Use of Task-Relevant Spoken Word Stimuli in an Auditory Brain-Computer Interface

By

Jeremy Burnison

Submitted to the Graduate Neuroscience Program and the
Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Committee members

Dr. Jonathan Brumberg, Chairperson

Dr. Kostas Kokkinakis

Dr. Robert Fiorentino

Dr. Mark Chertoff

Dr. Randolph Nudo

Date defended: March 03, 2017

The Dissertation Committee for Jeremy Burnison certifies
that this is the approved version of the following dissertation :

Use of Task-Relevant Spoken Word Stimuli in an Auditory Brain-Computer Interface

Dr. Jonathan Brumberg, Chairperson

Date approved: March 03, 2017

Abstract

Auditory brain-computer interfaces (aBCI) may be an effective solution for communication in cases of severely locked-in, late stage ALS (Lou Gehrig's disease) and upper spinal cord injury patients who are otherwise not candidates for implanted electrodes. Feasibility of auditory BCI has been shown for both healthy participants, (Hill et al., 2004), and impaired populations (Sellers and Donchin, 2006). (Hill et al., 2014) found similar BCI performance in healthy participants and those with locked-in syndrome in a paradigm comparing words to pure tone stimuli. Additional BCI research has explored variations to augment P300 signals for use in speller paradigms, including more meaningful auditory stimuli (Klobassa et al., 2009; Furdea et al., 2009; Simon et al., 2014). It has been recognized in these studies that end users would much prefer natural sounds over a repeated tone stimulus. All of these systems required an association of sound with target stimuli, typically enforced by a visual support matrix. These systems would not be usable by the target end users of an auditory BCI. Attempts to remove the need for visual referencing by investigating a BCI system with serial presentation of spoken letter streams as stimuli (Hoehne and Tangermann, 2014) or spoken words (Ferracuti et al., 2013) has met with limited success but presents a potential high speed communication solutions. The present study highlights a method of using BCI task relevant spoken word stimuli to eliminate visually presented references. By utilizing spoken word stimuli, a BCI system could utilize a range of stimuli equivalent to the size of the users vocabulary and provide faster communication output than spelling systems. As a control, spoken word stimuli that have no task specific relevance are also tested. Stimuli audio-spatial cues have shown significant improve-

ments in aBCI performance (Käthner et al., 2013; Schreuder et al., 2011). The present study specifically evaluates the potential improvements to BCI performance of semantic and audio-spatial relevance by eliciting auditory oddball P300 responses to task relevant directional stimuli (spoken words: ‘front’, ‘back’, ‘left’, ‘right’). Participants completed several trials of a motivational game with directionally relevant targets over two experimental sessions. Offline analysis of training data was accomplished to evaluate the impact of stimulus characteristics on BCI performance. Questionnaire results on workload, motivation and system usability accurately reflected participant’s BCI performance. A behavioral button press study was utilized to further investigate the influence of spatial cues used in the paradigm, but also highlighted differences in the semantic relevance of the stimuli. Behavioral results correlated with BCI performance. The results of this study indicate task relevant stimuli are a viable option for eliminating artificial and visual stimulus references. This study’s results highlight several considerations for future auditory BCI studies, including: classifier selection, hearing threshold importance, aid of behavioral correlates to BCI performance and use of spatially separated spoken word stimuli.

Acknowledgements

I would like to thank a number of family and friends for their support during my academic pursuit of this degree. I would like to thank the countless students I have instructed and my own instructors that have help me experience so many wonderful moments of learning at the University of Kansas. I would like to make a special note of appreciation to the Life Span Institute and the Speech Language and Hearing department for giving me an academic home and a body of students and faculty to call my colleagues and friends. I thank Dr. Mabel Rice for provided me with opportunities to discuss my work with her and many other powerful minds, as well as many enlightening discussions on academic excellence as well as KU basketball. I would like to thank Dr. Honglian Shi and Dr. Yomna Badawi for guiding me through my cellular level neuroscience research experience and being there for me as I transitioned out of industry back into academia. I must thank Becky Harris for invaluable help with a thousand unforeseen needs. I'd like to thank the many members of the Society of Neuroscience Kansas City chapter for making me an executive member and letting me participate in many wonderful education and outreach programs. I'd like to thank my committee members in their guidance through not only this dissertation work but in pursuit of academic goals I hadn't even formulated yet. Each of them have opened my eyes to the inner working of achieving a doctorate degree and the modern complexities surrounding scientific research. They are all shining examples of how best to be a successful academic. My mentor, Dr. Jonathan Brumberg, graciously brought me back to Kansas, taught me all that I needed to know about BCI and answered so many questions very few people could. He has been a good friend and confidant throughout

this journey.

My parents, Howard and Diane Burnison were very supportive through these last several years and have made my life so much easier than it would have been otherwise. They along with my sister, Alisha, have have done everything they can to help me through several recent hardships. I couldn't have done this without their help on many levels. I would like to thank my grandfather, William Edward Schesser for the inspiration to do great things and take great care of the ones you love. Unfortunately, we lost he and my grandmother, Grace, during this process.

I would like to make a special thanks to Seth Lewin for camaraderie in academia and in almost all other aspects of life. Our friendship was crucial in keeping us both sane through the first years of graduate school. Johana Bravo has supported me, emotionally, in ways I didn't even know I could use support. She has brought me lunch, encouraged by continuation of the program and given me hope and motivation when I needed it most. She brings me incredible joy like no one else ever has. My son, Noah, gave me the initial motivation to return to the Midwest and so often reminds me of my own love of science and technology. I hope that what I've learned here can inspire and aid him in his life's pursuits.

Contents

List of Figures

List of Tables

Nomenclature

ASSR Auditory Steady State Response

BCI Brain-Computer Interface

CLIS Complete Locked-In Syndrome

ECoG Electrocorticography

ERPs Event Related Potentials

ICA Independent Components Analysis

LDA Linear discriminate analysis

RT Reaction Time - Time for participant to respond to stimulus. (ms)

SMR Sensory-Motor-Rythm

SNR Signal-to-Noise Ratio

SWLDA Step-wise linear discriminant analysis

ALS Amyotrophic Lateral Sclerosis

CP Cerebral Palsey

UI User Interface - the BCI system presented to the user

Chapter 1

Introduction

The condition of quadriplegia and mutism is known as locked-in syndrome (LIS) and is characterized by complete paralysis of the voluntary motor system with intact cognition and sensation Plum and Posner (1972). The etiology of LIS is associated with brainstem stroke, but may also arise from progressive neuromotor disorders such as amyotrophic lateral sclerosis (ALS) (Bauer et al., 1979). Patients with progressive motor neuron diseases gradually lose the ability to voluntarily control their muscles and will eventually require assistive technology to aid communication.

Augmentative and alternative communication (AAC) devices have provided many novel and effective ways to interpret what little output patients can still produce. This technology provides opportunity for diverse communication with or without the aid of a caregiver or speech and language pathologist. As they begin to show signs of speech and language impairments, patients with motor impairments master the skills needed to control these devices with their remaining motor abilities. As the disease progresses, patients need to learn and adapt to new input methods that match their changing sensory, cognitive, and motor ability. For example, initially a touch screen could be used to select communication options and later eye tracking may allow for reliable cursor control on the device.

Brain-computer interfaces (BCIs) are an emerging technology that uses recordings of brain activity to allow a user to control a computer program. BCIs are generally intended for individuals with profound neuromotor impairments that may result in paralysis (quadriplegia or hemiplegia) and / or the loss of articulate speech (anarthria). For example, Complete LIS leaves an individual completely unable to communicate or interact with the world around them. This horrific condition is highlighted by Brumberg and Guenther and deserves continued focus from BCI researchers and clinicians (Guenther and Brumberg, 2013). These systems can be beneficial to clinical patients by controlling their environment, locomotion or even provide communication output (Wolpaw et al., 1998, 2000; McCane et al., 2015). Individuals with cerebral palsy (CP), may also benefit from BCI owing to their severe neuromotor and speech impairments.

The prevalence of brain stem stroke leading to LIS is so low that no prevalence data is available (Smith and Delargy, 2005). In the US, there are approximately 20,000 people living with ALS at any given time (The ALS Association). A study on clinicopathology of LIS patients found that only 3.4% of the sample population reached CLIS status (Hayashi et al., 2016). Using this limited information, it may be estimated that approximately 680 CLIS patients may be living in the US, unable to communicate. The instances of LIS or complete LIS are likely not enough to promote the investment of medical device companies or communication device companies to develop products for this population.

1.1 Brain-Computer Interface for Communication

One of the primary and original applications for BCI was proposed by Wolpaw and colleagues to replace communication for those that had no other means (Wolpaw et al., 1991; Kübler et al., 2001). Despite decades of research this original application has yet to be fully integrated into clinical practices. BCI has a history of neuroscience and engineering development, but the future of BCI must also incorporate clinical experiences and practices currently in use for individuals with ALS, CP, and LIS.

The overall motivation for development or improvement in BCI technology is to eventually realize a clinically effective system (Wolpaw et al., 2000; Vaughan et al., 2006). BCI communication may allow more freedom and self-reliance than these patients could achieve otherwise, but in some cases this may be the only solution for any interpersonal communication. As these diseases progress, eye movement and eyelid control may become labored or unreliable, which can negatively impact visual BCI performance. In these cases, an auditory BCI paradigm may be the best option available. Auditory brain computer interfaces are one possibility to potentially re-open the door of communication for many individuals and could provide an enormous positive impact on their quality of life.

The present study uses lessons learned from AAC device development to provide a firm foundation for future communication replacement using an auditory BCI. The design and configurations of AAC devices comes from years of speech pathologist work and research and individualized considerations for each AAC user. By moving to a user interface that reflects systems that already exist and are already used by both patient and clinicians, we can leap over one huge hurdle of general clinical acceptance. The cost of developing and implementing a BCI user interface for communication is also ameliorated by this approach. There are still many hurdles to overcome to meet performance standards and ease of use of a widely accepted clinical BCI, but this approach overcomes several of them.

The research covered in this dissertation is motivated by clinical considerations that have been somewhat absent in the BCI research community. Past BCI research and current developments are summarized with a focus on non-invasive BCI systems intended for communication replacement. A BCI system for clinical use is proposed with a heavy influence from clinical practices of augmentative and alternative communication. Some unique features of such a BCI are investigated in study participants without any neurological impairments in both BCI and associated behavioral experiments. The research reported informs on future BCI development of a purely auditory BCI for communication using spoken word stimuli.

1.1.1 Principles of BCI

Brain-computer interface (BCI) was first conceptualized by Vidal (1973), but has since become a reality and is a quickly growing field of research in electrical and biomedical engineering, psychology, neuroscience, as well as clinical rehabilitation fields such as speech-language pathology, physical therapy, and occupational therapy. BCI systems developed in research labs may provide evidence that communication through BCI is possible but additional research is needed to bring the technology to widespread clinical acceptance. Understanding how these systems operate and their current capability is a first step in uncovering how they may be improved for clinical use.

Operation of any BCI can be broken into four steps (Schalk et al., 2004). First, for any BCI paradigm, the user completes a mental task. By completing this task, specific neural activity in the brain is produced. Second, accurate and reliable measurements of that brain activity are recorded. Third, brain activity recordings are interpreted as a decision or intention of the BCI user. Finally, the BCI produces the desired output (i.e. moving a wheelchair forward), providing feedback to the user on how the system interpreted the brain activity. Each of these steps is highly dependent on the others, making the development of an effective BCI a challenging but rewarding endeavor.

The study presented here utilizes non-invasive EEG recording for BCI control of communication outputs. This class of BCI will be the primary focus of discussion on BCI, however, many examples and concepts presented apply to other types of BCI systems.

1.1.1.1 Producing and Measuring Brain Activity

BCI systems use a number of technologies for measuring brain activity. Signals from invasive techniques like electrocorticography ECoG or microelectrodes represent single neurons (action potentials or single-unit recordings) or groups of neural firings (multi-unit recordings and local field potentials). A number of invasive techniques have demonstrated very reliable control of BCI systems (Moritz and Fetz, 2011; Schalk and Leuthardt, 2011). These technologies require surgery.

Non-invasive approaches monitor brain activity from outside the skull and are intended for patients that are unwilling or unable to undergo invasive surgery. Non-invasive techniques like

EEG, fMRI or fNIRS must consider indirect measures of the composite activity of millions of synchronized neurons. Signals measured with non-invasive measurements are often noisier and less reliable than invasive technologies.

Depending on the BCI task, the user's brain activity may be endogenous or exogenous. Endogenous signals are produced from internal intentions of the BCI user. A class of neural signals called sensory-motor rhythms (SMR) are produced by the BCI user when imagining limb movements or other motor outputs. These signals are initiated by the user or may be cued by the experimenter but are produced by entirely internal thought and are therefore classified as endogenous.

Exogenous signals are typically produced by processing of presented stimuli. Modification of attention to these stimuli is one of the primary ways exogenous signals can convey the BCI user's intention. A rapid serial visual presentation (RSVP) paradigm visually presents a set of symbols to the user, one at a time. The user has a specific symbol in mind and when that symbol is presented, a well characterized, amplified positive deflection around 300ms is elicited in the user's EEG. By processing the visual stimuli with some intention, the P300 signal can be used to identify the desired symbol.

1.1.1.2 Interpreting Brain Activity

The P300 is an example of an event related potential (ERP). An ERP is simply an (averaged) deflection in the EEG time course which is time locked to an attended stimulus, provoking a neural response. Identifying which stimulus exhibits the most pronounced P300 signal allows the BCI to interpret a discrete selection out of multiple options.

In other paradigms, a more continuous control of the BCI can be interpreted. The SMR is represented by spectral power of the EEG signal at specific frequency bands termed mu (8-12 Hz) or beta (18-25Hz). By modifying imagined limb movement and/or sensation, a BCI user can output a continuously variable spectral power within these frequency bands. This continuous signal lends itself to effective control of computer cursor position or movements.

The method of interpreting brain activity depends on how it is produced and measured. Interpreting ERP's is often accomplished with discrete linear models or machine learning algorithms (Krusienski et al., 2006) which classify each stimulus presentation into a category. Typically, binary categories represent either target or non-target stimulus. By identifying the 'target' stimulus the BCI can interpret the user's decision.

Collection of sufficient training data to produce a reliable model is required to account for considerable noise in non-invasive measurements. The user's intention is known by the BCI during training, so that brain activity can be characterized for each possible BCI selection. This trained model is then used to interpret future brain signals as specific BCI outputs. This model is the link between the brain's activity and control of the computer program.

1.1.1.3 Control a computer program

All BCIs aim to interpret neural activity so that reliable control of a computer program is achieved. The system may act as an extension of the individual's body by providing feedback to the BCI user. Limb movements provide tactile, proprioceptive and visual feedback of the neuromotor activity in the cortex, improving dexterity and coordination over time. A BCI can provide cortical activity information to the user through multiple sensory modalities, which can improve the level of control of a BCI with practice. This phenomenon highlights the widespread possibilities of BCI applications that plastically change neural activity utilized by the BCI system.

With some BCI's, the plastic changes of brain activity alone may bring therapeutic benefits. Stroke rehabilitation applications are gaining momentum and many researchers are now focused on BCI's effectiveness in recovering motor function after stroke. Neurorehabilitation BCIs have utilized SMR signals to feedback cortical motor neuron activity, aiding in recovery of hand and arm mobility (Ramos-Murguialday et al., 2013). These same signals can control a robotic limb, control cursor locations, or direct a powered wheelchair. The interface is key to providing feedback to the BCI user and this feedback may serve a number of purposes. Brain-computer interface directed control of neural activity has been shown in non-human primate studies (Moritz and Fetz, 2011). Success in human control of a non-invasive, EEG based BCI systems has also been demonstrated

(Wolpaw and McFarland, 2004). BCI systems utilizing this type of therapy are already being produced commercially, including Recoverix by g.tec, which has been used in the rehabilitation of upper limb motor skills after relatively few training sessions in stroke patients with severe paralysis.

A BCIs for communication replacement might take on many forms. Literate patients will likely be able to spell or select words in order to convey ideas and maintain interpersonal relationships. Discrete selection paradigms will likely use ERP brain signals to select from a number of possible items. Often, an ERP signal is used to make discrete letter selections for spelling applications, but this study will explore selection of words.

1.1.2 Communication BCI in Research

The design of a communication replacement BCI determines how the BCI user expresses themselves. Non-invasive BCI research has centered around spelling systems in order to provide users with communication output as diverse as their own speech. Continued communication is key to a good quality of life, so optimizing communication output from the BCI is a paramount goal of BCI research and development.

1.1.2.1 RSVP and SMR Spelling Systems

Many EEG features and ERPs can be used to control a computer program though the BCI approach, but the P300 signal is the most common one employed in communication replacement BCIs. A spelling task is the most frequently used as well. In such a paradigm, the P300 ERP is elicited by identification of a target letter in an oddball paradigm presentation and is present in healthy and impaired populations that BCI research targets (Sellers and Donchin, 2006). An oddball paradigm is one in which a set of stimuli are presented in a random order with a particular target stimulus presented less often than other non-target stimuli.

The P300 ERP is reliably elicited across a wide range of paradigms making it advantageous for BCI applications. A survey of P300 spellers highlights the influences of system parameters on BCI usability and its widespread applicability to healthy users (Guger et al., 2009). The P300 designation indicates that the signal is a positive deflection (P) in the EEG voltage approximately 300ms after the onset of the target stimulus. The timing and scalp location of maximum P300 amplitude can vary depending on several factors, including: presentation scheme, the mode and clarity of the stimulus, mental state of the study participant and EEG recording parameters (namely ground and reference electrode locations).

The interpretation step of the BCI system aims to identify differences in P300 amplitudes due to target or non-target stimulus presentation. Recent P300 feature extraction algorithms consider the entire ERP waveform in machine learning context and do not focus solely on the P300 amplitude. BCI performance in oddball paradigms have further shown to rely heavily on ERP features other than the P300 amplitude (Halder et al., 2013; Hill et al., 2014).

While new algorithms, user interfaces and parameters for P300 Speller BCIs have been optimized through systematic research, the P300 spelling task of attending to the target letter, has remained largely the same (Krusienski et al., 2006). One development in visual P300 spelling paradigms has come in the way of efficient letter presentation using language models (Orhan et al., 2011), improving time to selection of the desired letter. These systems identify which letters are most likely to be selected next, according to those language models, and increase the presentation frequency of those letters. This approach may speed up the rate of accurately selecting desired letters.

Spelling systems using SMR signals have been developed as well and include clever ways of optimizing letter selection. The virtual keyboard also uses SMR signals to select letters on an asynchronous timeline (Scherer et al., 2004). Imagined foot movements scroll letters through a left and right selection box and the corresponding imagined hand movements select the desired letters. The Hex-o-Spell method uses binary selection of groups of letters to reduce both the accuracy needed in the continuous output and the number of decisions to make a letter selection (Blankertz et al., 2006).

1.1.2.2 Visual Grid Spellers

The visual P300 speller has been a popular paradigm for communication BCI in research (Farwell and Donchin, 1988; Donchin et al., 2000). The grid spelling paradigm has served as the benchmark task of communication BCI development in terms of performance (Cecotti, 2011). While RSVP presents a single letter at a time, the grid spelling systems present all letters simultaneously. Highlighting of grid rows and columns in a randomized fashion is used to elicit an amplified P300 signal when the letter of focus is highlighted. Grid spelling systems produce faster spelling rates, because the target letter can be presented (highlighted) more frequently without biasing a specific letter.

Many visual P300 BCI user interfaces have used black and white spelling grids or single letter presentation. Bi-color chromatic flickering or highlighting has shown some improvements in performance (Takano et al., 2009). See Figure 1.1 for an example of a visual P300 grid spelling display. The 3rd row is highlighted in this figure. Once the letter of focus is highlighted multiple times, the BCI system would make an accurate selection.

Figure 1.1: P300 Grid speller example Farwell and Donchin (1988)

Language models can also be used to optimize the highlighting of likely letters in the gird speller system. Ma et al. (2012) found that optimizing letter flashing, according to a statistical language model, reduced character selection time by over 50%.

BCI speller design considerations have improved the speed of letter selections while maintaining accuracy (Blankertz et al., 2007). All of these developments have resulted in minor gains in accuracy and speed of communication.

1.1.2.3 Auditory spellers

While visually evoked ERP's are reliable and well researched, there is a clinical need for a system that doesn't rely on visual acuity and attention (Nijboer et al., 2008b). For those patients with strong visual control, eye tracking technology is well developed and is a widely used clinical tool to augment or replace communication output for paralyzed patients. Patients may benefit from RSVP spellers if gaze control is limited. Patients with poor eye or eyelid control may have difficulties with eye tracking systems and these systems do not work reliably in all lighted or any unlit environments (e.g., outdoors). Visual BCI systems are unlikely to ever be more effective than eye tracking, so for many potential BCI users an auditory BCI may be the best solution.

An auditory brain computer interface (aBCI) is a system that uses auditory stimuli to elicit a neural response, which is detected, classified and used to convey an intended decision. The P300, outlined previously, is also elicited by auditory oddball stimulus presentation schemes. Grid spelling techniques may still be employed by having a set of auditory stimuli that correspond to the rows or columns of the grid of letters. By making a selection of row and then of column the intended letter can be selected (Cai et al., 2012; Käthner et al., 2013; Schalk et al., 2004; Schreuder et al., 2011).

Improvements in Stimuli A myriad of auditory stimuli have been employed in an effort to move away from the less pleasant tone stimuli typically used in auditory ERP studies. Simon et al. (2014) conducted a pre-study found a set of animal sounds were the most discriminable stimuli from a group of 5 different environmental sound sets. A comparison of visual and an auditory P300 based speller using simple environmental sounds (i.e. 'Thud', 'Chime') showed it was able to perform similarly to the visual paradigm after 11 sessions (Klobassa et al., 2009). A study comparing tones, spoken and sung syllables found that the tone stimuli showed lower classification accuracy than syllables (Höhne et al., 2012). Researchers have used paradigms where spoken letters were presented as auditory stimuli showing these stimuli were viable (Horki et al., 2015; Sellers and Donchin, 2006) . These studies support the use of more natural auditory stimuli for both performance and acceptance by potential BCI users.

Auditory only BCI Visual references or visual support matrices are used extensively in auditory BCI grid spelling systems (Kübler et al., 2009; Klobassa et al., 2009; Furdea et al., 2009). Visual references allow the user to maintain a continuous association of auditory stimulus and grid row or column to be selected. While this allows any set of auditory stimuli to be used in the P300 oddball paradigm, it forces either memorization of letters associated with each auditory stimulus or a visual reference. Memorization of letter-stimulus associations may increase working memory requirements and has a negative impact on the ERP used for BCI control (Pratt et al., 2011). The blind or those suffering from complete LIS would be unable to utilize a visual reference in an auditory BCI.

In an 'auditory only BCI' the user is provided auditory instruction, auditory stimuli to control the system, and auditory feedback of BCI selections. No visual reference, stimulus or instructions would be critical to the use of the system. Such systems are intended for blind individuals, those with poor visual acuity, limited eye control and/or no other means of conventional communication output.

Auditory Steady State Response (ASSR) paradigms have served as auditory only BCI in multiple research studies, (Hill and Schölkopf, 2012; Hill et al., 2012; Halder et al., 2010). Typically, one auditory stream is played in the BCI users right ear and the other in the left ear. Attention modulation to one of two concurrent auditory streams allows only binary decisions to be made in the ASSR paradigms. Binary decisions are common in communication of severely paralyzed patients and may provide a comfortable transition to BCI but would provide a very slow communication rate (Higashi et al., 2011; Kanoh et al., 2010).

Spatial Cues Just as ASSR used lateralization of auditory streams to aid the user in attending to one sound stream, separation of auditory stimuli in other paradigms can aid the aBCI user's attention to a target stimulus. In attempts to speed up spelling rates in an auditory only BCI the Charstreamer paradigm used rapidly presented spoken letters in multiple asynchronous sound streams (Hoehne and Tangermann, 2014). Two sound streams were played in the right and left ear independently and the third stream was played in both ears. While this audio was presented over headphones this presentation utilized human spatial hearing to give the perception that streams were coming from the left, right and in front of study participants.

Spatial cues have been identified as beneficial for performance of auditory oddball BCI performance (Schreuder et al., 2010; Käthner et al., 2013). Separating stimuli by sound source location provides an additional auditory cue to the stimuli allowing improved attention to a target stimulus. These studies used tones with noise to enhance the perception of sound source location.

Auditory BCI research has provided evidence that spatial cues and natural stimuli are beneficial to BCI performance. Auditory only systems have been developed but suffer from slow spelling rates or poor performance. Utilizing spoken word stimuli would allow an auditory only BCI to communicate ideas with a single selection instead of the many required to spell a word with spelling systems.

Speech stimuli have been utilized in several studies already mentioned. Recent studies have used common word stimuli in oddball paradigms with success (Ferracuti et al., 2013; Kleih et al., 2015). With a number of studies succeeding in utilizing speech like or spoken word stimuli for auditory BCI, future research should continue to explore speech driven auditory only BCI for clinical use.

1.2 A Proposed BCI for Clinical Use

While improvements in the functionality of BCI systems are being realized in research, considerations for user experience and effectiveness in daily communication have been overlooked. Considering clinical acceptance and BCI features optimal for daily communication replacement, a BCI is proposed.

The proposed BCI would use words or ideas as selection items instead of letters for spelling, limiting the output to frequently communicated ideas but allowing communication to occur faster and with less effort. Discrete selections will be made by characterizing target class EEG features in the time domain. The system should be auditory in both stimulus presentation and in feedback and instruction to the user. The control of an AAC device provides a well-developed BCI user interface that has already achieved clinical acceptance. These features provide the primary guidelines for developing a clinically focused BCI.

Motivation from AAC It is very important to ensure that user-centered design practices are incorporated into new BCI research. The BCI society has formally recognized the need for clinician and patient involvement in the research and development activities surrounding BCI (Kubler et al., 2006), and some clinical researchers are already paving the way for inclusion of clinicians and patients to be full partners in BCI development (Peters et al., 2016). BCI could aid in activities of daily living of the BCI user (Suyama, 2016).

BCI control of an AAC device highlights many advantages in clinical acceptance. Most electronic AAC devices output synthesized or recorded speech, and typically provide a selection of words and phrases rather than just letters. These audio outputs are meant to communicate to individuals around the patient, but could be used as stimuli for aBCI paradigms and also serve as a means of informing the user without intact vision of available selection items.

Many AAC device 'page sets' represent a collection of communication items, customized by the user along with a speech language pathologist or caregiver. A communication item often resides in a categorical, nested menu, which allows for a large number of possible communications with very few item selections. For example, a menu of 5 different categories with 6 items in each category would allow 30 different options to be selected with 2 trials. In most auditory grid spelling applications two selections, a row and a column, are required to select a letter. With some AAC devices menus include an option to move the user into a spelling mode with individual letter selection. This spelling mode could utilize any BCI spelling system that works best for that specific AAC-BCI user.

Many patients with neurodegenerative disease may use an AAC device in the earlier stages of the disease progression. By using BCI control of this device they will already be very familiar with much of the system and method of communication, as will their friends, family and caregivers. Utilizing BCI control of an AAC device may be the easiest transition for such patients entering complete LIS.

Clinical Need for Auditory Only BCI BCI represents a complex and likely expensive clinical tool that may be somewhat unreliable and slow in terms of communication output. For patients that still retain some motor control or eye movement, physical switches, caregiver interpretation of motor output or eye tracking are likely to be less complicated, less expensive, more reliable and less prone to environmental factors than BCI. When patients completely lose motor output and intact vision, auditory only BCI may be the only communication option available. In this way, auditory only BCI represents a more likely clinical tool than vision based BCI or aBCI requiring vision. Such systems might also be necessary for blind or severely visually impaired patients.

Auditory only BCIs may reduce user fatigue compared to visual BCIs. In an ASSR study an ALS patient with good visual acuity was quoted as saying “my eyes get tired, but never my ears“ (Hill et al., 2014). This study featured a purely auditory BCI, where ERP eliciting stimuli and other task instructions were presented through auditory means.

Another use-case for auditory BCI would be controlling a communication device while using vision for other tasks. When controlling wheelchair movement it would be beneficial for BCI users to use their vision to monitor their surroundings instead of attending to visual stimuli. An aBCI could be used while users look at the person they are communicating with instead of looking at a computer screen. This would fit a very natural communication paradigm and may improve social interaction.

Currently auditory BCI has demonstrated lower performance in terms of percent accuracy and information transfer rate (ITR) compared to similar visual P300 spelling systems (Sellers and Donchin, 2006; Kleih et al., 2015). However, studies conducted over multiple sessions show aBCI accuracy can reach that of visual systems with training (Nijboer et al., 2008b; Klobassa et al., 2009). BCI researchers frequently report the number of participants able to use the BCI system above a threshold (70%), indicating that the BCI system proposed in a study is feasible for clinical use for communication (Kübler et al., 2001). Many auditory BCI systems have demonstrated meeting this requirement but improvements in speed and improved accuracy are desired.

Auditory only BCI communication replacement candidates can greatly benefit from or may even require this paradigm modality. A method to greatly increase speed of communication would greatly benefit the clinical acceptance and usability of these systems.

Word vs. Letter Selection BCI A system of selecting whole words or even phrases in a series of nested menus would greatly increase the time to communicate those available communication outputs. While developments in BCI for communication focus primarily on spelling applications, some patients may prefer or require this alternative means of communication. This system allows a patient to convey more complex ideas in much less time than spelling, but provides a limited set of outputs.

While full-word selection should not replace spelling applications, it should be recognized as a useful paradigm to any potential BCI user (as it is for AAC). The arrangement and configuration of such page sets could be customized and re-arranged with little user training and no new memorization of stimulus-item association.

For regular clinical use, improvements in the user interface are needed in order to aid attention, motivation and ease of use. Improvements in the user interface (UI) will not only help to promote clinical acceptance of such devices, but if carefully designed, would aid in the proficiency of the fundamental cognitive tasks required for users to operate the system.

Percent accuracy reflects how often a correct selection is made by the BCI system while ITR combines both speed and accuracy to describe the amount of information conveyed per second. Depending on the aBCI paradigm, an item or letter selection can require many seconds of stimuli presentation. ITR is used to compare performance across various paradigms that may favor either speed of selection or percent accuracy. Expressed in units of bits per minute, this measure considers time to selection, how many selections are available, and percent accuracy achieved. aBCI grid spelling systems require at least two item selections for a single letter to be chosen. RSVP systems require several letters to be presented. Depending on how a nested menu is arranged, hundreds of possible communication items could be chosen in the same number of selections as would be required to spell a single word. By utilizing a system where more complex ideas can be conveyed the slower auditory paradigms may become more accepted and useful in clinical systems.

Spoken Word Stimuli Spoken Word Stimuli provide an intuitive means of selecting whole words or concepts in a BCI paradigm. The proposed BCI would be an item selection BCI using spoken word auditory stimuli in order to control an AAC device a patient is already familiar with. By utilizing spoken word stimuli, a number of advantages can be realized.

It is anticipated that classifying EEG of spoken word stimulus may be more complex and variable and yield poorer accuracy than tone or environmental sound stimulus. Several studies have already used spoken letters, syllables and words with some success. These studies have used BCI systems with binary decision or spelling output and have yielded slow communication rates. Selecting words will allow for fewer BCI selections to convey frequently communicated ideas.

While spoken word stimuli have been used in auditory BCI research, the full benefits in communication replacement have not been explored. By utilizing spoken word stimuli, a clinical BCI for communication replacement may exhibit many benefits including:

- Presenting natural and comfortable auditory stimuli
- Reducing the cognitive load of memorizing stimulus/decision associations.
- Utilizing the well characterized P300 EEG signal and oddball paradigm.

- Providing a purely auditory system with no visual reference required.
- Easily integrating BCI with AAC devices, and
- Allowing for rapid communication of frequent ideas

Motivation Factor User-centered BCI design has the potential to improve user motivation, which past research has correlated to improved performance (Käthner et al., 2013; Nijboer et al., 2008b). Some studies show that patients with impairments sometimes outperform healthy subjects (Piccione et al., 2006), and it is hypothesized that motivation may play a large role in this result. Attentiveness and focus also influence EEG signals and it is likely that motivation improves BCI performance through these two cognitive state characteristics. For healthy participants, there is no real-life benefit from performing well in these studies. For patients that may benefit from BCI use in the future, aiding research and development of BCI may provide a very motivating scenario. Researchers have found that grouping participants by motivation doesn't show any significant effects on performance between groups (Kleih and Kübler, 2013). Between subject variability in BCI performance is typically high and other factors besides motivation may play a larger role.

While the impact of motivation on performance needs further study, it should be accepted that a lack of motivation would not be beneficial to BCI use. Future studies on BCI user interfaces should consider motivating aspects but not at the expensive of other factors of BCI performance.

1.3 Investigating Stimulus Relevance

In the proposed BCI, a multitude of auditory stimuli must be effectively classified by the BCI for a diverse set of communication ideas to be expressed. It may be that standard auditory stimuli may allow for much more reliable BCI classification accuracy and overall performance. This would require a stimulus-communication item reference that would eliminate one of the benefits of decision relevant stimuli. It is important to also understand how using a set of standard spoken word stimulus over relevant stimuli might impact BCI performance.

This is the first aBCI study to measure the effect of task relevant stimuli on BCI performance. Optimization of auditory BCIs using task or decision relevant stimuli is expected to reduce cognitive workload, reduce or eliminate required training periods and aid in clinical acceptance of auditory only BCI. This study aims to quantify any BCI performance enhancement that can be realized with decision relevant stimuli.

Research should continue to investigate how the unique characteristics of such a system will influence BCI operation and which features might be considered for future clinical BCI design. Stimulus relevance will be investigated in a directional task to allow for association of stimuli spatial cues and decision to add another layer of relevance to the stimuli. A behavioral experiment is included to further investigate how spatial separation of stimulus and stimulus presentation parameters will influence optimal attention modulation by the BCI user. These behavioral measures are anticipated to correlate with aBCI performance, so this hypothesis will also be tested.

In order to maintain engagement, the BCI task in the current study engages participants in a game-like environment, with multi-trial goals and a colorful and interesting visual reference. Although the aim of the study is to investigate aspects of spoken word stimuli and their influence on BCI performance, improved performance through a motivating task is also considered. The vast majority of BCI research uses spelling tasks to gauge performance but this study introduces an intuitive non-spelling task, more reflective of the AAC – BCI paradigm proposed thus far.

BCIs must be customized for each user, by taking into consideration: residual motor output, visual acuity, hearing loss, and experience with AAC devices. Cost of the BCI system, maintenance and operation of the system and availability and training of caregivers may influence the optimal solution for a patient. The preferences of a patient may lead to use of a BCI like the one proposed but a host of other options, including spelling tasks, are also likely to be clinically desirable.

Chapter 2

A Novel aBCI to Investigate Spoken Word Stimuli Relevance

2.1 aBCI Design Rationale

This section details the process of developing the auditory BCI system used in this study, along with a discussion of its motivation and experimental design. Much of the rationale is based on existing theoretical frameworks, but practical considerations influenced the study design as well. A summary of the scope and goals of the study is initially given, followed by the rational for developing the BCI system and study design.

2.1.1 BCI Requirements

The motivation of this study is to investigate aspects of spoken word stimuli that might prove useful for an auditory only BCI for communication. The first aim is to test the impact of semantic relevance of the words, as this characteristic would be critical to the flexible and intuitive benefits of such a system. The second aim is to understand how audio-spatial cues benefit auditory BCI performance. The design utilized also aims to incorporate motivating factors and engaging interfaces to optimize performance of the system.

The Requirements of the BCI in this study are:

System Should incorporate features that have exhibited optimal performance in past BCI studies including:

1. Feature extraction and classification approach
2. Comprehensive EEG acquisition

Task

1. must be non-spelling
2. Selections must have some spatial relevance
3. must be engaging and motivating to optimize performance

Stimuli

1. optimal presentation rate for performance
2. must have semantic relevance and include control stimuli without task relevance
3. must be presented in an oddball paradigm to elicit P300 like ERP signals
4. Must include spatial cues as this has shown performance enhancement

2.1.2 BCI System Design

The BCI designed for this study took into account all of the requirements defined in the previous section. The BCI system will utilize a professionally designed EEG acquisition system developed by BCI experts, g.tec. Feature extraction and classification approaches are modeled after those used in past BCI research. The user interface and stimulus presentation are novel to the BCI literature and are designed to test stimulus characteristics inherent in an auditory only, word selection BCI using spoken word stimuli.

This study aims to uncover the influence of the of semantic and spatial relevance of spoken word stimuli on BCI performance. Piccione et al. (2006) employed a directional task that required a ball icon to be moved to a target location. This unique auditory BCI helped inspired the use of a directional task, allowing spatial and semantic relevance of the target stimulus to be tested simultaneously. By using a directional task in a game like environment, audio source location cues (spatial cues) of the stimuli will hold relevance to the task itself. The impact of semantic relevance will be tested by using 'directional' and 'non-directional' word stimuli.

The functional components of the BCI system are first described and then the details of the study itself are described.

2.1.2.1 BCI System Components

Any BCI system can be separated into five of major components. Each of these components will be defined and discussed below:

1. Stimulus Presentation
2. Signal Acquisition
3. Signal Processing
4. Feature Extraction
5. Feedback Mechanism

Stimulus Presentation in this study is primarily comprised of the auditory presentation of spoken words from various speakers positioned around the head of the participant. It also includes the visual presentation of the target sound's location, as well as the concept of the PacGame described later in more detail. In many BCI systems the processing of a stimulus evokes EEG signals or ERPs that are used to control the BCI. Presenting well controlled and consistent stimuli with well characterized ERP features is key mechanism of BCI functionality. The participant's task is simply to attend to the target stimulus and ignore all other stimuli as best they can.

For this study, well controlled recording of spoken word stimuli in a sound dampening environment with professional audio equipment was accomplished. Each auditory stimulus is post-processed to ensure auditory features of the stimuli are as equivalent as possible.

Signal Acquisition is accomplished here with 62 channels of EEG recording, with monopolar ground and left ear recording reference electrode. In whatever manifestation of neuronal activity recording done, it is important that the recording device monitor activity of the area of the brain expected to produce the signal of interest. This illustrates how stimulus presentation and signal acquisition for a given system are linked and should be designed with one another in mind. The EEG channels utilized here extensively cover all areas of the cortex anticipated to produce useful signals for BCI control. Utilization of numerous channels enables sophisticated post processing of the signal to eliminate spatially centralized non-brain activity, user motion artifacts.

Other features of the acquisition hardware aided in elimination of additional sources of noise. Active electrodes were utilized in this study to aid in additional elimination of environmental noise. A high input impedance EEG amplifier was used to again eliminate signals induced in the recording hardware. Additional steps to eliminate non-brain activity from the recorded EEG signals can be accomplished through signal processing.

Signal Processing steps often including spectral filtering and selection of EEG features based on the expected cortical activity generated by the presented stimulus. These steps are specific to the signal of interest and are applied to eliminate confounding variability and environmental noise, which is a major challenge in EEG signals. Whether applied after the EEG recording is complete or in real-time, as the EEG is being recorded, these steps modify the raw voltage measurements at the scalp.

ERP signals are often averaged together to help improve signal to noise ratio (SNR). In the present BCI design, presenting each stimulus 15 times per trial allows random variation in the signal to be potentially cancelled out. Again, it should be noted how appropriate signal processing is dependent on the stimulus presentation and signal acquisition details of the overall system.

In auditory oddball paradigms, a number of stimuli are played in random sequence with one stimulus acting as the 'target'. A single trial includes several randomized sequences of the stimuli. The user attends to and anticipates the presentation of the target stimulus by keeping count of the number of 'target' presentations they hear throughout a trial.

Feature Extraction is a process of identifying the differences in EEG signals that can be used to control a BCI. Training the BCI classifier characterizes the differences in EEG signals resulting from target stimulus presentation and those occurring due to non-target stimuli. In training sessions, recording EEG signals for each participant during stimulus presentation allows differences in target vs. non-target EEG traces to be identified and quantified for that specific BCI user. The mathematical representation of these differences in EEG defines the BCI classifier. In online sessions, this classifier is applied to the continuously recorded EEG and decisions about which auditory stimulus most closely represents the target category can be made. This decision is conveyed as the BCI user's choice and that result is feedback to the user through some user interface.

Feedback Mechanisms are often accomplished by the same features of the system accomplishing the stimulus presentation, but this may not always be the case. Visual, auditory and/or tactile feedback inform the BCI participant what outputs the BCI system has produced. The BCI user is not consciously aware of the neural activity that produced the BCI output, but giving feedback of the system's choice allows the participant to create those associations. This connection of unconscious cortical activity to the BCI user's consciousness through artificial feedback is hypothesized to induce neural plasticity and potentially provide an avenue for physiological recovery of function.

Feedback during the online sessions of the present study was accomplished by the movement of a computer screen icon in the direction corresponding to the 'target' sound. An auditory feedback also indicated if the selection was correct or not. See Section 2.1.4 for more information on the BCI paradigm. In SMR paradigms continuous feedback is essential and is shown to rapidly allow users to change their mental tasks to produce the desired outputs (Wolpaw et al., 2002). In the present system, feedback occurs only at the end of a trial and may not allow for significant adaption by the user within a single session.

2.1.3 Participants

As in the majority of BCI studies (Furdea et al., 2009; Halder et al., 2010; Nijboer et al., 2008b; Piccione et al., 2006; Klobassa et al., 2009; Hoehne and Tangermann, 2014; Schreuder et al., 2011), healthy participants are initially utilized to investigate the feasibility and advantages of a BCI system. Healthy participants do not fatigue as quickly and can complete many trials, are easily recruited, provide clear feedback and don't require many special considerations during experimentation. Feasibility of BCI approaches in healthy patients has been shown to generalize to impaired populations showing similar outcomes, albeit often with reduced performance (Sellers and Donchin, 2006; Kübler et al., 2009; Nijboer et al., 2008a; Simon et al., 2014). This highlights the need for improved methodologies and the importance of continued research with the disadvantages of clinical populations in mind.

2.1.4 The Task

The BCI task used in the present study is similar to all auditory P300 based tasks. The participant is instructed to count the number of presentation of the target stimulus in an oddball paradigm. The auditory oddball paradigm is characterized by serial presentation of multiple stimuli, in random order, where a specific 'target' stimulus is presented much less frequently than the other distracting stimuli. The presentation of anticipated and rarely occurring target stimulus elicits an amplified P300 signal in the EEG compared to the non-target stimulus presentation. In the ubiquitous BCI spelling systems the participant attends to a letter in a specific visual location or a specific auditory stimulus corresponding to a letter and counts the number of presentations they recognized. Counting is a strategy to maintain attention on that target over the repeated presentation of all the other stimuli.

The PacGame user interface uses the task described to move a pac-man icon in one of four direction in a 5x5 grid. Four spoken words are presented from four different locations around the participant, each corresponding to a direction the icon would move. Details of the PacGame user interface are covered in Section 3.1.3.1.

2.1.5 Spoken Word Stimuli

Testing the feasibility of spoken word stimuli in an auditory BCI paradigm is most informative if the ultimate clinical system is considered. In an auditory BCI the spoken word stimulus allows the user to have an intuitive association with item selection and the target stimulus. This eliminates the need for visual reference and flexibility in item selection options. One of the primary aims of this study is to test how this intuitive connection between the meaning or semantics of the stimulus and the item selection influences the BCI operation.

Directional words ['front', 'back', 'left', and 'right'] were used to represent the intended direction of the PacGame icon and the stimulus location with respect to the seated participant. The control group for this experiment utilizes spoken words that do not have any semantic relevance to the directional task. Non-Direction words ['joy', 'while', 'care', 'doubt'] were chosen that meet the following criteria:

- English words with high linguistic frequency similar to the direction words being used.
- Should have similar duration, syllables and phonemic content as the direction words.
- Abstract words that could not be easily associated with a direction, location or object.
- Non-action words as these could be associated with motion in a specific direction.

All word chosen are monosyllabic and common words in American English. Selecting high frequency English words is one of the easier tasks as these are words that are most likely to come to mind for any native English speaker. Words dealing with time and emotions meet the third and fourth criteria. All eight words were matched in intensity, pitch and duration using Praat software. Each stimulus was recorded in a sound booth and then processed to maintain similar acoustic characteristics.

Because of the acoustic complexity and high degree of variability of human speech, acoustic characteristics for all words could not be perfectly matched. While an envelope of intensity can have equivalent energy between words, the profile of intensity over the utterance of 'front' and

’back’ cannot be identical due to the differences in proper pronunciation of the two words. By limiting acoustic variance between words, the amount of variability in neural processing that occurs as the participant hears each stimulus is limited.

Previous research has shown EEG signals are more variable between participants when using words than when utilizing more simplistic auditory stimuli, like pure or complex tones (Hill et al., 2014; Hoehne and Tangermann, 2014). The activation of a diverse set of cortical areas related to the lexical representation of these spoken word stimuli surely makes for a more difficult challenge in reliably differentiating the resulting EEG traces. The feasibly of using this category of auditory stimulus with previously vetted feature extraction techniques is one of the major outcomes of this study.

2.1.6 Classifier

The term classifier, here, refers to the method of selecting of a possible BCI output based on ERPs from the EEG signal. With most ERP classifier approaches the EEG data recorded during a stimulus presentation is separated into spatial and temporal features. The spatial component is comprised of several different EEG channels placed in different locations on the scalp for head and earlobes. In the present study, 62 monopolar active electrode channels serve as spatial locations of EEG. EEG voltage at each of these 62 locations were recorded at 256 Hz sampling frequency.

In order to differentiate signals produced during stimulus presentation EEG data is segmented into epochs which are time aligned to the onset of the stimulus presentation. In order to improve signal to noise ratio (SNR) of ERP signal detection, several epochs can be collected by presenting multiple sequences of each stimulus over several trials. Using many trials, a mathematical model can be generated for each spatial and temporal feature to describe the statistical differences in target and non-target EEG signals.

EEG data collection in research environments often includes a large number of EEG channels (spatial locations) and is capable of high data collection rates (up to 1000 Hz). Consequently, there may often be a huge number of temporal/spatial features in a BCI training set (In the present design: 62 channels x 0.6 sec x 256 Hz x 15 reps x 4 stimuli x 32 trials).

Researchers often reduce this feature set by decimating the temporal stream of data, which also eliminates some noise sources in the process. Decimating includes two steps. First the data is low pass filtered at a frequency equal to the sampling rate divided by the decimation factor. Second, the data is downsampled to this rate by selecting evenly spaced values over time. This process reduces the size of the feature set.

Selecting of EEG channels that reflect large P300 deflections in a given paradigm may allow for a reduction of spatial features. These could be identified visually after processing and plotting target and non-target averages for a given participant (Schreuder et al., 2011). The step-wise linear discriminate analysis (SWLDA) algorithm automates feature selection by adding each feature, one at a time, to a linear discriminant model and determining the features that provide significant increases in predictive power for categorizing target and non-target stimuli. The order in which to add features to the linear model is randomized to remove effects of bias on the specific spatial location or temporal feature. This data centric approach may suffer from overfitting and a lack of generalization for future data collection, but has been shown to yield good accuracy results in P300 BCIs (Krusienski et al., 2008).

Many of the P300 BCI systems in recent research have utilized SWLDA (Simon et al., 2014; Käthner et al., 2013; Furdea et al., 2009). This approach was also used in the present study's aBCI pilot trials. Regularized linear discriminant analysis (RLDA) model was compared offline and found to yield improved accuracy, so RLDA replace the SWLDA classification approach for the final study design.

2.1.6.1 BCI2000

A pilot study attempted to implement a step-wise linear discriminant analysis (SWLDA) approach using BCI2000 software, developed and maintained by the National Center for Adaptive Neurotechnologies (Schalk et al., 2004). SWLDA is programmed into the core BCI2000 environment, which can be used to control a P300 based BCI.

This experiment's user interface and stimulus presentation software was initially constructed in Python and it was expected that BCI2000 could provide EEG classification and interpretation along with this interface. Unfortunately, after many attempts to use multiple BCI2000 extensions, incompatibility between the EEG acquisition hardware, the User interface, and BCI2000 eliminated this option.

Custom Python Classifier Another pilot attempt was made, using a custom BCI classifier algorithm constructed in Python to ensure compatibility between PacGame interface and EEG acquisition. Much of the SWLDA approach was previously coded in Python by Collin Stocks (PY3GUI, `https://github.com/collinstocks/Py3GUI`). This publicly available set of scripts from GitHub was used to guide development of an online classifier. EEG data analysis tools developed for MATLAB in the Speech and Applied Neuroscience Lab were used to process the EEG data and generate the SWLDA classifier model. The PacGame software would then load the classifier model to provide real-time decoding and control of the BCI interface. See Appendix: pacgame_decoder.py for details on the decoder programming.

Increasing presentation rate Pilot participants suggested the presentation rate was somewhat slow, losing the participant's focus and engagement in the task. The stimulus onset asynchrony (SOA) was initially set to 750ms and was able to be decreased to 400ms, giving no significant silence period between stimuli. The number of presentations increased from 10 to 15 per trial as a result of faster presentation.

2.1.6.2 Online EEG Processing

The final online processing pipeline for incoming online data included filtering, downsampling, segmenting of data, a baseline correction step and then passing the pre-processed data to the online classifier. The data for an entire trial was collected before starting the online pre-processing.

Filering and Downsampling A forward-backward 1 to 51.2 Hz band pass filter was applied to ensure zero phase shift would occur in the data. A factor of 5 downsampling was accomplished by selecting only every 5th time data point. The upper limit of 51.2 Hz in the bandpass filter was determined by dividing the original sampling rate of 256 Hz by the downsampling factor of 5 (256 Hz/5 = 51.2 Hz).

Segmenting The start of a trial's data collection in the PacGame program was adjusted to begin before the target presentation instead of after, allowing a few seconds of data collection buffer to reduce the filter edge effects on the EEG corresponding to actual trial stimulus presentation. Using the bi-directional filter and using 12 points of zero padding on either end of the data the filter edge effects were further minimized. The computer sound and parallel port channels were recorded simultaneously with EEG data to mark each stimulus onset. These flags were used to segment epochs 100ms before and 800ms after each stimulus onset. Parallel port values indicated which stimulus was being presented and whether or not it was a target or non-target stimulus. Baseline correction averages data points collected 100ms before to the onset of sound stimulus [-100 0]ms and subtracts this value from all data points in the recorded epoch window.

2.1.6.3 RSLDA

In initial pilot runs suggested that target vs. non-target average ERP plots might differ significantly between each spoken word stimulus. In order to help minimize the variation of modeling a target sound over 8 different spoken word stimulus presentations, separate classification models were generated for the Direction and Non-Direction word sets. This was completed for sessions using both SWLDA and RLDA.

Regularized sub-class linear discriminant analysis (RSLDA) completes the RLDA method for each unique target stimulus (Hohne et al., 2014; Höhne et al., 2016). Splitting the binary target vs. non-target model into separate classification models based on each stimulus as a target, as opposed to each grouping of stimuli (Direction/Non-Direction), might be advantageous and segregate sources of variation in the training data. This technique was evaluated in offline analysis along with RLDA and SWLDA to compare BCI performance between classifier approaches (Section 4.2.3).

2.1.6.4 Dynamic Stopping

Dynamic stopping is a technique used in a BCI system to evaluate the confidence of a decision after each sequence of stimuli is presented. If the confidence of that decision meets a pre-determined selection criteria prior to the end of the stimulus sequences, the system will stop the stimulus presentation and output the decision to the BCI user. This allows a dynamic number of presentations of the stimuli set for each trial and optimizes the time to selection on a trial-by-trial basis.

In the current paradigm fifteen sequences of four words are presented during each trial. Optimizing the number of presentations needed to make a correct decision could speed up communication rates. By making a confident decision after, for example, six or seven presentations the BCI system can work faster and require less participant effort as the system and user improve their accuracy.

The potential benefit of implementing the dynamic stopping routine in the present system was investigated with offline analysis. BCI percent accuracy of decisions made after one through all fifteen sequences per trial can be found in Figure 4.8. The ITR performance is also reported for different number of sequences, which accounts for the shorter time to selection that results from requiring fewer sequences. ITR considers accuracy and time to selection. By reducing the number of required sequences to make an accurate selection, the ITR may increase.

2.1.7 Questionnaires

One major design consideration for the BCI was to ensure the user interface was highly motivating. Participants' self-reported level of motivation, and reports of workload and system usability were collected in past auditory BCI studies (Käthner et al., 2013; Simon et al., 2014). Factors of self-reported motivation, workload and system usability were tested in this study for their influence on performance. Analysis of motivation, workload and usability are invaluable for user-centered clinical BCI development and should be included in future studies.

2.1.8 aBCI analysis

The number of trials that were correctly classified as the intended choice of the participant divided by the total number of trials attempted gives the percent accuracy for each group of trials completed. BCI system performance was estimated using the training data collected and estimating accuracy using 10-fold cross validation on each individual sub-trial as well as 2-fold cross validation on full-trial aggregates. Subject, session and word set condition performance were compared to uncover the influence of these factors.

Investigation into the morphology of ERPs utilized by the aBCI classifiers was investigated through participant specific grand-average plots of target and non-target stimulus presentations. Topographical plots over the course of stimulus presentation indicated specific electrodes of interest that present with maximal voltage fluctuation over the course of stimulus presentation. Stimulus specific grand-average plots of target and non-target informed on the variance in morphology across spoken word stimuli.

ITR was calculated, per Wolpaw et al. (1998), for each sequence within a trial. ITR averaged across participant and for select participants is plotted for each sequence in Figure 4.9 on page 66. This metric along with accuracy reflects the capability of the tested aBCI system.

Self-reported workload or cognitive load was assessed using the NASA-TLX questionnaire giving a weighted score out of 100 with 100 being the highest possible workload for a task. Categories of workload are given individual scores and these results highlight the level and type of workload demand perceived by the participant while utilizing the aBCI system. Correlation between questionnaire results and BCI performance are computed and summarized in 4.4.1.

The impact of spatial cues on BCI performance was expected to be reflected in the difference between 'centralized' stimuli (those coming from the front and back speakers) and lateralized stimuli (sounds coming from the left and right speakers). During the BCI sessions of this study, experimenter observations as well as participant reports suggested no consistent difference in performance between centralized and lateralized stimuli existed. This was confirmed through statistical comparison included in 4.2.6.1.

2.2 Word Recognition Task

The Word Recognition (WR) task was developed and utilized here to further investigate the impact of spatial cues in the BCI task. It was confirmed that BCI percent accuracy results did not reflect an influence from stimulus spatial cues. A behavioral test was employed to precisely evaluate spatial cue influence on attending and recognizing target stimuli without the added complexity of collecting EEG. No influence of semantic relevance of stimuli on aBCI performance was uncovered, so both Direction and Non-Direction word sets were tested in this behavioral task as well.

Spatial cue information was modified for three different presentation conditions to evaluate spatial cue impact on behavioral performance. The task was designed to be similar to the BCI system in that four words were serially presented with one acting as the target. Participants were instructed to press a button when they heard a target word. The ability of the participants to recognize the target word presentation is expected to influence the morphology of the ERP measured in the BCI experiment (influencing BCI performance), as well as the reaction time in pushing a button (behavioral measure). Therefore, a correlation was computed between BCI accuracy in the aBCI task and behavioral measures (reaction time and accuracy) in the Word Recognition Task.

2.2.1 Presentation Condition

The first condition tested was one that most closely matched the BCI trial presentation. Four different spoken word stimuli were each played from a different speaker around the participant in several random sequences. Each stimulus was always played from the same speaker. This condition provides a strong sound source location cue for the target stimulus. This condition was termed the BCI condition.

In the second condition the spoken word stimuli all came from the same speaker, in front of the participant, forcing the participant to identify the target word through acoustic cues of the word itself and no longer rely on spatial cues. This condition was termed the NoCues condition. Since spatial cues were hypothesized to be the primary cue used in the BCI paradigm, it was expected that the NoCues condition would yield poorer button push performance than the BCI condition.

In the third condition, target and non-target words came from randomly selected speakers for each sequence. This condition applies spatial cues to each stimulus as in the BCI condition, but because the spatial location of each spoken word varies during a trial a participant cannot solely utilize stimulus location to identify the target. Participants will again need to rely more on the acoustic characteristics of the target word itself. This condition was termed the Dynamic condition.

Random speaker/location assignment is similar to the presentation scheme in a past aBCI study using spoken word stimuli (Ferracuti et al., 2013). For directional words, the meaning and sound location may be congruent or match up in some sequences, enhancing recognition of the target stimulus. In other sequences the stimulus meaning and location may be incongruent, disrupting the participants focus and decreasing performance. Reaction time measures between congruent and incongruent target stimuli presentations were also compared.

2.2.2 Attention and Fatigue

BCI classification typically benefits from additional stimulus presentations per trial; additional data creates a better, more generalizable classification model. It is possible that some or all participants become fatigued during trials of longer duration and may lose focus on the target. This could increase noise and reduce the signal of interest in the recorded EEG used to train the BCI classifier, as well as noise created during online use of the BCI.

The impact of the number of sequences of the four words presented in a given trial was also tested in the WR experiment. Fewer sequences within a trial may reduce fatigue-induced errors in the behavioral experiment and reflect improvements in focus that might enhance signals used by the BCI classifier. A reduction in SNR may also result from reduced stimuli presentations but overall advantages of reduced trial lengths can be investigated using the ITR metric.

In the aBCI paradigm an orange arrow serves as a visual reference for the target. The target stimulus was presented at the beginning of each trial in the WR experiment, as it was in the aBCI experiment, but no visual indicator of the target was ever present on the screen. The WR experiment therefore, requires a greater working memory requirement on participants, which may also increase fatigue and reduce performance.

2.2.3 Reaction Time and BCI performance

Recognition of a target word within a randomized sequences of auditory stimuli was the goal for the participants of both the PacGame and WR experiments. The experimenter is informed of this target stimulus recognition via EEG signals in the PacGame task and by button press in the WR task. A more precise behavioral measure is used in the WR task to further investigate the influence of stimulus presentation parameters on target stimulus recognition. Reaction time (RT) is a behavioral measure that summarizes the time of several sensory, cognitive, motor, and attention processes (Carlson et al., 1983).

Most participants of the aBCI study contributed to the Word Recognition study so correlation within participants could be tested. The results of this behavioral study were expected to correlate with BCI accuracy but strong correlations weren't expected, because many other factors influence RT and BCI performance.

This experiment will inform on a framework for testing stimuli presentation parameters of novel BCI systems with simpler behavioral measures independent of EEG signals. For P300 oddball driven BCI studies, significant correlation could provide support for collecting behavioral results to inform on optimal stimulus presentation parameters. Behavioral studies may yield more definitive results with less effort from both researchers and participants. Using behavioral studies would allow for more BCI system features to be tested independent of other BCI system influences like EEG acquisition and classifier approach.

Chapter 3

Study Protocols

The methods used in this study follow typical clinical practice, past literature on auditory Brain Computer Interfaces and lessons learned through iterative pilot studies. The following chapter outlines the specific details of executing the final experimental study design. Pilot study results are mentioned as they informed on final study design. All study procedures were approved by the Institutional Review Board of the University of Kansas and all participants provided their informed consent prior to engaging in study activities.

3.1 PacGame BCI System and Protocol

The finalized BCI system used for the study conducted here is composed of several hardware and software components. The following list highlights the major components.

1. EEG equipment
   (a) G-tec Gamma cap
   (b) 62 g.SCARABEO and or g.LADYbird active EEG electrodes (g.tec)
   (c) 2 g.GAMMAearclip Ag/AgCl electrodes (g.tec)
   (d) 1- g.SCARBEOgnd, 1- g.SCARABEO 'Z' electrode (g.tec)

(e) g.tec Hiamp with 64 active channels and 16 passive channel recording capability

(f) g.HEADbox Active

(g) g.TRIGbox

2. Audio Equipment

   (a) Mackie 1202-VLZPro Sound board

   (b) Motu Ultralite mk3 external sound card

   (c) Crown D-75A ampliphier

   (d) 2x Tannoy 6” Passive Nearfield Monitors (Left/Right)

   (e) Bose Video Room*mate* Monitors (Front/Back)

   (f) microphones

   (g) MAICO MA40 audiometer - hearing screening

   (h) Brüel & Kjær G-4 Type2250 sound level meter

3. Computers and connections

   (a) Intense PC - EEG data acquisition

   (b) Windows 7 Dell Desktop - PacGame UI

   (c) Custom serial port connection to g.TRIGbox

   (d) Custom Parallel port connection to g.TRIGbox

   (e) Custom audio line-in(s) to g.TRIGbox

   (f) ASUS VG248 video monitors

4. Environmental

   (a) A temperature controlled, electrically shielded anechoic chamber

3.1.1 Screening Participants

Participants completed a hearing screening to ensure similar dichotic hearing between both ears. No strict criteria for hearing thresholds were implemented so that hearing loss (if present) could be investigated as an influencing factor on BCI performance of lateralized stimuli.

Participants also completed a screening questionnaire to ensure they met critical inclusion criteria, including:

- No pacemaker
- No metal implants in the body
- Fluent American English speaker
- No history of severe mental disease or brain injury

Potential participants were then scheduled for their first and if possible second session. Participants completed screening procedures and provided informed consent to participate in the study. See Appendix A.3 for screening form details. After reviewing the screening sheet the experimenter conducted the hearing screening. If any criteria for inclusion were not met then the participant was dismissed.

3.1.1.1 Hearing Screening

Participant sat in a comfortable chair inside a sound proof room along with the experimenter and the audiometer. The participant faced away from the experimenter and the audiometer so no visual cue of the tones being played would influence the test results. Calibrated headphones were placed on the participant's ears and the test began. The participant was instructed to raise their left or right hand whenever they heard a two-tone train in the corresponding ear.

A MAICO MA40 audiometer was used to test the lowest intensity tone that could be identified by the participant in each ear. A set of tones at the frequencies: 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz were tested. Audiograms were recorded for each participant and are included in the tabulated data in the Appendix A.5. For each tone a volume of 0dbSPL was initially presented in a two-pulse train in both of the participants ears. The intensity was raised 5dB and the two-pulse train repeated until the participant was able to identify the tone and raised their hand. Once the tone was identified the intensity was reduced 5dB once again to ensure that intensity was the actual minimum value that could be identified. The intensity was then recorded on paper corresponding to the frequency and ear tested. This procedure was repeated for both ears and all previously listed tone frequencies.

Criteria

The primary reason for conducting the hearing screening was to ensure that participants did not have significant hearing loss or have significant loss in one ear as opposed to the other over the frequencies important for speech. The decision to reject participants based on hearing loss was revoked once the study began in anticipation that notable hearing thresholds may influence BCI results in a measurable way and could be characterized. For this reason, no participants were rejected due to hearing loss.

3.1.2 EEG setup

After the hearing screening the participants remained seated in the sound proof room where a 64 channel EEG cap was placed on the head of the participant. Alignment of the cap to anatomical features was accomplished. A chin strap held the cap in place. A clip from the EEG lead sleeve to the back of the cap helped reduce imbalanced tension on any single or group of EEG electrode leads.

MATLAB (The MathWorks, Natick, MA) Simulink models were used for data acquisition from the g.HIamp. These models included impedance measurement modules that were used to investigate the relative impedance of all electrodes with respect to the ground and a Z-electrode use for this active electrode impedance measurement. Electrolytic gel was injected beneath each electrode and the tip of the syringe was used to move hair out of the way so that good contact between the scalp and electrode surface was achieved with the gel. Color coded indicators for each electrode in the Simulink impedance module signaled to the researchers that <30KOhm relative impedance had been reached.

Once all 62 EEG and the two ear clip reference electrodes exhibited this minimum impedance the impedance module was closed. The g.Hiamp module was then opened and acquisition settings were loaded for the training data collection. The amplifier settings included recording of trigger channel data, 256 Hz sampling rate, 8-point buffer and Butterworth notch filters applied at 58-62 Hz to remove power line noise.

3.1.3 aBCI User Interface Software

The PacGame user interface begins with a prompt to enter the participants ID as well as session ID. When running the 'online' blocks of trials the participant ID is used to identify the classifier file used for decoding and for naming log files written out during online trials. This participant ID also contained a designation for condition and session as different decoder model weights were used for each condition and session within each participant. The main menu appeared after hitting enter on the session ID prompt input screen.

See Figure 3.1 for a view of the PacGame Menu screen. On this screen icons can be clicked with a mouse cursor to enter a routine that presents a single block of either training or online trials of conditions 1, 2 or 3. Condition three was tested in pilot runs and utilizes pure tone stimuli. It was found that participants were not able to easily associate the tones with direction or location and would not serve as good controls to past studies that utilized such stimuli. A help button was available to be clicked to view the instructions to the participant and information about each BCI condition (1,2,3). A button labeled 'Dyn' turned off and on the dynamic stopping feature for online sessions, but this feature was never fully implemented into the software. Clicking the 'Quit' button closed the program.

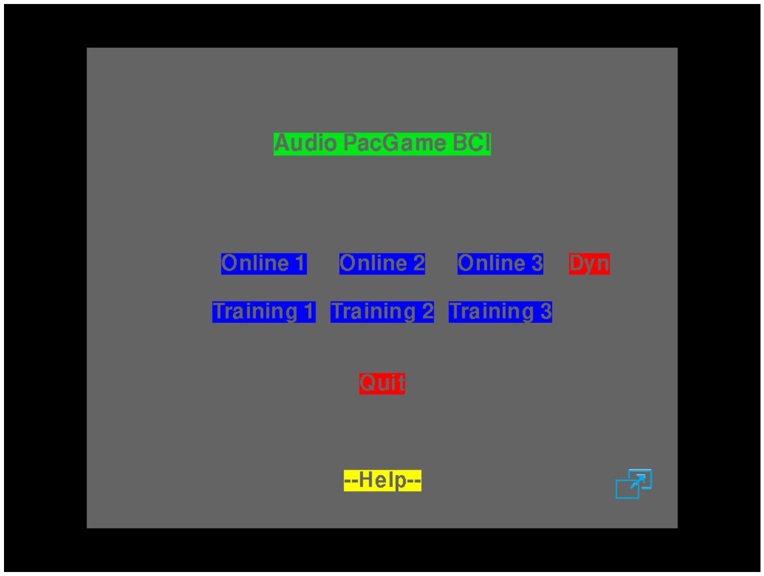


Figure 3.1: PacGame Main Menu

3.1.3.1 PacGame BCI Task

The PacGame BCI paradigm is imbedded in a game like interface where a little yellow face icon, called Pac, is positioned in the middle of a 5x5 grid of spaces. See Figure 3.2 for visual reference. A cherry icon was placed at one of the four corners of the grid. The objective was to move Pac to the cherry location. For each trial the aBCI classifier identifies one stimulus/direction as the mostly likely target. The classifier decision will move Pac one grid space in the associated direction. Once Pac reaches the cherry, the icon will move back to the center and the cherry icon will move to another corner of the grid. The four trials required for the Pac icon to reach the cherry icon will be referred to as a run. Four runs are accomplished with the cherry icon in each corner to complete one block of trials. With this configuration, each block included 4 trials with each direction/stimulus acting as the target, resulting in 16 trials total.

The instructions to the BCI user are typical of an auditory oddball task. Each trial consisted of the presentation of the target stimulus, presented twice, with a 1 second pause in between, before the trial began. Next, a rapid serial presentation of all four auditory stimuli, each played from a specific speaker placed around the seated participant, began. Each of the four stimuli were played in a random order before they were repeated again in another random sequence. Each trial included

15 randomized sequences of the 4 stimuli. There was no pause or interruption between sequences. The participant was instructed to count the number of target presentations that occurred throughout the trial. The strategy suggested was to ignore all other non-target stimuli for that trial and focus on the target word.

The target presentation also includes the visual presentation of an orange arrow indicating the relative source location of the target sound stimulus. The arrow above the grid pointing up indicates the target stimulus will be played from a speaker positioned directly in front of the participant at 0º azimuth. Arrows on the left or right side of the grid pointing in the left or right direction will play from speakers positioned at ±90º azimuth respectively. An arrow displayed below the grid pointing down indicated the target sound is played from a speaker positioned behind the participant. Two seconds after the second target stimulus presentation the series of auditory stimuli is initiated.

Figure 3.2: Example of PacGame interface

3.1.3.2 Training Trials

With the EEG system ready for acquisition, the experimenter changed the monitor in the sound proof booth to receive input from the Dell Desktop where the experimental presentation software was running. Microphones in the experiment control area were wired through the sound board and the experimenters voice was heard in the left and right speakers in the sound booth. A microphone placed behind the participants monitor allowed the participant to be heard by the experimenter, listening through headphones outside the booth.

After giving the participant a single trial demonstration of the PacGame task the block was restarted and EEG acquisition was initiated. After any questions were answered for the participant the training trials began. Participants completed two blocks of the BCI PacGame task for each condition to account for training data.

During training sessions Pac will move in the direction of the arrow after each trial. The target direction will be one that will bring Pac closer to the target, cherry location. For example, if the cherry was placed in the bottom left corner as is seen in the first image of Figure 3.2, the four trials will include two trials of left speaker targets ('care' or 'left') and two trials of rear speaker targets ('while' or 'back'). Which of the two possible directions is selected as the target for each trial is randomized by the program.

Once Pac reaches the cherry, the Pac icon is returned to the center grid square and the cherry is relocated to another corner of the grid. With completion of subsequent runs, the cherry will move to each of the four corners in a randomly generated order. Once Pac has reached each corner one block of trials is complete. This results in completion of four trials for each stimulus acting as target for a total of 16 trials. Two blocks of training trials were completed for each word set condition.

During stimulus presentation of the training blocks the PacGame software set parallel port and serial port signals to indicate which stimulus was being presented. These signals along with audio channel signals, indicating when an auditory stimulus was presented to the participant, were recorded along with the EEG data by the Hiamp via the g.StimBox.

Thresholds for these 'trigger' channels were manually set by the experimenter to allow for audio triggers to appear with each stimulus presentation at minimum volume. This aided the timing accuracy of stimulus onset and to minimize silence periods in the audio trigger channels during word production. These trigger channels were used by the model generation script to segment the EEG data into appropriate target and non-target stimulus presentation segments. See Section 2.1.6.2 for additional details of EEG preprocessing and decoding algorithm approach.

3.1.4 Stimuli

The stimuli were spoken word stimuli presented in a 400 ms long sound file recorded and post recording processed in Praat software. Stimuli were recorded with an AKG head mounted microphone and Motu Ultralite_mk3 external sound card connected to a Dell desktop computer running Window 7. These stimuli were adjusted to have the time of voicing begin as early as possible in the 400 ms long sound file. No silence or breaks are intended between stimulus presentation during the trial so each word was spoken at a rate to reach as close to 400 ms duration as possible.

Each stimulus intensity was scaled to match the recorded envelope intensity at the location of the participant's seated position. With stimuli played from their respective speakers, intensity was recorded with a Brüel & Kjær sound level meter positioned at expected head height and location of study participants, equidistant from all four speakers. The average intensity over the 400 ms duration of each stimulus was adjusted to meet approximately 65db SPL. Pilot participants were queried on any differences in stimulus intensity. No perceptible intensity imbalanced was reported.

Table 3.1 reports the duration of voicing of each auditory stimulus used in the BCI paradigm in seconds. See Appendix A.1 for additional details of each acoustic stimuli including waveform, spectrogram and pitch and intensity contours.

Table 3.1: Stimuli Durations (seconds)

Front	Back	Left	Right	Joy	Care	While	Doubt
0.3929	0.4375	0.4277	0.3899	0.3903	0.3961	0.4037	0.3073

Stimulus files were played through four output channels of the Motu sound card, each channel fed to one of the four speakers placed around the participant in the sound-treated chamber. Each sound file is a four-channel file with silence on all channels but the one corresponding to the appropriate speaker. Figure 3.3 shows the spoken word stimuli and the location around the participant for each condition.

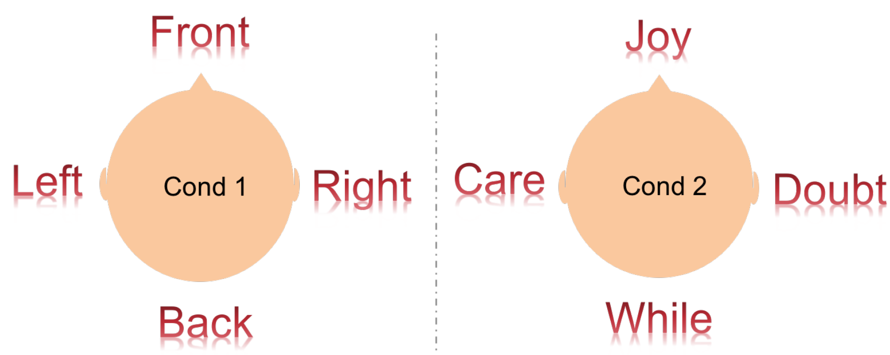


Figure 3.3: Stimuli Locations

3.1.5 Model Generation

After completion of training trails for both Direction and Non-Direction stimuli groups, EEG recordings were stored and processed by a MATLAB script to pre-process and compute the linear discriminate analysis (LDA) weights for the online decoder. See Appendix A.7 for the script code.

3.1.5.1 Preprocessing

The raw EEG data was initially zero-phase high-pass filtered above 1.0 Hz. This data was run through a blind source separation algorithm termed to compute an independent components analysis (ICA) with FastICA with parameters to segregate independent components that have high correlation to the EEG electrodes closest to the participant's eyes. This algorithm presented several candidate independent components potentially representing eye movement. Scalp topographical plots and full session recorded data traces were inspected and independent components were manually selected for removal.

After ICA rejection, the reconstructed time series data was low pass filtered at 51.2 Hz again using a zero-phase shift technique. The EEG data was then down-sampled by a factor of 5, and all trigger channel data and time stamp vectors were also aligned to this processed dataset.

EEG time segments or epochs were segregated by aligning time zero to the onset of the auditory trigger channel during aBCI trials. The parallel port value at the time of auditory trigger onset provided a label for each sound presentation with which target and non-target epochs were identified.

Epochs that contained a value of greater than ±150 microvolts were rejected from the dataset in expectation that a large muscle or motion artifact has influenced this piece of EEG signal and the values are not reflective of neural activity due to the BCI task.

The RLDA model was trained using algorithms developed by the Berlin BCI group and described previously (Hohne et al., 2014).

A single block includes four trials for each stimulus to serve as the target sound. Completing 2 blocks per condition gives 8 trials of EEG data, or 120 target presentations for each stimulus. A total of 480 target presentations and 1440 non-target presentations utilized by the decoder to fit the linear discriminate model for a given condition/session. A classifier file was written by the MATLAB script to the experimental computer. This file was then loaded by the PacGame software to set parameters for the online decoder.

While the classifier was being generated, the participant filled in the form for the payment system during the first session and read details on the NASA-TLX workload survey in the second session.

3.1.6 Online Trials

The participants then completed two rounds of online trials. The original intention was to complete an entire block of trials, but because of very poor online performance and limited time only two rounds were complete for both conditions. This still typically resulted in sixteen trials total as each online round is limited to eight trials. The minimum number of correctly classified trials to reach the cherry and complete the task was four, however very few participants reached the cherry in eight trials so the round ended after eight trials. This ended the EEG portion of the session.

3.1.7 Session Balancing

The EEG-BCI protocol for both session 1 and session 2 are identical. The order of completing Non-Direction words ['care', 'joy', 'doubt', 'while'] or Directional words ['right', 'left', 'front'. and

’back’] was balanced across the sessions and also across participants. Even numbered participants completed thirty-two trials of Direction words first and odd completed Non-Direction words first in both training and online paradigms. This order was switched for all participants in Session 2.

3.1.8 Interviewing Methods

At the start of each session the participant was asked to rate their current level of mood or motivation on a continuous scale from 1 to 10. The number 1 being a “bad mood” or “Extremely unmotivated” and 10 being a “good mood” or “Extremely motivated”. This visual analog scale (VAS) rating will be referred to as the participants self-reported level of motivation.

Next, the hearing screening was conducted in the anechoic chamber. The anechoic chamber was also where the EEG recording took place. During Session 2 the participant was again asked to complete the VAS for motivation after signing the consent form. The EEG protocol was again conducted and afterwards two more questionnaires were completed.

The first questionnaire was the NASA-TLX. This questionnaire is meant to identify the type and level of workload on the participant while they are attempting to use the BCI.

The survey is composed of two stages of reporting workload. First the participant was asked to select between two of six different workload categories that were most important to the participant’s experience of workload. All combinations of binary comparison of the 6 categories were presented and the participant’s selections were tallied. Weightings based on the tallies for each of the 6 categories were made. The participant then rated the six categories on a 20 point, unnumbered scale from Very Low to Very High.

A descriptive prompt is provided for each category. For example, rating of Temporal Demand is elicited by the prompt, “How hurried or rushed was the pace of the task?” The 20-point scale rating was converted to a 0-100 scale by 5 point increments and was multiplied by the weightings to give a score for each workload category. See Appendix A.4 for worksheets used for this survey.

The participants then completed the System Usability Scale (SUS) survey which asks ten questions, rated on a 1 to 5 scale from Strongly Disagree to Strongly Agree respectively. The overall score for usability was computed with a formula provided with the SUS survey. See Appendix A.5 for SUS survey materials.

The NASA-TLX and SUS ratings are completed independently for Directional and Non-Directional word sets. Only one set of weightings for the NASA-TLX workload categories are used for both Condition 1 and 2 ratings.

3.2 The Word Recognition Task

The word recognition task tested three different stimulus presentation paradigms, similar to those presented during the BCI task. In this experiment, the participant is again seated in the sound proof booth, approximately 1 meter from a computer monitor and positioned equidistant from four speakers surrounding them orthogonally. Direction and Non-Direction words are again tested and balanced across participants in order of completion.

There are several differences in the Word Recognition (WR) task and presentation from the aBCI paradigm, however. In the WR task there is no PacGame visual presentation, instead a fixation cross is presented on the screen. No EEG cap is placed on the participant, instead the participant's attention to the target sound is evaluated by pressing a button each time they hear that trial's target sound. Each trial's target word is randomly assigned beforehand, instead of relating to the direction the Pac icon would go to reach the cherry. Each of the four words are used as the target once before being used as target again. In addition, each trial is randomly assigned the number of sequences it will utilize.

3.2.1 Audio Presentation Conditions

The 'BCI' presentation paradigm or condition is one that most closely matches the word presentation scheme of the aBCI task. A target presentation plays the target word from its respective speaker twice and the random sequences begin. In this paradigm, a set of four words is played from one of a set of four speakers at random, just as was done in the BCI task. During the trial presentation, each of the four words are randomly presented before repeating, just as in the BCI paradigm. Sixteen trials were completed with Direction words or the Non–Direction words before switching this condition and repeating the trials with the other set of words. There was no orange arrow or other visual reminder of which stimulus was the target, so minor working memory influences are potentially more present in this task.

In the second word recognition task presentation condition the spatial cues are removed from the auditory stimuli. Here all the parameters of presentation and instructions of the task are equivalent except that the presentation of the targets and all stimuli are presented from the speaker placed directly in front of the participant. This condition will be referred to as the NoCues condition.

The third word recognition task again alters the location of sound presentation and will be referred to as the Dynamic condition. Here all four speakers are again used, however, a given word stimulus was not always presented from the same speaker.

The target stimulus in the Dynamic condition was first played twice from the front speaker and then random sequences of the four words are again played with the variable number of sequences as described before. The target word and the three other words of the current condition will play from a randomly selected speaker during each sequence. In each sequence, each of the four words will be played from a different speaker, so that each speaker plays exactly one word per sequence. See Figure 3.4 for an example of the first two sequences of a trial in which the word 'Right' is the target and the sequences include playing the target word from the back and left speakers. During a trial, the target sound will play an equal number of times from each speaker. This again requires an interval of four trials per condition to balance these stimulus specific conditions.

3.2.2 Number of Sequences per Trial

The number of sequences or presentations per trial of each of the four words is not set at 15 as it was in the BCI paradigm. The influence of the number of sequences in a trial was tested by running trials with 4, 8, 12 or 16 sequences of stimuli presentation.

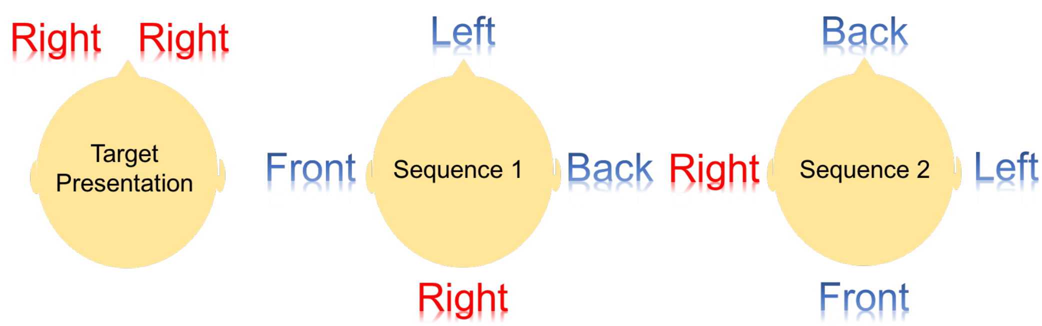


Figure 3.4: Dynamic Condition Example

A trial with each of these number of sequences was run before repeating a trial with a given count. Each of the target words were used in a trial with each of the four different counts. In the Dynamic condition, each of the four words is played from each of the four speakers an equal number of times. Balancing the location presentation of each stimulus in the Dynamic condition requires a multiple of four sequences per trial, so that is why these numbers of sequences were tested. Using four different target words with four different counts gives us 16 trials per presentation condition and per word set condition (Directional or Non-Directional words).

3.2.3 Protocol

The order of presentation conditions completed for the word recognition task was randomly assigned to each participant. BCI, NoCues, and the Dynamic condition were completed with 16 trials per word set condition for all participants. Completion of all possible condition orders requires 12 different experiment condition sets. Three different presentation conditions yield six presentation orders. Multiply this by the two different orders of completing the word set condition. What improvements in task performance result from practice or what influence fatigue has on the participant.

As was done with the aBCI tasks, the EEG amplifier and g.TRIGbox recorded the onset of each auditory stimulus, a signal indicating which stimulus was presented and if the stimulus was a target or not. The g.TRIGbox also recorded the timing of the participants button press. The data collected through these trigger channels was then analyzed to determine what the reaction time for each stimulus presentation was. A percent accuracy of button press was determined by detecting how many target stimuli did not get a button press.

Participants were instructed to keep their eyes open and to maintain fixation on a fixation cross to mimic the BCI instructions. It was also stressed to the participant to press the button immediately after the target stimulus was presented. Participants were allowed to press the button in whatever way they felt most comfortable.

A brief survey was completed by each participant after the experiment. This survey queried the participants for information about the easiest and most difficult conditions regarding presentation order, stimulus sequence, word set condition or any aspect of the task they found difficult. Participants were also asked to estimate the number of target stimuli they missed pressing the button for each presentation condition.

Chapter 4

aBCI Study Results

The results of testing the aBCI on several healthy participants is detailed here with reference to additional results in the Appendix. The initial focus is on the general performance of the system and the effect of word condition on performance. A number of analyses are completed to describe or discount influence of many other variables such as BCI session, the sequence of auditory presentation within a trail or the order of Direction vs. Non-Directional completion.

4.1 Participants

Participants in the current study were primarily college aged young adults with self-reported normal or corrected to normal vision, and no history of neurological disorder, disease or injury. In total 22 participants signed consent and participated in at least some portion of the study. Recruitment was primarily accomplished through word of mouth, flyers and referrals from participants in other studies. The hearing screening, thorough checks on equipment settings, and immediate scheduling of both sessions for eligible participants helped mitigate unnecessary loss of participant data.

Four participants were included in the pilot study, but the pilot paradigm was changed to remove SWLDA classification approach and speed up presentation from 750ms inter-stimulus interval (ISI) to 400ms ISI. Pilot results are reported as it is still informative in reflecting the paradigm used. The majority of results are reported with a focus on the data collected from 16 subsequent participants that completed 2 sessions of the final aBCI study design.

For the sixteen participants included in the study the age, handedness and gender of each is reported in Table 4.1. All participants included in the study reported their race and ethnicity as White, not Hispanic. Additional health history information regarding compatibility with EEG data collection and neurophysiological condition was also collected. No participants were rejected based on screening information.

Table 4.1: Participant Information

Participant	age	handedness	gender
05	22	Left	female
06	26	Right	female
07	29	Right	female
09	26	Right	female
10	29	Right	female
12	21	Left	female
13	25	Right	female
14	21	Left	female
15	20	Right	female
16	27	Right	male
17	25	Right	male
18	18	Right	female
19	22	Right	male
20	59	Right	female
21	26	Right	female
22	19	Right	female

The sample of participants was heavily biased by female gender. Females have exhibited larger P300 amplitudes in Käthner et al. (2013) and other studies cited by this author. However, Oliver-Rodríguez et al. (1999) found male P300 amplitudes were larger in a study to compare gender differences from affective stimuli. A study comparing ERP BCI performance has seen no influence from gender in similar studies Schreuder et al. (2011). A Mann-Whitney test on the difference between male and female BCI performance as described in 4.2.2, suggests there was no difference between gender ($W = 247$, p-value = 0.4531).

A list of languages that participants were familiar with is included in Table 4.2 along with the number of participants that reported having some familiarity with the language.

Table 4.2: Participant Language Familiarity

Languages	Spanish	ASL*	Hebrew	Mandarin	Japanese	German	Italian
# of Participants	8	3	2	2	1	1	1

*ASL - American Sign Language

4.2 BCI % accuracy

The number of trials that were correctly decoded divided by the total number of trials attempted gives the percent accuracy for any group of trials completed. Percent accuracy is the primary metric of performance for the BCI system. The feasibility of the system to robustly and correctly select the items the user is intending to is evaluated with the accuracy metric. It is important to understand the aspects of the system that yield the highest accuracy and which features could benefit from modification.

Understanding the influence of each system feature requires computing accuracy from several different approaches. Investigation into the accuracy of the BCI system will include estimating percent accuracy using 10-fold cross validation on each individual sub-trial as well as full-trial application where 2-fold cross validation will be used. Comparing single-trials to full-trial accuracy will highlight the benefit of averaging multiple single-trials together. Across Session accuracies tell us how well the models generated will generalize over multiple EEG sessions. All percent accuracies for offline data analysis were computed using RSLDA classifiers.

Comparing accuracies across word sets, individual target stimuli, and participants illustrates variability resulting from use of spoken word stimuli. The importance of spatial cues was investigated by identifying any imbalance in performance between stimulus location. Specifically, the difference in performance of centralized ('front' and 'back') speakers and ('left' and 'right') speakers was used to evaluate the performance benefits of spatial cues.

4.2.1 Sub-Trial Percent Accuracy (offline)

The initial approach taken to test the predictive power of the model generated from the aBCI training data was to run cross-validation on the resulting model. Cross-validation splits the training data into parts, using one subset (A) to fit the linear model. Another subset (B) will act as input to the model to generate predicted outcomes. The process was repeated by varying what data comprises subset A and B and an average of predicted outcomes was used to estimate offline percent accuracy. Accuracy estimates can be computed with many variations in the data subsets on which the model is trained or tested.

A 10-fold cross-validation technique was used to predict the percent accuracy for every stimulus presentation to be correctly categorized as a target or non-target. 10-fold cross-validation splits training data into 10 random noncontiguous, but equally sized sets. A model was generated from one of the 10 sets of data and that model was tested on the other 9 subsets. Accuracy was estimated for each of the 10 subsets and the average accuracy across all these iterations is reported.

Results for each participant can be found in Table 4.3a. In Table 4.3b. the accuracy score was aggregated over each Session and Condition.

It may be noted that the majority of participants, sessions and conditions yield similar accuracies in the mid 60s. This result comes from the classification of every single utterance as target or non-target reflecting Type I and Type II errors. This result could also be termed single-trial accuracy and are often reported as such (Blankertz et al., 2011; Hill et al., 2014; Hohne et al., 2014). A very consistent mid 60's percent was somewhat encouraging as sub-trial accuracy reflects a difficult decoding problem for BCI. The BCI system proposed here intends to aggregate 15 trials and their classification scores together in order to improve accuracy performance beyond what was reported here. The training data used in this calculation was also 1/10th of that expected to be used in the real-time BCI system.

The uniqueness of each participant's accuracy can be seen in Figure 4.1 as each box-plot considers the 10-fold cross validated percent accuracy for each session and condition completed by that participant. These boxplots show the results for the first four pilot participants as well as the final study design results. A horizontal line indicates the BCI research fields standard for minimum performance with a BCI. Four participants (aB07, aB21, aB9, and aB10) were able to achieve >70% accuracy in at least one session/condition with this measure of offline accuracy.

Table 4.3: Offline sub-trial Accuracy (10-fold cross validation)

a.

Participants	Mean %Accuracy
05	63.346
06	65.401
07	67.864
09	66.302
10	64.336
12	62.318
13	66.562
14	65.313
15	63.438
16	62.174
17	66.380
18	65.026
19	68.438
20	67.6330
21	74.010
22	67.930

b.

Session	Condition	% Accuracy
1	Direction	65.452
2	Direction	66.357
1	NonDirection	66.535
2	NonDirection	65.755

4.2.2 Full-Trial Percent Accuracy (offline)

The BCI system was designed to make predictions of target vs. non-target categorization for a given stimulus class based on the median score of 15 sequences of each stimulus. In order to compute the expected accuracy of the system offline, an approach to use cross-validation but compute decisions based on all 15 sequences was used.

When splitting data for cross-validation but evaluating full trials, it is most valid to retain all stimulus presentations from a full trial in the tested data subset. It is also important to retain balanced number of trials where each of the stimuli serve as target, in order to capture all EEG variability due to stimulus differences. Instead of randomizing the sequences that go into model training and tested data subsets, here a 2-fold cross-validation design was accomplished by splitting the test and training sets into the two different blocks completed during the training session. Each block was composed of sixteen full trials where each of the stimuli act as the target in four trials. For each percent accuracy reported the percent correct reflects an aggregate of trials trained on Block 1 and tested on Block 2 as well as trials trained on Block 2 and tested on Block 1, resulting in thirty-two cross-validated trials.

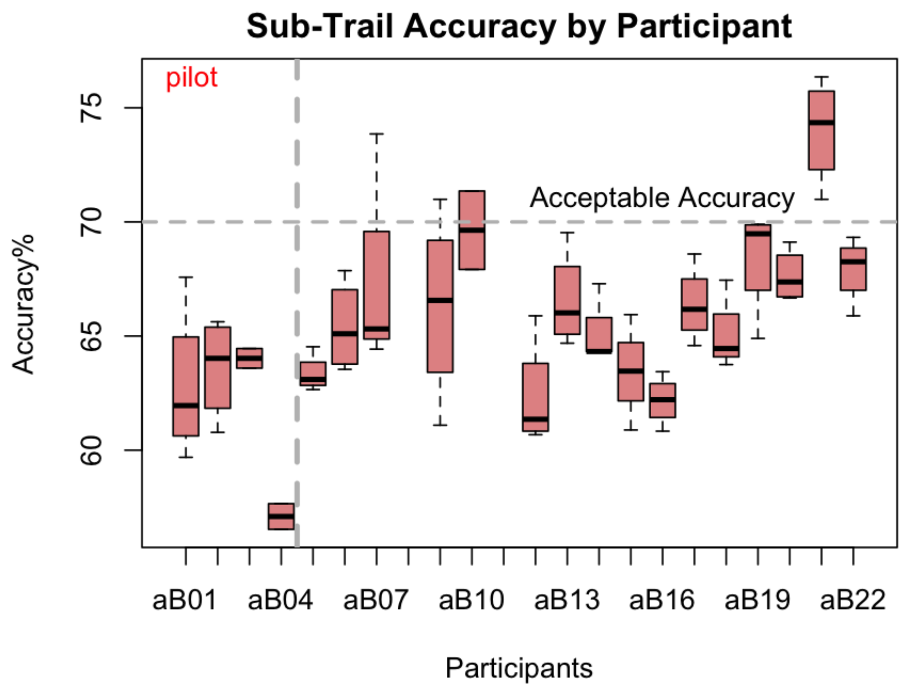


Figure 4.1: Offline Sub-trial Accuracy

Accuracies aggregated for each participant are included in Figure 4.2. Pilot data and horizontal line at 70% accuracy is also included as in Figure 4.1. It should be noted the general increase in accuracy over sub-trial calculations was likely due to the fact that accuracies are based on 15 sequences instead of just one and that more data was used to generate the classifier models. Twelve of the sixteen final study design participants that exhibit at least one session/condition that resulted in >70% accuracy using RSLDA decoder as opposed to the four participants that met this criterion using the sub-trial 10-fold cross validation result. The sizable improvement in full-trial accuracy over single trial supports the use of full-trial decoding in such a system.

Table 4.4 highlights some of the best performances in regard to offline percent accuracy. While these best-performer results are often highlighted in the BCI literature it should also be noted that a very high number of participants achieved minimum offline performance in this study with no training or accurate feedback.

The accuracy between sessions was not found to differ dramatically. Figure 4.3 on page 59 shows the similarity of between Session 1 and Session 2, although Session 2 exhibited less variance

Participant	Accuracy	Session	Condition
aB22	84.38%	1	Non-Direction
aB19	87.5%	1	Direction
aB20	87.5%	1	Non-Direction
aB22	87.5%	1	Direction
aB19	90.62%	2	Direction
aB21	90.62%	1	Non-Direction
aB21	90.62%	2	Direction
aB21	96.88%	1	Direction

Table 4.4: Top Offline Accuracies

(session 1 - 353.51, session 2 - 272.11). A paired t-test shows these two distributions are not different from one another (t = -0.2026, df = 59.001, p-value = 0.8401). Difference between session would highlight the effect of experience or training with the BCI system.

A paired t-test of accuracy between conditions was not significantly different (t = 1.368, df = 28, p-value = 0.1822). 4.4 shows the similarity of performance between the Direction and Non-Direction word trails. This result suggests that semantic relevance was not an influencing factor on performance.

The order in which each condition was completed was balanced across participant and within participants to balance any effects of training or experience with the BCI. A paired t-test between the offline RSLDA accuracies of the first condition in a session and the second condition within a session, regardless of condition, was not significant (t = 0.023078, df = 59.849, p-value = 0.9817), proving the order of completion did not have an appreciable effect on BCI performance.

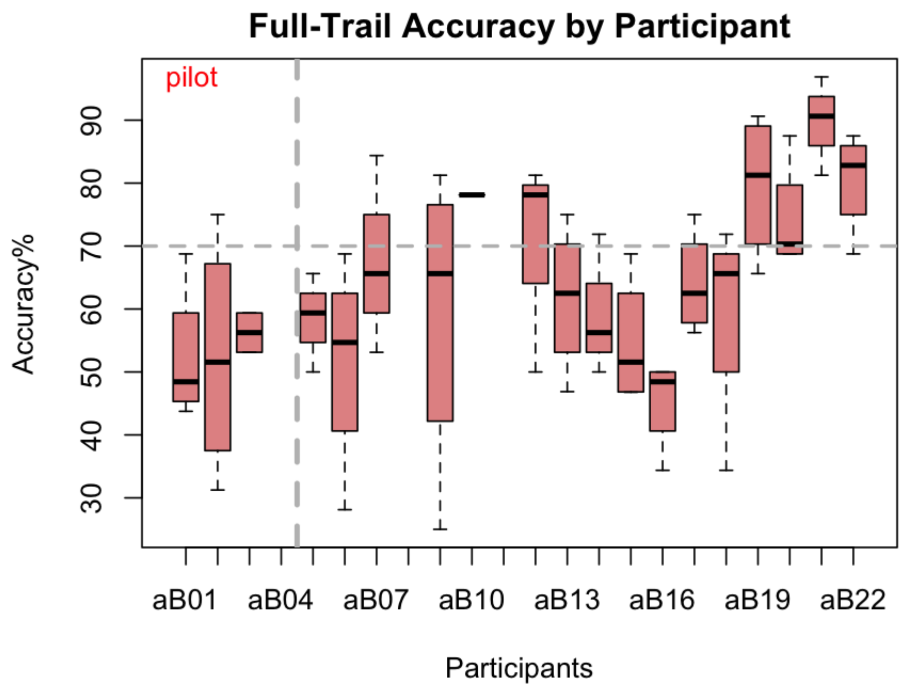


Figure 4.2: Offline % Full-TrialAccuracy: Participant

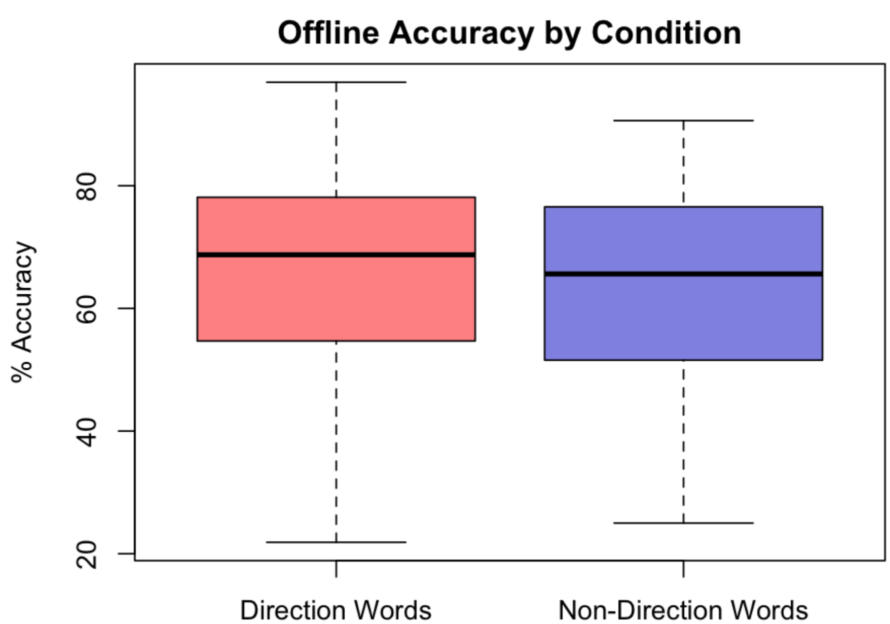


Figure 4.4: Offline % Accuracy: Condition

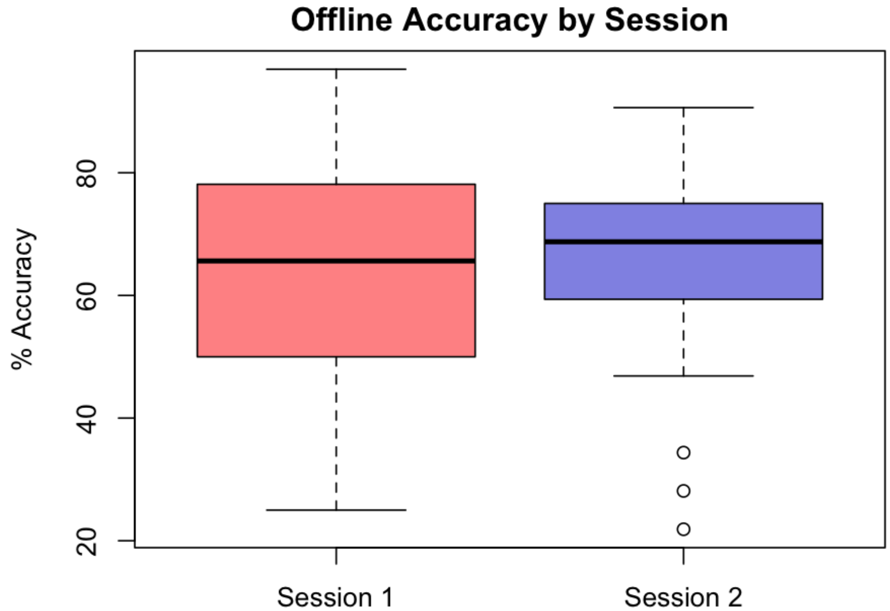


Figure 4.3: Offline % Accuracy: Session

4.2.3 Classifier Comparison

When comparing the full-trial classification results using RLDA and the SWLDA it was discovered that RLDA yielded higher accuracy. RSLDA was also tested and yielded results very similar and often better than RLDA. See offline percent accuracy estimates for these three classifier methods aggregated across participants, sessions and conditions in Figure 4.5. A plot is provided for Pilot subjects data as well as the subsequent 16 participants in the final study design. For all offline analysis reported the RSLDA approach was utilized. An omnibus ANOVA test of effect of accuracy found a significant main effect of classifier (F=20.544 p-value=9.48e-09). The results in accuracy using RSLDA over SWLDA were significant (t = 5.7667, df = 114.82, p-value = 6.943e-08).

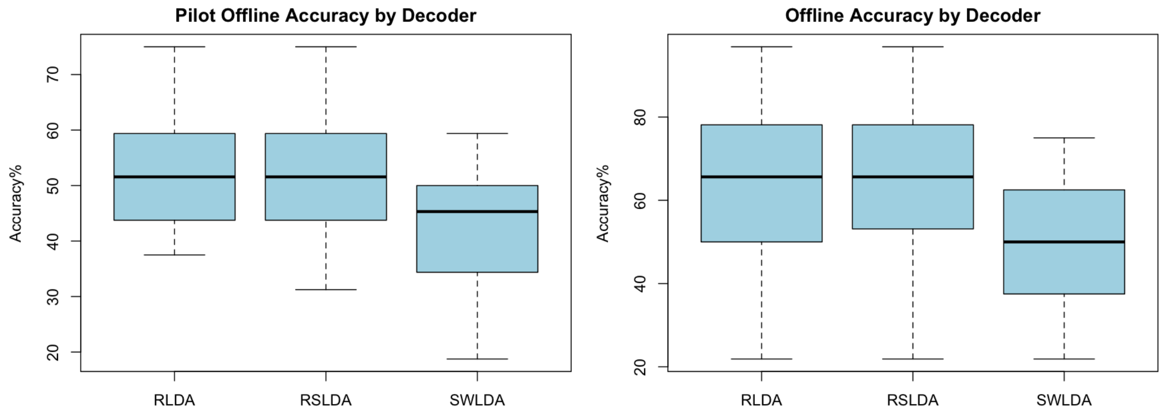


Figure 4.5: Offline Classification by Decoder

4.2.4 Accuracy Across Sessions

In BCI systems, the training session data is used to generate a model used in online trials to classify EEG features and provide output to the BCI interface. Differences in EEG cap placement and cognitive states of alertness produce variance in EEG signals from the BCI user across sessions, requiring new training data to be collected and a new classifier model to be generated each session. A good deal of data is necessary to generate a reliable and flexible model so training a classifier each time an EEG cap is worn is time consuming and bothersome in research or clinical application. Often researchers test the usefulness of past session training data in hopes that the signals produced by the user and extracted by the system are generalizable enough to work well in a variety of conditions including different days and different EEG cap applications.

Across-session accuracy was calculated by averaging of the percent accuracy from training a model on session 1 data and testing on session 2 data and vice versa within a participant. The accuracies by participant include calculations across both sessions in both word set conditions, since separate models are made for each word set.

Figure 4.6 shows that across-session accuracies are similar to those found by cross-validation within a session and are not statistically different between word set conditions ($t = 0.96945$, $df = 53.004$, p-value = 0.3367) or by session used for training ($t = 0.053397$, $df = 53.971$, p-value =

0.9576).

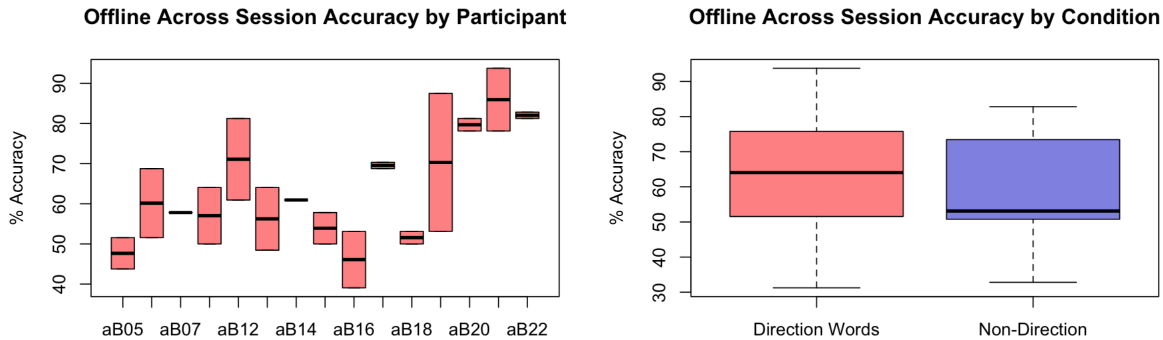


Figure 4.6: Offline % Accuracy: Across Sessions

The across-session accuracy calculation was most similar to the 2-fold full-trial accuracy calculation. Accuracies collected across session and within session were found to be very similar, indicating a substantial generalization of the models created in the aBCI system. Although the SWLDA classifier was found to be less accurate than the RSLDA classifier in offline within-session analysis, it was found to be statistically equivalent in generalizing across sessions. Figure 4.7 compares the mean full-trial, across session, accuracy of each participant utilizing both of SWLDA and RSLDA classification methods. A t-test of these distributions yields no significant difference in means (t = 0.67738, df = 27.997, p-value = 0.5037).

Unlike within session accuracy, the classifier models utilized here use the full training dataset of a given session, doubling the training data used to generate the model. It may be that additional data fed into the model may benefit the SWLDA classifier. Because the RSLDA generates a unique model for each target stimulus, it utilizes one quarter of the data used by SWLDA. In terms of required data, RSLDA may be at a disadvantage.

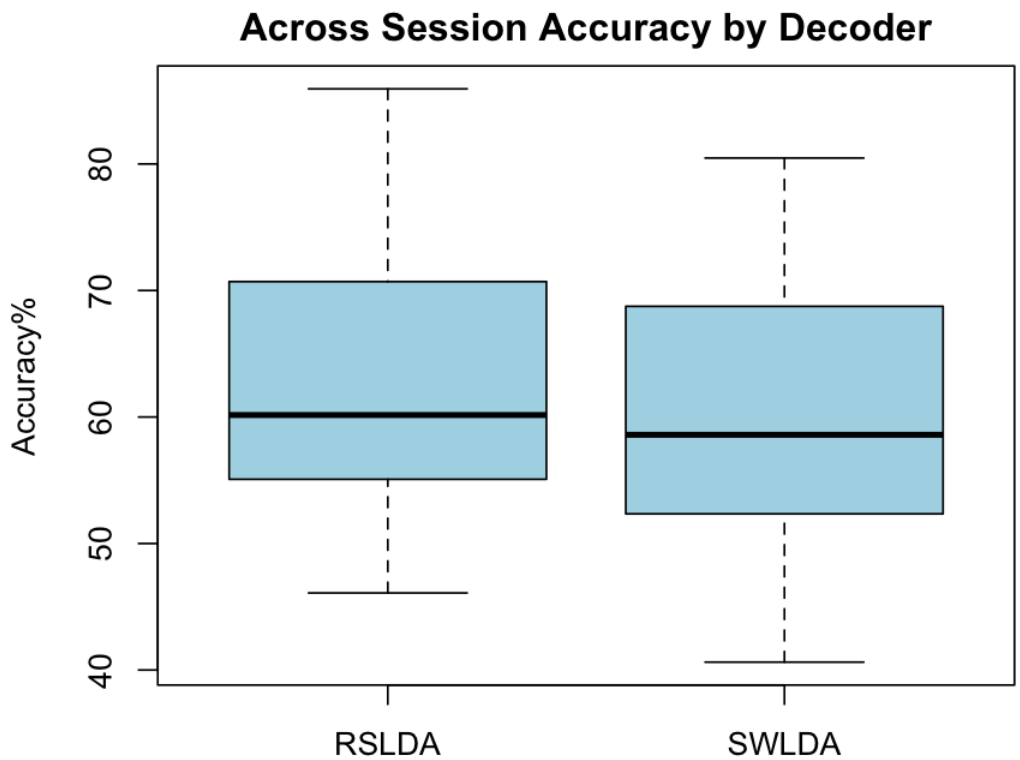


Figure 4.7: Across Session Accuracy Decoder Comparison

4.2.5 Accuracy by Sequence

The use of dynamic stopping has been tested using RLDA previously Schreuder et al. (2013) and should be considered in future versions of the current aBCI if evidence suggests this would enhance the average time of making an accurate selection. The potential benefit of dynamic stopping was evaluated by looking at the accuracy of utilizing fewer than were presented sequences or presentation of each stimulus.

RSLDA was used in an offline dynamic stopping approach. In a RSLDA model weights for classifying each stimulus are defined. During a trial, each stimulus presentation was assigned a classification score. This score indicates how likely the stimulus was a target or non-target for that specific spoken word. After a number of presentations of each stimulus, scores for each stimulus sub-class are combined. The median instead of the mean of these scores was utilized as it is a more representative statistic when outliers are present. The median score of all sub-classes are compared and the one which is closest to the target direction (in this case the most negative) was the decided stimulus sub-class. If that stimulus was the intended target then the classification was correct. Classification was completed for each sequence of the four stimuli being presented, producing an estimate of accuracy for each sequence within a trial.

Figure 4.8 shows accuracies obtained by using 2-fold cross-validation of full-trials using the RSLDA decoder. Each point represents the mean accuracy across all trials, participants and sessions, if that number of sequences was used to make the BCI decision. Error bars represent 2 standard errors of the data for each sequence.

Accuracies by sequence are also plotted for select participants. Participants 19 and 21 achieved some of the best accuracies, while participants 6 and 16 some of the worst. For participant 16 a performance ceiling was reached at about 8 sequences while participant 21 benefited from at least 13 sequences.

Sequences greater than thirteen may not provide better BCI performance in the current aBCI paradigm. Accuracy at fifteen iterations was less than 70%, on average, so any optimization of parameters to improve accuracy should be implemented in future designs to reach this threshold. The balance of accuracy and time to present numerous stimuli gives an overall rate of selection. Information transfer rate (ITR) is a metric that can be used to evaluate BCI performance taking into account accuracy, time to selection and the number of selection items possible.

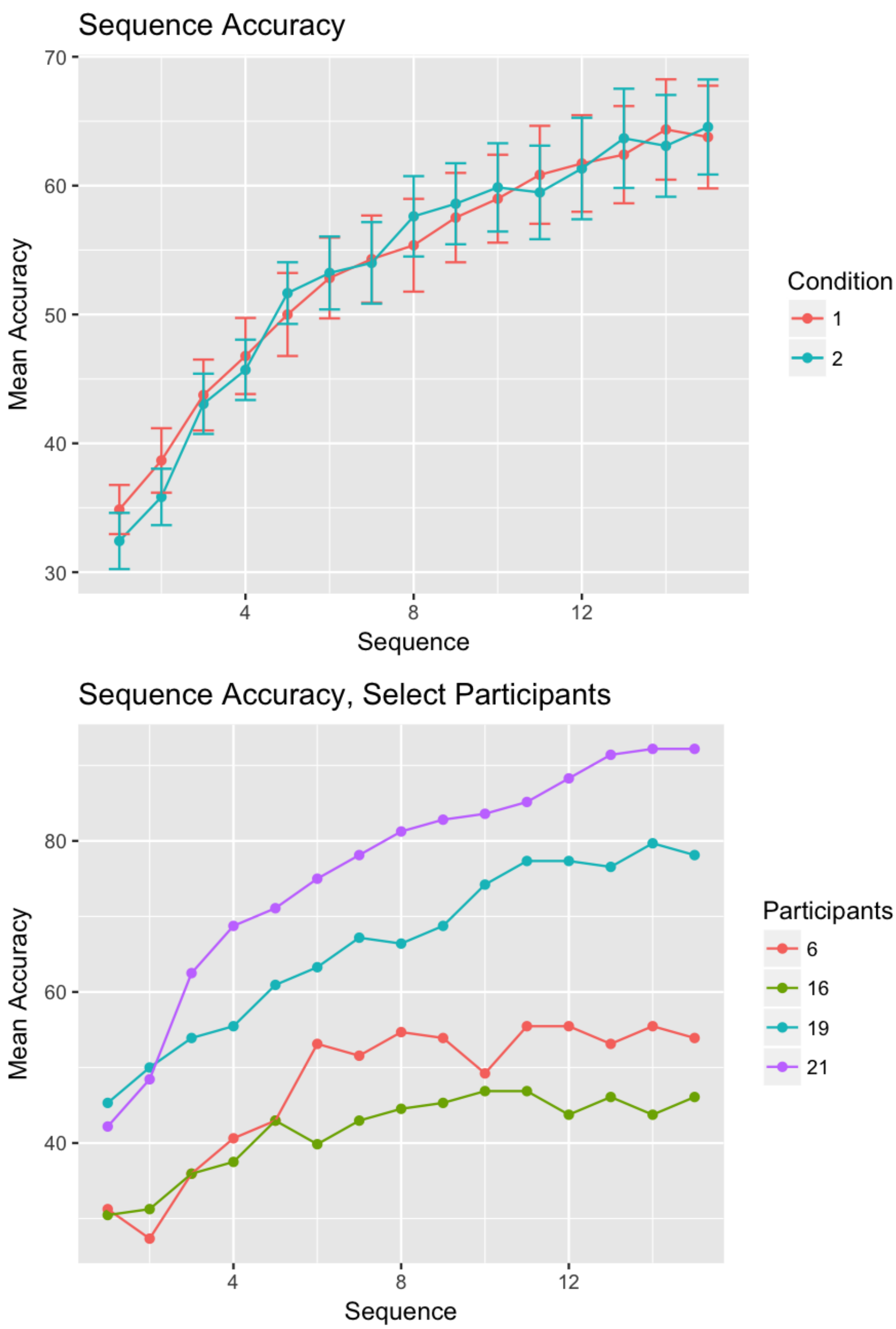


Figure 4.8: Accuracy by Sequence

The calculation for Information transfer rate (ITR) starts with computing number of bits or binary pieces of information generated per trial (Wolpaw et al., 1998). In Equation 4.1, the letter B signifies the bit rate in bits per minute, N is the number of stimuli or possible targets and P is the probability of correctly selecting the intended target.

$$B = log_2 N + P log_2 P + (1-P)\, log_2 \left(\frac{1-P}{N-1}\right) \tag{4.1}$$

The ITR is then calculated with the formula below, were B is the bit rate defined above and V is the rate of selections or trials per minute.

$$ITR = B \cdot V$$

Average ITR for all participants plateaued near the 15 sequences mark in both Figure 4.9 and Figure 4.8. Fewer sequences in this system would not appear to improve ITR. Looking at ITR for select participants shows that those that performed well were able to improve performance through multiple sequences while those that performed poorly did not see their accuracy improve with additional data added to the decision process. Accuracy improvements were not present between fourteen and fifteen sequences for participant aB19 and aB21, so minor drop in ITR occurred between these points. The maximum ITR achieved (8.48bits/min) by participant aB22 was at sequence 13. The worst maximum ITR was achieved by participant aB11 at sequence 10 (0.930 bits/min).

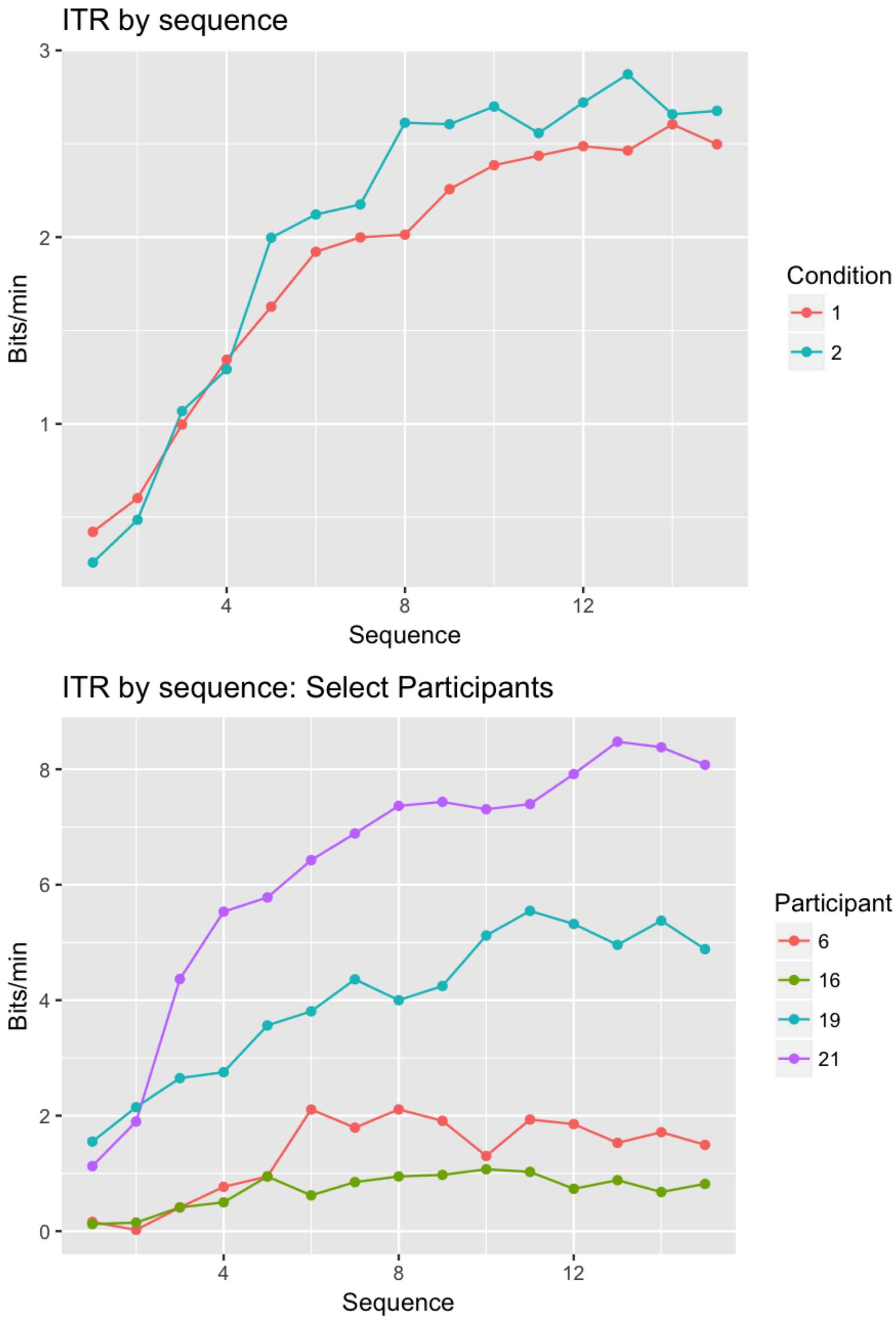


Figure 4.9: ITR by Sequence

4.2.6 Accuracy by Stimulus

Several participants commented that a certain direction or word was more prominent or distracting than others. Front and back stimuli location confusion was anticipated but not profoundly observed by the participants or through the data. One pilot participant did voice that they could not identify that the word 'while' was actually coming from the speaker behind them and perceived it came from the front speaker. Throughout other pilot and official study design trials it was expressly explained that the words 'joy' would come from the front speaker and 'while' from the rear speaker before starting training blocks of the Non-Directional condition.

Participant comments on the Non-Directional words often centered around the emotional feelings elicited by the word 'doubt' or 'joy'. These emotion words were selected based on the proposed difficulty of associating an emotion with a spatial direction. Emotional or meaningful words for participants may induce some influence on attention originating from the semantics of these two stimuli. This influence was expected to be highly individualized, and was not explored further.

Figure 4.10 summarizes the mean RSLDA 2-fold cross-validated accuracy by target stimulus across all participants and sessions. The target words spoken are listed below each bar. The direction the sound comes from is indicated by the bar color. The blue bars, coming from the front speaker, exhibit the lowest Accuracy of either condition.

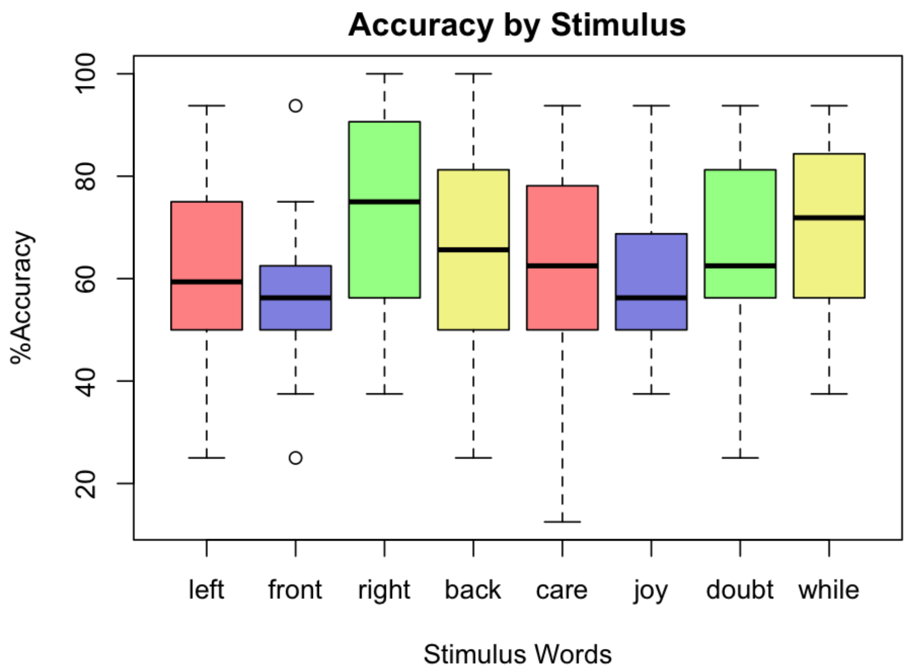


Figure 4.10: Accuracy by Stimulus

4.2.6.1 Spatial Salience

In order to investigate how spatial separation of stimuli influence aBCI performance, a comparison of lateralized ['left', 'right', 'care', 'doubt'] and centralized ['front', 'back', 'joy', 'while'] sounds was made. Lateralized words or those coming from the left and right of the participant in either Directional or Non-Directional conditions was anticipated to have more spatial salience and potentially provide improved BCI accuracy. A pairwise t-test of full-trial RSLDA accuracy by subject, session and condition between lateralized targets and centralized targets yielded no significant differences (t = 0.015817, df = 494, p-value = 0.9874). An ANOVA of accuracy on left, right or centralized sounds didn't show a significant effect either (F = 1.4717 p-value = 0.2335).

Individual preferences for left, and right lateralized sounds may be pronounced at the individual level and effectively washed out in aggregated results across participants. Figure 4.11 shows stimulus specific percent accuracy of each participant grouped into left ['left', 'care'], right ['right', 'doubt'] lateralized stimuli or central ['front','back', 'joy','while'] stimuli. For all but two participants, one of the lateralized groups has the highest accuracies. For 11 of the 16 participants the right lateralized words are the highest accuracy, reflecting the highest accuracy for the word 'right' shown in Figure 4.10. Two of the three left handed participants, aB05, and aB12 had their highest and tied for highest accuracies with left side targets. The third left handed participant, aB14 still showed minor preference for right lateralized sounds. Participant aB07 showed left side preference but is right handed, however this subject showed higher hearing thresholds in the right ear as shown in Table 4.5. Minor hearing loss in the right ear may explain some left side preference in this participant. These results suggest handedness and binaural hearing thresholds play some role in lateralized aBCI stimuli performance.

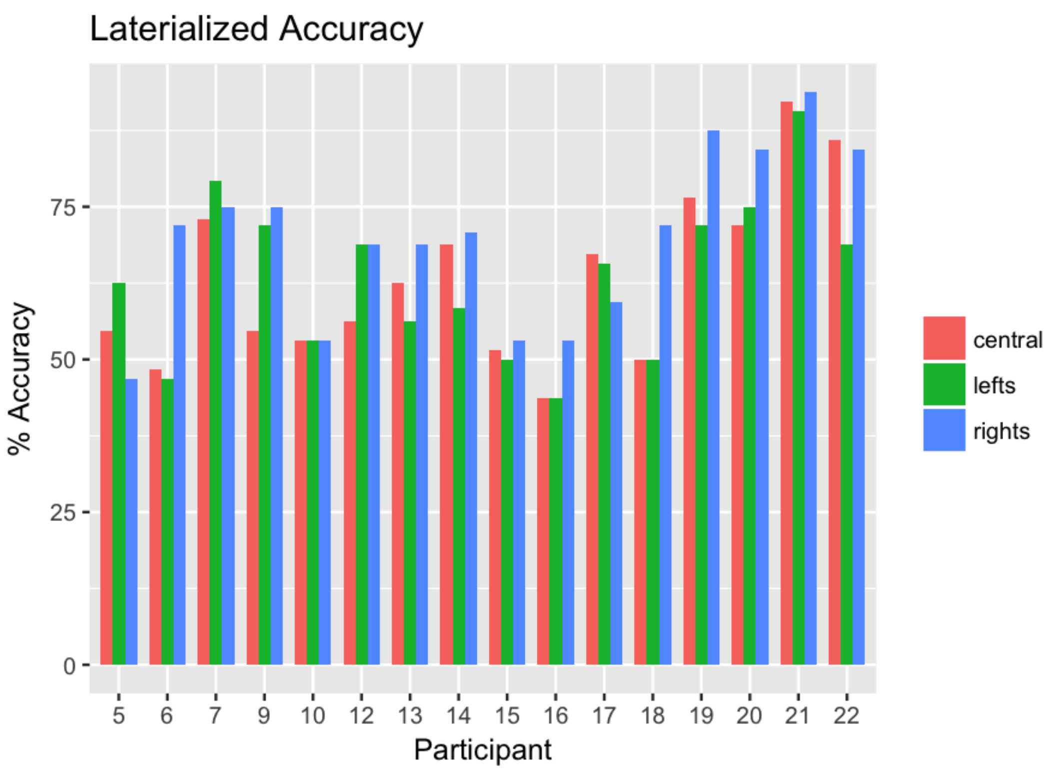


Figure 4.11: Lateralized Accuracies

4.2.6.2 Hearing Loss and Lateralized Performance.

There was hope to investigate if sizable differences in thresholds between ears might yield a difference or preference for stimuli played from the left or right location. For this reason a strict requirement on perfect hearing in study participants was not enforced. This concept could be tested by presenting a given stimulus at different volumes to perfect hearing listeners, but a full spectrum reduction in volume is not the same as frequency specific, asymmetric hearing loss, so effects could differ.

Differences between thresholds at the specific frequencies tested in the hearing screening were computed. The thresholds in the right ear were subtracted from those in the left ear and any instance of an absolute value of greater than 10dB in this difference is listed in Table 4.5. Positive values in the right most column of this table indicate higher thresholds in the right ear and may induce some preference or increased performance with left side presented target stimuli as demonstrated by participant aB07. Negative values indicate higher thresholds or increased hearing loss in the left ear. Participants aB06, aB20, aB19 all showed better performance on right side presented target stimuli as well as displaying some relative hearing loss on the left side.

Participant	Frequency(Hz)	Right-Left(Thresholds)
aB06	4,000	-15
aB07	250	15
aB07	500	15
aB07	1,000	20
aB17	6,000	-25
aB17	8,000	-20
aB19	4,000	-20
aB19	6,000	-15
aB20	4,000	-15

Table 4.5: Lateral Difference in Hearing Thresholds

The instances of notable hearing threshold differences and coincidence with BCI performance of lateralized stimuli indicates some relationship on a within-participant basis. While this study was not optimized to uncover the specific correlation of hearing thresholds and aBCI performance, multiple cases in these results support the notion that hearing loss, as well as handedness, may influence aBCI performance using spatially separated auditory stimuli.

4.2.6.3 Target Confusion

aBCI accuracy estimated by target stimulus gives us some indication on which directions were potentially the most difficult. It may also be informative to explore which direction or stimulus was selected instead when the target was not chosen. Figure 4.12 shows 2 confusion matrices indicating the percent of trials each word was chosen given a specific target word. The x-axis indicates the target word and the y-axis indicates the word chosen. The percentages indicate how often for that Target word the Winner word was chosen. The graph on the left includes accuracies for the Direction Words and the graph on the right the Non-Direction words.

The highest confusion for the Direction words was 'right' as a winner when 'left' was the target, at more than 17% of the time. The confusion of 'left' being selected when 'right' was the target was only 9.85% showing the strong preference for the stimulus 'right' between the two. These two direction are likely to be the most salient and produce the most confusion or distraction when not serving as the target. The top left four numbers in the Direction Words chart indicate how often left and right were selected when neither were the target. It should be noted all of these numbers are higher than all numbers in the bottom right quadrant where we see the percentages for 'front' and 'back' being selected when neither was the target. This highlights a minor preference for the lateralized stimuli but this effect was not found in 'care' and 'doubt', the lateralized words in the Non-Direction set.

The front and back directions were anticipated to have a high level of confusion between each other but at an average of 12.71% confusion they are not confused much more often than any other pair and less than the left / right presented stimuli confusion at 13.45%. The Non-Direction words 'while' and 'joy' represent the back and front locations in the second set of stimuli. They saw a confusion of 13.475% on average, while 'doubt' and 'care' saw an average confusion of 11.915%. These results oppose the trend seen in the Direction words, indicating that this spatial cue may not be detrimental and that acoustic and semantic features of the words themselves may have larger influences on salience and distraction.

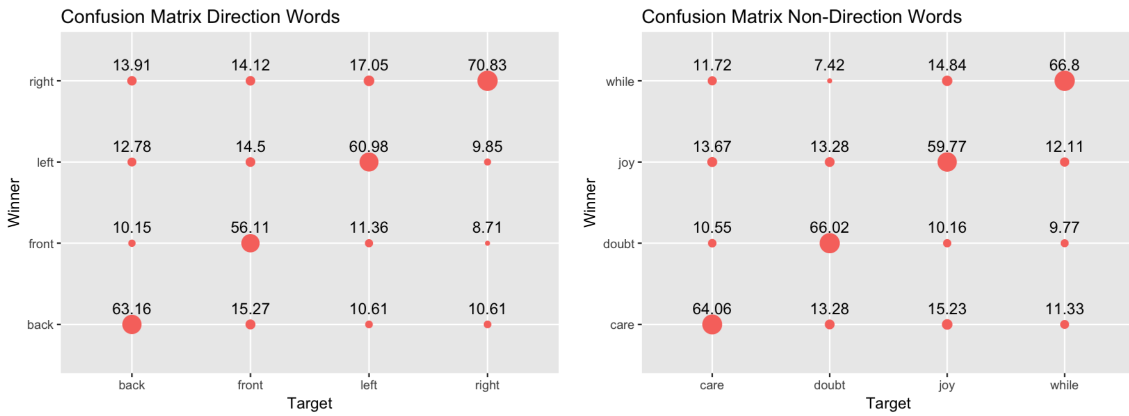


Figure 4.12: Confusion Matrix

4.2.7 Online Results

Online sessions utilized the PacGame RSLDA classifier to select the intended stimulus and corresponding direction. Although a few participants exhibited somewhat promising online results the majority of online sessions yielded near chance control of the system (26.185% average across all data). Online results for the RLDA classifier using a single session or 32 trials of training data for each condition was reported in Figure 4.13. This figure plots a smoothed density estimate line plot of the Direction and Non-Direction conditions. The variance in Online accuracy averaged for each participant, session and condition was 1.84%.

Because the percent accuracies of online trials were far less than offline analysis some differences in these two classification approaches must still remain. For this reason, no conclusions will be gleamed from the online results.

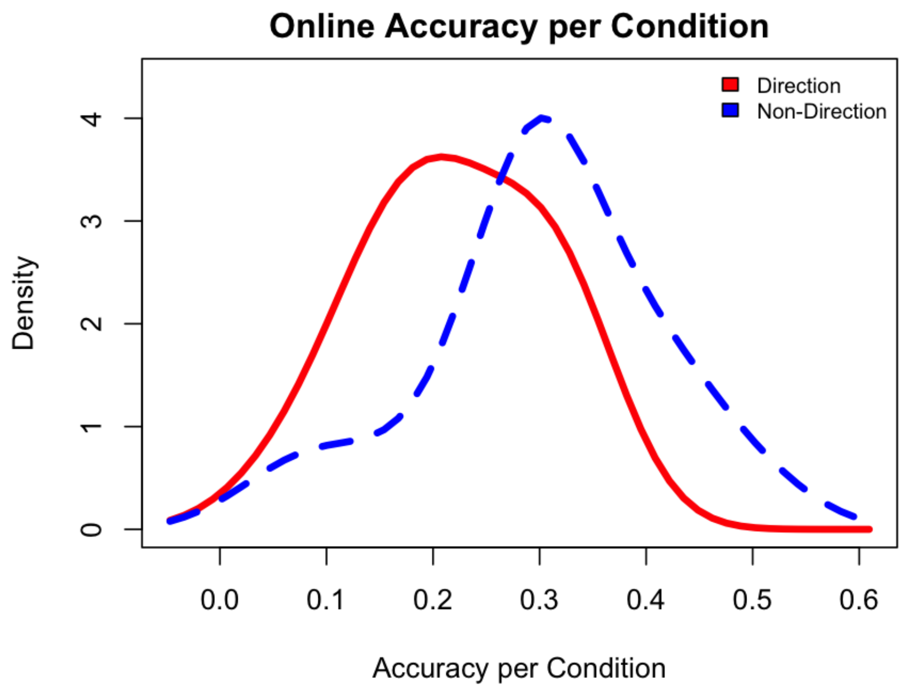


Figure 4.13: Online Accuracy by Condition

4.3 Waveform Analysis and P300 amplitude

Grand-average EEG waveforms of each stimulus for Target and Non-target averages across all participants and sessions are presented in Figure 4.14. The shaded regions around each average trace represent 95% confidence intervals over all the stimulus presentations averaged into the plots. The Target traces in blue present a slightly more negative deflection than Non-Target traces around 100ms after stimulus onset. Differences between stimuli in the positive deflection around 200ms and decay thereafter are not as pronounced.

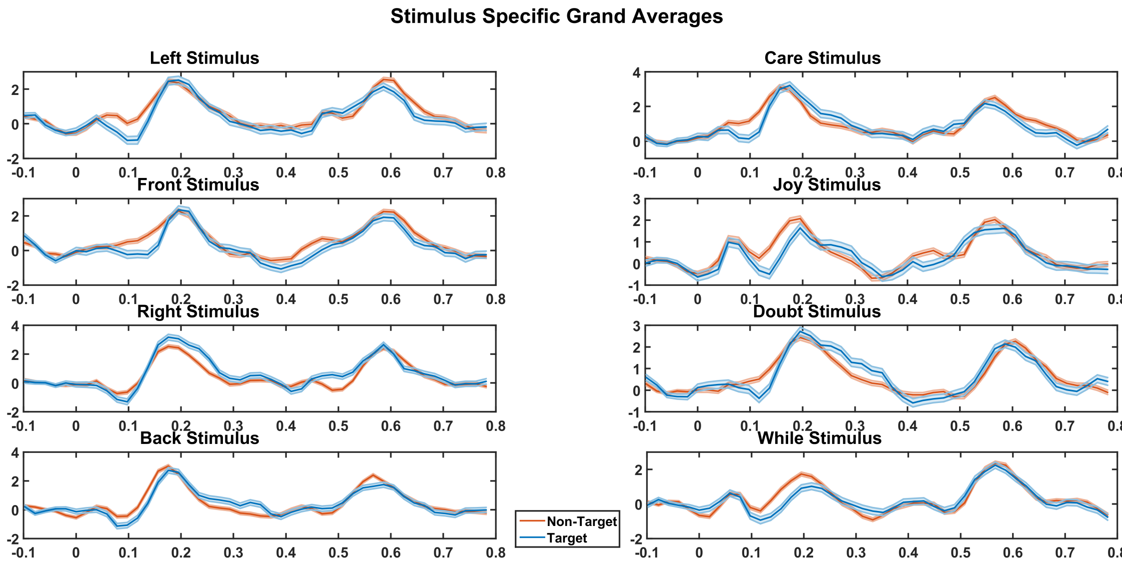


Figure 4.14: Grand Average by Stimulus and Condition

Participant averages were investigated to identify patterns in the ERP signals that might fit with past observations of speech stimuli in BCI paradigms. Figure 4.15 illustrates the approach taken. Initially the grand average of all Stimulus presentation aligned epochs are plotted in a series of topographical plots (or scalp map) of the time course of the epoch, showing, by color, the relative EEG amplitude at different locations on the scalp over time. Yellow and orange indicate more positive potentials at a location while green and then blue indicate more negative scalp voltages. Each topoplot's color scheme is normalized for the range of voltages present at that time point. The time after stimulus onset represented by each topographical plot is noted above the scalp map.

The time series topoplot was inspected to identify what central electrode might see the largest amplitude deflections during a stimulus presentation. In Figure 4.15 a positivity over fronto-central electrodes grows and dissipates around 100ms after stimulus onset. Then a more central positivity persists until approximately 240ms. A frontal negativity appears before another positivity around 350ms.

After selecting between FCZ, CZ, CPZ or PZ based on the topoplots, a grand average plot of all Target and Non-Target stimuli (across conditions and sessions) was made. The blue, Target traces typically show a pronounced positive deflection at ~200ms over non-Targets that persists to

300 to 400ms after stimulus onset, depending on participant.

The differences between the target and non-target traces are the basis for the linear classification method. The differences are modeled at each time point and at all electrode locations on the scalp. The RSLDA method was utilized for offline data presented in this report, which generates a model of the target/non-target differences for each target stimulus. To illustrate what the RSLDA classification method would focus on, the target and non-target traces for each target stimulus are also included in the example figure. While similar morphology appears in all stimulus plots, differences between stimuli are apparent, especially in the differences between target and non-target. White space between the traces indicates a lack of overlap in the 95% confidence interval shaded regions and a likely strong predictor of difference that will be utilized by the classifier to identify target stimuli.

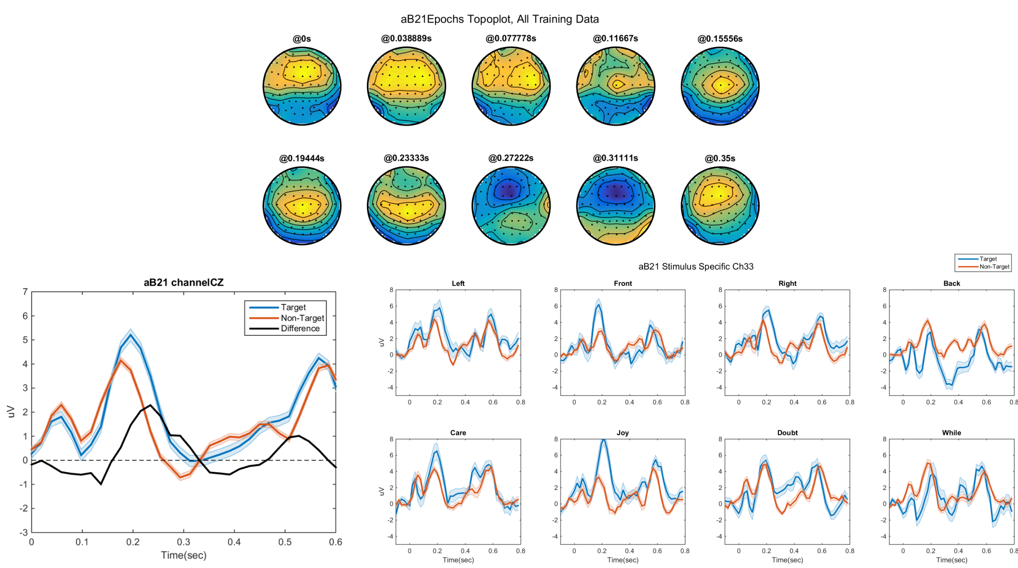


Figure 4.15: Participant aB21 ERP example

Figure 4.16 illustrates the same topographical plots and grand average ERP comparisons as Figure 4.15 but of participant aB16. Graphics of these kind for all participants can be found in the Appendix A.3. While general timing of relative positive and negative deflections was very similar between these two participants, as can be seen in the Grand average plots, some differences in location of most pronounced positivity and negativity was apparent in the scalp maps. Participant

21 in the first figure was one of the best offline RSLDA percent accuracy performers while aB16 was one of the worst. The noise or additional oscillations apparent in the grand average and stimulus specific ERP traces for aB16 may illustrate some of the reasons for lower percent accuracy performance in the aBCI. The 95% confidence regions of the stimuli specific plots appear much wider for aB16 than aB21, illustrating the increased variance in aB16's training data.

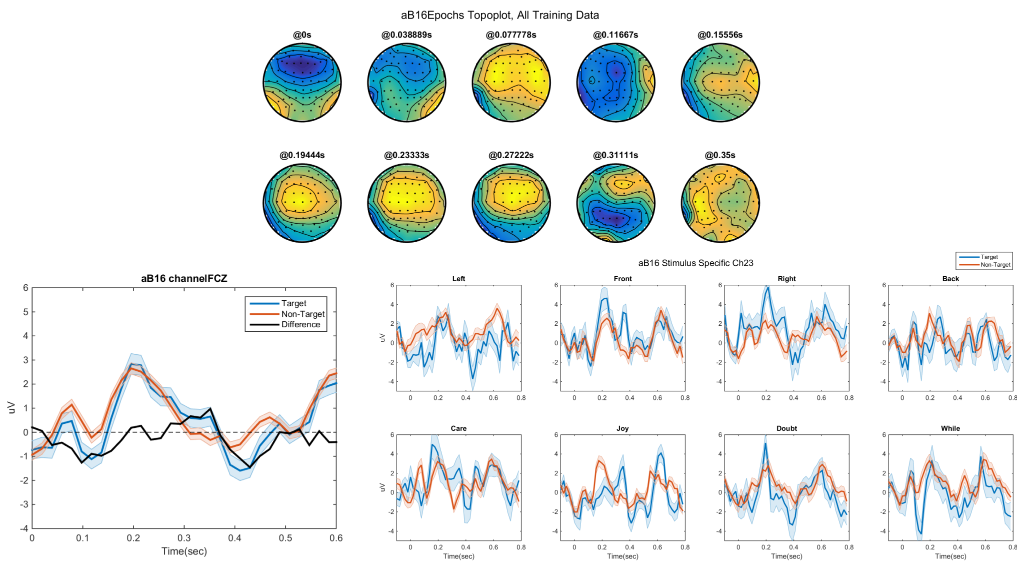


Figure 4.16: Participant aB16 ERP example

4.4 Questionnaire Summaries

A self-reporting of motivation was given at the start of each session and 2 questionnaires were completed at the end of session 2. These same self-reports have been completed in past research and may give some indication of the sources of success or failure arising from individual participants or the BCI system as a whole. Many of these measures were quite consistent across participants, indicating they are a result of the system design and not wholly based on individual preference.

4.4.1 NASA-LTX

The NASA-Task Load Index (NASA_TLX) overall score was negatively correlated with the RSLDA percent accuracy score., indicating increased self-reported workload corresponded to lower aBCI accuracy. The NASA_TLX data doesn't fit a normal curve or pass the Shapiro-Wilks test for normality (W = 0.92158, p-value = 0.0008884) so a non-parametric measure of correlation is reported yielding a very weak correlation (Spearman's rho = -0.2757801). NASA-TLX survey was completed for both Direction and Non-Direction word conditions, but no significant difference was found between the participant's overall ratings between conditions (t = -0.20909, df = 25.975, p-value = 0.836).

Figure 4.17 shows the mean and 2 standard deviation error bars for the weighting of the individual categories of workload included in the NASA_LTX. These weightings come from the selection of one category between each possible pair of workload sources. The maximum and minimum weighting for each would be 5 and 0 respectively.

Participant comments and explanation during administration of the survey yielded some insight into the cause of the resulting weightings. Physical demand was selected the least often, likely due to no perceived physical task associated with the aBCI. Mental demand was weighted the highest, while at least some combination of Frustration, Effort and Performance related workload seemed to be present for most participants.

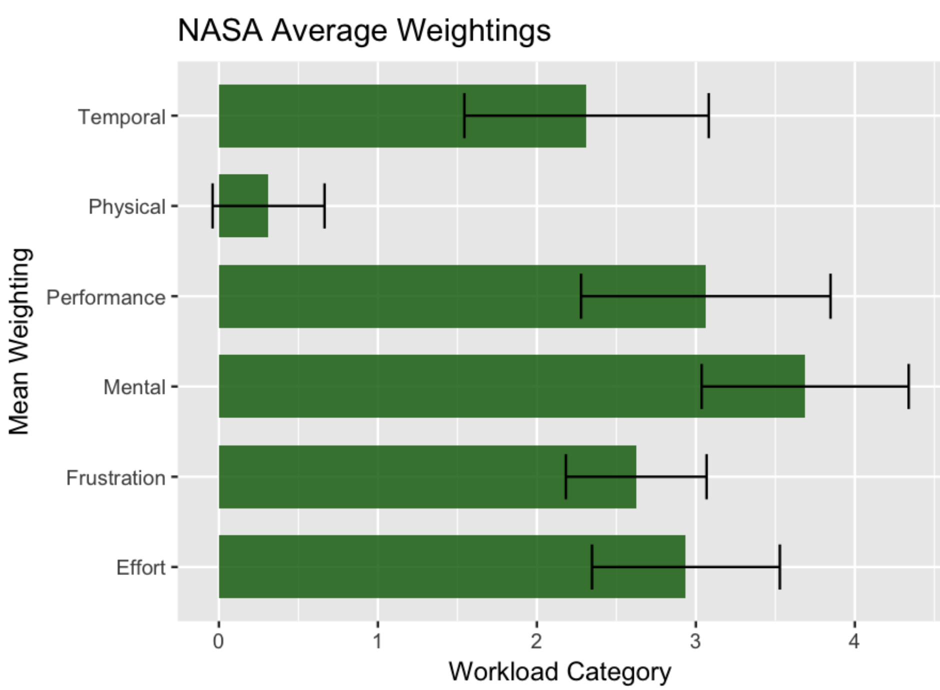


Figure 4.17: NASA weightings

4.4.2 SUS

The global average for the System Usability Scale is a score of 76, a typical score for most systems evaluated with this scoring system. The overall averaged score for the PacGame aBCI was 64.3, indicating that the participants were less satisfied with the system than most systems evaluated with this scale. In most cases, participants voice disappointment in the poor online results. Because the online tests yielded very low accuracies, completion of a round (Pac reaching the cherry icon) occurred very seldom.

The SUS rating was found to pass the normality test (W = 0.96666, p-value = 0.08097) and did weakly correlate with aBCI percent accuracy (Pearson's r = 0.2797689). The online feedback participants received was not a good indication of performance. However, participants may perceive internal clues of how well they are attending to the target stimuli and accomplishing the BCI task, reflecting somewhat accurate self-reports.

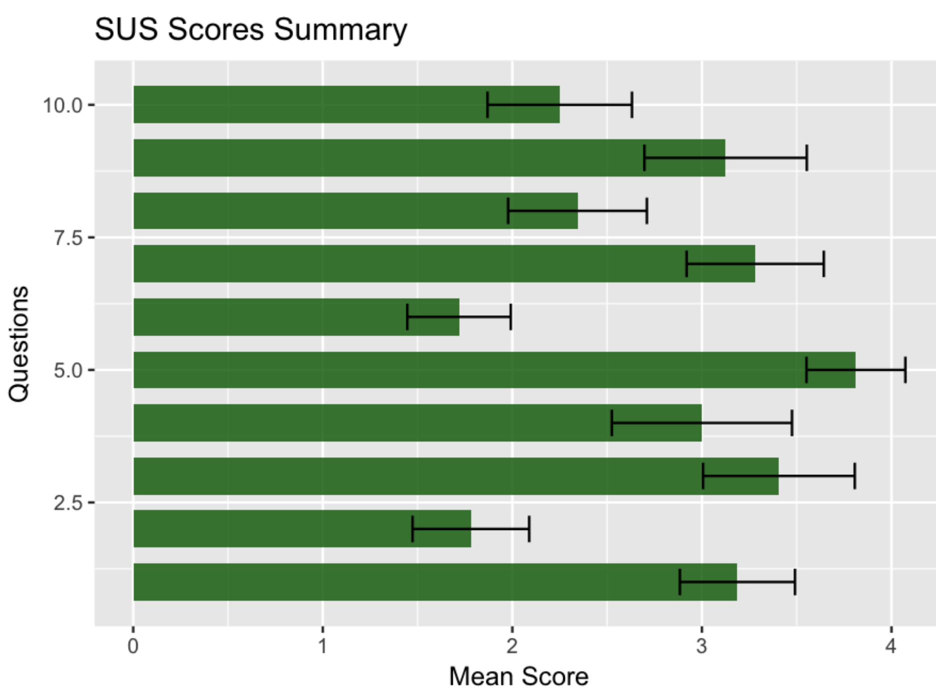


Figure 4.18: SUS Score Summary

4.4.3 Motivation

In Käthner et al. (2013) a correlation of self-reported motivation and P300 amplitude was found, however a Pearson correlation was used which assumes normality in the distributions tested. Motivation, in this study, doesn't pass the Shapiro-Wilks test for normality (W = 0.93932, p-value = 0.003529) and so a rank order, non-parametric, Spearman correlation was calculated (rho = 0.1928773).

Motivation was self-reported and should be viewed as relative to each participant. Motivation's influence on BCI performance by relating the difference in motivation and BCI accuracy between sessions. Figure 4.19 plots the change in motivation vs the change in RSLDA offline percent accuracy between sessions. Session 2 motivation and accuracy were subtracted from Session 1 motivation and accuracy respectively. The blue dashed line shows the significant regression line indicating, inverse an relationship between change in motivation and change in accuracy. Each point is labeled with the participant's id number.

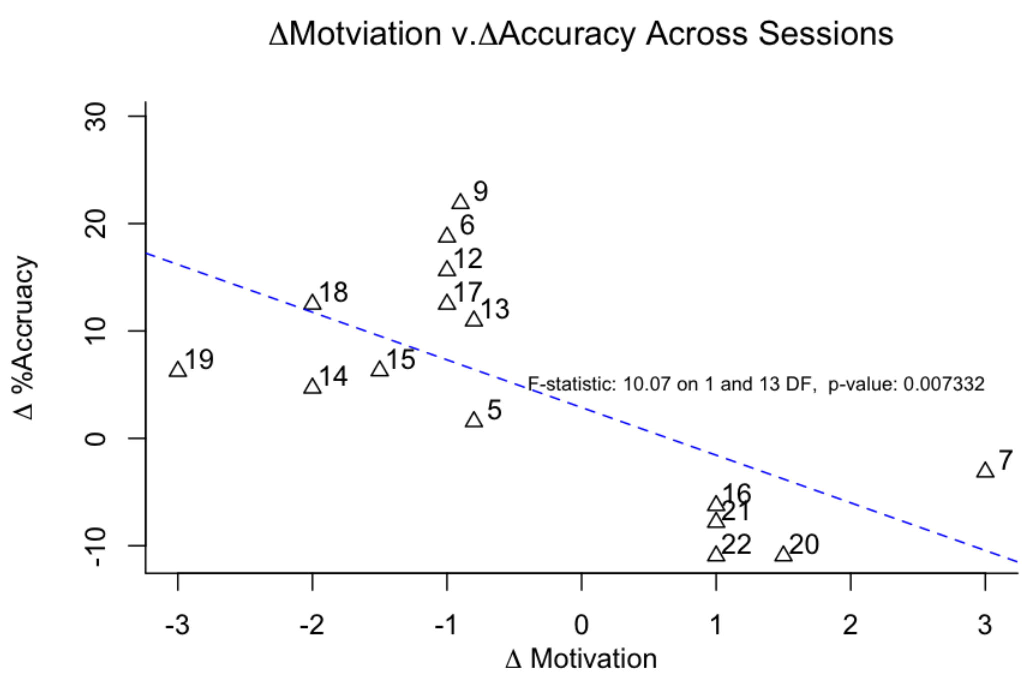


Figure 4.19: Change in Motivation and Accuracy Across Sessions

4.5 BCI results Summary

The BCI system displayed a high level of performance in a few participants and was found to produce generalizable models that could be used across EEG sessions. While participant performance varied greatly, the average accuracy of the system did not greatly deviate between sessions, stimuli or word sets. In analyses where additional training data was used to generate the linear model improved accuracy was estimated. Use of a regularized linear classification method was shown to be significantly better than SWLDA.

Spatial relevance was tested by looking at performance of target words from specific auditory source locations, but found no statistical differences. However, stimulus specific analysis highlighted the importance of hearing thresholds in performance of lateralized stimuli. This would suggest hearing to be tested and accounted for in future auditory BCI studies utilizing spatially separated auditory stimuli.

Questionnaire results tended to mathematically align with the offline aBCI accuracy. Although online feedback did not give clear indication of their performance, participants seemed to understand how well they were doing with the task. The importance and reliability of self-report in these types of research environments is critical to efficient development of useful BCI systems. Self-reports of motivation trended opposite that of aBCI performance across sessions.

Investigation of ERP waveforms in all participants gave a thorough account of what EEG is elicited by spoken word stimuli in such a paradigm. These results qualitatively reflect performance across participants through differences in target/non-target traces. Timing of ERP difference in target/non-target do not suggest a positive deflection at or after 300ms was the primary predictor of BCI performance, which has also been found in other aBCI studies (Halder et al., 2013). ERP variation seen across participants and across stimuli suggest stimulus specific models, like RSLDA, may be beneficial with spoken word stimuli sets, however dramatic differences in ERP morphology between stimulus were not identified.

Chapter 5

Word Recognition Results

5.1 Purpose

The overall effect of spatial separation, the difference in performance between central and lateralized stimulus, was expected to be significant in the aBCI, but it was not. An additional behavioral experiment was undertaken to further investigate the usefulness of spatial separation of stimuli. The Word Recognition task (WR) is, as the name implies, a task of recognizing a target word amongst distractor words presented as they were in the aBCI task. A button press is used to signal to the experimenter the participant has identified the target word has been played.

Fifteen healthy participants completed three different stimulus presentation conditions including the BCI, NoCues and Dynamic conditions outlined in Section 5.3.1. Each presentation condition included the the Direction and Non-Direction word sets used in the aBCI study. Reaction time and accuracy measures further investigated spatial cues, the number of presentations in a trial, and specific stimuli influences on recognition of the target word in the experiment's participants.

All but one WR participant completed some portion of the aBCI study, although one WR participant was dropped from the aBCI study. In attempts to extend the WR results to aBCI, correlation of reaction time (RT) and aBCI accuracies was computed from results of 13 participants.

5.2 RT and Button Press Accuracy

The performance metrics of the Word Recognition Task are the reaction time (RT) and the percent accuracy of pressing the button after target presentation. Using no timing restrictions, participant's percent accuracy was near ceiling in most cases and additional attempts to identify missed or very late button presses was implemented into the automated reaction time analysis.

Reaction time calculations were accomplished by computing the difference in the recorded time of first button press directly following the onset of a target stimulus. If this time difference was greater than 2 seconds, the target presentation was recorded as a miss. If this time was less than 100ms than it is expected that the button press was in anticipation of the target or a result of a late button press from a previous target. One hundred millisecond RT was chosen as the low-end cutoff as it is referenced as a rare and nearly physiologically impossible reaction time. Only highly trained athletes have exhibited lower reaction times in less complicated stimulus recognition paradigms (Pain and Hibbs, 2007). Reaction time outside ±2SD (standard deviations) from the grand mean is often filtered out in Psychology research. With the higher level of variability resulting from these tasks, this approach would have resulted in a negative reaction time low-end cutoff and a high-end cutoff at ~830ms. The upper bounds of reaction time was calculated by adding 3 times the interquartile range (IQR) to the grand median of all reaction time measurements, resulting in an upper limit of 638.3ms. This value is well outside the reaction time IQR of each participant. The median reaction time for all button presses within the finalized boundaries [100ms 638.3ms] range was 307.5ms.

Reaction times within zero to 100ms are tabulated by participant in Table 5.1, along with presses occurring after the upper limit threshold. The late hits category also includes button presses greater than 2 seconds that were originally recorded simply as misses. Percent accuracy scores reflect the number of reaction times that fall within acceptable bounds divided by the total number of target presentations. There were 960 target presentation for each subject, including all three presentation conditions and two word set conditions. There was no penalty included in accuracy for additional button presses during the experiment.

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Early Hits	2	10	1	1	11	7	3	18	7	14	46	46	8	11	2
Late/Miss	*49*	*132*	*14*	*17*	*19*	*89*	*36*	*97*	*121*	*17*	*112*	*299*	*91*	*82*	19
Total Outliers	51	142	15	18	30	96	39	115	128	31	158	345	99	93	21

Table 5.1: Misses and Anticipation by Participant

The relationship of accuracy and reaction time is typically intertwined. Strategies of the participant will vary in preference for rapid reaction time or accuracy. In the current paradigm, the rate of stimuli presentation may limit the ability of the participants to decide whether to bias quick reaction time or accuracy. The correlation between the accuracy and reaction time was expected to be negative as increases in reaction time should allow for accurate button presses and vice versa. As the reaction time and accuracy are not normally distributed a Spearman's rho correlation was computed. By pairing mean reaction times and accuracies for each participant, presentation and word set condition a significant rho value of -0.6527137 (S = 175180, p-value = 9.781e-12) was found. This strong negative correlation indicates good performers likely had both short (fast) reaction times and high accuracies and that poor performers had long reaction times (slow) and low accuracies. Good correlation suggests that either measure is a good indicator of performance in the task.

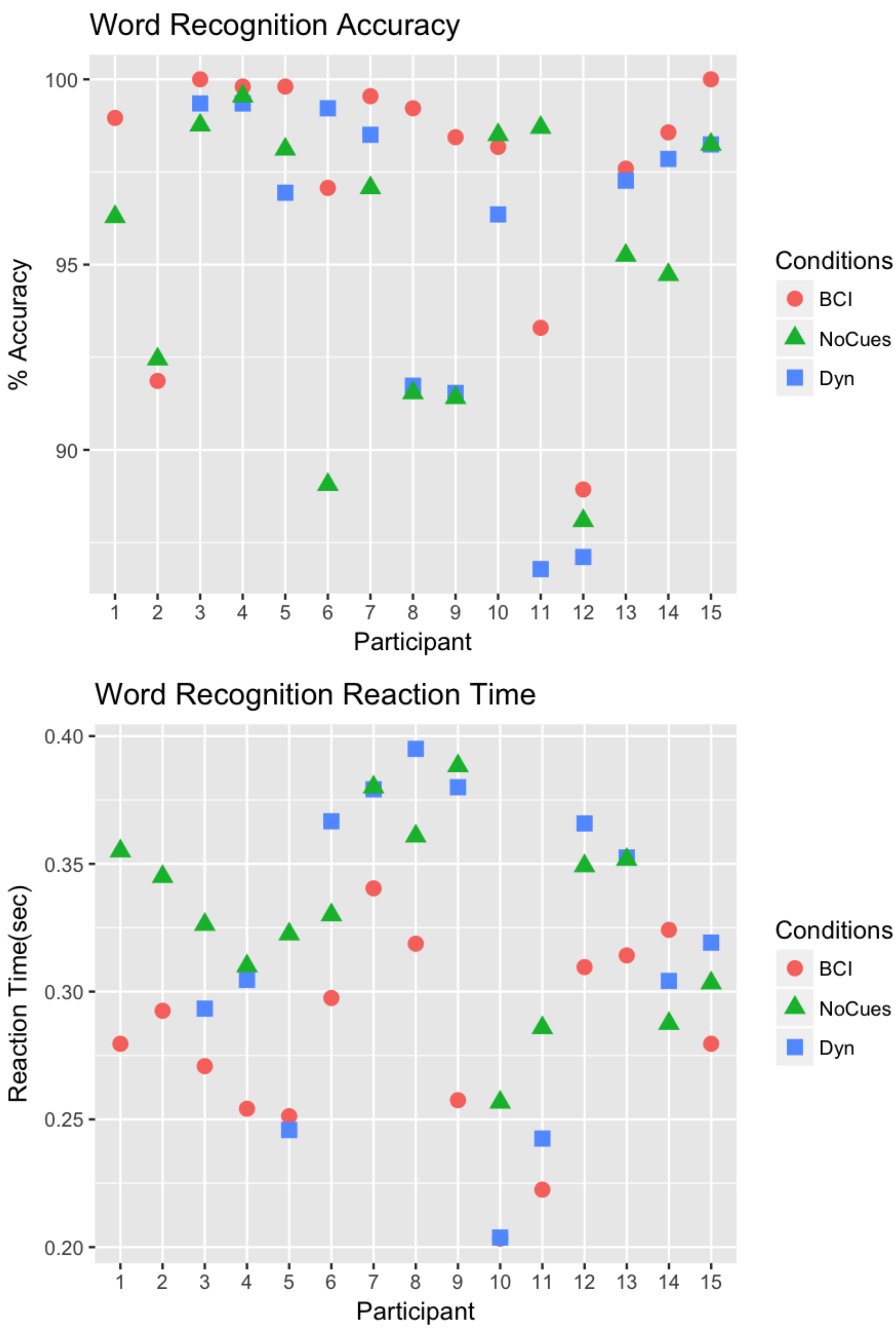


Figure 5.1: Accuracy and RT by Participant

Figure 5.1 illustrates the variability in participant's Accuracy and RT. The percent accuracy and median RT are illustrated for each presentation condition. The BCI condition consistently yields the highest accuracy and the the shortest reaction time for nearly every participant. The NoCues condition represents the slowest RT for most participants.

5.3 Effects of Condition

In order to account for several influencing factors in the reaction time, a linear mixed effects model was generated with the main effects of presentation condition and word set condition along with interaction of these conditions with their respective order of completion for each participant. A main effects of number of sequences and a random effect of participant was also included. The F-table in Table 5.2 on page 85 shows that all of these independent variables are significant predictors of log10 RT except for the interaction between word set and order of completion of the two word sets.

	numDF	denDF	F-value	p-value
(Intercept)	1	12321.00	997.75	0.00
Presentation Condition	2	12321.00	354.32	0.00
Order	2	12321.00	47.36	0.00
Word Set	1	12321.00	17.35	0.00
Word Order	1	12321.00	11.56	0.00
Sequences	1	12321.00	62.96	0.00
Presentation Cond:Order	4	12321.00	40.57	0.00
Word Set:Word Order	1	12321.00	0.13	0.71

Table 5.2: Controlling for Order of Presentation

5.3.1 Presentation Condition

The presentation conditions were designed to investigate the impact of spatial cues as they were employed in the aBCI paradigm. Overall, the Reaction time between the presentation conditions showed significant results in a one-way ANOVA of $\log_{10}$ RT (F = 276.9, p-value = < 2.2e-16). These results not only give some indication of the importance of spatial cues in P300 oddball presentation schemes but in speech recognition as a whole. The presentation was at a reasonable speaking rate and while the words are few and repeated at random the paradigm may have implications and influences tied to general speech recognition. Multiple comparisons between conditions confirms the RT differences between each of the presentation conditions. See Table 5.3 for statistical comparison between presentation conditions with Tukey corrected p-values.

The lack of spatial cues in NoCues condition made the task more difficult for some participants. The median RT for the BCI condition was 51.7ms faster than the NoCues condition. Nearly every participant exhibited equal or better accuracy RT in the BCI condition than in NoCues. The Dynamic condition had the greatest variation within condition as illustrated in Figure 5.2. When questioned which of the conditions was hardest, 10 of the participants chose the Dynamic condition and 4 chose the NoCues condition. Participant 11 commented that the Dynamic condition was the easiest and that BCI was the most difficult despite their Dynamic condition having the worst performance in RT and accuracy.

Linear Hypotheses:	Estimate	Std. Error	t value	Pr(>|t|)
NoCues-BCI==0	0 0.070764	0.003036	23.310	<2e-16
Dyn-BCI==0	0.043216	0.003160	13.677	<2e-16
Dyn-NoCues==0	-0.027548	0.003210	-8.582	<2e-16

Table 5.3: Presentation Conditions Multiple Comparisions

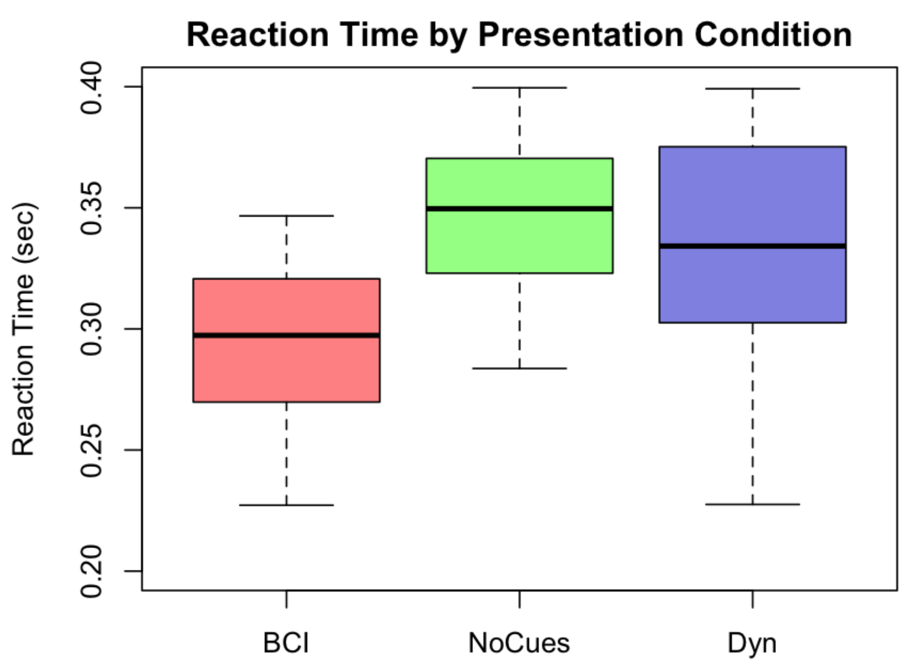


Figure 5.2: RT by Presentation

5.3.2 Sequence Count

Using a generalized linear model the equation for RT predicted by number of sequences returns the following equation (stderror=0.000234, t-value=6.601, p-value=4.25e-11).

$RT = 0.001545 * sequences + 0.302$

The fewer number of target sounds to identify in a trial, the faster the reaction times were, indicating that some fatigue during the longer trials may influence reaction times. For every sequence, RT increased by 1.545ms, according to the resulting model. The variance in RT reduces as the number of sequences goes up, but this may be due to the larger number of reaction time measures recorded as the number of sequences increases.

The accuracy of the aBCI should improve with additional sequences as was shown in Section 4.2.5. Reaction time demonstrated the opposing effect of fatigue within a trial. The benefit of additional trials must be balanced with increasing fatigue in the BCI system in order to reach optimal ITR.

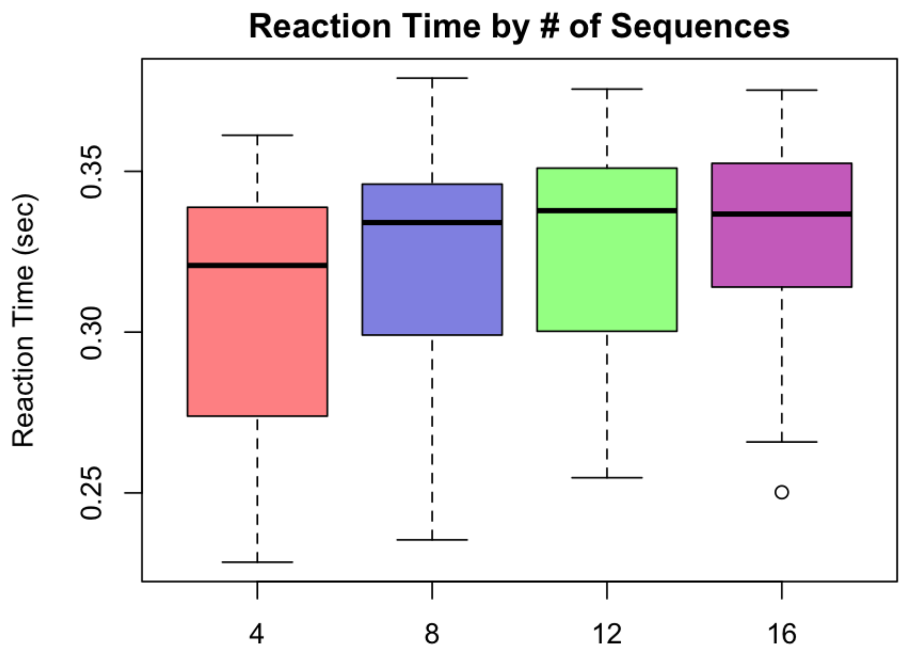


Figure 5.3: RT by Sequence Count

5.3.3 Word Sets

In the aBCI experiment no significant difference in performance was found between word sets. One hope for the WR experiment was to uncover differences in semantic and spatial relevance that were not measurable in the noiser EEG signal. The difference in RT between word sets for each presentation condition was investigated.

In the 'BCI' presentation condition there is a semantic connection between each word stimulus and its spatial location. In the other two conditions, participants must utilize the acoustic differences in the stimuli and cannot rely solely on spatial cues. In order to uncover the influence of Word Set on RT, a multiple comparison of word set RT within each presentation condition was completed. The $\log_{10}$reaction time difference between word sets within the BCI condition was significantly different than zero (t-value=2.662, p-value = 0.0231) well as within the Dynamic condition (t-value=6.264, p-value=<1e-06).

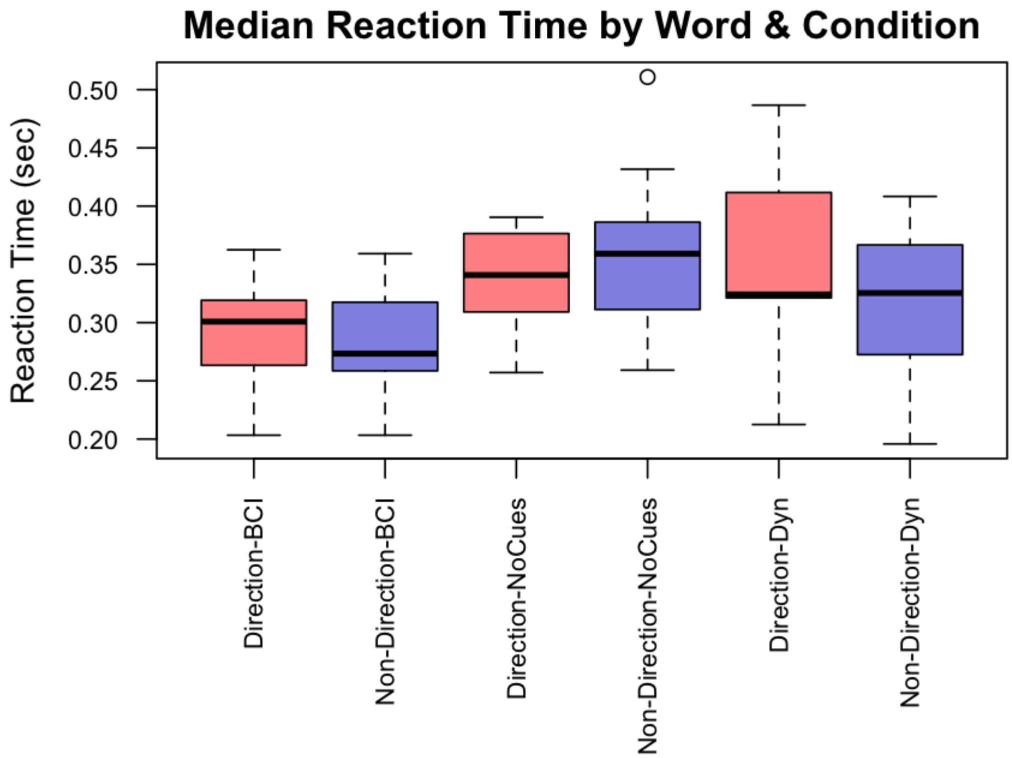


Figure 5.4: Word and Condition Median RT

The NoCues condition did not see significant differences between word set reaction times (t-value=-2.075, p-value=0.1098). A further comparison of the effect of word set is accomplished, taking into account each stimulus.

5.4 Location and Stimulus Specific Performance

The word 'front' exhibited the best reaction time of the Directional words in both the BCI and NoCues condition. For the BCI condition this was the stimulus that the participant was oriented towards, so this may have given an advantage. In the NoCues condition this word is the only one where sound location and word meaning match up. The other stimuli likely create an incongruence and slow RT.

In the BCI condition there is a better reaction time for the Non-Direction words that are lateralized ('care' and 'doubt') than centralized ('joy', 'while') (t = -4.7457, df = 2048.9, p-value

= 2.221e-06). This may indicate that lateralized sounds provide more salience, but not when semantic relevance is included in the distractors. This highlights another example of the Directional stimuli providing benefits of relevance for the targets, but also detriments by making the distractors relevant as well.

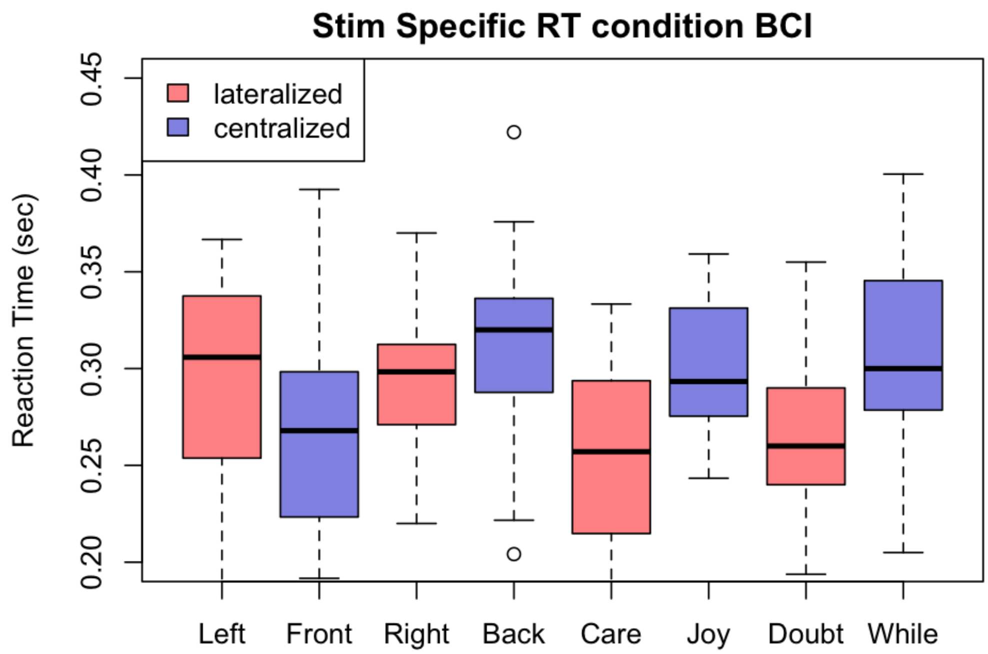
In the NoCues condition the lateralized words did poorer. Since these words may be expected to come from the left and right speakers instead of the front speaker, there seems to be an effect. This concept is again highlighted by the fact that the word 'front', which corresponds to the presented stimulus location, exhibited the best reaction time. The Non-Direction words in the NoCues case don't show such a separation between stimuli. ANOVA tables of $\log_{10}$ RT predicted by stimuli word set and specific stimuli are included in Table 5.4, showing that the significance of the word set in the BCI condition. In the No cues condition the word set is not a significant effect.

BCI	Df	Sum Sq	Mean Sq	F value	Pr(>F)
targ	1	0.07	0.07	2.56	0.1099
word	1	1.80	1.80	61.43	5.642e-15
targ:word	1	0.04	0.04	1.28	0.2586
Residuals	4683	136.91	0.03		

NoCues	Df	Sum Sq	Mean Sq	F value	Pr(>F)
targ	1	0.13	0.13	3.73	0.0535
word	1	0.05	0.05	1.32	0.2509
targ:word	1	0.02	0.02	0.67	0.4114
Residuals	4562	160.62	0.04		

Table 5.4: ANOVA of word and stimulus on RT

These results present several cases that indicate the directional meaning of the words is a factor in RT measures.



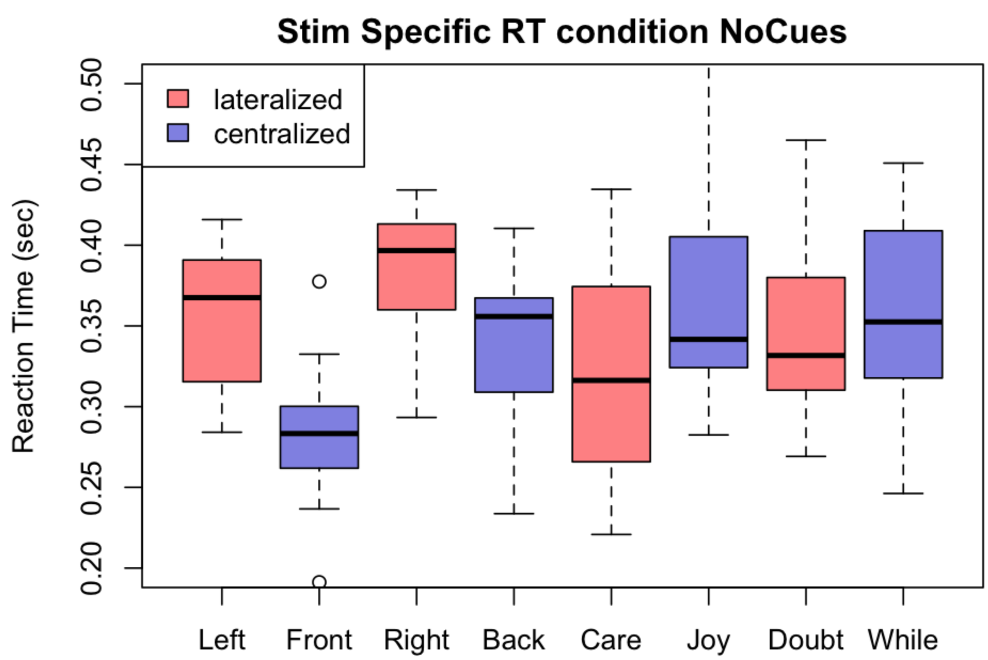


Figure 5.5: Stimulus RT

Incongruence Events Incongruence is the condition where features of a stimulus contradict one another. In the classic Stroop test a color word is presented on a screen, for example 'green'. If the color of the letters in the word 'green' are in fact red, for example, then there is an incongruence with that stimuli. In the Dynamic condition the location of the Directional words may have congruence (i.e. 'left' coming from the left speaker) or an incongruence ('right' coming from the front speaker). Many participants voiced difficulty with incongruent stimulus, such as the word left coming from the right speaker, during the Dynamic condition trials. This incongruence of location and word meaning only occurs in the Directional word set in the Dynamic condition. The influence of RT between congruent and incongruent target stimuli in this condition was tested with a pairwise t.test of $\log_{10}$ transformed RT, but was not found to be significant ($t = -0.48867$, $df = 843.38$, $p\text{-value} = 0.6252$).

It has been documented that individual strategies of navigation influence the incongruence negativity, an EEG artifact present during incongruent auditory stimuli presentation with spatial cues (Buzzell et al., 2013). Individual differences may determine how semantic and audio-spatial relevance influences BCI performance, creating an inconsistent effect across subjects and conditions. Figure 5.6 shows the median RT for congruent and incongruent target stimuli in the Dynamic condition. In most participants, the differences in RT are negligible. Participant 8 saw faster times with congruent stimuli but participants 9 and 12 RT were faster for incongruent targets.

5.4.1 WR and aBCI Precent Accuracy

The ability to identify a target stimulus in these paradigms should impact button press reaction and influence P3 and other ERPs used by the BCI. Therefore, the relationship of BCI accuracy and reaction time is expected to correlate. Correlation between RSLDA offline aBCI percent accuracy and mean RT for each participant and word set was computed using Spearman's rho and was found to be significant. See Table 5.5. This table includes correlations of participant aBCI accuracy and RT including breakdown of correlation with each presentation condition of the WR experiment.

When comparing correlations of aBCI accuracy and RT of the BCI or Dynamic Word Recognition conditions, the correlation is similar to the overall. The RT from the NoCues condition, however, does not show a significant correlation with aBCI percent accuracy. Spatial separation between stimuli has been shown to be an important factor of performance in the RT performance of

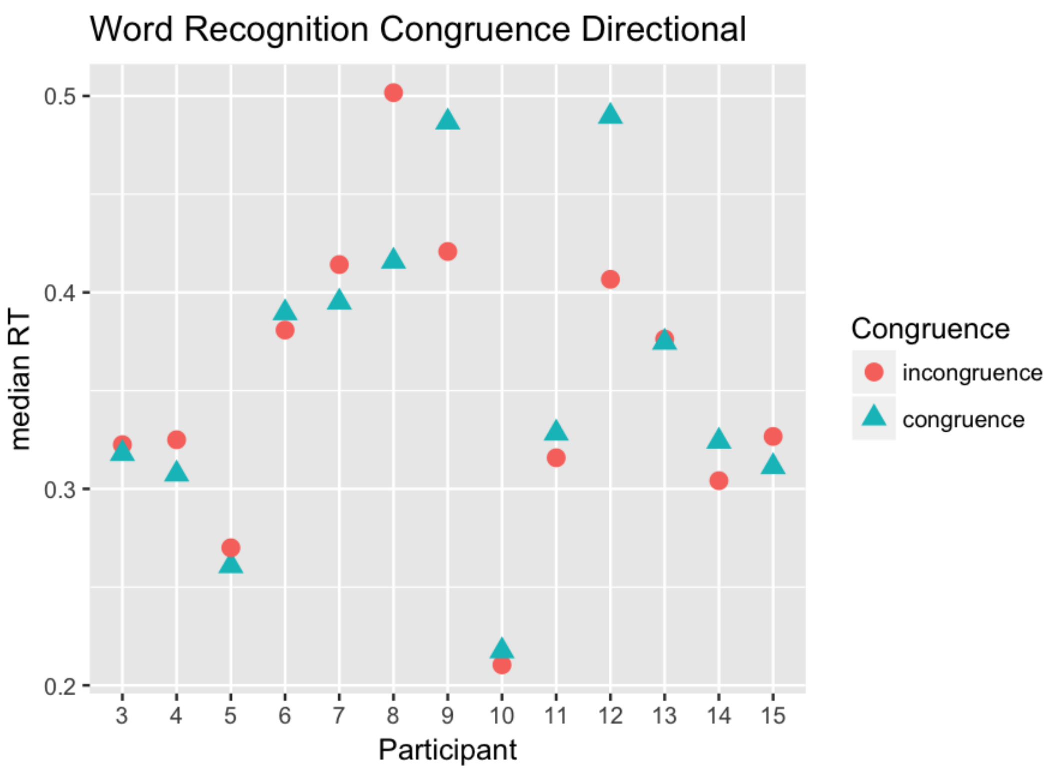


Figure 5.6: Congruence RT performance

the WR experiment and is critical in the relationship with the aBCI experiment performance. From this relationship, it may be concluded that other factors that influence the WR-RT result would also impact the aBCI performance in a similar fashion.

aBCI % accuracy Cor	rho	S	p-value
All WR-RTs	-0.4198087	4152.9	0.03275
WR-BCI	-0.4138412	4135.5	0.03558
WR-NoCues	-0.1500774	3364	0.4643
WR-Dyn	-0.3963695	4084.4	0.045

Table 5.5: Correlation of WR and aBCI accuracy

Chapter 6

Discussion

6.1 Discussion of Study Results

With comparison to other BCI systems tested the offline performance showed similar results. This system utilized spatially separated spoken word stimuli as has been featured in very few BCI paradigms until now (Ferracuti et al., 2013; Kleih et al., 2015). These results support the feasibility of a multi-class auditory BCI for communication replacement using word stimuli with maximal spatial separation. With improvements in online decoding and optimization of presentation parameters, such a system should be tested with patient populations.

The investigation into spatial relevance lead to interesting findings regarding preference for specific lateralization of stimuli that may be influenced by individual's handedness and relative hearing thresholds. RT measures proved that spatial cues were significantly beneficial in the WR experiment.

RT was found to correlated with aBCI performance, giving strong support that behavioral studies can inform on BCI paradigms. The influence of stimulus presentation parameters may be better elucidated through behavioral measures than through measuring BCI performance.

6.1.1 Semantic and Spatial Relevance.

6.1.1.1 Semantic

Semantic relevance of spoken word stimuli is a key feature in the auditory only clinical BCI proposed in Chapter 1. This study introduced a novel BCI design that aimed at testing how the relevance of auditory stimulus to the BCI selection could impact or influence the performance of the BCI. The Directional word set's meaning corresponded to the audio-spatial features of the stimulus and related to the task of moving an on-screen icon in the related direction. The semantic relevance was tested through the use of a control group of Non-Directional, semantically irrelevant words. The performance between the stimuli sets was very similar, therefore, the results of this study do not suggest relevance of spoken word stimuli has influence on the performance of the system.

Participants rated the two sets of spoken word stimuli in the aBCI equivalently on ratings of workload (NASA_TLX) and system usability (SUS) and typically voiced no difference in the difficulty between word sets. The participant's task is simply to count the number of target stimuli as they are presented. Comments from study participants completing the BCI tasks, word recognition tasks and pilot trials indicate that in the BCI task audio-spatial cues were primarily used. The auditory cues and semantics of the words appear to be ignored and not critical in the task when stimuli are clearly spatially separated. However, in the Word Recognition task experiment the influence of spatial separation was found to supports the notion that this cue is valuable in an oddball paradigm.

In the WR experiment, RT differences between word sets was found to be significant, within conditions that had spatial separation. In both the BCI and Dynamics condition the Non-Direction words exhibited faster reaction times, indicating that processing of relevant stimuli may be confounding in the WR experiment. In the NoCues condition RT was not significantly different between word sets, however, the only word that held semantic/spatial relevance was the word 'front', since all stimuli came from the front speaker. The word 'front' exhibited the fastest reaction time of all Direction and Non-Direction stimuli for the NoCues condition (Figure 5.5), suggesting that congruence in semantics and spatial location was able to improve word recognition. One hypothesis on why the BCI and Dynamic conditions show improved RT in Non-Direction stimuli may be that the Directional non-target words more effectively distracted participants in recognizing a directional target word. This could be further explored by utilizing Directional target words presented with Non-Directional, non-target stimuli. This suggest that in the BCI paradigm competeing influences from semantic and spatial relevance may have played a role in BCI performance.

Semantic meaning inherent in the stimulus provides intuitive association to the BCI user and may eliminate the need to memorize association of stimulus and BCI item selection. Auditory only BCI are likely most beneficial to users that couldn't use visual reference, so use of such stimuli is critical to realize a useful tool for many LIS patients.

6.1.1.2 Spatial

While several past studies have highlighted the benefit of spatial cues (Schreuder et al., 2011; Käthner et al., 2013; Schreuder et al., 2010; Ferracuti et al., 2013), a recent auditory BCI spelling system did not find inclusion of spatial cues in stimuli advantageous (Baykara et al., 2016). In the BCI experiment we did not find significant performance differences between lateralized or centralized groups of stimuli. While minor differences were seen in performance between different auditory locations in some participant, other potentially influencing factors were identified. Spatial separation of auditory stimuli may lend to variations in aBCI performance that relate to hearing thresholds, handedness or other lateralized perception preferences.

The total effect of spatial separation in word recognition is explored through comparison of the BCI and NoCues condition in the Word Recognition experiments in 5.3.1. Reaction time differences between the BCI and NoCues conditions suggest audio-spatial features of stimuli allows much easier identification of the target stimulus. Even in the confounding circumstance of the Dynamic-WR condition, having spatial cues in the stimuli improved RT and accuracy over the NoCues condition.

In a past aBCI paradigm using spoken word stimuli the location of specific stimuli was not held constant as in this aBCI task (Ferracuti et al., 2013), while in past aBCI using tone based stimuli the sound source locations were held constant. Faster RT in the BCI condition over the Dynamic condition suggest optimal performance is likely to result from a constant stimulus sound source location. Although confirmation of these anticipated effect should be confirmed in an aBCI experiment this behavioral result should be considered by BCI researchers.

The word 'right' had the highest accuracy by a noticeable margin (mean = 70.8%). The next highest was 'while' at 66.8%. In the Non-Direction word condition, the right side presented word 'doubt' had a fairly high accuracy as well (66.02%). It may be that words played in the right ear might have some additional salience due to the laterality of language processing in the left hemisphere. Right handed participants may have a preference for stimuli originating from their right side. The words 'right' and 'doubt' both had the shortest duration in their respective stimuli groups, which may influenced their recognition in some way.

'Joy' and 'front' are the least successful targets regardless of distractor. The front speaker acting as the target direction may be the most susceptible to distraction due to participants having to resist head orienting response encouraged by distractor stimuli. Also, front presented stimuli may be the most easy to perceive and so less effort goes into attending to that sound. Reduced effort or focus may have had detrimental effects on EEG resulting from front presented stimulus targets. In the WR-BCI condition the 'front' stimulus target exhibited the fastest median RT for all the Direction words. This contradictory result may highlight the variable influence of stimuli presentation on aBCI accuracy and RT measures. Factors enhancing EEG signals to be used by the BCI may not always improve reaction time measures in a similar task, and vice versa.

Investigating the stimulus specific RT measurs showed no consistent difference in lateralized vs. centralized stimulus except in the Non-Directional word set. The words 'care' and 'doubt' (lateralized) showed faster reaction times than 'while' and 'joy' (centralized) in the BCI condition. However, this was also true in the NoCues condition suggesting that the acoustics of words themselves may be the source of this effect and not the spatial separation present in the BCI condition.

6.1.2 ERP findings and BCI performance

A positive peak near 200ms was consistently present across many participants and stimuli. A negative peak near 100ms was also often present as well. The timing variation in peak negativity around 100ms may partly be due to the difference in attack and voice onset time inherent in the different constant sounds that begin each spoken word stimulus. See Appendix A.1 for more detail on the rise in intensity of each stimulus noted by the yellow intensity curves in the spectrogram of each auditory stimulus. Topographic maps of activity showed large positive and negative deflection in centralized electrodes over the ERP windows of stimulus presentation.

The ERP plots of stimuli and participant averages showed that a consistent difference in target and non-target stimuli responses occurred between 200 and 300ms. These differences persisted out to 400ms in some participants. For example, participant aB18 exhibited a more negative deflection around 300ms in the target EEG response as compared to non-target EEG responses in nearly all of the spoken word stimuli. In some participants more pronounced differences in target and non-target response could be identified in the 100 to 200ms range. When comparing the ERP signals elicited by spoken word vs beep stimuli in an auditory oddball presentation it has been found that patterns are far less consistent between subjects (Hill et al., 2014, 2004).

The notable ERP features described here do not fit the classic definition of P300 ERP. In the most similar paradigms to those used in the present study, it is not the P3 signal that differentiates target and non-target ERP traces (Halder et al., 2013).

In a few participants, including aB07 and aB19, a central negativity persisted much later into the ERP window than other participants. This difference did not seem to coincide with poor performance as these two participants exhibited better performance than most.

Mean accuracy was consistent in across-session accuracy and within session cross validation 10-fold. Different amounts of training data. lending to the generalizable use of training data using the method of this particular system. This also indicates that a small amount of training data proved good performance but that a full sessions worth would result in some improvements to % accuracy.

Means of BCI accuracy were not statistically improved across the 2 sessions, however, the variance in session 2 data suggests some improvement in participants performing poorly in the first session. It is expected that listening to spoken word stimuli is so natural that participants acclimate to the task quickly and reach ceiling performance with little BCI experience. It's also possible that this study included far too little experience with the system to see improvements over sesssions. In Jeremy Hill's study of word and tone stimuli, two sessions was not sufficient to foster improvements in performance (Hill et al., 2014). In a comparison of visual and auditory BCI, researchers found that auditory BCI accuracy was statistically less than visual BCI until the 11th session (Klobassa et al., 2009). Five sessions of training was able to increase an average aBCI ITR from 0.17 bits/min to 3.08 bits/min for five motor impaired end-users (Halder et al., 2016).

6.1.3 Questionnaires were accurate predictors of offline performance

More frequent selection of Temporal demand in NASA-TLX sources of workload weighting comparisons likely stemmed from perception that word stimuli presentation was at a rapid rate. This varied between participants. While most participants felt it was a comfortable pace, some may have found the rate somewhat rapid and might have hindered performance. Performance and Frustration were often rated highly and it is suspected that the poor online classifier performance likely influenced this with many participants.

Every participant's change in motivation between sessions was opposite in sign to their change in aBCI percent accuracy. Relative motivation between sessions exhibited a statistical relationship with BCI performance but this relationship was counterintuitive, indicating that decreased motivation increased BCI performance. Past reports on motivation showed correlations between P300 amplitude and motivation (Käthner et al., 2013).

Participants gave great insight through the experiment development and during the study. It is not only critical to have BCI user involvement for clinical applications but during system development so that proper prioritization can be given to design aspects influencing performance and usability. Utilizing questionnaires, as were employed in the current study, will allow BCI system developers to compare systems of varying design in terms of ergonomics and user preference.

6.1.4 Decoding Approach

Three EEG classification techniques were evaluated offline with the aBCI training data collected during the experiment. The SWLDA, and RSLDA decoding approaches were of primary focus for offline analysis and RSLDA proved to be consistently superior. SWLDA did provide lower accuracy in nearly every analysis, including those in across session accuracy.

SWLDA likely suffers from overfitting, where model weights reflect all the variance present between target and non-target classes and may include non robust features. Such a model may not hold true with data collected under slightly different circumstances, making it less generalizable. The regularization employed in the RLDA or RSLDA approaches more appropriately weights small differences in the target/non-target groupings. Regularization penalizes redundant terms that cause high covariance in the many EEG signals.

RSLDA may lack over fitting but may also require less data than the SWLDA approach. For smaller training sets it may be that SWLDA would do worse, as it seemed to in the 2-fold within session full trial offline accuracy measures as compared to the cross session accuracy.

The RSLDA method utilizes four models in this study, one for each stimuli in a condition, using 1/4th the data used for the RLDA. If data for a given stimulus is dramatically different than other stimuli the RSLDA method might better classify each sound as target or non-target. If target and non-target EEG traces are different in similar ways, regardless of the stimulus representing both classes, than the increased data used to fit the single model in RLDA may yield improved results. RSLDA yielded improved or equivalent performance to RLDA classification for every session, condition and subject. For more diverse spoken word stimuli RSLDA is likely to yield improved performance over RSLDA.

A dynamic stopping feature is not likely to benefit the performance of the current system. ITR and accuracy calculations by sequence indicated that accuracy below 70% would result from fewer than fifteen sequences for many participants. For the best performing participants only one or two sequences less than 15 resulted in higher ITR. With increased training it is possible users of the current aBCI may improve accuracy and general performance to allow dynamic stopping to aid in increasing ITR of the system.

6.1.5 Word Recognition

RT in the word recognition task significantly correlated with BCI performance. This type of behavioral test may be a reliable indicator of auditory-oddball-task BCI performance. Analysis of RT and button press Accuracy metrics did show effects of spatial separation, while evaluation of this in the aBCI paradigm was not sensitive enough to highlight differences.

Number of sequences, the order of word set and the word set itself were all found to be significant predictors of RT. This highlights the sensitive nature and powerful metric that button-press reaction time represents. This research tool, primarily utilized in Psychology, may serve as a useful performance indicator in BCI development in both research and clinical application.

Faster reaction times were seen in the Non-Direction words when spatial cues were used (BCI and Dynamic conditions). The Direction word set may have provided more distracting non-target stimuli than in the Non-Directional word set. When spatial cues were removed (NoCues), the Directional word produced faster reaction times although the differences between word set was not significant. This effect of semantic relevance was not uncovered by the aBCI experiment but indication from the WR experiment suggested that differences exist between the perception of spoken words with and without contextual relevance.

Many participants commented that the WR task was difficult and that they were challenged in at least some of the conditions. Most participants were able to identify instances where they had missed a target presentation and which conditions, word sets, or words they did better or worse on. While many participants found various aspects of the task difficult performance was homogenous overall. Participant 12 demonstrated consistent outlier performance in accuracy and RT for all conditions. Participant 12 admitted to having an attention deficit diagnosis after completing the study which is likely to explain this result. Participant 10 exhibited much faster reaction time averages in all conditions than the other participants. This participant mentioned just having come to the lab after exercising so may have been in a heightened physical and mental state.

While correlation in performance of the current aBCI and behavioral experiments were found, this may not apply in all cases of behavioral and BCI paradigms. It is likely other auditory oddball paradigms will see similar correlation, however, many parameters such as number of unique stimuli, presentation rate, and diversity in acoustic profiles of the stimuli may have large influences on aBCI and button press results that don't equate. Additional comparison of behavioral indicators and BCI performance should be accomplished to further understand this relationship in the wide array of BCI system configurations already present in the research field.

Future development of BCI presentation schema should include a behavioral correlate using RT and accuracy as metrics. Slower rate of stimulus presentation and changes in spatial separation suggested in the aBCI should be first optimized through behavioral experiments that can be completed with far less complication and more precise results.

6.1.6 Slower presentation rate

The presentation rate used in the WR was rapid enough to challenge participants and tease apart the differences in nearly all tested parameters. This rate was likely a detriment to the PacGame aBCI performance. In the WR-BCI task, two of the 15 participants were able to achieve 100% accuracy with the button press. The average WR-BCI accuracy for all 15 participants was 93.90%. While this is a high accuracy some participants likely missed recognizing the target stimulus multiple times. The presentation of stimulus in an aBCI paradigm should not be so challenging that target presentation could be missed.

Faster presentation rates likely require aBCI users to utilize basic sensory processing to discriminate and identify target stimuli. Slower presentation rates might allow more cognitive processing to occur before the next stimulus is presented allowing more complex neural processing to be used in identification of target stimuli. In the present study it is expected that participants primarily attended to audio-spatial cues of the stimuli. The processing of interaural time and level differences (ITD, ILD) begin as early in the auditory pathway as the superior olivary complex. Identification of acoustic characteristics defining the spoken words will be processed much later in auditory cortex. Semantics of the words would be accomplished even later in areas within and outside of the temporal lobe.

The rate of auditory presentation is likely to influence EEG features in terms of timing and morphology. In past research it has been found that the P300 amplitude does not predict performance Halder et al. (2013) but can be correlated with earlier and later ERP features. In this study the largest difference in target and non-target auditory evoked potentials (AEP) occurred well after 300ms. The later ERP features may result from the much slower presentation rate of 960 SOA. In contrast, earlier (100ms, 200-300ms) ERP features showed differences in target/non-target averages in the present study. In the present study the SOA was 400ms. Similar target/non-target ERP features were found in Höhne et al. (2012)where tones and syllable stimuli were used. The rate of presentation in this study was 135ms SOA more closely matching the present study. Various timing in auditory ERP differences can be found as can stimuli presentation rates.

It has been documented that presentation rate of visual stimuli influences timing of ERP features Krusienski et al. (2006). The influence of the SOA and number of distractors has also been investigated thoroughly by Gonsalvez and Polich (2002). A slower auditory presentation rate was found to improve BCI performance in Käthner et al. (2013). This study utilized tones in noise so additional investigation into optimal presentation rates with spoken word stimuli should be considered. With new paradigms and/or stimuli these parameters could be optimized through behavioral experiments like the WR task.

6.2 Future Directions

6.2.1 Stimulus and Presentation

While presentation rate using spoken word stimuli has yet to be optimized, a number of other parameters may also benefit from optimization. The duration, intensity and pitch (or gender) of the stimuli are very difficult to equate between different spoken words but standardizing these characteristics is likely to aid BCI performance. However, optimizing discriminability of stimuli is also important and could be accomplished with participant feedback as was done in Simon et al. (2014) but a behavioral measure like RT may provide more useful quantitative metrics. Use of multiple talkers or other highly discriminable acoustic features may prove ideal for aBCI use.

Use of synthesized words would allow for greater control of most acoustic characteristics of the stimuli and would be a likely solution for a flexible clinical tool. Synthesized stimuli would allow generation of any word desired by the BCI user without needing a produce a processed, standardized, high quality recording. Acoustics of the newly generated word could be processed in a more automated way meet any standards deemed beneficial.

The number of unique stimuli in a set could also be increased from four, in attempts to increase ITR, improve ERP signals due to more rarity in the target stimulus and provide more options in the BCI protocol.

aBCI research has primarily used front field presentation, which is well known to provide optimal spatial discrimination. All stimuli presented in the front field and/or through headphones. Headphones represent a more likely clinical scenario and front field would be a better comparison of more published aBCI systemsKäthner et al. (2013); Schreuder et al. (2010). Headphone presented stimulus should utilize ILD and ITD as well as HRTF Ferracuti et al. (2013) information to optimize the perceived spatial separation of auditory stimuli. Pneumatic Headphones would present stimuli with minimal influence on EEG signals, provide a quiet environment, and ensure spatial cues would be independent of head position.

Emotional words 'joy' and 'doubt' appeared to be very salient to participants in the present study. Influencing Emotional characteristics could be tested in a behavioral study and confirmed in an aBCI protocol. Familiar faces are used in visual spellers, so it may be that more emotion provoking spoken words would enhance aBCI performance.

6.2.2 BCI system components

Functional system improvements may also benefit a future auditory BCI design. RLDA was found to be effective but test of online performance in future studies would provide additional support for RLDA use in clinical aBCI. Continued comparison of RSLDA, RLDA and SVM decoding approaches may elucidate an optimal method for spoken word stimulus. In Ferracuti et al. (2013) SVM was utilized and although a different measure of performance was reported, similar number of participants were able to achieve >70% full-trial accuracy.

Word stimuli elicited fairly consistent ERP morphologies across word stimuli. Classification models capable of differentiating target vs. non-target regardless of the target stimulus may be capable of classifying words that were not included in classifier training. If classifiers could be developed that would extend to untrained word stimuli a BCI for word selection would be drastically more flexible than a system that required training of all possible stimulus target options. This would be a major achievement in the development of the BCI system proposed in Chapter 1.

6.2.3 Study Design

In past aBCI research, multiple sessions showed that auditory BCI could reach the performance levels of visual BCI techniques. Multiple sessions using an optimized auditory only spoken word stimuli BCI could highlight benefits and/or challenges inherent in using this class of BCI in a daily use clinical setting.

A Purely auditory system, with only a fixation cross as was used in the WR task. A similar presentation scheme with a question-answer paradigm may provide a means of engaging participants in a purely auditory BCI with real-life communication context. By presenting a question sentence to the participants and providing a number of possible answer words, the BCI user could select an internally generated answer. This may be somewhat analogous to the free spelling mode used in some BCI speller research studies (Simon et al., 2014; Kleih et al., 2015).

6.3 Clinical Acceptance and Translational Research

Entering the realm of translational research requires several considerations beyond that of basic science research. Namely, the end users of the developing technology should be consulted to ensure successful acceptance and satisfaction of the technology's intended functionality. If the BCI system developed through basic science research is highly functional in a laboratory environment but does not meet the requirements of its intended benefactors, then there is still a great deal of work to do.

Using a BCI for communication requires a highly flexible system that can work, reliably, in everyday living environments and must be useable by the user and their caregivers. Systems typically run by engineering and neuroscience researchers are likely not well suited for patients and clinicians. Researchers in the field have become aware of the importance of clinical acceptance of BCI systems and have begun to push for inclusion of clinical researchers to partner with engineering and basic science researchers to provide relevant clinical direction to research study designs Kubler et al. (2006); Peters et al. (2016).

Neuroscience, engineering and clinical collaboration in the BCI research field is a necessary step in moving these critical scientific discoveries out of the labs and into the real world where they can make a real difference in someone's life. The user of a BCI may use the device for all of their activities of daily living Suyama (2016). It may be the only way for them to communicate and interact with the world around them. How they do this must be highly personalized in order for the system to be utilized and for it to achieve true communication replacement.

Clinician and end user feedback on design and development will continue to shape clinical BCI research. While advancements in software and hardware of the systems have been achieved, the proof of concepts have occurred in well controlled environments, do not completely reflective of real-world use. Signal acquisition, BCI paradigm and user interface will need to be customized and optimized of for every clinical BCI user.

6.3.1 Custom Signal Acquisition

The technology behind BCI use has been developed and thoroughly tested in highly controlled environments. While BCI research technology has proven to be feasible for the intended user population, the systems tested would not be suitable for daily use by an impaired individual. Experienced EEG technicians and expensive research grade EEG systems are required for high quality brain signal acquisition.

Most studies, like this one, employ rigorous electrical noise canceling environments and signal processing techniques to ensure clear and reliable EEG or EcOG signals. EEG system manufactures have had some success creating acquisition systems that can provide the signal quality needed for BCI use in mobile, real-life scenarios. Some researchers have even taken it upon themselves to develop tools that will take the field much closer to acceptance by the end user (Debener et al., 2015; De Vos et al., 2014).

Hardware developments like these may allow the acquisition of brain signals for BCI control in real life environments a possibility. This has been the aim of EEG technology advancement for many years. Dry electrodes and mobile EEG acquisition provide the convenience likely to be required for any clinical daily use of a BCI to be effectively implemented. While an ideal clinical system of hardware is not readily available, the basic technology development needed is nearing the level where such a device is feasible. Cost of these devices is also a necessary consideration and some researchers and open source EEG enthusiasts are making large strides to make open access inexpensive hardware and software for BCI development, (**i.e** OpenBCI).

6.3.2 BCI approach (Paradigm and classifier)

Very complex algorithm development and mathematical model training should be optimized for specific BCI system users. Various EEG signals may be more prominent in certain BCI users so this may challenge a BCI design to utilize different types of EEG signals depending on the users neurophysiology, capacity for attention and working memory as well as other cognitive and executive functions. Machine learning techniques do attempt to provide this flexibility, however simple linear models often provide maximal performance in most BCI research.

Clinicians must be trained to evaluate a patients ability to use various BCI paradigms and classification approaches. By developing a decision tree pathway and efficient BCI success evaluations this may be possible .

It has been documented that inter-subject variability in EEG representing speech sounds does not allow for a grand averaged model between subjects O'Sullivan et al. (2014). All BCI studies utilize subject and session specific training so extending this to target stimulus specific model generation is not a burdensome step. This concept was accomplished in the present study with minimal effort. The results of this study suggest stimuli specific classifier models may be beneficial when using spoken word stimuli.

6.3.3 User Interface

This study utilized a custom user interface that presented a motiving game to the aBCI user. User interfaces (UI) should be highly customized to the users preferred communication schemes. Influences from optimal BCI paradigms and stimulus choices may restrict UI configurations but effective use of the BCI should be highly prioritized based on the user's specific need.

The proposed clinical BCI system discussed in Chapter 1 was suggested to incorporate the use of an existing AAC device that would work best for the candidate BCI user. Utilization of spoken word stimulus to directly select full word or phrase communication selections would greatly enhance BCI-Utility, a metric proposed in a BCI-based AAC review (Thompson et al., 2013). This metric, not yet well defined by the BCI research field, would highlight the improvements in aBCI performance and communication output achieved by the proposed paradigm.

References

G. Bauer, F. Gerstenbrand, and E. Rumpl. Varieties of the locked-in syndrome. *Journal of Neurology*, 221(2):77–91, 1979. doi: 10.1007/BF00313105. URL `http://dx.doi.org.ezproxy.bu.edu/10.1007/BF00313105http://www.springerlink.com.ezproxy.bu.edu/content/k6134772574vp427/`.

E. Baykara, C. A. Ruf, C. Fioravanti, I. Käthner, N. Simon, S. C. Kleih, A. Kübler, and S. Halder. Effects of training and motivation on auditory P300 brain-computer interface performance. *Clinical Neurophysiology*, 127(1), 2016. ISSN 18728952. doi: 10.1016/j.clinph.2015.04.054.

Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Michael Schroeder, John Williamson, Roderick Murray-Smith, and Klaus-Robert Müller. The Berlin Brain-Computer Interface presents the novel mental typewriter Hex-o-Spell. *3rd International BCI Workshop and Training Course, Graz, 2006, Graz, Austria(2006).*, pages 2–3, 2006. doi: 10.1.1.66.7603. URL `http://eprints.pascal-network.org/archive/00002463/{%}5Cnhttp://eprints.nuim.ie/1786/`.

Benjamin Blankertz, Matthias Krauledat, Guido Dornhege, John Williamson, Roderick Murray-Smith, and Klaus-Robert Mßüller. A Note on Brain Actuated Spelling with the Berlin Brain-Computer Interface. *Universal Access in Human-Computer Interaction. Ambient Interaction*, 4555/2007:759–768, 2007. doi: 10.1007/978-3-540-73281-5_83. URL `http://dx.doi.org/10.1007/978-3-540-73281-5{_}83http://www.springerlink.com/content/03hr3j4812561652/`.

Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus Robert Müller. Single-trial analysis and classification of ERP components - A tutorial. *NeuroImage*, 56(2): 814–825, 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.06.048. URL `http://dx.doi.org/10.1016/j.neuroimage.2010.06.048`.

George a. Buzzell, Daniel M. Roberts, Carryl L. Baldwin, and Craig G. McDonald. An electrophysiological correlate of conflict processing in an auditory spatial Stroop task: The effect of individual differences in navigational style. *International Journal of Psychophysiology*, 90(2):265–271, 2013. ISSN 01678760. doi: 10.1016/j.ijpsycho.2013.08.008. URL `http://dx.doi.org/10.1016/j.ijpsycho.2013.08.008`.

Zhenyu Cai, Shoji Makino, Takeshi Yamada, Tomasz M. Rutkowski, and Information Engineering. Spatial Auditory BCI Paradigm Utilizing N200 and P300 Responses. *arXiv preprint arXiv: ...*, pages 1–22, 2012. URL `http://arxiv.org/abs/1210.2944`.

Jerry S. Carlson, C. Mark Jensen, and Keith F. Widaman. Reaction time, intelligence, and attention. *Intelligence*, 7(4):329–344, 1983. ISSN 01602896. doi: 10.1016/0160-2896(83)90008-9.

Hubert Cecotti. Spelling with non-invasive Brain-Computer Interfaces - Current and future trends, 2011. ISSN 09284257.

Maarten De Vos, Katharina Gandras, and Stefan Debener. Towards a truly mobile auditory brain-computer interface: Exploring the P300 to take away. *International Journal of Psychophysiology*, 91(1), 2014. ISSN 01678760. doi: 10.1016/j.ijpsycho.2013.08.010.

Stefan Debener, Reiner Emkes, Maarten De Vos, and Martin Bleichner. Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific reports*, 5, 2015. ISSN 2045-2322. doi: 10.1038/srep16743.

Emanuel Donchin, Kevin M. Spencer, and Ranjith Wijesinghe. The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engi-*

neering, 8(2):174–179, 2000. ISSN 1063-6528. URL `http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=847808`.

L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988. ISSN 00134694. doi: 10.1016/0013-4694(88)90149-6.

Francesco Ferracuti, Alessandro Freddi, Sabrina Iarlori, Sauro Longhi, and Paolo Peretti. Auditory paradigm for a P300 BCI system using spatial hearing. In *IEEE International Conference on Intelligent Robots and Systems*, 2013. ISBN 9781467363587. doi: 10.1109/IROS.2013.6696453.

A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübßler. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, 46 (3):617–25, may 2009. ISSN 1540-5958. doi: 10.1111/j.1469-8986.2008.00783.x. URL `http://www.ncbi.nlm.nih.gov/pubmed/19170946`.

Craig L Gonsalvez and John Polich. P300 amplitude is determined by target-to-target interval. *Psychophysiology*, 39(3):388–396, 2002. ISSN 0048-5772. doi: 10.1017/S0048577201393137.

Frank H. Guenther and Jonathan S. Brumberg. Unchained mind. *ASHA Leader*, 18(1), 2013. ISSN 10859586.

Christoph Guger, Shahab Daban, Eric Sellers, Clemens Holzner, Gunther Krausz, Roberta Carabalona, Furio Gramatica, and Guenter Edlinger. How many people are able to control a P300-based brain-computer interface (BCI)? *Neuroscience Letters*, 462(1):94–98, 2009. ISSN 03043940. doi: 10.1016/j.neulet.2009.06.045.

S. Halder, M. Rea, R. Andreoni, F. Nijboer, E. M. Hammer, S. C. Kleih, N. Birbaumer, and a. Kübler. An auditory oddball brain-computer interface for binary choices. *Clinical Neurophysiology*, 121(4):516–523, 2010. ISSN 13882457. doi: 10.1016/j.clinph.2009.11.087. URL `http://dx.doi.org/10.1016/j.clinph.2009.11.087`.

S. Halder, I. Käthner, and A. Kübler. Training leads to increased auditory brain-computer interface performance of end-users with motor impairments. *Clinical Neurophysiology*, 127(2), 2016. ISSN 18728952. doi: 10.1016/j.clinph.2015.08.007.

Sebastian Halder, Eva Maria Hammer, Sonja Claudia Kleih, Martin Bogdan, Wolfgang Rosenstiel, Niels Birbaumer, and Andrea Kübler. Prediction of Auditory and Visual P300 Brain-Computer Interface Aptitude. *PLoS ONE*, 8(2), 2013. ISSN 19326203. doi: 10.1371/journal.pone.0053513.

Kentaro Hayashi, Yoko Mochizuki, Ryoko Takeuchi, Toshio Shimizu, Masahiro Nagao, Kazuhiko Watabe, Nobutaka Arai, Kiyomitsu Oyanagi, Osamu Onodera, Masaharu Hayashi, Hitoshi Takahashi, Akiyoshi Kakita, and Eiji Isozaki. Clinicopathological characteristics of patients with amyotrophic lateral sclerosis resulting in a totally locked-in state (communication Stage V). *Acta Neuropathologica Communications*, 4(1):107, 2016. ISSN 2051-5960. doi: 10.1186/s40478-016-0379-3. URL `http://dx.doi.org/10.1186/s40478-016-0379-3`.

H. Higashi, T. M. Rutkowski, Y. Washizawa, A. Cichocki, and T. Tanaka. EEG auditory steady state responses classification for the novel BCI. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2011, pages 4576–4579. IEEE, aug 2011. ISBN 978-1-4577-1589-1. doi: 10.1109/IEMBS.2011.6091133. URL `http://www.ncbi.nlm.nih.gov/pubmed/22255356`.

N J Hill and B Schölkopf. An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli. *Journal of neural engineering*, 9(2):026011, apr 2012. ISSN 1741-2552. doi: 10.1088/1741-2560/9/2/026011. URL `http://iopscience.iop.org/1741-2552/9/2/026011/pdf/1741-2552{_}9{_}2{_}026011.pdf`.

N. Jeremy Hill, Aisha Moinuddin, Ann-Katrin Häser, Stephan Kienzle, and Gerwin Schalk. Communication and Control by Listening: Toward Optimal Design of a Two-Class Auditory Streaming Brain-Computer Interface. *Frontiers in Neuroscience*, 6(December):1–14, 2012.

ISSN 1662-4548. doi: 10.3389/fnins.2012.00181. URL http://www.frontiersin.org/Neuroprosthetics/10.3389/fnins.2012.00181/abstract.

N Jeremy Hill, Erin Ricci, Sameah Haider, Lynn M McCane, Susan Heckman, Jonathan R Wolpaw, and Theresa M Vaughan. A practical, intuitive brain-computer interface for communicating 'yes' or 'no' by listening. *Journal of neural engineering*, 11:035003, 2014. ISSN 1741-2552. doi: 10.1088/1741-2560/11/3/035003. URL http://www.ncbi.nlm.nih.gov/pubmed/24838278.

N.J. Hill, Thomas Navin Lal, K. Bierig, N. Birbaumer, and Bernhard Schölkopf. An Auditory Paradigm for Brain–Computer Interfaces. 2004. doi: 10.1.1.136.7814. URL http://eprints.pascal-network.org/archive/00000407/.

Johannes Hoehne and Michael Tangermann. Towards user-friendly spelling with an auditory brain-computer interface: the CharStreamer paradigm. *PloS one*, 9(6):e98322, jan 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0098322. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4041754{&}tool=pmcentrez{&}rendertype=abstract.

Johannes Hohne, Benjamin Blankertz, Klaus Robert Muller, and Daniel Bartz. Mean shrinkage improves the classification of ERP signals by exploiting additional label information. *Proceedings - 2014 International Workshop on Pattern Recognition in Neuroimaging, PRNI 2014*, (3):2–5, 2014. doi: 10.1109/PRNI.2014.6858523.

Petar Horki, Daniela S Klobassa, Christoph Pokorny, and Gernot R Müller-Putz. Evaluation of healthy EEG responses for spelling through listener-assisted scanning. *IEEE journal of biomedical and health informatics*, 19(1):29–36, jan 2015. ISSN 2168-2208. doi: 10.1109/JBHI.2014.2328494. URL http://www.ncbi.nlm.nih.gov/pubmed/25014972.

Johannes Höhne, Konrad Krenzlin, Sven Dähne, and Michael Tangermann. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *Journal of Neural Engineering*, 9(4):045003, 2012. ISSN 1741-2552. doi: 10.1088/1741-2560/9/4/045003. URL http://www.ncbi.nlm.nih.gov/pubmed/22831919.

Johannes Höhne, Daniel Bartz, Martin N. Hebart, Klaus Robert Müller, and Benjamin Blankertz. Analyzing neuroimaging data with subclasses: A shrinkage approach. *NeuroImage*, 124:740–751, 2016. ISSN 10959572. doi: 10.1016/j.neuroimage.2015.09.031.

Shin'ichiro Kanoh, Ko Ichiro Miyamoto, and Tatsuo Yoshinobu. A brain-computer interface (BCI) system based on auditory stream segregation. *Journal of Biomechanical Science and Engineering*, 5(1):32–40, 2010. ISSN 18809863. doi: 10.1299/jbse.5.32.

Sonja C Kleih and Andrea Kübler. Empathy, motivation, and P300 BCI performance. *Frontiers in human neuroscience*, 7(October):642, 2013. ISSN 1662-5161. doi: 10.3389/fnhum.2013.00642. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3797970{&}tool=pmcentrez{&}rendertype=abstract`.

Sonja C. Kleih, Andreas Herweg, Tobias Kaufmann, Pit Staiger-S??lzer, Natascha Gerstner, and Andrea K??bler. The WIN-speller: A new intuitive auditory brain-computer interface spelling application. *Frontiers in Neuroscience*, 9(OCT):1–11, 2015. ISSN 1662453X. doi: 10.3389/fnins.2015.00346.

D. S. Klobassa, T. M. Vaughan, P. Brunner, N. E. Schwartz, J. R. Wolpaw, C. Neuper, and E. W. Sellers. Toward a high-throughput auditory P300-based brain-computer interface. *Clinical Neurophysiology*, 120(7):1252–1261, 2009. ISSN 13882457. doi: 10.1016/j.clinph.2009.04.019.

D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. Toward enhanced P300 speller performance. *Journal of Neuroscience Methods*, 167(1):15–21, 2008. ISSN 01650270. doi: 10.1016/j.jneumeth.2007.07.017.

Dean J Krusienski, Eric W Sellers, François Cabestaing, Sabri Bayoudh, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. A comparison of classification techniques for the P300 Speller. *Journal of neural engineering*, 3(4):299–305, 2006. ISSN 1741-2560. doi: 10.1088/1741-2560/3/4/007.

Kubler, Mushahwar, Hochberg, and Donoghue. BCI Meeting 2005—Workshop on Clinical Issues and Applications. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14 (2):131–134, 2006. ISSN 1534-4320. doi: 10.1109/TNSRE.2006.875585. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1642752`.

A Kübler, B Kotchoubey, J Kaiser, J Wolpaw, and N Birbaumer. Brain–computer communication: Unlocking the locked in. *Psychological bulletin*, 127(3):358–375, 2001.

Ivo Käthner, Carolin a. Ruf, Emanuele Pasqualotto, Christoph Braun, Niels Birbaumer, and Sebastian Halder. A portable auditory P300 brain-computer interface with directional cues. *Clinical Neurophysiology*, 124(2):327–338, 2013. ISSN 13882457. doi: 10.1016/j.clinph.2012.08.006.

Andrea Kübler, Adrian Furdea, Sebastian Halder, Eva Maria Hammer, Femke Nijboer, and Boris Kotchoubey. A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Annals of the New York Academy of Sciences*, 1157: 90–100, 2009. ISSN 00778923. doi: 10.1111/j.1749-6632.2008.04122.x.

Rui Ma, Navid Aghasadeghi, Julian Jarzebowski, Timothy Bretl, and Todd P Coleman. A stochastic control approach to optimally designing hierarchical flash sets in P300 communication prostheses. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 20(1):102–12, jan 2012. ISSN 1558-0210. doi: 10.1109/TNSRE.2011.2179560. URL `http://www.ncbi.nlm.nih.gov/pubmed/22203722`.

Lynn M. McCane, Susan M. Heckman, Dennis J. McFarland, George Townsend, Joseph N. Mak, Eric W. Sellers, Debra Zeitlin, Laura M. Tenteromano, Jonathan R. Wolpaw, and Theresa M. Vaughan. P300-based brain-computer interface (BCI) event-related potentials (ERPs): People with amyotrophic lateral sclerosis (ALS) vs. age-matched controls. *Clinical Neurophysiology*, 126(11):2124–2131, 2015. ISSN 18728952. doi: 10.1016/j.clinph.2015.01.013.

Chet T Moritz and Eberhard E Fetz. Volitional control of single cortical neurons in a brain-machine interface. *Journal of neural engineering*, 8(2):025017, 2011. ISSN 1741-2560. doi: 10.1088/1741-2560/8/2/025017.

F. Nijboer, E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N. Birbaumer, and A. Kübler. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119(8):1909–1916, 2008a. ISSN 1388-2457. doi: DOI:10.1016/j.clinph.2008.03.034. URL `http://www.sciencedirect.com/science/article/B6VNP-4ST4634-1/2/35c7038cc19f1eebaf01759a430f0426http://www.sciencedirect.com/science/article/pii/s1388-2457(08)00239-3`.

Femke Nijboer, Adrian Furdea, Ingo Gunst, Jürgen Mellinger, D. J. McFarland, Niels Birbaumer, and A. K\\"ubler. An auditory brain-computer interface (BCI). *Journal of Neuroscience Methods*, 167(1):43–50, 2008b. ISSN 0165-0270. doi: DOI:10.1016/j.jneumeth.2007.02.009. URL `http://linkinghub.elsevier.com/retrieve/pii/S0165027007000726`.

J C Oliver-Rodríguez, Z Guan, and V S Johnston. Gender differences in late positive components evoked by human faces. *Psychophysiology*, 36:176–185, 1999. ISSN 0048-5772. doi: 10.1017/S0048577299971354.

Umut Orhan, Deniz Erdogmus, Brian Roark, Shalini Purwar, Kenneth E. Hild, Barry Oken, Hooman Nezamfar, and Melanie Fried-Oken. Fusion with language models improves spelling accuracy for ERP-based brain computer interface spellers. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 5774–5777, 2011. ISBN 9781424441211. doi: 10.1109/IEMBS.2011.6091429.

James a O'Sullivan, Alan J Power, Nima Mesgarani, Siddharth Rajaram, John J Foxe, Barbara G Shinn-Cunningham, Malcolm Slaney, Shihab a Shamma, and Edmund C Lalor. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral* *cortex (New York, N.Y. : 1991)*, pages 1–10, jan 2014. ISSN 1460-2199. doi: 10.1093/cercor/bht355. URL http://www.ncbi.nlm.nih.gov/pubmed/24429136.

Matthew T. G. Pain and Angela Hibbs. Sprint starts and the minimum auditory reaction time. *Journal of Sports Sciences*, 25(1):79–86, 2007. doi: 10.1080/02640410600718004. URL http://dx.doi.org/10.1080/02640410600718004. PMID: 17127583.

Betts Peters, Aimee Mooney, Barry Oken, and Melanie Fried-Oken. Soliciting BCI user experience feedback from people with severe speech and physical impairments. *Brain-Computer Interfaces*, 2621(April):1–12, 2016. ISSN 2326-263X. doi: 10.1080/2326263X.2015.1138056. URL http://www.tandfonline.com/doi/full/10.1080/2326263X.2015.1138056.

F. Piccione, F. Giorgi, P. Tonin, K. Priftis, S. Giove, S. Silvoni, G. Palmas, and F. Beverina. P300-based brain computer interface: reliability and performance in healthy and paralysed participants. *Clinical Neurophysiology*, 117(3):531–7, mar 2006. ISSN 1388-2457. doi: 10.1016/j.clinph.2005.07.024. URL http://www.sciencedirect.com.ezproxy.bu.edu/science/article/B6VNP-4J5T8DM-1/2/12836ec2953224897245c6694efa7717http://www.sciencedirect.com.ezproxy.bu.edu/science?{_}ob=ArticleURL{&}{_}udi=B6VNP-4J5T8DM-1{&}{_}user=489277{&}{_}rdoc=1{&}{_}fmt={&}{_}orig=search{&}{_}sort=d{&}{_}docanch.

F. Plum and J. B. Posner. The diagnosis of stupor and coma. *Contemporary Neurology Series*, 10:1–286, 1972. ISSN 0069-9446. URL http://www.ncbi.nlm.nih.gov/pubmed/4664014.

Nikki Pratt, Adrian Willoughby, and Diane Swick. Effects of working memory load on visual selective attention: behavioral and electrophysiological evidence. *Frontiers in human neuroscience*, 5(June):57, 2011. ISSN 1662-5161. doi: 10.3389/fnhum.2011.00057.

Ander Ramos-Murguialday, Doris Broetz, Massimiliano Rea, Leonhard Läer, Özge Yilmaz, Fabricio L. Brasil, Giulia Liberati, Marco R. Curado, Eliana Garcia-Cossio, Alexandros Vyziotis, Woosang Cho, Manuel Agostini, Ernesto Soares, Surjo Soekadar, Andrea Caria, Leonardo G.

Cohen, and Niels Birbaumer. Brain-machine interface in chronic stroke rehabilitation: A controlled study. *Annals of Neurology*, 74(1):100–108, 2013. ISSN 03645134. doi: 10.1002/ana.23879.

Gerwin Schalk and Eric C. Leuthardt. Brain-computer interfaces using electrocorticographic signals. *IEEE Reviews in Biomedical Engineering*, 4:140–154, 2011. ISSN 19373333. doi: 10.1109/RBME.2011.2172408.

Gerwin Schalk, Dennis J. McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R. Wolpaw. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004. ISSN 00189294. doi: 10.1109/TBME.2004.827072.

Reinhold Scherer, Gernot R. Müller, Christa Neuper, Bernhard Graimann, and Gert Pfurtscheller. An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate. *IEEE Transactions on Biomedical Engineering*, 51(6):979–984, 2004. ISSN 00189294. doi: 10.1109/TBME.2004.827062.

Martijn Schreuder, Benjamin Blankertz, and Michael Tangermann. A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue. *PLoS ONE*, 5(4), 2010. ISSN 19326203. doi: 10.1371/journal.pone.0009813.

Martijn Schreuder, Thomas Rost, and Michael Tangermann. Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. *Frontiers in Neuroscience*, 5(OCT):1–12, 2011. ISSN 16624548. doi: 10.3389/fnins.2011.00112.

Martijn Schreuder, Johannes Hoehne, Benjamin Blankertz, Stefan Haufe, Thorsten Dickhaus, and Michael Tangermann. Optimizing event-related potential based brain-computer interfaces: a systematic evaluation of dynamic stopping methods. *Journal of neural engineering*, 10(3):036025, 2013. ISSN 1741-2552. doi: 10.1088/1741-2560/10/3/036025. URL `http://www.ncbi.nlm.nih.gov/pubmed/23685458`.

Eric W. Sellers and Emanuel Donchin. A P300-based brain-computer interface: initial tests by ALS patients. *Clinical Neurophysiology*, 117(3):538–48, mar 2006. ISSN 1388-2457. doi: 10.1016/j.clinph.2005.06.027. URL `http://www.sciencedirect.com.libproxy.mit.edu/science/article/B6VNP-4J6WR2K-3/2/cd9c7c185c58d60bca5d97c565aebb18http://www.sciencedirect.com.libproxy.mit.edu/science?{_}ob=ArticleURL{&}{_}udi=B6VNP-4J6WR2K-3{&}{_}user=501045{&}{_}rdoc=1{&}{_}fmt={&}{_}orig=search{&}{_}sort=d{&}{_}doc`.

Nadine Simon, Ivo Käthner, Carolin A Ruf, Emanuele Pasqualotto, Andrea Kübler, and Sebastian Halder. An auditory multiclass brain-computer interface with natural stimuli: Usability evaluation with healthy participants and a motor impaired end user. *Frontiers in human neuroscience*, 8:1039, jan 2014. ISSN 1662-5161. doi: 10.3389/fnhum.2014.01039. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4288388{&}tool=pmcentrez{&}rendertype=abstract`.

Eimear Smith and Mark Delargy. Locked-in syndrome. *BMJ (Clinical research ed.)*, 330(7488):406–9, feb 2005. ISSN 1756-1833. doi: 10.1136/bmj.330.7488.406. URL `http://www.ncbi.nlm.nih.gov/pubmed/15718541http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC549115`.

Takayuki Suyama. A Network-Type Brain Machine Interface to Support Activities of Daily Living. (9):1930–1937, 2016. doi: 10.1587/transcom.2016SNI0002.

Kouji Takano, Tomoaki Komatsu, Naoki Hata, Yasoichi Nakajima, and Kenji Kansaku. Visual stimuli for the P300 brain-computer interface: A comparison of white/gray and green/blue flicker matrices. *Clinical Neurophysiology*, 120(8):1562–1566, 2009. ISSN 13882457. doi: 10.1016/j.clinph.2009.06.002.

David E. Thompson, Stefanie Blain-Moraes, and Jane E. Huggins. Performance assessment in brain-computer interface-based augmentative and alternative communication. *BioMedical Engineering Online*, 12(1), 5 2013. ISSN 1475-925X. doi: 10.1186/1475-925X-12-43.

Theresa M. Vaughan, Dennis J. McFarland, Gerwin Schalk, William A. Sarnacki, Dean J. Krusienski, Eric W. Sellers, and Jonathan R. Wolpaw. The wadsworth BCI research and development program: At home with BCI. In *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, volume 14, pages 229–233, 2006. ISBN 1534-4320. doi: 10.1109/TNSRE.2006.875577.

J J Vidal. Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2(1):157–180, 1973. doi: 10.1146/annurev.bb.02.060173.001105. URL `http://dx.doi.org/10.1146/annurev.bb.02.060173.001105`. PMID: 4583653.

J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller. EEG-based communication: Improved accuracy by response verification. *IEEE Transactions on Rehabilitation Engineering*, 6(3):326–333, 1998. ISSN 10636528. doi: 10.1109/86.712231.

J R Wolpaw, N Birbaumer, W J Heetderks, D J McFarland, P H Peckham, G Schalk, E Donchin, L a Quatrano, C J Robinson, and T M Vaughan. Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):164–173, 2000. ISSN 1063-6528. doi: 10.1109/TRE.2000.847807.

Jonathan R Wolpaw and Dennis J McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17849–54, dec 2004. ISSN 0027-8424. doi: 10.1073/pnas.0403504101. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=535103{&}tool=pmcentrez{&}rendertype=abstract`.

Jonathan R. Wolpaw, Dennis J. McFarland, Gregory W. Neat, and Catherine A. Forneris. An

EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3):252–259, 1991. ISSN 00134694. doi: 10.1016/0013-4694(91)90040-B.

Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain-computer interfaces for communication and control. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 113(6):767–91, 2002. ISSN 1388-2457. doi: 10.1016/S1388-2457(02)00057-3. URL `http://www.ncbi.nlm.nih.gov/pubmed/12048038`.

Appendix A

Appendix

A.1 Spoken Word Stimuli

For each stimulus utilized in the PacGame and Word Recognition experiments a graphic from Praat shows the detail of the recorded spoken word acoustic stimulus. The top graph is the sound waveform and bottom graph is the spectrogram of the .wav file. The x-axis for both graphs is time and the two graphs are horizontally aligned in time. The y-axis in the waveform is voltage and the spectrogram is in Hertz. The blue line in the spectrogram represents a pitch contour and the yellow line an intensity contour.

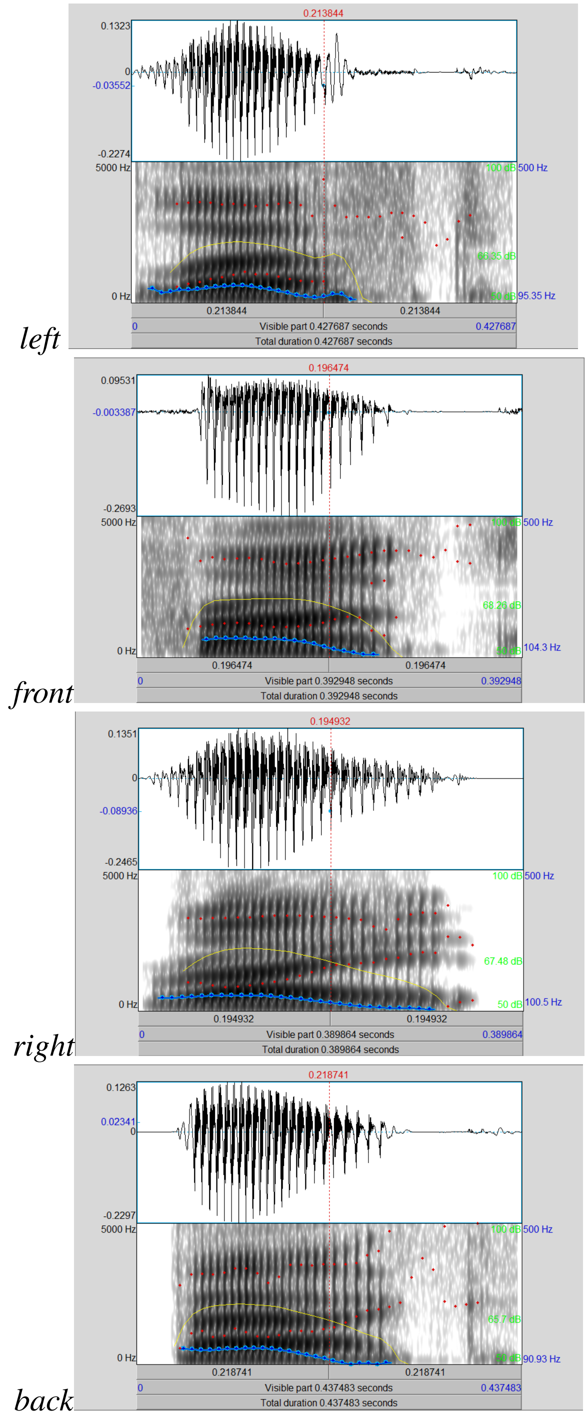


Figure A.1: Acoustic Graphics of the Directional Spoken Word Stimuli care

joy _

doubt

while

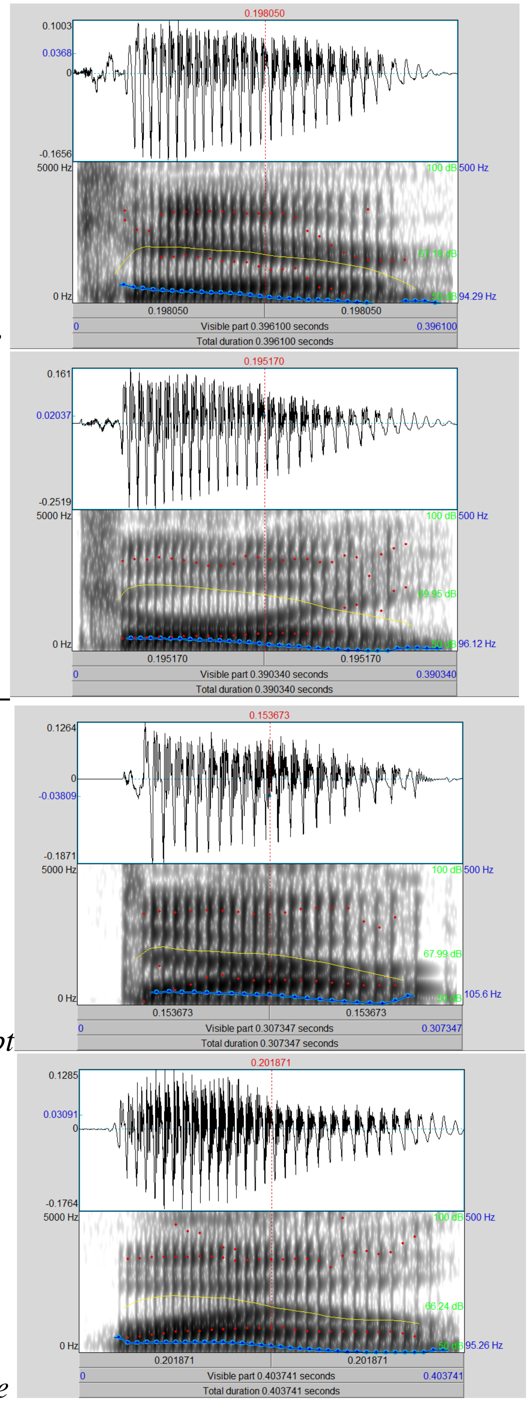


Figure A.2: Acoustic Graphics of the Non-Directional Spoken Word Stimuli

A.2 Questionnaire Materials

The information sheet filled out by each participant during the screening process is included below. Demographic information as well as questions concerning study eligibility are included.

Phone / Email Screen

____ Age ____ Gender ____ Handedness (right/left)

Please identify with one of the following ethnic categories (choose one):

□ Hispanic or Latino, □ Not Hispanic or Latino

Please indicate all of the following racial categories you most identify with (may check more than one):

□ American Indian or Alaska Native, □ Asian, Native Hawaiian or Other Pacific Islander

□ Black or African American, □ White

____ Native speaker of American English

____ Familiar with Language other than American English; if so, explain:

____ History of claustrophobia (explain procedure)

____ Injury to head/hands that could impede task performance

____ History of serious head injury; if so, explain:

____ History of neurological disorder(s) (including stuttering) or seizure disorder (including epilepsy); if so, explain:

EEG / BMI Compatibility

Do you have/have you had:	**YES**	**NO**	**IF YES, Please Explain**
History of Head Trauma	____	____	______________________________
Electrical or Magnetic Implant	____	____	______________________________
Cardiac Pacemaker	____	____	______________________________
Neurostimulator	____	____	______________________________
Implanted Pumps	____	____	______________________________

Last Updated: 6/19/2012

Figure A.3: Screening Sheet

The NASA_TLX is composed of two sections. The first is the weighting of each source of workload as accomplished by comparison of each. The sheet these weighting selections are tallied on and the sheets where participants indicate the ratings for each source of workload are included in Figure A.4 below.

Appendix D.

Subject ID: ____________ **Date:** ____________

SOURCES-OF-WORKLOAD TALLY SHEET		
Scale Title	*Tally*	*Weight*
MENTAL DEMAND		
PHYSICAL DEMAND		
TEMPORAL DEMAND		
PERFORMANCE		
EFFORT		
FRUSTRATION		

Total count = ____________

(NOTE - The total count is included as a check. If the total count is not equal to 15, then something has been miscounted. Also, no weight can have a value greater than 5.)

Appendix E.

Subject ID: ____________ **Task ID:** ____________

WEIGHTED RATING WORKSHEET			
Scale Title	*Weight*	*Raw Rating*	*Adjusted Rating* (Weight X Raw)
MENTAL DEMAND			
PHYSICAL DEMAND			
TEMPORAL DEMAND			
PERFORMANCE			
EFFORT			
FRUSTRATION			

Sum of "Adjusted Rating" Column = ________

WEIGHTED RATING =
[i.e., (Sum of Adjusted Ratings)/15]

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure A.4: NASA_TLX

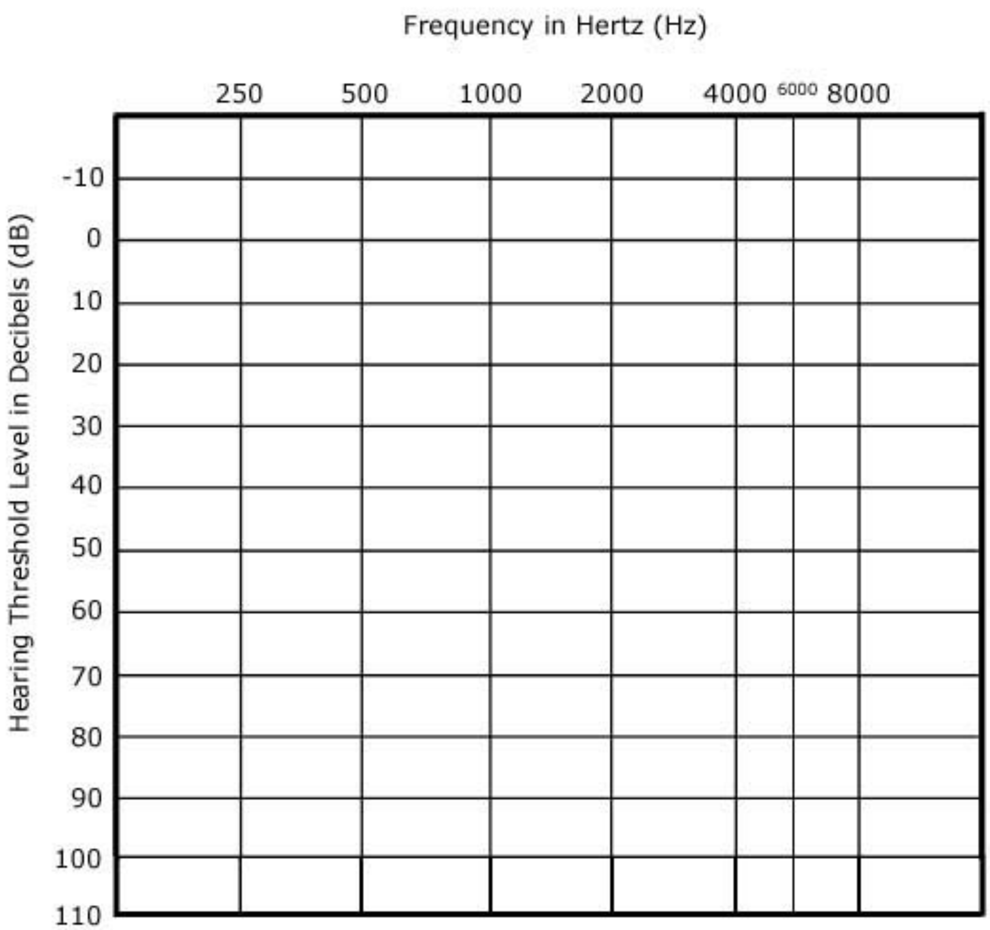
Subject ______________ Session: ______________

VAS - Rate your current mood and motivation (mark anywhere on the line you feel is appropriate)

Extremely Unmotivated Extremely Motivated

1 - - - - - - 2 - - - - - - 3 - - - - - - 4 - - - - - - 5 - - - - - - 6 - - - - - - 7 - - - - - - 8 - - - - - 9 - - - - - - 10

Bad mood Good Mood



VAS and SUS Questionnaire – SANLAB 2016

Figure A.5: VAS and Audiogram

Subject ______________ Session: ______________

SUS – Condition 1

1. I think that I would like to use this system frequently.

Strongly Disagree 1	2	3	4	Strongly Agree 5

2. I found the system unnecessarily complex.

Strongly Disagree 1	2	3	4	Strongly Agree 5

3. I thought the system was easy to use.

Strongly Disagree 1	2	3	4	Strongly Agree 5

4. I think that I would need the support of a technical person to be able to use this system.

Strongly Disagree 1	2	3	4	Strongly Agree 5

5. I found the various functions in this system were well integrated.

Strongly Disagree 1	2	3	4	Strongly Agree 5

6. I thought there was too much inconsistency in this system.

Strongly Disagree 1	2	3	4	Strongly Agree 5

7. I would imagine that most people would learn to use this system very quickly.

Strongly Disagree 1	2	3	4	Strongly Agree 5

8. I found the system very cumbersome to use.

Strongly Disagree 1	2	3	4	Strongly Agree 5

9. I felt very confident using the system.

Strongly Disagree 1	2	3	4	Strongly Agree 5

10. I needed to learn a lot of things before I could get going with this system.

Strongly Disagree 1	2	3	4	Strongly Agree 5

VAS and SUS Questionnaire – SANLAB 2016

Figure A.6: System Usability Scale Rating Form

Word Recognition Task Questionnaire

Participant ________________ Condition Order ________________ Talkers ________________

1. Which set of words was hardest? Direction Non-Direction
2. Which condition was hardest? (Circle One) BCI No_Cues Dynamic
3. Which was easiest? (Circle One) BCI No_Cues Dynamic
4. How many target sounds did you miss all together?

   BCI _________ No_Cues ________ Dynamic _________

5. Did you ever forget which sound was the target? Was there a condition or set of words that this happened more with?
6. Did you ever close your eyes? Did you every look somewhere other than the focus cross? Why?
7. What situations (combination of words, condition, presentation order made it the hardest to identify the target and hit the button?)

Figure A.7: Word Recognition Questionnaire

A.3 Waveform Analysis by Participant

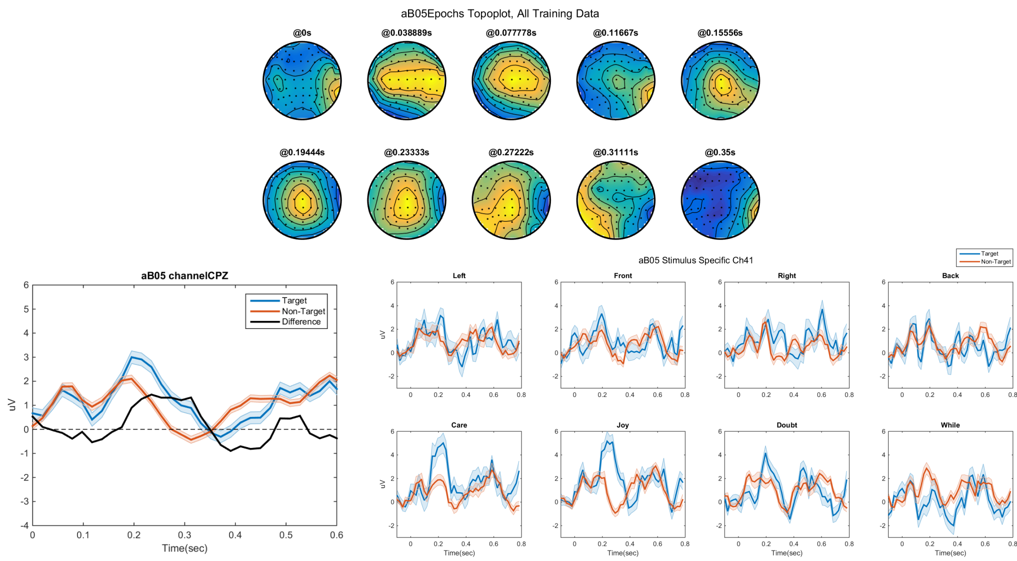


Figure A.8: Participant aB05 ERP example

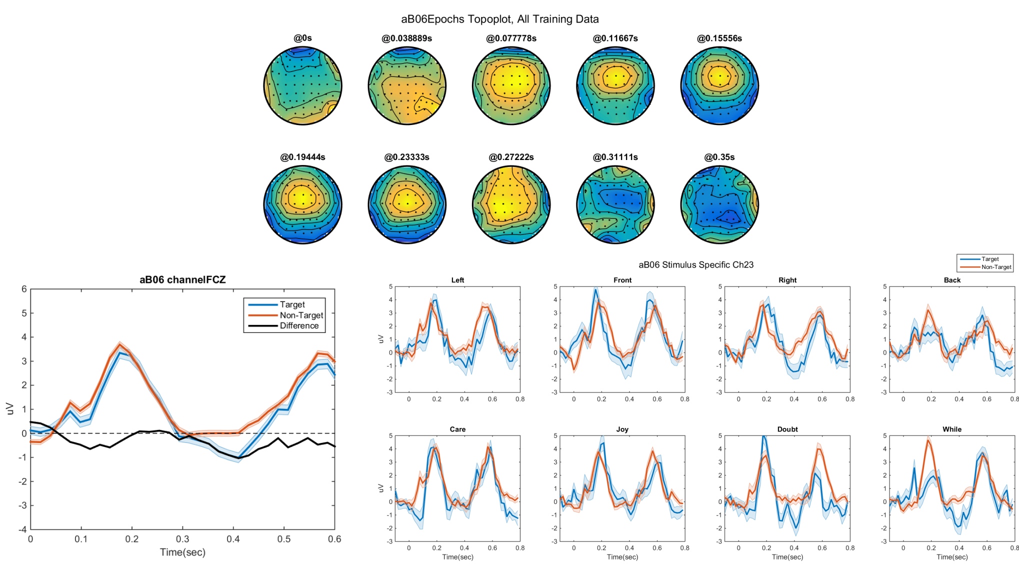


Figure A.9: Participant aB06 ERP example

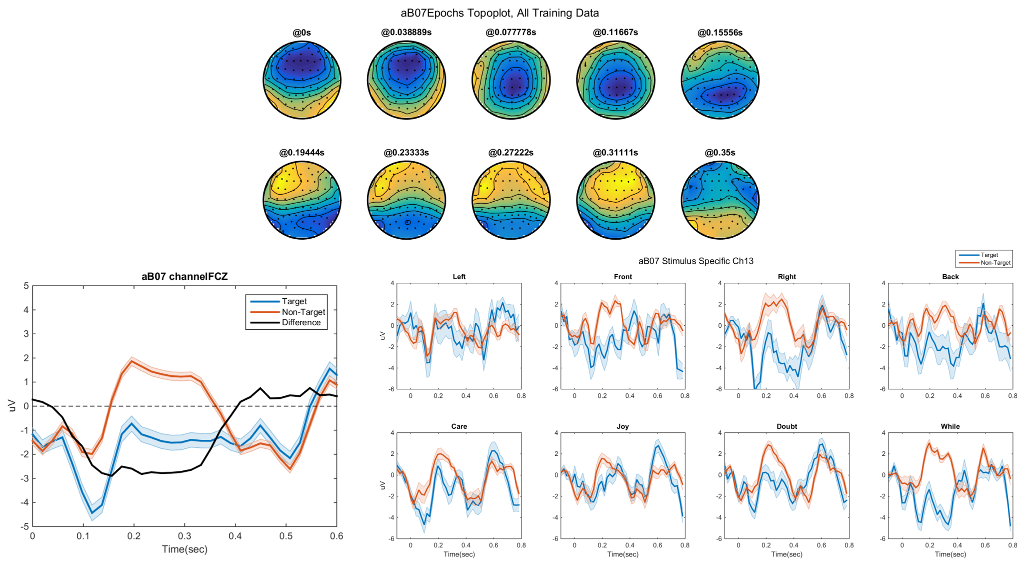


Figure A.10: Participant aB07 ERP example

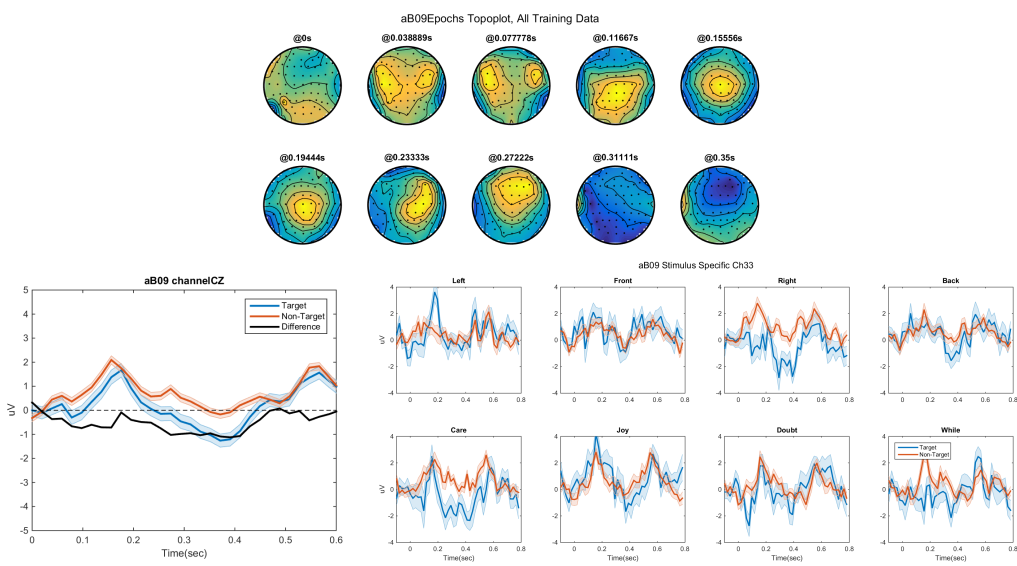


Figure A.11: Participant aB09 ERP example

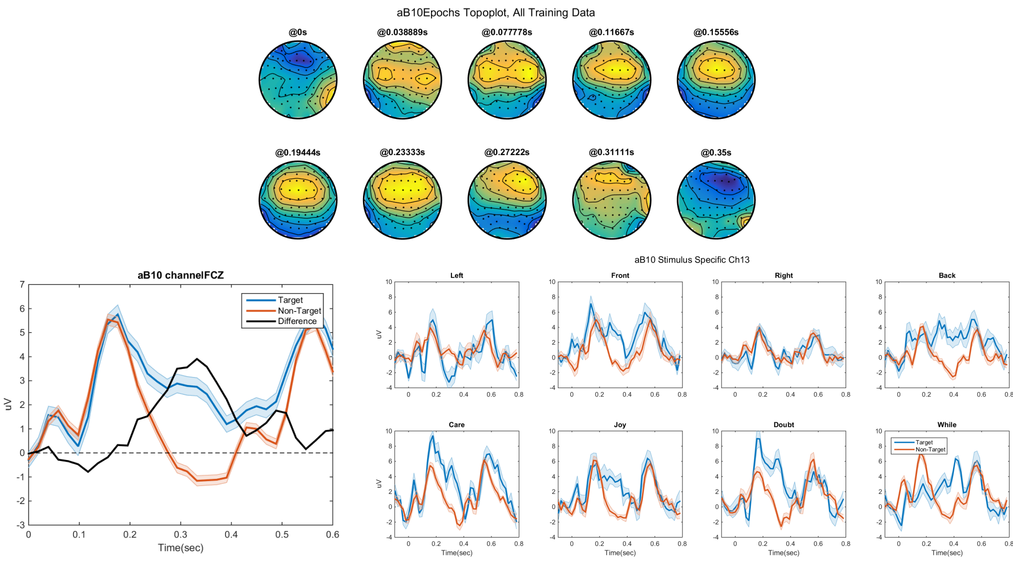


Figure A.12: Participant aB10 ERP example

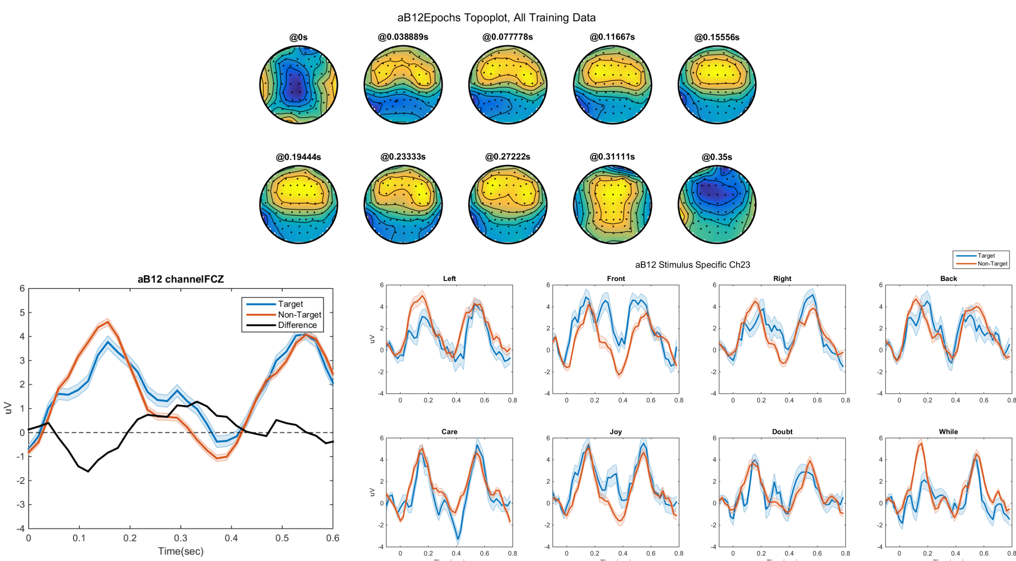


Figure A.13: Participant aB12 ERP example

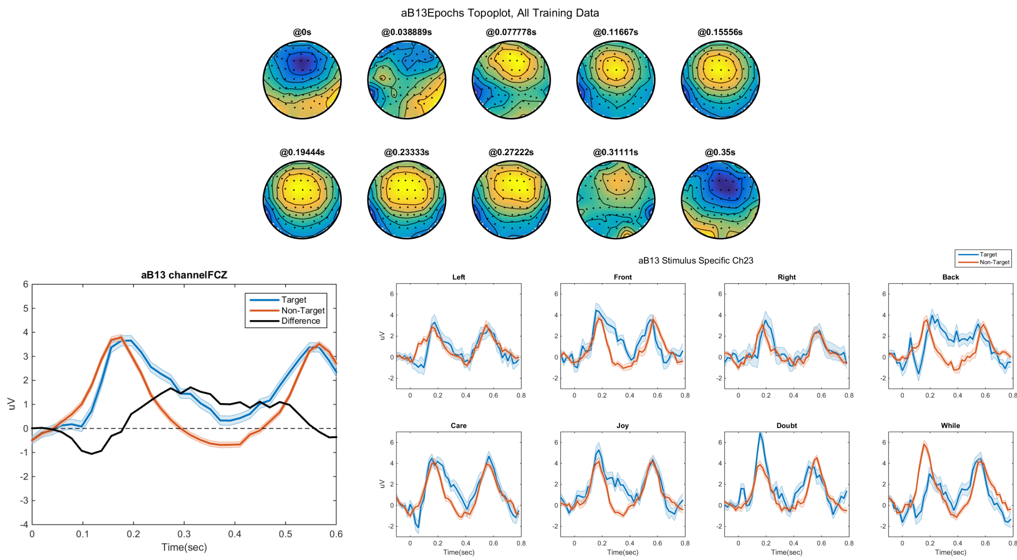


Figure A.14: Participant aB13 ERP example

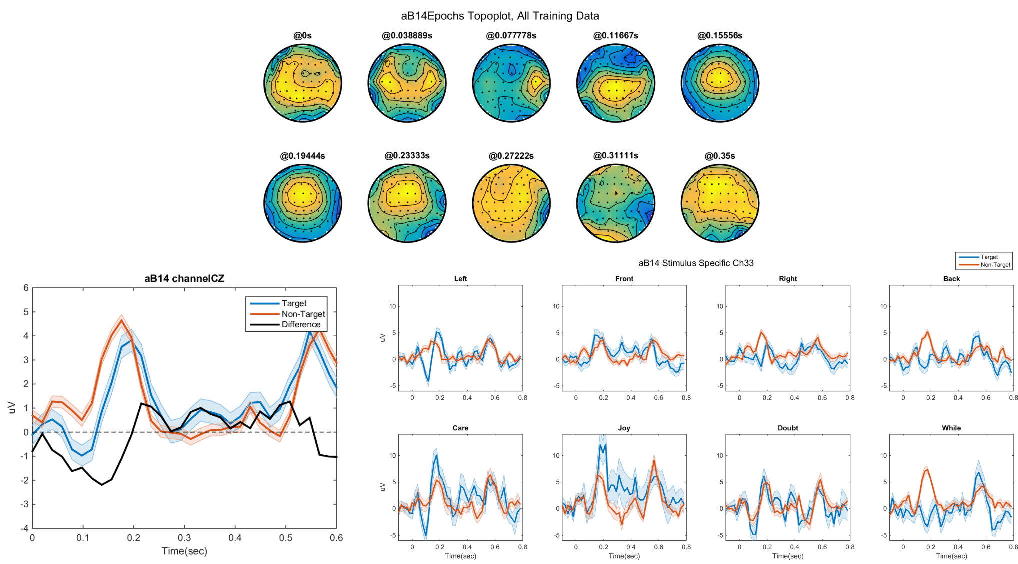


Figure A.15: Participant aB14 ERP example

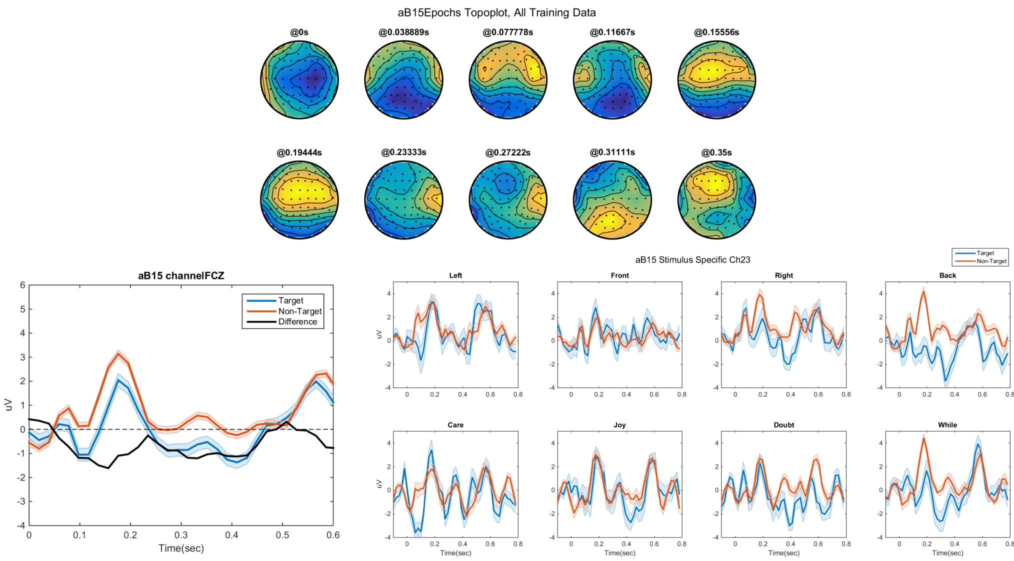


Figure A.16: Participant aB15 ERP example

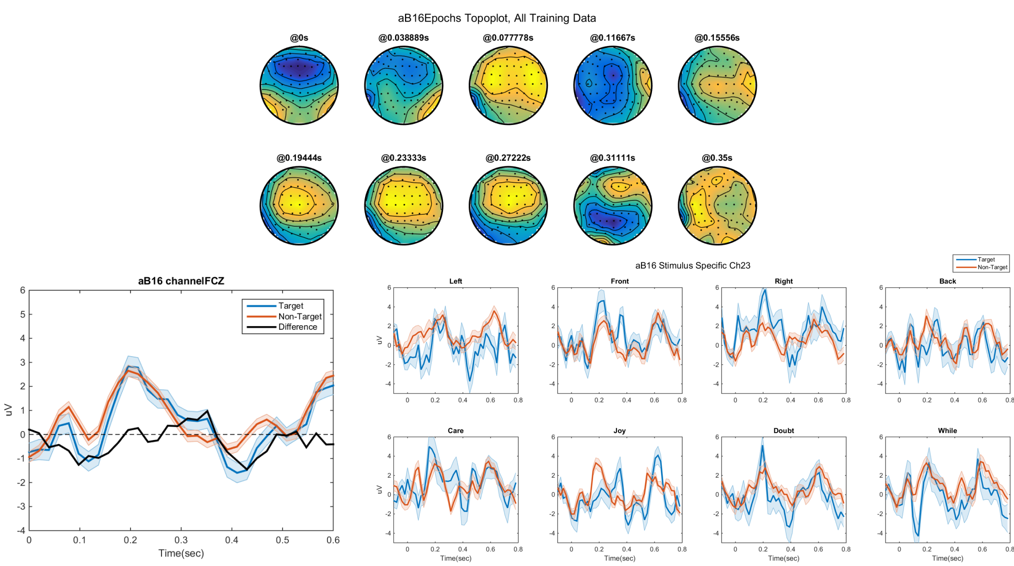


Figure A.17: Participant aB16 ERP example

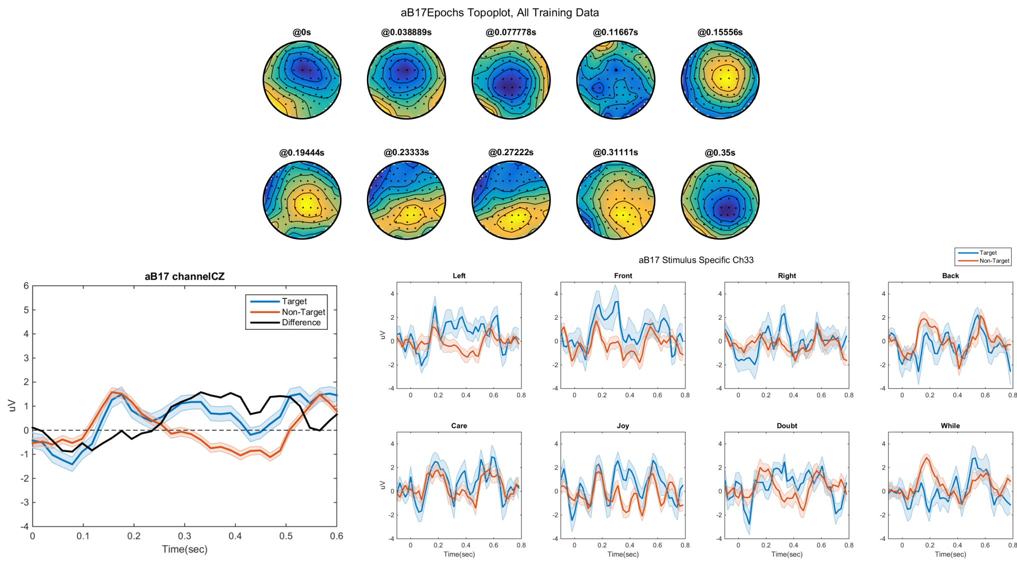


Figure A.18: Participant aB17 ERP example

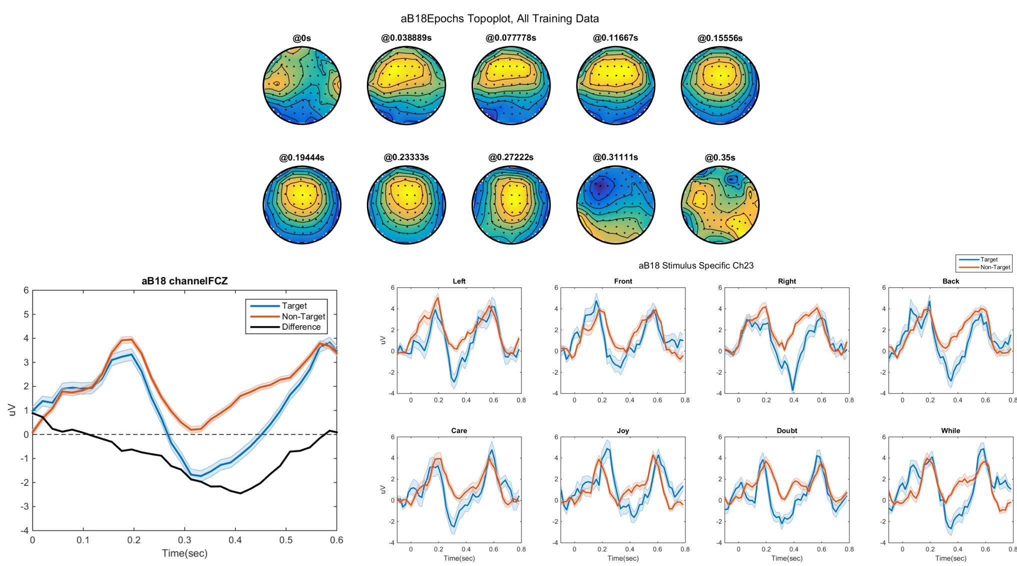


Figure A.19: Participant aB18 ERP example

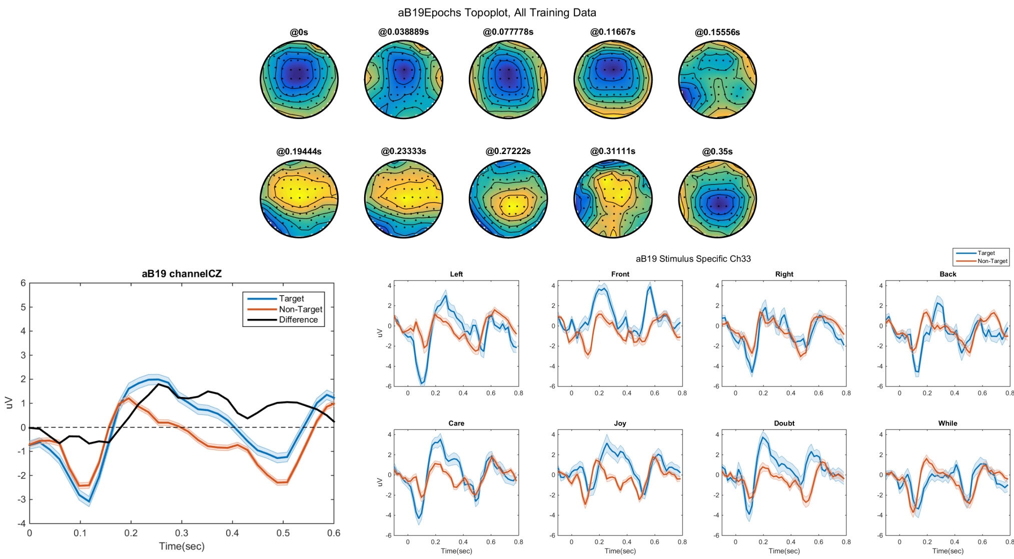


Figure A.20: Participant aB19 ERP example

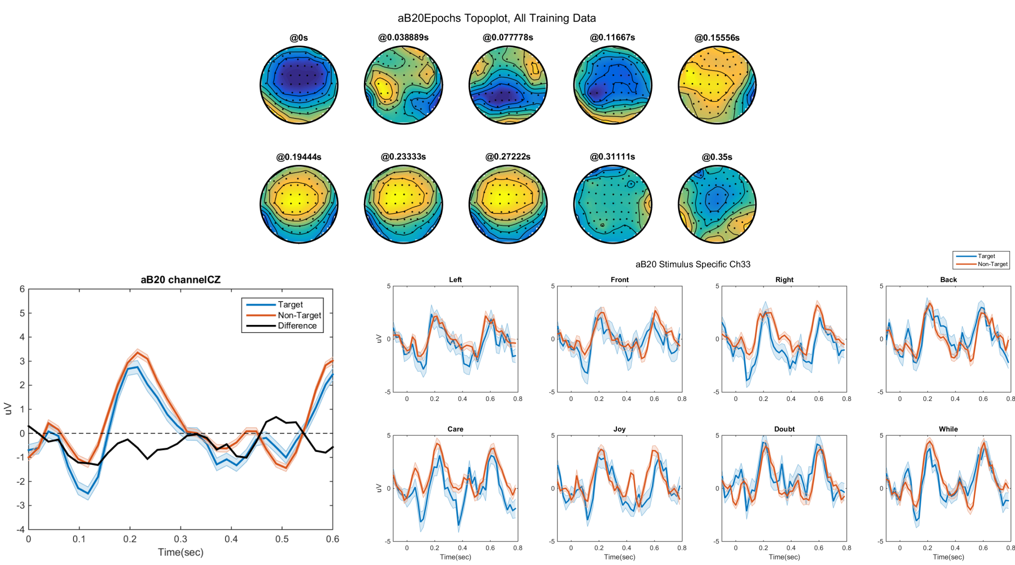


Figure A.21: Participant aB20 ERP example

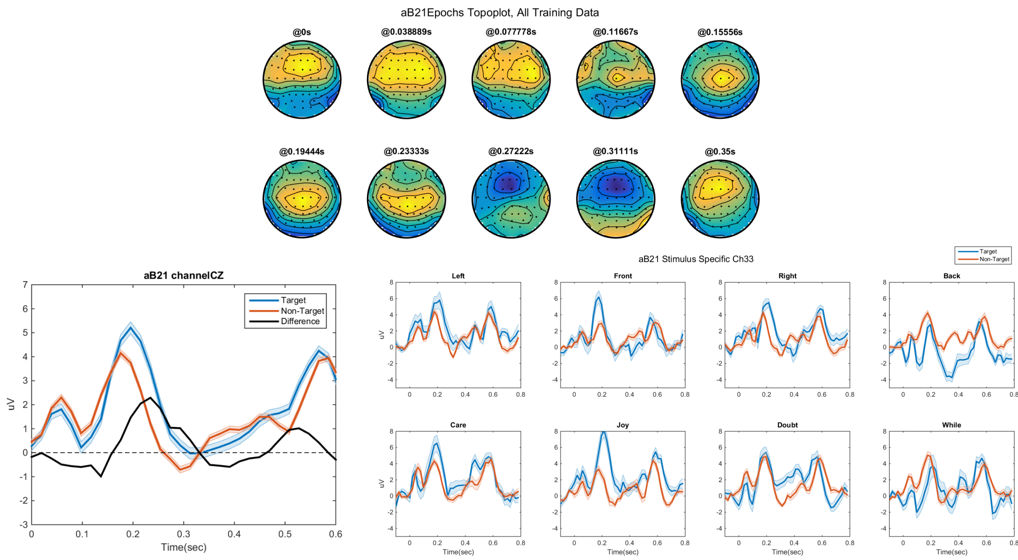


Figure A.22: Participant aB21 ERP example

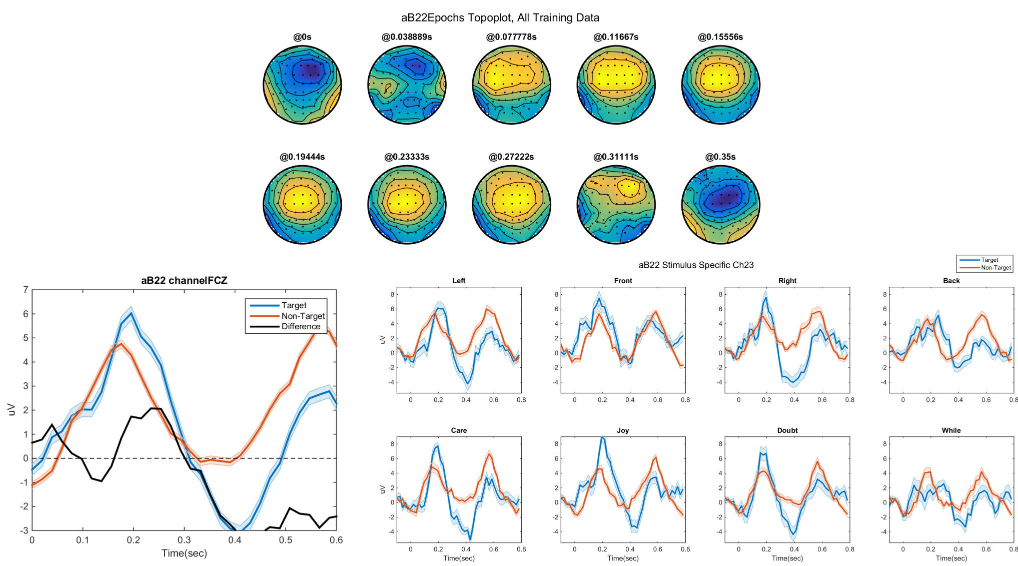


Figure A.23: Participant aB22 ERP example

A.4 Additional Results

A.5 Data Profile

Although several precautions were taken to prevent loss of data from participants, some unexpected events, technical difficulties and human error resulted in partial or total loss of data. What specific

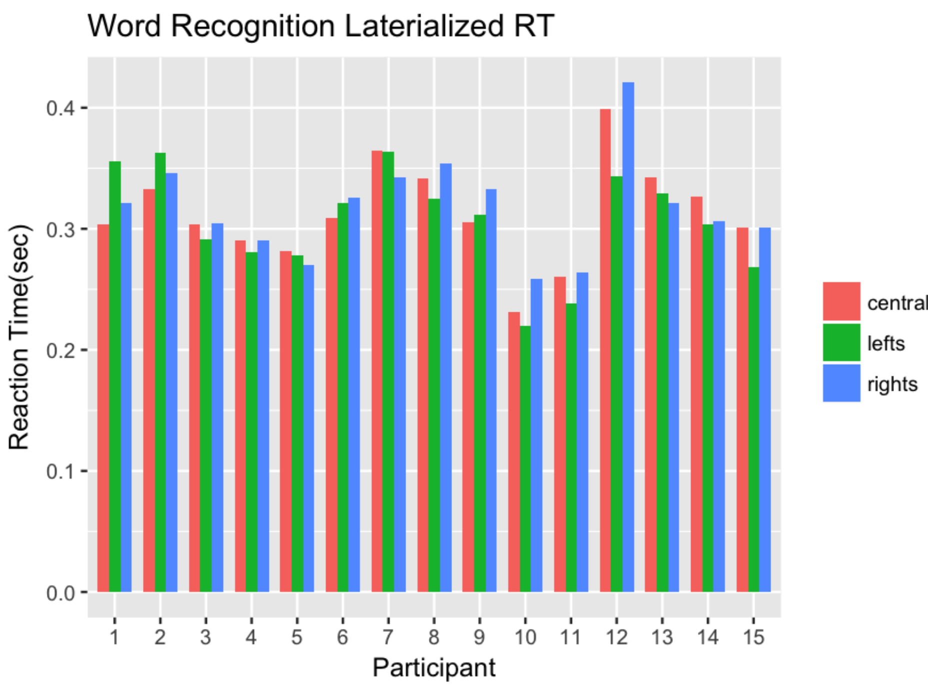


Figure A.24: WR lateralized Performance by Subject

data is not included in the following aggregated results will be outlined here but may be included in specific sections when noteworthy.

EEG data collected was interrupted and the raw data files were corrupted and unusable. for subject aB07, session 2, condition 1 or Direction words. Participant aB14, session 1, condition 2 or Non-Direction words also had a corrupted file. No online trials were completed for those specific session/conditions, as no data to train a decoding model was available. Because the training sequence takes the majority of the session time, in the instance of a corrupted training data file, there is little time to rerun the training set. Extending the session time would violate the consent form signed and increase fatigue. Conducting additional training trials would also provide the participant additional experience with the aBCI and potentially bias their results. For these reasons additional training sessions were not completed.

Participant aB10's second session of data collection was missing some of the sound onset trigger information due to hardware setup or malfunction. This precluded online trials and any offline data analysis including 'Across Session' testing.

The RLDA online decoder was not finalized until the second session for participant 6, therefore, no online aBCI data was collected for participant 5, and the first sessions for participants 6 and 7. The NASA-TLX was not completed or lost for participants 6 and 7.

Two participants that completed session 1 were dropped from the study completely. Participant 8 was dropped due to an inability to schedule the second session and technical issues with data files from their first session. Participant 11 did not meet all criteria to be included in the study and was not scheduled for a second session.

The WR task had 15 participants with 2 participants that did not have aBCI results to compare RT and aBCI accuracy.

Participants WR1 and WR2 do not have Dynamic condition results due to technical errors in data acquisition.

A.6 PacGame Code

A git repository of the PacGame software python code can be found here: https://epinasty@bitbucket.org/epinasty/

Git is a software versioning system that can you can learn more about here: https://git-scm.com/

Python installation and several dependencies are needed to run this software. This software was written in Python 2.7 syntax. The user interface primarily relies on pygame library that is not supported on macOS.

https://www.python.org/downloads/

A.7 Decoder Training Script: MATLAB

The following script was run after collection of aBCI training data to generate a RLDA classifier weight file. This file would be loaded into the PacGame software to be used in online classification.

% Process aBCI script

% another version of a scrip meant to generate a decoder file for the

% PACGame software. This version works with a commit put into the PacGame

```
% git repository branch test_data 10/17/16. This branch was used to debug the decoder.
% decoder_mat_comparison.m was used to achieve this.
% This scrip will generate a classifier for the following parameters, will not include
% dynamic stopping thresholds but will work best with online decoding.
starttime = 10;
forder = 4;
decR = 5.; %going to make about 51.2Hz sampling(20ms)
badpar = [];%1,32:34];
ewin = [-0.1 0.8];
base = [-0.1 0];%[-.5 0];
addpath('E:\\programs\\matlab\\eeglab13_4_4b\\');
%eeglab %need to run and close this for topoplots
pres= 15*4; %number of presentations per trail
%load in and process Symulink files
[dfiles,dpath] = uigetfile({'*.mat', 'EEG data files (*.mat)';'*.*','All files (*.*)'},...
'Pick EEG data files','multiselect','on');
if isnumeric(dfiles) & dfiles == 0
error('User canceled')
end
if ~iscell(dfiles)
dfiles = {dfiles};
end
subject = dfiles{1}(1:8);
raw = eegdata(fullfile(dpath,dfiles),[62 2 0 5],starttime);
raw.readlocs('L:\\eeg.locs'); %fullfile(pname,fname));
raw.fs = round(raw.fs);
fs = raw.fs;
```

```
flt = raw.filter(1,forder,'filtertype','high','filtermethod',@filtfilt,'logging',true);
%badchannels
%flt.averagebadchannel({'C6'});
% % Semiautomated removal of artifacts
[ic, A, W] = eegdata.dobss(flt,'fastica');
% Other options include: 'fastica', 'runica', 'infomax', 'sobi', 'amuse'
a_ic = eegdata.selectartifactic(flt, ic, eegdata.montage.eog, A,...
'thresh',[0.05,2]);% can include a fourth parameter A for manual selection of ICs, otherwise
use automatic criteria
flt.data = eegdata.removeartifactic(a_ic, ic, A);
flt = flt.filter(raw.fs/decR,forder,'filtertype','low','filtermethod',@filtfilt,'logging',true);
% downsample
dec = flt.copy();
dec.data = downsample(dec.data',decR)';
dec.trig(2,:) = dec.trig(1,:)+dec.trig(2,:)+dec.trig(3,:); %new snd trigger novisual situation
dec.trig = downsample(dec.trig',decR)';
dec.time = downsample(dec.time',decR)';
dec.fs = dec.fs/decR;
% trig correction
indx = find(dec.trig(2,:)>1);%fix instances of 2 values
dec.trig(2,indx) = 1;
dec.parsetrigs([5,2], 'logging',true);
%dec.parsetrigs(5, 'logging',true, 'offset',3);
figure, plot(dec.time, dec.trig([5,2],:))
itol = find(dec.evts.tdur < 0.085); %event's caused by overlap of sound and next par trigger
dec.evts =dec.rejectevents(itol); %this should leave 640 trials/training block
fprintf('Number of Trials: %d \n',length(dec.evts.tdur))
```

```
epoch = dec.copy();
epoch.epochdata(ewin,base,'logging',true);
test = epoch.copy();
parth = 'E:\\data\\eeg\\aBCI\\Training\\processed\\';
test.save(strcat(parth,subject,'.mat'),'abci');
disp(strcat('file saved \',parth,subject,'.mat'));
%% Reject large amplitude epochs
reject = expdata.threshold(epoch,150,'logging',true);
epoch = epoch.rejectepochs(reject(1:end-1)); %errors because numtrials is now in event struct
doesn't matter
%% model weight setup
modwin = [0,0.6]; %time window of data we'll create the model for.
mtimes = find(epoch.etimes>modwin(1) & epoch.etimes<modwin(2)); % && find(epoch.etimes<0.61)];
%time points considered for model
mch = 1:62;%channels considered for the model
z=randperm(size(epoch.vepochs,3));%array of random trial indecies
[ch,tpts,trials]=size(epoch.vepochs(mch,mtimes,z)); %use this randomized subset of trials
%arrays to pass model fit function
targs_z = [squeeze(epoch.evts.label(z)>16);squeeze(epoch.evts.label(z)<16)]; %
flat_ep_z = reshape(epoch.vepochs(mch,mtimes,z),[ch*tpts,trials]);
weights =train_RLDAshrink(flat_ep_z,targs_z);
pth='E:\\data\\eeg\\aBCI\\classifiers\\';
save(strcat(pth,subject,'_rlda2_classifier.mat'),'weights','decR','ewin','pres','base','modwin','fs','forder');
disp('classifier saved')
```