | 6 | 6 | 3 | 4 | 4 |
| 10 | 4 | 7 | 6 | 7 | 6 | 5 | 1 |
| 10 | 5 | 7 | 3 | 7 | 6 | 3 | 3 |
| 11 | 1 | 5 | 5 | 3 | 3 | 5 | 6 |
| 11 | 2 | 6 | 5 | 4 | 3 | 6 | 7 |
| 11 | 3 | 5 | 6 | 3 | 3 | 6 | 7 |
| 11 | 4 | 6 | 5 | 5 | 5 | 6 | 6 |
| 11 | 5 | 6 | 3 | 5 | 5 | 6 | 7 |
| 12 | 1 | 6 | 7 | 2 | 1 | 3 | 6 |
| 12 | 2 | 3.5 | 7 | 4 | 1 | 7 | 7 |
| 12 | 3 | 4 | 5 | 5 | 5 | 1 | 7 |
| 12 | 4 | 5 | 6 | 5 | 3 | 1 | 7 |
| 12 | 5 | 5 | 5 | 4 | 3 | 7 | 7 |
| 13 | 1 | 6 | 5 | 5 | 6 | 6 | 7 |
| 13 | 2 | 6 | 2 | 3 | 3 | 7 | 7 |
| 13 | 3 | 6 | 5 | 5 | 5 | 7 | 7 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 13 | 4 | 5 | 5 | 2 | 3 | 7 | 7 |
| 13 | 5 | 5 | 4 | 3 | 4 | 7 | 7 |

**Table D.4: Post-test questionnaire results**

| | | | post-study | | | |
|---|---|---|---|---|---|---|
| Participant number | day | 7 | 1 | 2 | 3 | 4 |
| 1 | 1 | 1 | | | | |
| 1 | 2 | 1 | | | | |
| 1 | 3 | 1 | | | | |
| 1 | 4 | 1 | | | | |
| 1 | 5 | 1 | 3.5 | 3 | focus better | counting |
| 2 | 1 | 1 | | | | |
| 2 | 2 | 1 | | | | |
| 2 | 3 | 1 | | | | |
| 2 | 4 | 1 | | | | |
| 2 | 5 | 1 | 1 | 1 | no | say letter |
| 4 | 1 | | | | | |
| 4 | 2 | 3 | | | | |
| 4 | 3 | 4 | | | | |
| 4 | 4 | 2 | | | | |
| 4 | 5 | 7 | 2 | | focus better | counting |
| 5 | 1 | | | | | |
| 5 | 2 | 5 | | | | |
| 5 | 3 | 5 | | | | |
| 5 | 4 | 6 | | | | |
| 5 | 5 | 5 | 4 | 3 | focus better | counting |
| 6 | 1 | 5 | | | | |
| 6 | 2 | 6 | | | | |
| 6 | 3 | 7 | | | | |
| 6 | 4 | 7 | | | | |
| 6 | 5 | 6 | 3 | 4 | no | read letter silently |
| 7 | 1 | 5 | | | | |
| 7 | 2 | 5 | | | | |
| 7 | 3 | 5 | | | | |
| 7 | 4 | 5 | | | | |
| 7 | 5 | 5 | 1 | 1 | introspective | counting |
| 8 | 1 | | | | | |
| 8 | 2 | 2 | | | | |
| 8 | 3 | 2 | | | | |
| 8 | 4 | 1 | | | | |
| 8 | 5 | 1 | 5 | 5 | no | counting |
| 9 | 1 | 7 | | | | |
| 9 | 2 | 7 | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 9 | 3 | 7 | | | | |
| 9 | 4 | 7 | | | | |
| 9 | 5 | 7 | 5 | 5 | no | singing abc's |
| 10 | 1 | 7 | | | | |
| 10 | 2 | 3 | | | | |
| 10 | 3 | 6 | | | | |
| 10 | 4 | 3.5 | | | | |
| 10 | 5 | 3.5 | 6 | 6 | no | playing song in head, changing notes when flash |
| 11 | 1 | 5 | | | | |
| 11 | 2 | 5 | | | | |
| 11 | 3 | 5 | | | | |
| 11 | 4 | 5 | | | | |
| 11 | 5 | 5 | 5 | 4 | no | counting |
| 12 | 1 | 7 | | | | |
| 12 | 2 | 7 | | | | |
| 12 | 3 | 7 | | | | |
| 12 | 4 | 7 | | | | |
| 12 | 5 | 7 | 3 | 3 | no | counting |
| 13 | 1 | 3 | | | | |
| 13 | 2 | 5 | | | | |
| 13 | 3 | 3 | | | | |
| 13 | 4 | 3 | | | | |
| 13 | 5 | 3 | 2 | 2 | no | counting |

**Table D.5: Post-study questionnaire results**

## D.2 Two session study

The subject recruitment materials were substantially the same for this study as for the training effects study. The primary difference was that there were only two sessions instead of five.

### D.2.1 Stimulus presentation order

We need 2 prompts each day for them to produce their own phrases:

day 1

Favourite colour
favourite food

day 2

favourite animal
favourite month

So the sessions, then, are as follows:

headset fitting
short phrase to orient to the task ("bird" is always good)
phrase 1 "cherry blossoms"
phrase 2 "jumping fish"
phrase 3 "raging rivers"

train model (do pre-test questionnaire)

phrase 4 "crazy cats"
prompt 1 (write answer then spell)

short break (leave to subject how long, but at least 2 minutes)

prompt 2 (write answer then spell)
phrase 5 "double rainbow"

revert to untrained model

phrase 1 "cherry blossoms"

## Day 1 questionnaire

Participant number:

Pre-test questions (actually answered as a break between tasks)

1) Do you generally concentrate well? 1- no, not at all, 7- yes, always
2) How do you feel today? 1- horrible, 7-fantastic
3) Do you have any conditions that could affect today's test? Neurological conditions, ADHD, etc. yes, no, prefer not to say (discuss with experimenter if yes)
4) What is your gender? Male, female, other or prefer not to say
5) How much did you sleep last night? 1- less than 5 hours, 7- 8 or more hours
6) How good a night's sleep was that for you? 1- horrible, 7- fantastic
7) Have you had caffeine today? 1- none, 7-7 or more cups of coffee (or equivalent)

8) Do you generally drink this amount of caffeine? 1- usually much less, 7-usually much more
9) Do you feel tense today? 1- totally mellow, 7- way stressed
10) Are you dehydrated? 1- not at all, 7- very dehydrated

Prompt #1 (complete when asked)

What is your favourite colour? ______________________________________

Prompt #2 (complete when asked)

What is your favourite food? ________________________________________

Post-test questions

1) How hard did you have to concentrate to make the system respond well? 1- not at all, 7 – concentrated very hard
2) Did it respond to your brainwaves well? 1- not at all, 7 – amazingly well
3) How hard was it to maintain attention? 1- easy, 7- very hard
4) Did you feel tired? Did your ability to pay attention fall off at some point in the study? 1- never, 7- frequently
5) Was the headset comfortable? 1- very uncomfortable, 7- very comfortable
6) Was the saline solution irritating? 1- very irritating, 7- not at all
7) Would you wear the headset in public? 1- no way, 7- sure, no problem

## Day 2 questionnaire

Participant number:

Pre-test questions (actually answered as a break between tasks)

11) Do you generally concentrate well? 1- no, not at all, 7- yes, always
12) How do you feel today? 1- horrible, 7-fantastic
13) Do you have any conditions that could affect today's test? Neurological conditions, ADHD, etc. yes, no, prefer not to say (discuss with experimenter if yes)
14) What is your gender? Male, female, other or prefer not to say
15) How much did you sleep last night? 1- less than 5 hours, 7- 8 or more hours
16) How good a night's sleep was that for you? 1- horrible, 7- fantastic
17) Have you had caffeine today? 1- none, 7-7 or more cups of coffee (or equivalent)
18) Do you generally drink this amount of caffeine? 1- usually much less, 7-usually much more
19) Do you feel tense today? 1- totally mellow, 7- way stressed
20) Are you dehydrated? 1- not at all, 7- very dehydrated

Prompt #1 (complete when asked)

What is your favourite animal? ______________________________________

Prompt #2 (complete when asked)

What is your favourite month? ________________________________________

Post-test questions

8) How hard did you have to concentrate to make the system respond well? 1- not at all, 7 – concentrated very hard
9) Did it respond to your brainwaves well? 1- not at all, 7 – amazingly well
10) How hard was it to maintain attention? 1- easy, 7- very hard
11) Did you feel tired? Did your ability to pay attention fall off at some point in the study? 1- never, 7- frequently
12) Was the headset comfortable? 1- very uncomfortable, 7- very comfortable
13) Was the saline solution irritating? 1- very irritating, 7- not at all
14) Would you wear the headset in public? 1- no way, 7- sure, no problem

## D.3 Memory attention study

### D.3.1 Stimulus presentation order

headset fitting
short phrase to orient to the task ("cat" was used for this study)
phrase 1 "jumping jacks"
phrase 2 "are not fun"
phrase 3 "in the heat"

train model

first card game
second card game

revert to untrained model

phrase 4 “jumping jacks”

### Pre-test Questionnaire

1) Do you generally concentrate well? 1- no, not at all, 7- yes, always

2) How do you feel today? 1- horrible, 7-fantastic
3) Do you have any conditions that could affect today's test? Neurological conditions, ADHD, etc. yes, no, prefer not to say (discuss with experimenter if yes)
4) What is your gender? Male, female, other or prefer not to say
5) How much did you sleep last night? 1- less than 5 hours, 7- 8 or more hours
6) How good a night's sleep was that for you? 1- horrible, 7- fantastic
7) Have you had caffeine today? 1- none, 7-7 or more cups of coffee (or equivalent)
8) Do you generally drink this amount of caffeine? 1- usually much less, 7-usually much more
9) Do you feel tense today? 1- totally mellow, 7- way stressed
10) Are you dehydrated? 1- not at all, 7- very dehydrated

## Post-test questionnaire

1) How hard did you have to concentrate to make the system respond well? 1- not at all, 7 – concentrated very hard
2) Did it respond to your brainwaves well? 1- not at all, 7 – amazingly well
3) How hard was it to maintain attention? 1- easy, 7- very hard
4) Did you feel tired? Did your ability to pay attention fall off at some point in the study? 1- never, 7- frequently
5) Was the headset comfortable? 1- very uncomfortable, 7- very comfortable
6) Was the saline solution irritating? 1- very irritating, 7- not at all
7) Would you wear the headset in public? 1- no way, 7- sure, no problem
8) Did the exercise where you were choosing cards and the exercise where you where selecting letters feel like the same kind of concentration? 1- no, very different, 7- identical

## Interview Questions

1) How was that? Do you have any general impressions of the activities you just did?
2) How hard did you feel you had to concentrate to get the system to respond the way you wanted it to?
3) Were you generally successful in getting it to respond as you wanted?
4) Did you feel tired by the end of the session?
5) How could the system have been made more engaging?
6) Were there any particular things about the system you found frustrating?
7) Could you see a more engaging game around these concepts being useful? If so, for whom and how?
8) Anything else to add?

## D.4 Ethics application

Ethics modifications occurred for each study, of course, but in the interests of brevity only the main application is presented here. The title page and preliminary sections containing personal information for the experimenter and supervisors have been omitted. All approvals and modifications are on file with the office of human research ethics under file number 14-097.

### K. Description of Research Project

1. **Purpose and Rationale of Research**

Briefly describe in non-technical language:
***Please use 150 words or fewer.***

5a. The research objective(s) and question(s)

Paying attention over a long period of time can be difficult, by using the ability to maintain focus as the determining factor in an onscreen game researchers hope to formulate techniques for training attentional capabilities. The brainwave readings provide an independent measure of attention, so they can help create a feedback loop where the user can train themselves to maintain attention to achieve greater results in the game. Researchers posit that training in one form of attention will benefit other measures of attention. The study aims to show both the ability of the system to discriminate between attentional states, and the effect of training on the individual user.

5b. The importance and contributions of the research

If the system can be trained to discriminate the user's attentional state, then the current generation of consumer hardware is sufficient to create feedback training software to improve attention. Although this study is in cognitively normal individuals to show attentional changes in very limited situations, one can imagine future therapeutic interventions for people with ADHD or other attention-related disorders. The price of these consumer-grade EEG devices is dropping from thousands of dollars to only a couple hundred dollars. As a result, if the quality of the devices is sufficient to measure actual brain correlates of attention then the opportunity to use this beneficially should not be overlooked.

5c. If applicable, provide background information or details that will enable the HREB to understand the context of the study when reviewing the application.

This research studies the brainwave changes involved in attention and is working towards the development of training protocols for improving attentional capacity. While some of the individual techniques have been validated on expensive lab hardware, the push of EEG hardware into consumer models raises the possibility of wider distribution of the neurofeedback approach.

---

## L. Recruitment

### 2. Recruitment and Selection of Participants

6a. Briefly describe the target population(s) for recruitment. Ensure that all participant groups are identified (*e.g., group 1 - teachers, group 2 - administrators, group 3 - parents*).

Participants will be students at the University of Victoria recruited through the Department of Psychology's voluntary research participation system or recruited for nominal payment through posters. Students in undergraduate courses in Psychology have the option of participating in research studies to earn extra credit in their courses, but are often in scarcer numbers in the summer so the option of recruiting participants by poster is also being described and prepared for. (Recruitment poster is attached, participants recruited in this fashion will be compensated with $20 per participant.)

6b. Why is each population or group of interest?

The purpose of this research is to examine the neural activities when participants are performing attention tasks in healthy young adults. University students fit this description and are a convenient population with which to work. In addition, participation in research studies such as these contributes to the educational experience of the students.

6c. What are the *salient* characteristics of the participants for your study? (*e.g., age, gender, race, ethnicity, class, position, etc*.)? List all inclusion and exclusion criteria you are using.

The characteristics of participants are representative of Psychology students at UVic: predominantly young adults, about 70% females, about 80% Caucasian and the remainder mostly Asian, and varied economic class (but mainly lower middle class to upper class). If recruiting from posters in ECS those characteristics will remain substantially unchanged.

6d. What is the desired number of participants for each group?

Including pilot subjects, this study will require approximately 40 participants
Pilot subjects will follow the protocol outlined in M.8b and will be used to calibrate the length of time stimuli are presented and to ensure that the instructions are clear and that reliable data is obtained from the hardware. Other than the fact that these initial participants will not be included in the final data analysis all other aspects of their participation, including consent and data retention, will be as described in this document.

6e. Provide a detailed description of your recruitment process. Explain:

i) List all source(s) for information used to contact potential participants (*e.g., personal contacts, listserves, publicly available contact information, etc.). Clarify which sources will be used for which participant groups:*

Students initiate contact with the study recruitment process by visiting the Department of Psychology's voluntary research participation system.

If the psychology department system fails to procure enough participants then students view a poster that leads them to initiate contact with the experimenter.

ii) List all methods of recruitment (*e.g., in-person, by telephone, letter, snowball sampling, word-of-mouth, advertisement, etc.) If you will be using "snowball" sampling, clarify how this will proceed (i.e., will participants be asked to pass on your study information to other potential participants?). Clarify which methods will be used for which participant groups.*

Potential participants visit a web site listing available research studies. Each study has a web page summarizing the purpose of and procedure used in the study. Students who opt to participate in a research study may earn extra credit in a psychology course (1 credit for each half hour of participation, with 2 credits generally corresponding to one percentage point in the final grade, depending on the course instructor).

If recruited via poster, participants view the attached poster and contact the experimenter.

iii) If you will be using personal and/or private contact information to contact potential participants (as stated above), have the potential participants given permission for this, or will you use a neutral third party to assist you with recruitment? *Note that this is not a concern when public and/or business contact information is used.*

No personal and/or private information will be used to contact participants.

iv) Who will recruit/contact participants (*e.g., researcher, assistant, third party, etc.) Clarify this for each participant group.*

Participants will be recruited by visiting a web site. No face-to-face interaction is involved. Information posted on the web site for a study is constructed by a research assistant and is approved by the faculty member who is chair of the Department of Psychology's Research Participation Committee. To encourage students to participate in research, the chair of the committee visits each Psychology 100A class in September to describe how the system operates and to answer questions that students may have about the system.

If recruited by poster, they participants will contact the experimenter to complete their sign-up into the study.

v) List and explain any relationship between the members of the research team (including third party recruiters or sponsors/clients of the research) and the participant(s) (*e.g., acquaintances, colleagues*). Complete item 7 if there is potential for a power relationship or a *perceived* power relationship (*e.g., instructor-student, manager-employee, etc.*). If you have a close relationship with potential participants (*e.g., family member, friend, close colleague, etc.)* clarify here the safeguards that you will put in place to mitigate any potential pressure to participate.

No power over relationship between the investigators and participants.

vi) In chronological order (if possible) describe the steps in the recruitment process. *(Include how you will screen potential participants where applicable).* Consider where in the process permission of other bodies may be required.

- Students visit a website listing potential studies and study details
- Students select the current study as of interest
- Students are contacted regarding scheduling a study session

## 3. Power Relationships (Dual-Role and Power-Over)

**If you are completing this section, please refer to the:**
**Guidelines For Ethics in Dual-Role Research for Teachers and Other Practitioners and the TCPS 2, Article 3.1 and Article 7.4.**

Are you or any of your co-researchers in any way in a power relationship, including dual-roles, that could influence the voluntariness of a participant's consent? Could you or any of your co-researchers potentially be *perceived* to be in a power relationship by potential participants? *Examples of "power relationships" include teachers-students, therapists-clients, supervisors-employees and possibly researcher-relative or researcher-close friend where elements of trust or dependency could result in undue influence.*

☐ Yes ☒ No ☐ Varies

If *yes* or *varies*, describe below:

i) The nature of the relationship:

ii) Why it is necessary to conduct research with participants over whom you have a power relationship:

iii) What safeguards (steps) will be taken to ensure voluntariness and minimize undue influence, coercion or potential harm:

iv) How will the power or dual-role relationship and associated safeguards be explained to potential participants:

| | |
|---|---|
| ☐ **Interviewing participants:**<br>☒ in-person<br>☐ by telephone<br>☐ using web-based technology (explain):<br>☐ Conducting group interviews or discussions (including focus groups) | ☒ **Attach draft interview questions** |
| ☒ **Administering a questionnaire or survey:**<br>☒ In person ☐ by telephone<br>☐ mail back ☐ email<br>☐ web-based* (see below)<br>☐ Other, describe:<br>*If using a web program with a server located in the United States (e.g., SurveyMonkey), or if there are other reasons that the data will be stored in the US (e.g., use of US-based cloud technology, sharing data with US colleagues, etc.), you must inform participants that their responses may be accessed via the U.S. Patriot Act. Please add the following to the consent form(s):<br>*"Please be advised that this research study includes data storage in the U.S.A. As such, there is a possibility that information about you that is gathered for this research study may be accessed without your knowledge or consent by the U.S. government in compliance with the U.S. Patriot Act. "* | ☒ **Attach questionnaire or survey:**<br>☐ standardized (one with established reliability and validity)<br>☒ non-standardized (one that is un-tested, adapted or open-ended) |

☒ **Administering a computerized task** (*describe in 8b or attach details*)

☒ **Observing participants**

*In 8b, describe who and what will be observed. Include where observations will take place. If applicable, forward an observational data collection sheet for review.*

| | |
|---|---|
| ☐ **Recording of participants and data using**:<br>☐ audio ☐ video ☐ photos or slides<br>☐ note taking ☐ flipcharts<br>☐ data collection sheet *(attach)* ☐ other: | ☐ Images used for analysis<br>☐ Images used in disseminating results *(include release to use participant images in consent materials)* |

☐ **Using human samples** (*e.g., saliva, urine, blood, hair*)

*Attach your Biosafety Approval, or your correspondence with the Biosafety Committee, to this application. Note that Research Ethics Approval is contingent on Biosafety Approval.*

☒ **Using specialized equipment/machines** *(e.g., ultrasound, EEG, prototypes etc.)* or other. *(e.g., testing instruments that are not surveys or questionnaires).* Please specify: EEG

☐ **Using other testing equipment not captured under other categories.**

Please specify:

☐ **Collecting materials supplied by, or produced by, the participants** *(e.g., artifacts, paintings, drawings, photos, slides, art, journals, writings, etc.)*

Please specify:

☐ **Analyzing secondary data** or secondary use of data (Refers to information/data that was originally gathered for a purpose other than the proposed research and is now being considered for use in research (*e.g., patient or school records, personal writings, lesson plans, etc.).*

☐ Secondary data involving anonymized information (Information/data is stripped of identifiers by another researcher or institution **before** being shared with the applicant*).*

| |
|---|
| ☐ Secondary data with identifying information (Data contains names and other information that can be linked to individuals,*( e.g., student report cards, employment records, meeting minutes, personal writings).*<br>*In item 8b describe the source of the data, who the appropriate data steward is, and explain whether (and how) consent was or will be obtained from the individuals for use of their data.* |
| ☐ **Other:**<br>Please specify: |

**Recruitment Materials Checklist:**

Attach all documents referenced in this section *(check those that are appended):*

☐ Script(s) – in-person, telephone, 3rd party, e-mail, etc.

☐ Invitation to participate *(e.g., Psychology Research Participation System Posting)*

☒ Advertisement, poster, flyer

☐ None; please explain why *(e.g., consent form used as invitation/recruitment guide)*

## M. Data Collection Methods

### 4. Data Collection

*Use the following sections in ways best suited to explain your project. If you have more than one participant group, be sure to explain which participant group(s) will be involved in which activity/activities or method(s).*

8a. Which of the following methods will be used to collect data? *Check all that apply.*

8b. Provide a sequential description of the procedures/methods to be used in your research study.
Be sure to provide details for all methods checked in section 8a. Clarify which procedures/methods will be used for each participant group. Indicate which methods, if any, will be conducted in a group setting. *List all of the research instruments and interview/focus group questions, and append copies (if possible) or detailed descriptions of all instruments. If not yet finalized, provide drafts or sample items/questions.*

**The participants will arrive at the location indicated for the experiment, be greeted by the experimenter and asked to go over the consent forms. After explaining the study and signing the consent forms, the participant will be asked to complete a paper-based pre-test questionnaire on attention and computer-related behaviour.**

**The participant will then have an EEG cap fitted to their head. Unlike the 64-gel coupled electrodes in the research EEG machines, this cap will have only 4-14 sensors that are saline coupled. Consumer-grade EEG has fewer fitting complications than research systems, as one would expect. Each electrode does still have to be precisely placed under hair, though, and sterile saline solution will be used to wet the electrodes until an onscreen program indicates sufficient conductivity has been achieved. Initial consent forms and electrode fitting are expected to take approximately 15 minutes.**

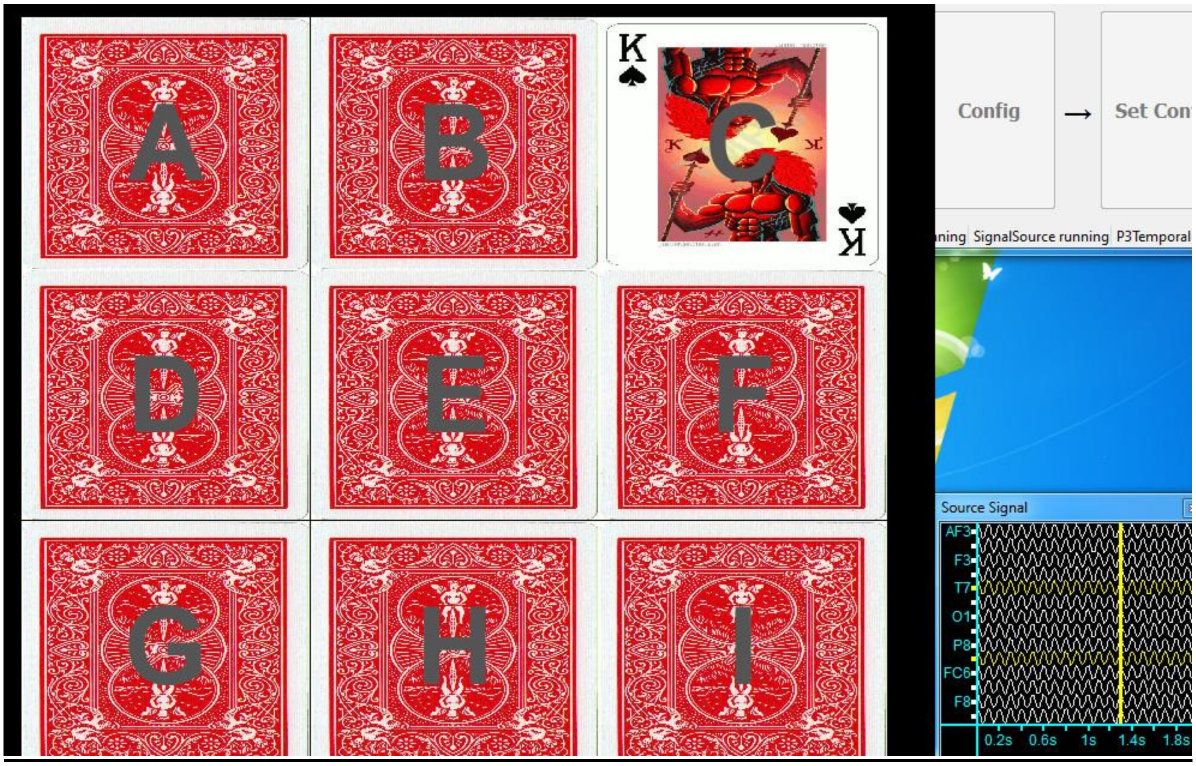
**The participant will then be asked to go through a series of training sessions intended to adapt the computer system to individual variations in the pattern of attentional focus. This training will consist of attending to onscreen stimuli in the order indicated by the training program. This portion of the study will take about 20 minutes.**

**After a brief rest while the experimenter loads the trained models into the system, the participant will spend 45 minutes in one of two experimental conditions. Participants in group A will play a memory game with the mouse. They will be shown cards and have to match pairs. The game will progress at the**

**same pace that will be asked of group B. Group B will play the same game, but instead of interacting with the mouse, attentional cues from the brainwave data will be used to identify the cards to be matched. This requires sustained attention during the session and each turn on the cards takes significantly longer than if the mouse were used.**

**The experimental conditions relate to the research question in that our hypothesis is that sustained attentional focus will lead to a training effect and cause better performance in the final attention test. The participants using the mouse (ie. those in group A) will experience a lower requirement for sustained attentional focus and will therefore derive less benefit from the training portion of the study than will those participants who need to keep sustained attention during the training tasks. No deception is being employed by not mentioning these experimental conditions because all participants are fitted with the EEG headset, use it for initial calibration, and use it for the final attentional test. Additionally, some training benefit is expected even from those participants who use the mouse during the middle portion of the study, just not as much training effect as with those from whom more sustained attention is required.**

**The following screenshot gives an example of the memory task, the individual cards are highlighted for a fraction of a second each and the EEG records the response potentials which will allow the system to determine which card the user is focusing their attention on. Since (at least in the experimental condition) the participants are using only their attentional focus to determine which card to select, this is where the "using your mind to control the computer" language on the recruitment poster comes from.**



**After the training game, the participant will be asked to do two test sessions to see how much better the computer prediction models have become, and how much the user behaviour has changed since the initial training session. This test phase will take approximately 20 minutes.**

**A final post-test questionnaire will be administered on paper to collect participants experiences with the system, along with a brief post-test interview. Participants will then be thanked and any final questions answered. This last portion of the study session is predicted to take less than 20 minutes for a total study time of less than 2 hours. (All questionnaires and interview questions are appended to this application.)**

**During the study, the experimenter will be present in the room observing the participant. The experimenter will record any issues of frustration with the software and the participant will be instructed to indicate verbally when the system deduces a different input than the participant intended (which is common during training of the system.) Other than during the post-test interview no formal recording of responses or coding of observational data will be performed. During the post-test interview responses will be recorded by hand and with audio recordings.**

8c. Where will participation take place for each data collection method/procedure? *Provide specific location, (e.g., UVic classroom, private residence, participant's workplace). Clarify the locations for each participant group and/or each data collection method*.

**The sessions will take place either in Cornett A056 or ECS 654.**

8d. For each method, and in total, how much time will be required of participants? *Clarify this for each participant group, each data collection method, and any other research related activities.*

**Each participant will devote 2 hours to the study, most of this to the onscreen training with a few minutes for the paper questionnaire and consent forms.**

8e. Will participation take place during participants' office/work hours or instructional time?

☒ No ☐ Yes. Indicate whether permission is required *(e.g., from workplace supervisor, school principal, etc.)* and how this will be obtained:

**Data Collection Methods Checklist:**

Attach all documents referenced in this section *(check those that are appended. Where draft versions are appended please ensure that final versions are submitted when available. If final versions differ significantly after you have obtained Research Ethics approval, you will need to submit a Request for Modification*:

☐ Standardized Instrument(s)

☒ Survey(s), Questionnaire(s)

☒ Interview and/or Focus Group Questions

☐ Observation Protocols

☐ Other:

---

## N. Possible Benefits, Inconveniences, and Risks of Harm to Participants

### 9. Benefits

Identify any potential or known benefits associated with participation and explain below. *Keep in mind that the anticipated benefits should outweigh any potential risks.*

☒ To the participant ☒ To society ☒ To the state of knowledge

By participating in this project, participants will add to their knowledge of how experimental psychologists conduct research and, more specifically, learn about empirical and theoretical issues pertaining to cognitive psychology. The debriefing will include the theoretical importance of the research, as well as examples of how the results will benefit and/or clarify our understanding of the issues under investigation. Therefore, participants will be given an opportunity to learn about a current topic in experimental psychology, and they may also gain a better understanding of their own cognitive processes. This research will also contribute to the general state of knowledge in the field and may have applied implications of value to society.

### 10. Inconveniences

Identify and describe any known or potential inconveniences to participants:
*Consider all potential inconveniences, including total time devoted to the research.*

The only foreseeable inconvenience associated with this research is the amount of time required to participate in the experiments. This will include the time required to travel to and from the location of the experiments as well as the length of the experiment itself. Participants will be fully informed as to the location and maximum duration (2 hours) of the experiment during the sign-up process and will thus be able to consider this information in deciding whether or not to participate. The length of the experiment will also be discussed when the informed consent form is reviewed.

### 11. Level of Risk

The TCPS 2 definition of “minimal risk research” is as follows:

> *“Research in which the probability and magnitude of possible harms implied by participation in the research is no greater than those encountered by the participant in those aspects of their everyday life that relate to the research.”*

Based on this definition, do you believe your research qualifies as “minimal risk research”?

☒ Yes it is minimal risk. ☐ No, it is not minimal risk.

Explain your answer with reference to the risks of the study and the vulnerability of the participants:

**University students are used to computer-based tasks requiring focused attention, the only additional element of this study is the use of an EEG headset to record neural potentials as they complete the assigned tasks.**

### 12. Estimate of Risks of Harm

Consider the inherent foreseeable risks associated with your research protocol and complete the table below by putting an X in the appropriate boxes. Be sure to take into account the vulnerability of your target population(s) if applicable:

| **Potential Risks of Harm** | **Very unlikely** | **Possibly** | **Likely** |
|---|---|---|---|
| i) Emotional or psychological discomfort, such as feeling demeaned or embarrassed due to the research | ☒ | ☐ | ☐ |
| ii) Fatigue or stress | ☐ | ☒ | ☐ |
| iii) Social risks, such as stigmatization, loss of status, privacy and/or reputation | ☒ | ☐ | ☐ |
| iv) Physical risks such as falls | ☒ | ☐ | ☐ |
| v) Economic risk (e.g., job security, salary loss, etc.) | ☒ | ☐ | ☐ |
| vi) Risk of incidental findings (*See Article 3.4 of the TCPS 2 for more information*) | ☒ | ☐ | ☐ |

| | | | |
|---|---|---|---|
| vii) Other risks: | ☒ | ☐ | ☐ |

**13. Possible Risks of Harm**

If you indicated in Item 12 (i) to (vii) that any risks of harm are *possible* or *likely*, please explain below:

13a. What are the risks? *(i.e., elaborate on risks you have identified above)*

Participants may become fatigued from viewing a computer screen for up to two hours. In studies involving recording of EEG activity, there are no risks associated with using the equipment. Participants may experience some scalp itchiness as a result of the sterile saline solution that is applied to the scalp at each electrode site.

13b. What will you do to try to minimize, mitigate, or prevent the risks?

Periodic breaks will be provided during which time the participant may rest his or her eyes. Fresh water will be provided to rinse the saline solution in case any discomfort occurs.

13c. How will you respond if the harm occurs? *(i.e., what is your plan?)*

The participant may terminate participation if rest breaks are not adequate to alleviate fatigue. If there is any evidence of sensitivity to the saline solution, the experiment will be terminated without penalty or loss of compensation.

13d. If you have indicated that there is a risk of Incidental Findings (vi) please outline your proposed protocol for information and/or action.

13e. If one or more of your participant groups could be considered vulnerable please describe any specific considerations you have built into the protocol to address this.

**14. Risk to Researcher(s)**

14a. Does this research study pose any risks to the researchers, assistants and data collectors?

**No risks to researchers are anticipated**

14b. If there are any risks, explain the nature of the risks, how they will be minimized, and how you will respond if they occur.

**15. Deception**

Will participants be fully informed of everything that will be required of them prior to the start of the research session?

☒Yes ☐ No *(If no, complete the Request to Use Deception form on the ORS website)*

## O. Incentives, Reimbursement and Compensation

16a. Is there any incentive, monetary or otherwise, being offered for participation in the research *(e.g., gifts, honorarium, course credits, etc.)*

☒ Yes ☐ No

If yes, explain the nature of the incentive(s) and why you consider it necessary. *Also consider whether the amount or nature of the incentive could be considered a form of undue inducement or affect the voluntariness of consent. Clarify which participant groups will be provided with which incentives.*

Participants will receive four credits toward their grade in a Psychology course. These bonus credits encourage and acknowledge students' participation in the research project.

If the psychology subject pool is insufficient in the summer, participants separately recruited will each be compensated with a $20 honorarium

16b. Is there any reimbursement or compensation for participating in the research *(e.g., for transportation, parking, childcare, etc.)*

☒ Yes ☐ No

If yes, explain the nature of reimbursement or compensation and why you consider it necessary. *Also consider whether the amount of reimbursement or compensation could be considered a form of undue inducement or affect the voluntariness of consent. Clarify which participant groups will be provided with which kind of reimbursement or compensation.*

There is no reimbursement planned other than the measures identified in 16a.

16c. Explain what will happen to the incentives, reimbursement or compensation if participants withdraw during data collection or any time thereafter *(e.g., compensation will be pro-rated, full compensation will be given, etc.)*

They will receive the credit without penalty or will be entered in the lottery.

## P. Free and Informed Consent

*Consent encompasses a process that begins with initial contact and continues through to the end of the research process. Consult Article 3.2 of the TCPS 2 and Appendix V of the Guidelines for further information.*

### 17. Participant's Capacity (Competence) to Provide Free and Informed Consent

*Capacity refers to the ability of prospective or actual participants to understand relevant information presented about a research project, and to appreciate the potential consequences of their decision to participate or not participate. See the TCPS 2, Chapter 3, section C, for further information.*

Identify your potential participants: *(Check all that apply.)*

| **Competent** | **Non-Competent** |
|---|---|
| ☒ Competent adults<br>☐ A protected or vulnerable population (*e.g., inmates, patients*) | ☐ Non-competent adults:<br>☐ Consent of family/authorized representative will be obtained<br>☐ Assent of the participant will be obtained (note that assent of the participant is always required) |

| | |
|---|---|
| ☒ Competent youth aged 13 to 18:<br>☐ Consent of youth will be obtained and parental/guardian consent is required, *due to institutional requirements (such as school districts) or due to the nature of the research (e.g., risks, etc.)*<br>☐ Consent of youth will be obtained, parents/guardians will be informed<br>☒ Consent of youth will be obtained, parents/guardians will *NOT* be informed<br>☐ Other, explain: | ☐ Non-competent youth:<br>☐ Consent of parent/guardian<br>☐ Assent of the youth will be obtained (note that assent of the participant is always required) |
| ☐ Competent children under 13 (*who are able to provide fully informed consent)*:<br>☐ Consent of child will be obtained and consent of parent/guardian will be obtained<br>☐ Other, explain: | ☐ Non-competent children *(young children and/or children with limited abilities to provide fully informed consent)*:<br>☐ Consent of parent/guardian<br>☐ Assent of the child will be obtained (note that assent of the participant is always required) |

**18. Means of Obtaining and Documenting Consent and/or Assent:**
*Check all that apply, consider all of your participant groups, attach copies of relevant materials, complete item 19:*

☒ **Signed** consent (*Attach consent form(s) - see template available )*

☐ **Verbal** consent (*Attach verbal consent script(s) - see template available*.)

**Explain** in 19 why written consent is not appropriate and how verbal consent will be documented.

☐ Letter of Information for **Implied** consent (*e.g., anonymous, mail back or web-based survey. Attach information letter, see template*)

☐ **Signed** or **Verbal assent** for non-competent participants *(Attach assent form(s), or verbal assent script(s)).*

**Explain** how verbal assent will be documented in 19.

☐ **Other** means. **Explain** in 19 and provide justification.

☐ Consent **will not be obtained**. See TCPS 2 Articles 3.5 and 3.7. **Explain** in 19**.**

☐ **Signed** consent from the parents/guardians for youth/child participants *(Attach consent form(s)).*

**Explain** how parents/guardians will provide informed consent for child/youth participants in 19.

☐ **Information letters** for the parents/guardians of youth/child participants *(Attach information letter(s)). If consent will not be obtained from parents/guardians and the parents/guardians will not be informed, explain why not in 19.*

**19. Informed Consent**

Describe the exact steps (chronological order) that you will follow in the process of explaining, obtaining, and documenting informed consent. Ensure that consent procedures for all participant groups are identified (e.g., group 1 - teachers, group 2 – parents, group 3 – students). Be sure to indicate when participants will first be provided with the consent materials *(e.g., prior to first meeting with the researcher?).* If consent will not be obtained, explain why not with reference to the TCPS 2 Articles 3.5 and 3.7.

Participants are recruited and through email schedule a time to come in for the study. The participant arrives at the study location, is greeted, and then immediately asked to read and signed the informed consent document. The participant is provided with a copy of the consent form and the experimenter will keep the signed copy on file as described in the data retention section of this document.

**20. Ongoing Consent**

*Article 3.3 of the TCPS 2 states that consent shall be maintained throughout the research project. Complete this section if the research involves interacting with participants over multiple occasions (including review of transcripts, etc.), has multiple data collection activities, and/ or occurs over an extended period of time.*

20a. Will your research occur over multiple occasions or an extended period of time *(including review of transcripts)*?

☐ Yes ☒ No

20b. If yes, describe how you will obtain and document ongoing consent. If consent procedures differ for each group or activity, please clarify each group or activity that you are referring to.

**21. Participant’s Right to Withdraw**

*Article 3.1 of the TCPS2 states that participants have the right to withdraw at any time and can withdraw their data and human biological materials.*

Describe what participants will be told about their right to withdraw from the research at any time (*i.e., who to contact and how*). If compensation is involved, explain what participants will be told about compensation if they withdraw. *If you have different participant groups and/or different data collection methods, clarify the different procedures for withdrawing as necessary.*

**As described in the consent document, participants have the right to withdraw at any time and the option of withdrawing their data.**

**22. What will happen to a person’s data if s/he withdraws part way through the study or after the data have been collected/submitted? If applicable, include information about visual data such as photos or videos.** *If you have different participant groups and/or different data collection methods, clarify the different procedures for withdrawing as necessary. Ensure this information is included in the consent documents.*

☐ Participant will be asked if he/she agrees to the use of his/her data. Describe how this agreement will be documented:

☒ It will not be used in the analysis and will be destroyed.

☐ It is logistically impossible to remove individual participant data (*e.g., anonymously submitted data*).

☐ When linked to group data (*e.g., focus group discussions*), it will be used in summarized form with no identifying information.

**Free and Informed Consent Checklist:**

Attach all documents referenced in this section *(check those that are appended):*

☒ Consent and Assent Form(s) – Include forms for all participant groups and data gathering methods

☐ Letter(s) of Information for Implied Consent

☐ Verbal Consent and Assent Scripts

---

## Q. Anonymity and Confidentiality

### 23. Anonymity

*Anonymity means that no one, including the principal investigator, is able to associate responses or other data with individual participants.*

23a. Will the participants be anonymous in the data gathering phase of research?

☐ Yes ☒ No

23b. Will the participants be anonymous in the dissemination of results *(be sure to consider use of video, photos)*?

☒ Yes
☐ Maybe. Explain below.
☐ No. If anonymity will not be protected and you plan to identify all participants with their data, provide the rationale below.

### 24. Confidentiality

*Confidentiality means the protection of the person's identity (anonymity) and the protection, access, control and security of his or her data and personal information during the recruitment, data collection, reporting of findings, dissemination of data (if relevant) and after the study is completed (e.g., storage). The ethical duty of confidentiality refers to the obligation of an individual or organization to safeguard entrusted information. The ethical duty of confidentiality includes obligations to protect information from unauthorized access, use, disclosure, modification, loss or theft.*

24a. Are there any limits to protecting the confidentiality of participants?

☒ No, confidentiality of participants and their data will be completely protected

☐ Yes, there are some limits to the researcher's ability to protect the confidentiality of participants (*Check relevant boxes below.)*

- ☐ Limits due to the nature of group activities *(e.g., focus groups):* The researcher cannot guarantee confidentiality
- ☐ Limits due to context: The nature or size of the sample from which participants are drawn makes it possible to identify individual participants *(e.g., school principals in a small town, position within an organization)*
- ☐ Limits due to selection: The procedures for recruiting or selecting participants may compromise the confidentiality of participants *(e.g., participants are identified or referred to the study by a person outside the research team)*

- ☐ Limits due to legal requirements for reporting *(e.g., legal or professional)*
- ☐ Limits due to local legislation such as the U.S.A. Patriot Act *(e.g., when there will be data storage in the United States).* When using USA based data instruments and data storage systems researchers are responsible for determining if this applies.
- ☐ Other:

24b. If confidentiality will be protected, describe the procedures to be used to ensure the anonymity of participants and for preserving the confidentiality of their data *(e.g., pseudonyms, changing identifying information and features, coding sheet, etc.) If you will use different procedures for different participant groups and/or different data methods be sure to clarify each procedure.*

**Participants are described only by subject number except on the consent forms which are to be kept in a locked file cabinet during the retention periods outlined in the consent documents. <u>All participants in this study will be tested individually rather than in groups. The observers mentioned in M.8b are members of the research team, nobody other than the principal investigator or his supervisors will be observing the sessions.</u>**

24c. If there are limits to confidentiality indicated in section 24a. above, explain what the limits are and how you will address them with the participants. *If there are different procedures for different participant groups and/or different data collection methods, be sure to clarify each procedure.*

## R. Use and Disposal of Data

### 25. Use(s) of Data

25a. What use(s) will be made of all types of data collected *(field notes, photos, videos, audiotapes, transcripts, etc.)*?

**The notes and electronic records from the study will be used in the preparation of journal and conference articles describing the research and inclusion in the experimenter's PhD thesis materials.**

25b. Will your research data be analyzed, now or in future, by yourself for purposes other than this research project?

☐ Yes ☒ No ☐ Possibly

25c. If yes or possibly, indicate what purposes you plan for this data and how will you obtain consent for future data analysis from the participants *(e.g., request future use in current consent form)*?

25d. Will your research data be analyzed, now or in future, by other persons for purposes other than explained in this application?

☐ Yes ☒ No ☐ Possibly

25e. If yes or possibly:

i) Indicate whether the data will contain identifiers when it is provided to the other researchers or whether it will be fully anonymous *(note that "fully anonymous" means* *that there is no identifying information, links, keys, or codes that allow the data to be re-identified).*

ii) How will you obtain consent from the participants for future data analysis by other researchers? *(If the data will be transferred in fully anonymous form, this request for future use can be made in the current consent form. If the data will contain identifiers or links/keys/codes for re-identification, consider requesting permission to contact the participants in the future, to obtain consent for the use of the data at that time)*.

**26. Commercial Purposes**

26a. Do you anticipate that this research will be used for a commercial purpose?

☐ Yes ☒ No

26b. If yes, explain how the data will be used for a commercial purpose:

26c. If yes, indicate if and how participants will benefit from commercialization.

**27. Maintenance and Disposal of Data**

Describe your plans for protecting data during the project, and for preserving, archiving, or destroying all the types of data associated with the research (*e.g., paper records, audio or visual recordings, electronic recordings, coded data*) after the research is completed:

27a. means of storing and securing data (*e.g., encryption, password protected computer files, locked cabinet, separation of key codes from raw data etc.*):

**Personally identifiable information will be kept in a locked file cabinet until destroyed. Electronic records without personally identifiable information will be retained as part of the thesis record.**

27b. location of storing data *(include location of data-storage servers if using web-based technology)*:

**ECS 654**

27c. duration of data storage *(if data will be kept indefinitely, explain why this is necessary and state whether the data will contain identifiers or links to identifiers)*:

**Relevant data for calculating the statistical results in the experimenter's PhD thesis will be retained, but not in a way that can be used to identify participants. <u>There is no intention to use this data for other studies, but in order to have the thesis be a complete record of the work, data for individual participants must form a part of that document (although not in any personally identifiable way.)</u>**

27d. methods of destroying or archiving data. If archiving data, please describe measures to secure or protect the data. If the archiving will involve a third party *(e.g., library, community agency, Aboriginal band, etc.)* please provide details:

**Data will be archived onsite in ECS 654.**

**28. Dissemination**

How do you anticipate disseminating the research results? (*Check all that apply)*

☒ Thesis/Dissertation/Class presentation

☒ Presentations at scholarly meetings ☒ Published article, chapter or book

☒ Internet (*Students: Most UVic Theses are posted on "UVicSpace" and can be accessed by the public*)

☐ Media (*e.g., newspaper, radio, TV*)

☐ Directly to participants and/or groups involved. Indicate how: *(e.g., report, executive summary, newsletter, information session):*

☐ Other, explain:

## S. Conflict of Interest

29a. Apart from a declared dual-role relationship (Section K, item 7), are you or any of the research team members in a perceived, actual or potential conflict of interest regarding this research project (*e.g., partners in research, private interests in companies or other entities*)?

☐ Yes ☒ No

29b. If yes, please provide details of the conflict and how you propose to manage it: